Masking and the Phase Response of the Auditory System

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Niall Andre Munson Klyn
Graduate Program in Speech and Hearing Science

The Ohio State University
2015

Dissertation Committee:
Lawrence L. Feth, Advisor
Eric C. Bielefeld
Eric W. Healy
Abstract

When a complex sound consisting of harmonic components is used as a masker for simple signals, the ability to simply detect a sound in that complex depends in part on power spectra of the sound, or signal, and that of the complex, or masker. This idea can be simply expressed in the ratio of power from the signal and that of the masker within a certain frequency range. If the phases of the harmonic complex are manipulated to create different waveforms, the amount of masking produced by the complexes can differ by a large amount. This difference in masking based on the phase relationships of the components is called the masker phase effect, or MPE.

The most prevalent explanation for this effect draws on models and simulations, and describes the MPE as the product of differentially modulated output waveforms from the basilar membrane. This research has commonly used so-called Schroeder maskers, which have similarly flat acoustic envelopes but are thought to produce different basilar membrane activation based on the rate of the instantaneous frequency sweep. This dissertation expands our understanding of the MPE over the course of three manuscripts. The first tests the role of the rate of the instantaneous frequency sweep in producing MPE with Schroeder maskers and maskers that share superficial features with Schroeder maskers. It also tests detection of spectrotemporally complex signals in order to provide an intermediate step between pure tone signal detection and speech recognition MPEs. The results from this study support the hypothesis that the MPE is dependent on the rate
of the instantaneous frequency sweep of the masker. It also shows MPEs of greater
magnitude for a 1 kHz signal and using harmonic complex maskers with phases set
according to estimates of traveling wave delay. The MPE for spectrotemporally complex
signals is shown to be similar to that for pure tones, and larger than what is reported for
spectrotemporally complex speech.

The second manuscript tests whether the MPE can be observed in an objective
threshold estimate, the wave V of the ABR. As the proposed location of the primary
physiological mechanism driving the MPE is the cochlea, the differences in masking
produced by the Schroeder maskers should be observable at all subsequent stages of
processing. Previous reports of the MPE in humans, however, have been limited to
behavioral data and one cortical MEG experiment. Thresholds for a 4 kHz tone burst
were estimated using wave V for twenty subjects in Schroeder maskers, and a significant
MPE (9.5 dB) was found. This is consistent with the anatomical loci of the MPE being
peripheral to the inferior colliculi. The smaller MPE than has been reported in behavioral
experiments may be a result of the short duration of the signal or overshoot.

The final manuscript details a new adaptive procedure for psychophysically
estimating the phase curvature of the basilar membrane. This procedure uses a small
number of trials at multiple phase curvature maskers and adapts the signal level until only
some maskers are ineffective. The effective maskers are discarded, new phase curvature
maskers are generated, and the process repeats. Simulations and a psychoacoustic
experiment provide data that this procedure can effectively estimate the phase curvature
in fewer trials than is necessary for the conventional procedure.
This document is dedicated to my children, Aram, Reijer, and Xochitl,

and to my wife,

Liesa

Thank you
Acknowledgments

Thanks are gratefully offered to the Department of Speech and Hearing for welcoming me and supporting me through my dissertation. The faculty, staff, graduate and undergraduate students have been integral to any success. Particular gratitude is due to my committee, Drs. Bielefeld and Healy, for their assistance and guidance in this process. Most of all, I would like to acknowledge the support of my advisor, Larry Feth. Without his support this would not have been possible.
Vita

1998........................................................Duluth East High School

2002........................................................B.A. Music, Classical Studies, Concordia College

2012........................................................M.A. Musicology, The Ohio State University

Fields of Study

Major Field: Speech and Hearing Science
# Table of Contents

Abstract ........................................................................................................................... ii

Dedication ......................................................................................................................... iv

Acknowledgments ........................................................................................................... v

Vita ....................................................................................................................................... vi

List of Tables .................................................................................................................... xii

List of Figures .................................................................................................................. xiii

Chapter 1: Introduction ................................................................................................. 1

1.1 Introduction ................................................................................................................ 1

1.2 Review of physiological transduction ...................................................................... 2

1.3 Masking ...................................................................................................................... 6

1.3.1 The power spectrum model .................................................................................. 6

1.4 Fluctuating maskers ................................................................................................. 10

1.4.1 Signal detection ...................................................................................................... 10

1.4.2 Fluctuating maskers for speech .......................................................................... 14

1.5 Masker phase effect ................................................................................................. 22
Chapter 3: Manuscript 2 - Masker Phase Effect in the Auditory Brainstem Response .... 83

3.1 Abstract .............................................................................................................. 83

3.2 Introduction ........................................................................................................ 84

3.3 Materials and methods ....................................................................................... 87
  3.3.1 Participants .................................................................................................. 87
  3.3.2 Stimuli .......................................................................................................... 88
  3.3.3 Procedures ................................................................................................... 90

3.4 Results ................................................................................................................. 92
  3.4.1 Amplitude .................................................................................................... 94
  3.4.2 Latency ......................................................................................................... 95
  3.4.3 Masked thresholds ...................................................................................... 96

3.5 Discussion ............................................................................................................ 98
  3.5.1 Latency and amplitude ............................................................................... 98
  3.5.2 Morphology .................................................................................................. 100
  3.5.3 Masked thresholds ...................................................................................... 101

3.6 Summary ............................................................................................................. 105

3.7 References ......................................................................................................... 106
List of Tables

Table 1. Pairwise comparison results from experiment 2. .................................................76
List of Figures

Figure 1. Twenty millisecond sections of maskers used in experiment 1. Arrows mark a discontinuity caused by repeating signals generated in the temporal domain, which is referenced in the general discussion. .......................................................... 54

Figure 2. Growth of masking for experiment 1. The figure shows the average overall level of a masker at signal threshold for the SBU signal in the top panel and the SBD signal in the bottom panel. Filled symbols are maskers with downward sweeping instantaneous frequencies (schr+, lin+, log+), open symbols are maskers with upward sweeping instantaneous frequencies (schr-, lin-, log-). Lines are simple linear regression to show approximate slope. Error bars represent one standard deviation. Symbols have been shifted along the abscissa to add clarity. .......................................................... 58

Figure 3. Mean MPEs for experiment 1. Error bars show standard deviation.............. 60

Figure 4. Growth of masking for experiment 2. Filled symbols are downward sweeping maskers ("+" maskers), open symbols are upward sweeping maskers ("-" maskers). Error bars are one standard deviation. Symbols have been shifted along the abscissa for clarity. .............................................................................. 66

Figure 5. Mean MPE for the signals in experiment 2. Error bars show standard deviation. .................................................................................................................................................. 68
Figure 6. Results from dynamic gammachirp filterbank for filter tuned to 4 kHz with an input of a 10 ms 4 kHz tone from 10-20 ms of each masker. The left column shows signals in the presence of maskers with downward sweeping instantaneous frequency (schr+, lin+, log+, don+). The right column shows signals in the presence of maskers with upward sweeping instantaneous frequency (schr-, lin-, log-, don-). Ellipses have been centered on the response to the 4 kHz pure tone. .......................................................... 73

Figure 7. Harmonic complexes ("Schroeder maskers") with phases set according to the formula proposed by Smith et al. (1986), an $f_0$ of 100 Hz, and components from 200-5000 Hz........................................................................................................................................... 89

Figure 8. Response waveforms from one subject to schr- (top) and schr+ (bottom) with a 4 kHz tone burst at 26 ms (vertical line). Arrows mark the wave V of the 80 dB condition. ........................................................................................................................................... 93

Figure 9. Average peak-to-trough amplitude of wave V as a function of signal level with regression lines. Error bars show standard deviation. Symbols are horizontally offset for clarity. ........................................................................................................................................... 95

Figure 10. Average latency as a function of signal level with regression lines. Error bars show standard deviation. Symbols are horizontally offset for clarity. ......................................................... 96

Figure 11. Waveforms from one subject to schr- and schr+ maskers without a signal. ... 97

Figure 12. A box plot of estimated thresholds for all subjects (n=20) using the presence/absence of wave V of the ABR. Boxes show first thru third quartiles with the 10th and 90th percentiles shown as whiskers. Data outside these percentiles are shown as individual dots........................................................................................................................................... 98
Figure 13. Flow chart of proposed adaptive procedure with the threshold value of 0.75 and four trials per block. Percent correct is abbreviated p(c). ................................................................. 117

Figure 14. Average absolute error as a function of trials per block with error bars at one standard deviation, based on 10,000 simulations of the adaptive procedure using five C values, a 3% inattention rate, and a threshold value of 1 (see text). Symbols have been offset along the abscissa for clarity. ............................................................................................................ 120

Figure 15. Average absolute error as a function of trials per block with error bars at one standard deviation, based on 10,000 simulations of the adaptive procedure using five C values, a 3% inattention rate, and a threshold value of n-1/n (see text). Symbols have been offset along the abscissa for clarity. ............................................................................................................ 121

Figure 16. Average absolute error as a function of trials per block with error bars at one standard deviation, based on 10,000 simulations of the adaptive procedure using five C values and a threshold value of n-1/n (see text). Symbols have been offset along the abscissa for clarity. ............................................................................................................ 122

Figure 17. Average absolute error as a function of trials per block with error bars at one standard deviation, based on 10,000 simulations of the adaptive procedure using 5 C values and a threshold value of 1 (see text). Symbols have been offset along the abscissa for clarity. ............................................................................................................ 123

Figure 18. Average number of trials needed to reach the stopping point for the adaptive procedure (ordinate) versus number of trials per block (abscissa) based on 10,000 simulations of the adaptive procedure using a 2% inattention rate (see text). Other
inattention rates did not significantly change the outcome. Error bars are one standard deviation. .............................. 124

Figure 19. Individual average thresholds estimated with the conventional procedure... 129

Figure 20. Averaged data from Figure 18 with error bars showing standard deviation. Fitted curves are sinusoids with four free parameters (see text). Symbols have been offset along the abscissa for added clarity. .......................... 132

Figure 21. Estimates of the C value that produced the minimum amount of masking for four subjects. Larger symbols are from the adaptive procedure with error bars at a standard deviation. Smaller symbols are from the conventional procedure; filled are the local minima (the lowest threshold tested) and open are the minima from a curve fit to the individual's thresholds. Symbols have been offset along the ordinate for clarity........ 133

Figure 22. Differences between an individual's adaptive estimates and local minimum (A) and fitted curve minimum (B)................................................................. 135
Chapter 1: Introduction

1.1 Introduction

The simplest possible sound - a sine wave - may be fully described by its frequency, amplitude, and phase. For complex sounds, such as most encountered outside of a lab or clinic, these same parameters must be known for all the components of the complex. For many years of scientific research into hearing, the human ear was thought to be "phase deaf;" insensitive to changes in the phase relationships of the components of complex sounds. Helmholtz's (1875) foundational work in hearing described timbre as being determined by "the number and relative strength of its partial simple tones, and in no respect on their differences of phase" (p. 184). This view held sway for most of a century, but it became clear in the mid-20th century that ignoring the phase relationship of sound stimuli was unwise. First, the interaural phase relationships of simple and complex sounds clearly play a role in the perceived location of the sound source (e.g., Licklider & Webster, 1949). Second, even when played to a single ear, altering the phase relationships of a complex sound can have a noticeable effect on the sound's timbre or sound quality (Licklider, 1957). Later research found that these phase relationships play a role in other phenomena, such as masking (Mehrgardt & Schroeder, 1983). The broad goal of this document is to add to our body of knowledge on the phase response of the auditory system, and in particular its role in masking. As is discussed below, the basilar
membrane's phase curvature has been implicated in the differential masking observed in previous research. The physiological manifestation of this phase curvature is the "frequency glide," which has been observed in animal models (Recio & Rhode, 2000; Shera, 2001; Summers et al., 2003). In the chapters that follow, we attempt to help bridge the gap between the physiological and computer models and the behavioral responses in humans.

Masking is of fundamental interest to hearing scientists and plays a significant role in our everyday lives. To paraphrase Tanner (1958), masking is the change in detectability of a sound in the presence of another sound. Masking extends beyond the laboratory, however, as anyone who has tried to hold a conversation at a loud party can attest. Mehrgardt and Schroeder (1983) were the first of many to show that systematically manipulating the phase of a masker can create differences in masking of up to 20 dB or more for simple signals. We will refer to this as the "masker phase effect" or MPE. The phase relationships of the partials of a complex sound do matter, but to understand the potential roles they may play in masking it may be useful to review some of what's known about the peripheral auditory system. Specifically, of interest is the point of transduction within the cochlea, when the vibrations of sound are first turned into the electrochemical energy of a nervous system.

1.2 Review of physiological transduction

As sounds progress from the external environment to the pinna, to the tympanic membrane, and through the ossicular chain, they undergo substantial filtering before reaching the oval window of the cochlea. This physical filtering results in selective
emphasis of mid-frequency bands and significant attenuation of low and high frequencies. At this point the movement of the stapes initiates a pressure wave in the fluid-filled cochlea, which propagates through the perilymph of the scalae vestibuli and tympani to the circular window. Most relevant to the sense of hearing is a slower, coupled wave consisting of a differential pressure wave in the endolymphatic fluid of the scala media and a corresponding displacement wave of the basilar membrane (Musiek & Baran, 2007; Patuzzi, 1996; Robles & Ruggero, 2001). This coupled wave is known as the "traveling wave", and travels from the base of the cochlea to the apex over the course of milliseconds (von Békésy, 1960). The basilar membrane itself extends laterally from the osseous spiral lamina of the modiolous out towards the spiral ligament on the lateral wall. The osseous spiral lamina resembles a spiral staircase that winds up from the base to the apex of the cochlea and gets progressively narrower as it rises (Musiek & Baran, 2007), whereas the basilar membrane gets progressively wider and thicker as it moves from the base to the apex (Musiek & Baran, 2007; Patuzzi, 1996; Slepecky, 1996). This thickening also corresponds to a decreasing stiffness gradient from base to apex, both accomplished through changes in the composition of the extracellular matrix of the membrane and the bordering mesothelial cells (Slepecky, 1996), and resulting in a significant change in impedance between the basal and apical end of the basilar membrane. The portion of the basilar membrane that is closest to the osseus spiral lamina is relatively rigid due to its proximity and the presence of the stiff pillar cells above it, whereas the more lateral portion of the basilar membrane is able move more freely (Musiek & Baran, 2007; Patuzzi, 1996). The mass and stiffness of the basilar membrane
includes another significant component, however, by way of the cells of the organ of Corti including the inner hair cells (IHC), outer hair cells (OHC), and support cells such as the Hensen, Deiters, and pillar cells.

This gradation of stiffness and mass along the length of the basilar membrane explains many of the passive features of its response to stimulation. The energy input via fluid compression by the stapes spreads via the slow fluidic pressure wave and the basilar membrane displacement wave (Patuzzi, 1996). The traveling wave in response to a single frequency stimulus encounters increasing amounts of compliance from the thickening and increasingly flaccid basilar membrane, resulting in a “build up” of amplitude as the wave slows down. At a particular point on the basilar membrane the wave “stalls”, and beyond that point the wave is fully dissipated. This process leads to a broad tonotopic mapping of the basilar membrane through its frequency dispersive qualities, where incoming acoustic energy has a corresponding place on the basilar membrane that responds most to a particular frequency (Patuzzi, 1996; Robles & Ruggero, 2001; Ruggero, Rich, Recio, Narayan, & Robles, 1997; von Békésy, 1960).

Moving to phase response of the basilar membrane, observing the movement of any given location on the basilar membrane shows evidence of a change in the instantaneous frequency of the response (Recio & Rhode, 2000; Recio, Rich, Narayan, & Ruggero, 1998). These "frequency glides" in mammals typically have an initial low frequency response which quickly moves towards the characteristic frequency and subsequently stabilizes. The rate of change of the instantaneous frequency may also be called the phase curvature for that location on the basilar membrane (Oxenham & Dau,
This is dependent on the characteristic frequency (CF) of the location on the basilar membrane, with lower CFs corresponding to longer delays in stabilization of the frequency response. Shera (2001a, 2001b) provided an explanation for cause of the frequency glides; he related the glides to the frequency dispersive properties of the traveling wave, and argued that the glide is a necessary consequence of the frequency-specific delay along the length of the basilar membrane. This frequency-specific delay is itself due to the physical hydrodynamics of the cochlea. The model presented by Shera was an excellent qualitative match for the frequency glides observed by other researchers in the upper frequency range; from roughly 1.5 kHz and up, termed the “scaling” region. Caution is warranted, he noted, in extrapolating to the apical portion of the cochlea where the change in duct size and the proximity of the helicotrema complicated the hydrodynamic predictions. Results supporting this explanation were found by Ramamoorthy et al. (2010), who used laser interferometry on passive guinea pig cochleae and compared their and other results to a more sophisticated hydrodynamic model. They found that modeling the fluid-structure coupling of the fluids of the scalae and the basilar membrane produced an excellent match to the observed physical behavior when the model included the physiological components of an actual cochlea; fluid compressibility, minor longitudinal coupling along the basilar membrane, and separate estimates of basilar membrane mass and stiffness. The evidence suggests that the frequency dispersive nature of the cochlea results in the frequency glide observed at the basilar membrane. This physiological behavior is believed to be the dominant determinant for the MPE. In order to discuss this phenomenon in more detail, we first turn to a brief overview of relevant
1.3 Masking

When two or more signals are presented to an ear, the detectability of those signals may be different than when they are presented independently. Some of this difference can be understood with reference to the physiological processes described above. Specifically, if a noise brings energy to the basilar membrane at the place that responds best to a particular pure tone signal, then the ability of the listener to detect that signal will depend in part on the relationship of the energy arriving from the signal and that arriving from the noise. This is a simple definition of the so-called "power spectrum model" of masking (Fletcher, 1940; Moore & Glasberg, 1987).

1.3.1 The power spectrum model

Formally in the power spectrum model, a masker may be represented by its power spectrum $N(f)$ when the power of the signal is $P_s$ and the auditory filter has a shape described by $W(f)$, then the power spectrum model is expressed as:

$$P_s = K \int_{-\infty}^{\infty} N(f)W(f)df,$$

where $K$ is a constant that represents the necessary signal-to-noise ratio at threshold (Moore & Glasberg, 1987; Patterson & Nimmo-Smith, 1980). This understanding of masking relies on a number of assumptions that may be simply deduced from the preceding formula: 1) the auditory system may be modeled as a series of overlapping bandpass filters. 2) Only one hypothetical auditory filter is relevant for detection, meaning that only noise that passes through that filter has any effect on detection and that the subject attends ideally to the filter with the maximum SNR. 3) The only signal and
masker parameters necessary to describe detection are the long-term power spectra of the signal and masker. As is true with any model, these assumptions serve to simplify a complex process in order to make it conceptually or mathematically tractable, and a result of this simplification is that the model is unable to account for phenomena that are not accurately described by its assumptions. Nevertheless, the model has successfully explained a wide variety of experimental results that date back nearly a century. An overview of some of these studies follows, and will allow a glimpse into the development of this model and its successes.

In 1924, Wegel and Lane pioneered experimentation on masking of one tone by another tone in one ear. Using a masker fixed at a single frequency and voltage, they amplified signals with frequencies from 400-4000 Hz until a difference was detectable in the masker. As the signal approached the masker in frequency the amount of masking increased, up until the signal was very close in frequency to the masker. At this point the masker and signal were close enough together in frequency to produce beats and thus improve the detectability and lower the observed masking. At the harmonics of the masker the same phenomenon may be observed, which is more pronounced at higher presentation levels. Egan and Hake (1950) confirmed and expanded on these results by using a narrow band of noise as a masker rather than a pure tone. This allowed for a smooth graph of the relationship between the amount of masking and the signal frequency, avoiding the "dips" observed by Wegel and Lane's (1924) use of a pure tone masker. These studies provide part of the evidence used to support the latter two assumptions (above) of the power spectrum model.
The first assumption listed above of the power spectrum model reflects research by Fletcher (1940), which presented masking data from pure tones and noise. Fletcher investigated the bandwidth of noise necessary to mask a signal, and related this bandwidth to physical distances on the basilar membrane. The signals and maskers were continuous, and Fletcher explicitly discarded the relevance of phase relationships in his introduction. Fletcher found that the portions of noise that exceeded the “critical” bandwidth did not contribute to masking, and he theorized that the masker components distant from the signal frequency were too far away on the basilar membrane to have any effect on the amount of masking. Noise within the critical bandwidth did affect the amount of masking, and as the bandwidth was reduced to less than the critical bandwidth for a particular frequency, the amount of masking observed would be similarly reduced. In other words, what Fletcher postulated was that the basilar membrane functioned as a series of bandpass filters, and that thresholds for a pure tone could be calculated based solely on the signal-to-noise ratio within a single filter: the power spectrum model.

Over the next several decades, a great deal of effort would be devoted to the refinement of Fletcher’s basic model. Estimating the critical bandwidth by way of 2-tone masking, masking by and of narrow bands of noise, and loudness summation produced results that were quite consistent over the range of methodologies employed (Scharf, 1970). One of the most comprehensive of these studies was that of Greenwood (1961a), who used pure tones masked by noise of varying widths and levels, as well as the inverse procedure – narrow bands of noise masked by pure tones – to obtain thresholds using the Békésy tracking method. The results are most clearly illustrated by a discussion of
Greenwood's experiment with noise maskers and pure tone signals. As the signal frequency approached the spectral center of a noise band that was less than or equal to the critical bandwidth for its center frequency, the masked threshold of the tone steadily increased until it reached its peak at the center frequency. Above that center frequency the threshold decreased. However, if the noise exceeded the critical bandwidth the masked threshold of the signal would reach a plateau over which the threshold no longer increased. At a sufficiently high frequency the threshold again decreased, producing masked thresholds that looked more trapezoidal than the triangular results of sub- and critical bandwidth noises. The level of the masked thresholds for sub- and critical bandwidths was proportional to the total power of the masker at low to moderate presentation levels. These results were consistent over the different experiments presented in the paper and were also largely consistent with others’ work. Furthermore, the data assisted Greenwood in more fully realizing Fletcher’s model relating the revised critical bandwidth to specific distances along the basilar membrane (Greenwood, 1961b).

The power spectrum model has been quite successful in predicting simple detection experiments, but as noted above, any simplification or assumption that is used to make a tractable model will necessarily limit the phenomena the model is able to represent. Even the earliest rigorous study of masking, that of Wegel and Lane (1924), recognized that while masking could occur by the interaction of stimuli and the basilar membrane it could also occur at higher levels in the auditory pathway. Moreover, sounds have duration and change over time. In order to explain any but the most basic detection studies it is necessary to combine the results of a short-term power spectrum model with,
at the least, some kind of temporal component, a decision which increases the complexity of the model substantially. The very appeal of the power spectrum model lies in its ability to simply represent relatively simple phenomena, such as the masking of fixed signals in steady state noise. Even restricting the discussion to masking, in order to model phase effects in masking, comodulation masking release, and modulation detection interference it is necessary to introduce complexity to the model by way of the inclusion of dynamic temporal information within and across “channels” in the auditory periphery.

1.4 Fluctuating maskers

1.4.1 Signal detection

Few sounds outside the laboratory do not change over time, but rather most fluctuate in level and spectral distribution of energy. The simplest expansion of the power spectrum model's detection experiment to include this fact would be to lower or raise the level of the masker at different points during its presentation. For this discussion the phrase "masking release" will be used, without specifically advocating for a particular interpretation of that phrase. Instead, in this discussion its use simply refers to any advantage in signal detectability, discrimination, or interpretation gained through modification of the masker.

Not long after Fletcher codified the power spectrum model, Miller and Garner (1948) tested detection of pure tones in fluctuating noise. Specifically, they used white noise to mask pure tones at octaves from 125-4000 Hz. The maskers were set so their continuous average level would be 40, 60, and 80 dB SPL, but the maskers were then cycled off and on for 25, 50, or 75% of the duration with interruption rates at 0.5, 2, 8,
32, 128, and 512 Hz. The subjects (the authors) controlled the level of the tone until they considered it to be at their masked threshold. By reducing the amount of time the noise was on, the overall level of the noise would be lowered and so masking would be reduced. However, the amount the masking was reduced far exceeded that predicted by that simple reduction (e.g., 3 dB in the case of a 0.5 noise/time ratio). At lower interruption rates, the difference in masking between the continuous and interrupted noise was as great as 45 dB for the 4000 Hz tone and 18 dB for the 250 Hz tone. As the rate of interruption increased, the difference in masking decreased, until nearly all benefit from the fluctuating masker was lost at an interruption rate of 128 Hz. As the total amount of time the noise was on and off was held steady within a single duty cycle, for Miller and Garner the relevant parameters were the duration of the silent interval and the frequency of the tone. Longer silent intervals allowed for easier detection, and lower frequencies needed longer silent intervals than higher frequencies to be equally detectable. Miller and Garner did not proposed specific mechanisms for the benefit from the fluctuating masker, but did tie their results to those described by Doughty and Garner (1947) in their experiments with necessary durations for the perception of pitch.

Buus (1985) offered an intuitively appealing explanation for the decreased thresholds in fluctuating maskers: "the auditory system somehow is able to detect the tone during periods of low masker energy as brief as a few ms." This idea has come to be known as "dip listening." Buus used two-tone and noise maskers in pure tone detection experiments, and in each case fluctuations in the temporal envelope of the masker led to a release of masking, up to 25 dB for his higher level masker. Slower envelope fluctuations
were more effective than faster ones, similar to Miller and Garner (1948), but for Buus' experiment the effect was robust up to the 160 Hz fluctuation rate. Dip listening can also help explain comodulation masking release (CMR) as first described by Hall, Haggard, and Fernandes (1984). In Hall et al.'s experiments, noise maskers were set such that the amplitude fluctuations within a critical band around the signal frequency had the same pattern as those outside of the critical band. Detection for these conditions was compared to detection with a random noise masker, where there would be no coherence in the temporal envelope across different bands. They found that the power spectrum model was sufficient to explain the masking produced by the random noise. However, adding energy (noise) outside the critical band could reduce the amount of masking if the noise had a similar temporal envelope to the noise within the critical band. In essence, the auditory system was using spectrally distant noise to listen in the dips of the auditory filter(s) that were tuned to the signal.

Bacon, Lee, Peterson, and Rainey, (1997) provided research that helped untangle the roles of within and across channel processes in the benefits of a sinusoidally modulated random noise masker. Bacon et al. presented data from different masker widths (0.5-4 ERBs and broadband) at different modulation depths (0, 0.75, 1), modulation rates (2-200 Hz), and masker levels (0-40 dB SPL spectrum level). Different combinations of masker parameters were tested with pure tone signals from 250-4000 Hz. The difference between the broadband and narrower band maskers was interpreted as CMR, which they argued arose from the contribution of across channel processes. The .5 or 1 ERB masker width would be primarily attributed to within channel processes.
Consistent with CMR research, broadband maskers produced a larger masking release in conditions that showed differences between the modulated and unmodulated maskers. The modulation depth had a surprising effect on the signal detection by their subjects, substantially - and significantly - reducing the masking release compared to the full modulation by as much as 30 dB. Their interpretation is that, for full modulation, the amount of masking produced in the dips of the masker is governed by the amount of forward masking produced by the previous peak (e.g., Widin & Viemeister, 1979). When the modulation depth is only 0.75, they argued that simultaneous masking dictates the amount of masking observed, an effect bolstered by further experiments in Bacon and Lee (1997). Modulation rate differentially affected the masking release observed; for maskers with a width of approximately 1 ERB, the modulation rate had little impact on amount of masking release (but see Peters & Hall, 1994 for a contrary result). For the broadband maskers the masking release decreased as the modulation rate increased until there was little to no masking release at modulation rates above 100 Hz. Finally, the effect of masker level also differentially affected ERB vs. broadband maskers; in particular the ERB masking release appeared to asymptote at 30 dB SPL, whereas the broadband masker showed no such tendency. Two explanations were offered, first that the broadband masker was subjected to more of an effect of suppression, effectively reducing the level of the broadband masker and preventing an increase in simultaneous masking in the dips of the modulated masker. The second explanation rested on one proposed by Plack and Oxenham (1996) that the masker was experiencing compression from the basilar membrane while the signal was low enough in level to be in a linear
region. As the signal level was increased it eventually reached the compressive region, but only for the ERB masker in their experiment. In the broadband masker the signal level at threshold may have stayed in the linear region, meaning the growth of masking did not asymptote.

One variable that is not constant between the experiments on masking release provided by fluctuating maskers is the modulator type. Square-wave, sinusoidal, raised sinusoidal, and mixed sinusoidal modulators have been used to change the envelope characteristics of the noise maskers, but it does not appear that the effects of different modulators have been systematically explored in signal detection.

1.4.2 Fluctuating maskers for speech

Rarely in the world outside a lab is hearing concerned with simple signal detection, so it seems prudent to give some consideration to the effect of fluctuating maskers in a more ecologically relevant task and using spectrotemporally complex signals, i.e., speech comprehension. The impact of fluctuating maskers on speech perception has been studied for the better part of a century (Licklider & Goffard, 1947; Miller, 1947; Miller & Garner, 1948; Miller & Licklider, 1950). Some of the earliest examples of masking release in published research investigated the effect of different amplitude modulation rates on the amount of masking observed for speech. Miller (1947) noted that the intelligibility of speech varied according to the rate of interruption of the masking noise, but a systematic investigation was first presented by Miller and Licklider (1950). Using phonetically balanced words, Miller and Licklider found that speech recognition at certain SNRs had a noticeable peak at a modulation rate of approximately
10 Hz. Irregular interruption of the noise presentation created only minor changes in their results. At lower modulation rates entire words may be presented in a silent or noisy interval, and at higher modulation rates the benefit accrued from modulation nearly disappears, similar to that described above for signal detection by Miller and Garner (1948). A 10 Hz rate of noise interruption was also found to produce masking release for spondees in NH listeners by Wilson and Carhart (1969). These results have been confirmed in studies using different modulators and duty cycles, though the precise peak in the modulation frequency for speech recognition may be closer to 8 Hz (Füllgrabe, Berthommier, & Lorenzi, 2006) or 20 Hz (Gustafsson & Arlinger, 1994) depending on other experimental variables such as the duty cycle, the signal type, and listener group.

Another relevant factor in speech recognition masking release is the level of attenuation for the masking noise during its “off” periods; the modulation depth. In the initial studies the noise was attenuated as much as was possible (e.g., by 80 dB in Miller & Licklider, 1950), but the importance of the depth of modulation and its interaction with presentation level, duty cycle, and signal type was appreciated by early researchers (Carhart, Tillman, & Johnson, 1966; Pollack, 1954). Perhaps the most systematic study of modulation depth was published more recently by Gnansia, Jourdes, and Lorenzi (2008), who used seven modulation depths ranging from 12.5% to 100% in a consonant identification task in sinusoidally modulated noise. The overall results showed a distinct increase in masking release at modulation depths above 50%. Even the smallest modulation depth (12.5%) produced significant masking release for consonants, however, corresponding to an SNR during the off portions of the masker of approximately -5 dB.
The general trend of increasing modulation depth producing larger masking release was consistent with other studies that have varied the depth of modulation (Carhart et al., 1966; George, Festen, & Houtgast, 2006; Gustafsson & Arlinger, 1994; Pollack, 1954). The benefit of a shallow modulation depth for speech does stand in contrast to results for signal detection, where less than a 75% modulation depth substantially degraded the benefit of modulating a noise (Bacon & Lee, 1997).

The effect of modulation depth is dependent in part upon the relationship of the time the noise is at its maximum and minimum, which with a simple square-wave modulated noise may be determined by the duty cycle. Several other modulators have been used in the study of masking release, however, and the choice of modulator type affects the amount of masking release observed. The noise may be modulated by a single sinusoid or several (Gustafsson & Arlinger, 1994), by regular or irregular square waves (Miller & Licklider, 1950), or by the extracted temporal envelope from one or more talkers (Bacon, Opie, & Montoya, 1998; Bronkhorst & Plomp, 1992; Festen & Plomp, 1990; Simpson & Cooke, 2005). Bacon et al. (1998) used noises modulated at 100% by a square-wave at 10 Hz, by the envelope from a single talker, or by the envelope of multi-talker (20 talker) babble, which allowed them to examine the relationships between these different modulation types and their subject groups. The square-wave modulated noise provided the largest masking release for sentence intelligibility, averaging approximately 15 dB for the normal hearing listeners and 5 dB for the hearing impaired listeners. Noise modulated by the envelope of multi-talker babble was indistinguishable in its effects from steady speech-shaped noise, and noise modulated by a single talker fell between these
two values. The periods of the masker in which the signal would be expected to dominate the combined waveform are increased in number and degree as the modulator is changed from multi-talker to single-talker, and increased again from single-talker to square-wave modulation.

The task and signals given to listeners also play a role in the observed masking release. In particular, the modulation rate that produces the largest masking release varies according to the type of target. When the task is sentence recognition, the best modulation rate for the noise tends to be higher (>10 Hz) than when the task is recognition of spondees or nonsense syllables (<10 Hz) (e.g., Bacon et al., 1998; Gustafsson & Arlinger, 1994; Miller & Licklider, 1950; Wilson & Carhart, 1969). Buss et al. (2009) tested the effect of rate for sinusoidally modulated noise on three different tasks that all used the same stimuli; 500 consonant-nucleus-consonant words were used in 3I-3AFC detection, 1I-3AFC identification, and a single interval open-set identification task. The masker was continuously presented at a constant level, and the presentation level of the words was varied to target the 50% accuracy point. Detection and 3AFC identification had the largest amount of masking release (15-20 dB on average) at low modulation rates (2.5 Hz) and increasing the rate steadily to 40 Hz decreased the amount of the masking release to ~10 dB. For open-set identification the opposite trend was present, with increasing amounts of masking release as the modulation rate changed from 2.5-40 Hz but with much less interaction overall and stayed between 5-10 dB on average. Buss et al. interpreted these results as a reflection of the different amount and quality of information necessary to successfully perform these tasks. Their interpretation was that
detection and small-set identification were facilitated by a single, relatively high-quality glimpse of the signal, whereas open identification was improved by multiple, lower-quality glimpses that were spread out over the entire word.

This idea of a "glimpse" is closely related to the dip listening hypothesis proposed by Buus (1985). Cooke (2006) formalized glimpsing in speech perception as part of his model of an automatic speech recognition system. Cooke’s glimpses are spectrotemporal portions of the masker and speech stimulus with a given SNR that are combined by the listener in order to extract the speech information. This basic model may help to explain how listeners are able to recognize speech in maskers that exhibit wideband fluctuation like square-wave modulation, as well as more complex modulation like checkerboard noise or speech. While the idea of dip listening does allow for different envelope features in different frequency bands, analyses that use glimpses tend to be even more focused on the quality of the target after some model peripheral processing, rather than of the acoustic signal itself.

More recently, some researchers have argued for an alternative model of understanding masking release, namely that masking release is afforded through a release from modulation masking rather than a release from energetic masking. It is not directly in opposition to glimpsing or dip-listening, but puts less focus on the ratio of energy in different frequency bands. It is clear that the amplitude modulations of speech or speech-like noise carry a great deal of information for listeners (e.g., Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Shannon, Zeng, & Wygonski, 1998; Turner, Souza, & Forget, 1995; Van Tasell, Soli, Kirby, & Widin, 1987). It is furthermore clear from
psychoacoustic studies that the detection or discrimination of amplitude modulation of simpler signals can be depressed by the use of maskers that have similar modulation rates (e.g., Bacon & Grantham, 1989; Moore & Jorasz, 1996; Takahashi & Bacon, 1992).

Kwon and Turner (2001) argued that there is every reason to suspect that masking release and modulation interference occur simultaneously in tests of speech recognition in amplitude modulated noise. In their first experiment, speech was noise-vocoded into 12 channels that were each low-pass filtered to remove envelope fluctuations above 4 or 16 Hz. These were presented at 5 dB SNR in wideband noise that was SAM at a frequency of 1.5, 3, 9, 25, or 40 Hz. If modulation interference occurred, it would be expected that masker modulations that were similar to those found in the noise-vocoded speech would degrade performance relative to other modulation rates. For their subjects, masking release was observed for the higher masker modulation rates, and the lower rates were not significantly different from the unmodulated condition. A second experiment used bandpass speech and noise for the same consonant recognition task, and presented one of the central speech bands (625-1562.5, 1000-2500, or 1500-3750 Hz) with a modulated (8 Hz) or unmodulated noise band. In this case, masking release was observed for two of the speech bands when the masker was in the same frequency region. When the masker band was further from the speech band, modulation interference of approximately 10% was observed. The third and fourth experiments were modifications of the first and second, respectively, but further degraded the speech signal with noise-vocoding and fewer speech channels. The results mirrored what was found in the first two experiments. Kwon and Turner interpreted their results as supporting the existence of modulation interference.
in speech, but noted that it was a small effect unless the inherent redundancy of the speech was particularly degraded. In this view, the modulated noise was perceptually fused with the target speech, providing inaccurate cues that reduced speech recognition (Bregman, 1994).

Stone, Füllgrabe, Mackinnon, & Moore, (2011) were able to provide much stronger evidence of modulation masking using tone vocoded speech mixed with noise in different ways. The first, which they termed "signal domain mixing," was essentially the same process used in previous masking release studies. The signal and masker were mixed prior to being vocoded, which resulted in the preservation of the noise's envelope in the output signal. The other mixing type, "envelope domain mixing," added the average RMS value of the noise to the speech signal in order to remove any amplitude fluctuations from the noise contribution to the mixture. These were compared using "steady" speech shaped noise, with an 8 Hz sinusoid amplitude modulation, or with the envelope of an interfering talker. For the signal domain mixed stimuli, masking release was observed that was similar to that reported elsewhere; the speech shaped noise produced the poorest performance, the mix using the envelope from a single talker afforded slightly better performance, and the sinusoidal amplitude modulated noise allowed the best performance for their subjects. Contrarily, when the noise's envelope fluctuations were removed, the steady noise provided the best performance. Modulating that noise with a sinusoid decremented performance significantly, and the noise with the envelope of the interfering talker was yet worse. Stone, Füllgrabe, & Moore (2012) expanded on these results but used speech - not vocoded - mixed with pure tones, low-
noise noise (LNN), and gaussian noise as maskers. The three maskers were chosen to provide three levels of amplitude modulations within the channels used to separate the speech signals. Each was tested with and without a sinusoidal amplitude modulation at 8 Hz. As in the previous experiment, masking release was observed for the noise masker. The LNN maskers showed less masking release, and the pure tone maskers did not show evidence of masking release. The basic argument put forward by these researchers and expanded in Stone and Moore (2014) is that masking release is not a product of energetic masking (i.e., suppression and swamping) being reduced in the low energy portions of the masker, but a release from modulation masking from the modulated waveforms created by the noise. In Stone and Moore (2014) they tested different modulation rates with pure tone maskers and the speech recognition threshold, and found that only a 1 Hz modulation rate provided significant masking release. Indeed, other modulation rates caused poorer performance relative to the steady maskers, which is consistent with an increase in modulation masking due to the introduction of envelope fluctuations in the maskers. The decrement in performance was quite small - on average only ~1 dB - and it's surprising given that the modulation rates (1-81 Hz) cover the most pertinent modulation rates of speech (Festen & Plomp, 1990; Rosen, 1992). These experimental results have had parallel support and theorization in recent models of speech transmission and recognition (Dubbelboer & Houtgast, 2008; Jørgensen & Dau, 2011; Jørgensen, Decorsière, & Dau, 2015; Jørgensen, Ewert, & Dau, 2013). It's worth noting that the amplitude modulation of the masker would be decreasing and increasing the level of the masker's modulations relative to the speech, yet has consistently shown an overall release
from masking in most studies. It appears these interactions have not yet been fully explored.

Though the cause of benefit for masker fluctuations in speech perception has been recently challenged, the challenge does not appear to be the relevance of the dips in the masker in revealing the speech signal. Rather, it is a challenge to the mechanism of masking during the "on" periods of the masker. This explanation's focus on the importance of information after some peripheral filtering rather than on the unfiltered acoustic waveform is a perfect segue into the primary focus of this dissertation.

1.5 Masker phase effect

As the discussion above shows, the acoustic waveform of a masker does not necessarily resemble the response of the basilar membrane at any given point. An excellent example of this is so-called Schroeder masking, first reported by Smith, Sieben, Kohlrausch, and Schroeder in 1986. The authors used a modified version of a formula first proposed by Schroeder (1970) to systematically vary the phase of the components of harmonic complexes. The formula is as follows:

$$\theta_n = -\pi n (n - 1)/N$$

Where $n$ is the harmonic number of the component and $N$ is the total number of components in the complex. The resulting acoustic waveforms have several interesting properties; first, they exhibit an extremely flat amplitude envelope with a very low peak factor. Second, by simply changing the sign in the equation, a time-reversed and inverted version of the waveform can be generated that has the same envelope. Lastly, the two waveforms (positive and negative) have linearly changing instantaneous frequencies, the
negative Schroeder complex (schr-) has a rising instantaneous frequency, whereas the positive Schroeder complex (schr+) has falling instantaneous frequency.

More interesting than the acoustic features of the waveforms are the mammalian responses to these complexes. In normal hearing listeners, the schr+ masker – with the downward instantaneous frequency sweep – produces significantly less masking of pure tones than the schr- masker. The difference in masking can be up to ~25 dB depending on the parameters of the test and the individual listener (Carlyon & Datta, 1997a, 1997b; Hoglund et al., 2013; Kohlrausch & Sander, 1995; Lentz & Leek, 2001; Oxenham & Dau, 2001b, 2004; Smith et al., 1986; Summers, 2000; Summers & Leek, 1998; Wojtczak & Oxenham, 2009a). Furthermore, significant differences in speech recognition have also been observed (Green & Rosen, 2013; Summers & Leek, 1998).

If considered from the perspective of the power spectrum model (Section 1.3.1 above), the two maskers have equal power spectra and would therefore be expected to produce equivalent amounts of masking. When the physiological processing (Section 1.2 above) is considered, it has been argued that the frequency dispersive nature of the cochlea causes differences in the nature of the output waveforms at any particular point on the basilar membrane. In other words, the two versions of the Schroeder masker produce differentially modulated (or "peaky" in the Schroeder literature) output waveforms, which has been described as a type of masking release (Gregan, Nelson, & Oxenham, 2013). From the start of Schroeder masking research, the underlying explanation has rested on the relationship between the phase curvature of the basilar membrane and that of the Schroeder maskers. Smith et al. (1986) proposed that the
downward-sweeping masker closely matched the magnitude but had an opposite sign of the phase-curvature of the basilar membrane at the location of the signal, resulting in a “peakier” internal waveform. In other words, the rising frequency glides in the basal end of the cochlea (Section 1.2) are counteracted by the falling instantaneous frequency of the schr+, and therefore the output of the filter contains large peaks and long valleys relative to the schr- case. These output waveforms of the basilar membrane permit listening in the dips of the schr+ maskers, something that is not possible to the same extent in the schr- masker. The two maskers are also likely differentially affected by the compressive nonlinearity of the basilar membrane. The signal in the dip portion of the schr+ masker would be amplified relative to that in the schr- masker (Smith et al., 1986). In addition, the peak portion of the schr+ would experience relatively less gain than the dips, which would reduce forward masking of the dips by the peaks (Gockel, Moore, & Patterson, 2002).

1.5.1 Animal models

Direct investigations of basilar membrane movement in response to Schroeder phase complexes have been pursued in chinchilla (Recio & Rhode, 2000) and guinea pig models (Summers, de Boer, & Nuttall, 2003). As these are the only examples of observation of the basilar membrane in response to these stimuli it is worth considering them, in particular the former, in a little more detail. Recio and Rhode (2000) used laser interferometry to observe the response of the basilar membrane of chinchillas to clicks, tones, and Schroeder maskers with and without an accompanying probe tone. Their results for clicks and tones substantiated the work of previous studies (e.g., Recio et al.,
but their results for the Schroeder maskers were novel. The formula used to construct the signals was identical to that found in Kohlrausch and Sander (1995), and they used an \( f_0 \) of 200 Hz and constructed the stimuli with all of the harmonics between 1000 and 20,000 Hz. The response to these waveforms was as predicted: \( \text{schr}^+ \) produced “peakier” waveforms than \( \text{schr}^- \), and the \( \text{schr}^- \) stimuli had higher RMS values than the \( \text{schr}^+ \) maskers. While the envelopes qualitatively appeared quite distinct in response to the positive and negative Schroeder complexes, a more empirical approach was used to quantify the differences. The authors computed the envelope and instantaneous frequencies of the basilar membrane responses to both stimuli. The duration of the envelope was found by observing the initial and final time the envelope exceeded 20% of its maximal value in one period of the masker waveform (i.e., 5ms), and then finding the time difference between these two points. As predicted, the \( \text{schr}^+ \) signals produced substantially shorter duration responses than \( \text{schr}^- \) signals. The instantaneous frequency response of the basilar membrane also showed differences between the two stimuli, with the \( \text{schr}^- \) signal inducing less accurate instantaneous frequency representation compared to the \( \text{schr}^+ \) signal. Finally, Recio and Rhode measured suppression of a tone in the presence of Schroeder maskers, though unlike many of the behavioral results previously discussed, the signal (8.1 kHz) was not present in the masker complex but was chosen to be close to the characteristic frequency of the recording site. This tone was presented at 36 dB SPL, and the level of the maskers were varied in 10 dB steps from 92-37 dB SPL except with a step of 5 dB from 92-87 dB SPL. The \( \text{schr}^- \) complex produced more suppression at every level, with a maximum difference
between it and the schr+ complex of ~10 dB. All of the results offered by Recio and Rhode were consistent with the hypothesis offered for psychoacoustic data; schr+ maskers produce peakier waveforms with lower RMS values and less suppression than schr- maskers at equal presentation levels.

Summers et al. (2003) carried out a similar experiment on the guinea pig, with results that were qualitatively consistent with those of Recio and Rhode. The responses recorded from schr+ signal were noticeably peakier than those elicited by the schr- signal, and interestingly the difference between the two signals diminished as the intensity increased. A similar result has been found in psychoacoustic data (Summers, 2000; Summers & Leek, 1998), where the difference in detection observable between the maskers is diminished as the level of the masker is increased. Summers et al. (2003) explained this variation by arguing that the phase curvature of the basilar membrane is flattened out as the level is increased.

A separate, indirect approach relies on the differences in auditory processing between humans and non-mammalian animals like birds. Bird data, though relatively scant, indicate a slightly positive phase curvature in the basilar membrane (Gummer, Smolders, & Klinke, 1987), as opposed to the negative phase curvature of humans, guinea pigs, and chinchillas (Robles & Ruggero, 2001). Leek et al. (2000) showed reduced masking and reduced phase effects in masking for budgerigars compared to humans using Schroeder-phase maskers. Dooling et al. (2001) expanded Leek’s results to two other bird species and also compared two electrophysiological responses, the cochlear microphonic (CM) and compound action potential (CAP), between these species.
and the Mongolian gerbil. The behavioral results for the birds supported those of Leek et al. (2000), and the CM and CAP responses qualitatively agreed with the behavioral responses; every bird species showed less masking difference to the Schroeder maskers than the gerbils, and, when present, the effect was in the opposite direction. In other words, negative Schroeder maskers were more effective than positive Schroeder maskers for birds, the opposite of human behavioral responses and gerbil electrophysiological responses. The lack of a large phase effect in birds was attributed by the authors to the critical ratios of the different species, with less interaction between masker components in a single basilar membrane filter being the norm for birds. Lauer et al. (2006) similarly found that the largest amount of masking was with a negative sign, but also used a more direct approach to estimating the phase curvature at the signal frequency by way of the Lentz and Leek (2001) modification to the Schroeder formula (discussed below). Using three different bird species, Lauer again found smaller phase differences than in humans, but unlike the previous studies found significant changes in the amount of masking due to phase differences in the masker. Like Leek, they argued that the different amount of phase masking effect between birds and humans can be explained by differences in the critical ratios, but that the same basic processes that govern phase masking differences in mammals must also be present in birds. These comparative behavioral data are therefore consistent with the role of dispersion on the basilar membrane in the processing of Schroeder maskers.

Another line of investigation has relied on recording far and near-field electrical potentials or magnetic fields in response to Schroeder signals. Unfortunately, there have
only been a few studies that have pursued this line of experimentation, and the majority have used animal models rather than human subjects. Dooling et al. (2001), mentioned above, is one of the first studies to use electrophysiological recordings with pure tone signals in Schroeder maskers, and recorded the CM and CAP from the round window of Mongolian gerbils and the scala tympani of budgerigars, canaries, and finches. The gerbils had a larger amplitude CAP in response to the schr-, whereas the birds had larger responses to the schr+ maskers. The CM responses to the maskers with a pure tone stimulus were consistent with the CAP response – gerbils showed a larger response to the stimulus in the presence of the schr+ masker than the schr-, and the birds generally showed the opposite trend but to a lesser degree.

That same year, Recio (2001) published recordings from the ventral cochlear nucleus (VCN) and auditory nerve (AN) of the chinchilla in response to Schroeder stimuli. The results from the VCN and AN corroborate and expand upon those found in the mechanical measurements taken on the basilar membrane; schr- maskers produced significantly higher firing rates and larger rate suppression than the schr+ maskers, and furthermore produced larger synchronization indices to the fundamental frequency of the masker in schr-. This synchronization to the fundamental of the masker was maintained to higher presentation levels for the schr- masker than the schr+ masker. As the level of the probe tone at CF is increased, the neuron shows less synchrony to the fundamental of the masker, indicating the fiber is instead responding to the signal frequency. The level at which this occurs is different from the two maskers, indicated by a horizontal shift of the function. The differences in rate suppression and threshold for synchrony indicate that the
negative Schroeder complex would be expected to be the more effective behavioral masker, consistent with the psychoacoustic data (e.g., Carlyon & Datta, 1997; Kohlrausch & Sanders, 1995; Summers & Leek, 1998; Smith et al., 1986). Recio also noted that the differences between the maskers depend on the CF of the neuron, with higher frequencies showing larger differences in the masker phase effect. There are two non-exclusive interpretations for this; first, Recio suggested that this difference may be due to a $0^\circ$ phase characteristic in the apex of the cochlea compared the base, and cited a simulation (not presented in the paper) using Cooper and Rhode’s (1996) data indicating the apex of the cochlea responds to positive and negative Schroeder complexes similarly. The second related to the rates of the instantaneous frequency sweep used in the study, as only two $f_0$’s (100 and 200 Hz) were employed, which resulted in a sweep rate of approximately 3.85 MHz/s. If the phase curvature of the basilar membrane is not constant over its length (Recio & Rhode, 2000; Robles & Ruggero, 2001), then the difference in the response to the maskers may have been due to a better replication of the basilar membrane phase characteristic of the upper frequencies. Changing the $f_0$ and thus the sweep rate of the frequency modulation might have resulted in a different pattern of results. A phase curvature of $0^\circ$ in the apical end of the cochlea is not the only possible explanation of the results, various nonlinear changes in curvature over the length of the basilar membrane may cause similar patterns.

1.5.2 Human behavioral responses

As the basilar membrane is physically inaccessible without debilitating surgery, human data on the MPE has been primarily behavioral. An indirect approach to testing
the hypothesis that the schr+ masker results in a peakier basilar membrane response than the schr- may be found by further consideration of the interaction between that response and the signal. If the response to the schr+ contains a peak of substantially greater amplitude than the valley, then a short-duration tone presented at different temporal locations in the masker should show a change in detectability as it moves from the peak to the valley of the basilar membrane response to the masker, also known as a masking period pattern (MPP) (Zwicker, 1976). The expectation would therefore be that the MPP for schr+ would contain greater fluctuations than that of schr-. This idea has received support in the data from multiple studies (Carlyon & Datta, 1997a, 1997b; Kohlrausch & Sander, 1995; Oh, 2013; Summers, 2000). Further consideration of the characteristics the basilar membrane allows for another prediction that may be tested behaviorally. The phase curvature of the basilar membrane is not constant over its length, but rather changes depending on the frequency (Dau, Wegner, Mellert, & Kollmeier, 2000; Shera, 2001a, 2001b). A Schroeder complex’s masking potential will therefore vary depending on its relation to the phase curvature of the basilar membrane around the signal, and varying the sweep rate or changing the signal frequency should have an effect on the masking. Lentz and Leek (2001) proposed a minor modification of Smith and colleague’s (1986) formula that allowed researchers to vary the sweep rate of a masker without altering the fundamental frequency or the number of components in the masker:

$$\theta_n = C\pi n(n - 1)/N$$

C is a scalar, which at -1 produces a schr- complex, at 1 produces at schr+ complex, and set to 0 corresponds to a sine or cosine-phase complex. Consistent with the expectation
laid out above, researchers have used this formula to tentatively map out the phase curvature over more of the length of the basilar membrane (Lentz & Leek, 2001; Oxenham & Dau, 2001b, 2004; Shen & Lentz, 2009).

To date there has been only one human study that has recorded changes in neural activity in response to Schroeder maskers, that of Rupp, Sieroka, Gutschalk, & Dau (2008). Using tones embedded in Schroeder-phase maskers they investigated auditory evoked field changes (MEG) to these stimuli and correlated these responses with psychoacoustic measures of partially masked loudness and the output of the auditory image model (AIM) by Patterson, Allerhand, and Giguere (1995). The positive deflection due to the signal onset (P1m) was larger in the schr+ masker than the schr- masker. The P1m is an early cortical component with a generator that has been localized to a position close to Heschyl’s gyrus in the auditory cortex. In a second experiment they compared the partially masked loudness period patterns of the signals to the MEG recordings and simulated auditory nerve findings from the AIM. The result from each experiment was significantly correlated with the others, which support the explanations offered for the behavioral data above.

1.5.3 Schroeder masking with speech signals

In Summers and Leek (1998), the authors examined the differential effect of the schr+ and schr- maskers with an f0 of 100 Hz and equal amplitude components between 200-5000 Hz on pure tone detection and speech recognition for small groups of NH and HI listeners. The speech recognition experiment used IEEE sentences spoken by a female talker and presented at average levels of 60, 70, and 80 dB SPL, with the harmonic
masker initially presented 20 dB higher than each sentence. When the sentence was reported by the subject with fewer than three of five keywords correct, the masker level was dropped 3 dB and the sentence was presented again, which continued until subject reported at least 60% of the keywords correctly. The final level was recorded as the threshold value for the sentence, and the average of 20 threshold estimations were used as the overall threshold. This was an unusual method for threshold estimation, and its possible errors or biases on the threshold estimates are unclear (Green & Rosen, 2013). It seems unlikely, however, that this method would differentially affect the maskers or the subject groups, so while the absolute thresholds may vary from those found by more traditional methods it would be reasonable to assume that any differences between maskers would not be attributable to the adaptive method. For the NH group in Summers and Leek’s experiment thresholds in schr+ maskers were 8-10 dB higher on average than in schr- maskers. For the HI listeners the average difference was approximately 0 dB, though one listener did exhibit a difference at the highest presentation level.

More recently, Green and Rosen (2013) expanded upon this research by including dynamically varying f₀ maskers, different spectral shapes, and three other phase relationships (random, cosine, and glottal source) in a speech recognition experiment with NH listeners. The authors used the more traditional SRT method of estimating threshold, and limited their maskers to 60, 70, or 80 dB SPL while varying their speech presentation level. As in Summers and Leek, they found a difference in the threshold depending on the phase of the masker, though the observed difference in the static f₀ case (the most comparable to the Summers and Leek paper) was only 1.5-5 dB. It is likely that
much of this difference was attributable to the lower presentation levels used in Green and Rosen, as Summers and Leek (1998) did show lower masker phase effect at lower presentation levels. Another possible cause of the difference was that Green and Rosen used a male talker for the IEEE sentences, whose $f_0$ averaged 115 Hz – very close to the $f_0$ of the Schroeder complexes at 100 Hz. As their first experiment showed that there were effects due to $f_0$, it is plausible that the similarity in $f_0$ between the male target speech and the masker somehow reduced the observable phase effect relative to that seen with a female talker used by Summers and Leek.

Regardless of the differences, both studies that tested speech recognition with Schroeder maskers produced smaller masker phase effects than has been demonstrated with the detection of pure tone signals. Summers and Leek (1998) proposed that this may be in part due to the smaller phase effect observed for lower frequencies (i.e., 1 kHz) in their experiment. In this hypothesis, the maskers produced more similarly peaky waveforms on the basilar membrane at frequencies relevant for speech perception. An alternative hypothesis also seems plausible if we consider the MPE simply as an alternative method to test masking release (e.g. Gregan et al., 2013). Modulating a broadband noise provides a release from masking (see section 1.4.2), and the physiological response to harmonic complexes is believed to result in different amounts of modulation at the level of an auditory filter. The MPE observed by Summers and Leek (1998) and Green and Rosen (2013) were within ranges reported for masking release from noise (e.g. Bacon et al., 1998; Festen & Plomp, 1990; Miller & Licklider, 1950). This picture is complicated somewhat by the difference in spectra between noise maskers
and Schroeder maskers, presumably allowing glimpses in spectral regions where the harmonics of the Schroeder maskers are resolved (Deroche, Culling, & Chatterjee, 2014). Furthermore, the modulation rate at the auditory filter due to the Schroeder maskers is believed to be the same as the $f_0$ of the complex, and the fundamentals used by Summers and Leek (1998) and Green and Rosen (2013) were higher than has been shown to be the most effective at producing masking release from noise (as discussed in section 1.4.2).

1.6 The auditory system's response to phase shifted harmonic maskers

According the explanations offered by researchers for the masker phase effect, the difference in the amount of masking observed with Schroeder maskers is caused by a similar mechanism as masking release from an amplitude modulated noise (e.g., Gregan et al., 2013). Unlike in modulated noise, however, the modulations with the Schroeder masker are not apparent in the acoustic waveform but are believed to manifest at the level of the cochlea. These modulations are the product of the interaction of the masker's phase curvature (i.e., the rate of change of the instantaneous frequency) and the phase curvature of the basilar membrane (represented physiologically in the frequency glide). The evidence for this hypothesis has come from the sources described above, but other experiments could shed further light on the human physiological response to phase shifted harmonic complexes - and thus more generally for complex sounds. Specifically, in Chapter 2, we test the underlying assumption of the interaction of the masker/basilar membrane phase curvatures through the use of maskers that share the relevant superficial features of Schroeder maskers. We use different frequency sweeps, or "chirps," to test the role of the changing instantaneous frequency on the ability of a complex sound to mask
signals. If the instantaneous frequency sweep is the relevant feature of the masker phase effect mechanism, we should be able to achieve similar results as long as that feature is maintained. Four of these novel maskers are derived from electrophysiological data and physiological models of the frequency dispersion of the cochlea, and, as such, may provide closer matches to the frequency glides thought to underlie the masker phase effect. In addition, two potential mechanisms for the smaller masker phase effect in speech were reviewed in the previous section; the first related to the spectrotemporal complexity of the signal and the second related to modulation masking. By testing detection of complex signals with these maskers, we can test the relevance of spectrotemporal complexity on the MPE.

As the primary physiological mechanism theorized for the MPE has been the basilar membrane's response to the masker and the effect has been observed at the cortical level electrophysiologically (Rupp et al., 2008) and behaviorally (e.g., Smith et al., 1986), a similar effect should be observable at lower levels of the auditory system. In Chapter 3 we test whether the masker phase effect can be observed in the "objective" threshold estimates provided by the auditory brainstem response (ABR). Given the similarity of the Schroeder complex to the electrophysiologically-derived chirps, we can make several predictions about the relative effectiveness of the Schroeder signals at evoking a response.

The chirps and Schroeder masking techniques arrive at different estimates of the frequency dispersion along the basilar membrane, and it is a long-term goal of this line of research to reconcile these estimates. As with many psychoacoustical experiments,
however, the dominant method of data collection for phase curvature estimates is quite time consuming (Oxenham & Dau, 2001b). Chapter 4 therefore presents a new adaptive method for psychophysically estimating the phase curvature of the auditory periphery. This method should substantially reduce the amount of subject time needed to provide an estimate without a significant increase in the number of assumptions made about the underlying processes.

Each of these chapters rests on the comprehensive background material presented above, but each is written as a separate manuscript. As such, each contains its own brief summary of the relevant research and questions, and provides results and discussions based on the experiments presented therein. Chapter 5 returns to the broader topic of the auditory system's response to phase-shifted maskers as revealed by the three manuscripts, and suggests lines of further productive research in relation to these findings.
1.7 References


2.1 Abstract

Schroeder-phase maskers are believed to interact with the phase curvature of the basilar membrane to produce significantly different amounts of masking depending on the direction and rate of the instantaneous frequency sweep. The present study tests several aspects of the masker phase effect (MPE); first, do maskers with similar instantaneous frequency sweep rates as Schroeder maskers produce similar amounts of MPE? Second, is there an increased MPE for maskers that exhibit non-constant phase curvatures hypothetically more similar to those inferred for the basilar membrane? Third, will spectrotemporally varying signals produce MPEs similar to those observed for pure tone signals, or smaller MPEs as have been observed with speech? Two experiments compared the masking produced by Schroeder maskers and chirps constructed in the temporal and spectral domains for pure tones and spectrotemporally varying signals. The results support the role of the sweep rate in reducing masking but question the larger masking produced by the upward sweeping maskers. Maskers with non-constant phase curvature were found to produce larger MPEs than Schroeder maskers only at 1 kHz. Spectrotemporally varying signals elicited MPEs similar in size to those observed for...
pure tones, and much larger than those reported for speech. These data further our understanding of the phase response of the cochlea.

2.2 Introduction

The amount of masking produced by harmonic complexes has been shown to depend, in part, upon the phase relationship of the components (Carlyon & Datta, 1997a, 1997b; Dau & Oxenham, 2001; Green & Rosen, 2013; Kohlrausch & Sander, 1995; Lentz & Leek, 2001; Mehrgardt & Schroeder, 1983; Smith, Sieben, Kohlrausch, & Schroeder, 1986; Summers & Leek, 1998). By systematically (e.g., Smith et al., 1986) or randomly (e.g., Mehrgardt & Schroeder, 1983) varying the phases of the components of the masker, the amount of masking can change by more than 20 dB. This phenomenon has been robustly demonstrated for detection of pure tone signals with normal-hearing listeners (Dau & Oxenham, 2001; Kohlrausch & Sander, 1995; Oxenham & Dau, 2001; Smith et al., 1986; Summers, 2001; Summers, de Boer, & Nuttall, 2003; Summers & Leek, 1998; Wojtczak & Oxenham, 2009), and a smaller effect has been found in speech recognition (Green & Rosen, 2013; Summers & Leek, 1998).

Smith et al. (1986) first provided evidence of this masker-phase-effect (MPE) using what have come to be called Schroeder maskers. These maskers are constructed based on a formula from Schroeder (1970) that systematically varies the phase of harmonic complexes to produce a signal with a very low peak-factor:

$$\theta_n = \pm \pi n (n - 1) / N$$

where $n$ is the harmonic number and $N$ is the number of harmonics in the complex. Switching the sign in the equation produces another waveform also with a flat acoustic
temporal envelope, but which is a time-reversed version of the first. The two maskers thus have identical power spectra but produce significantly different amounts of masking. Since Smith et al.'s first demonstration of MPE, it has been understood that the difference in masking is, in part, caused by the interaction between the masker and the basilar membrane. Specifically, it has been argued that the phase curvature of the masker interacts with the phase curvature of the basilar membrane, resulting in differentially "peaky" waveforms after auditory filtering. By "peaky" we mean the relationship of the peak amplitude to the average or rms amplitude of the waveform. The smaller amount of masking produced by the positive Schroeder masker is interpreted to mean it produces a peakier output waveform than the negative Schroeder masker (Kohlrausch & Sander, 1995; Oxenham & Dau, 2001b). This interpretation is bolstered qualitatively through auditory filter modeling (Kohlrausch & Sander, 1995; Oxenham & Dau, 2001; Smith et al., 1986) and is also supported by results from animal models (Dooling, Dent, Leek, & Gleich, 2001; Gleich, Leek, & Dooling, 2007; Lauer, Dooling, Leek, & Lentz, 2006; Recio, 2001; Summers et al., 2003). Following Lentz and Leek (2001), Oxenham and Dau (2001) varied the frequency sweep rate of the Schroeder masker in order to estimate the phase response of different auditory filters. They assumed that the masker that produced the smallest amount of masking for a particular frequency had a phase curvature that was equal in magnitude but opposite in sign to that frequency's place on the basilar membrane. The output waveform from this interaction was assumed to be maximally peaky, resulting in improved detection of the signal. Their results supported
the hypothesis that a human's basilar membrane's phase curvature is not constant over the range of frequencies they tested (250-8000 Hz).

There are other factors to be considered with the MPE in relation to speech recognition. While the MPE may be considered a form of masking release (Gregan et al., 2013), there are a number of differences between MPE and masking release from broadband noise. MPE using Schroeder maskers has been observed for speech recognition (Green & Rosen, 2013; Summers & Leek, 1998) but the amount of masking reported by the two studies was quite different, ranging from small (~1.5 dB for the lowest presentation level in Green and Rosen) to substantial (~10 dB in Summers and Leek). The differences between these estimates were attributed by Green and Rosen to the different threshold estimation procedures used in the two studies. In addition, Green and Rosen used a male talker, whereas Summers and Leek used a female talker. The difference in f_0 between the talker and the masker in the Summers and Leek study may have inflated the observed MPE relative to the Green and Rosen study. Green and Rosen (2013) furthermore tested Schroeder maskers with dynamic f_0's, and maskers with spectra shaped to more closely match that of speech. They found that using a dynamically changing f_0 slightly increased the observed MPE with speech, and that reshaping the spectra of the Schroeder maskers did not have significant effect on the MPE. Deroche et al. (2013) used harmonic maskers with either random or sine-phase relationships. They found no difference in the amount of masking produced by the two phase relationships when the masker had an f_0 of 100 Hz, but a significant release from masking in the sine-phase complex when the f_0 was 50 Hz. They also observed the opposite effect at 200 Hz,
where the sine-phase complex was more effective than the random-phase complex at masking speech. They argued in a follow-up paper that there is a trade-off in the effects of $f_0$ on speech recognition and simple noise detection (Deroche et al., 2014). As the $f_0$ is increased, the spectral distance between the harmonics is also increased, allowing listeners to use "glimpses" of the target in spectral regions where the harmonics of the masker are resolved. The change in $f_0$ also changes the spectral regions in which the phase relationships of the maskers are relevant for producing modulations. In other words, increasing the $f_0$ also raises the frequency range in which the masker's partials are believed to be interacting on the basilar membrane (Glasberg & Moore, 1990).

Furthermore, the differences in detection of noise bands in random and sine-phase complexes with an $f_0$ of 100 Hz were smaller than has been reported for pure tone detection, only 3.6 and 7.3 dB depending on the spectral profile (Deroche et al., 2014). The same masker parameters yielded no effect of phase on speech recognition (Deroche et al., 2013).

Even in regions where the masker's components are unresolved, then, the amount of modulation produced by phase-shifted complexes like Schroeder maskers will vary according to the phase curvature of the masker and the phase curvature of different characteristic frequency places on the basilar membrane. As the Schroeder maskers have a constant phase curvature and it is likely the human cochlea does not (Oxenham & Dau, 2001b), the size of the MPE may be dependent on the variation in the output waveforms over the range of the relevant auditory filters. An implication of Oxenham and Dau's results is that a positive Schroeder masker will more successfully match the magnitude of
the basilar membrane's phase curvature at some frequencies than at others. If a spectrally complex signal such as speech or broadband noise is used, then it is possible that the peakiness of the outputs of auditory filters will be more similar because the positive Schroeder masker will not optimally match the phase curvature of all of the relevant auditory filters. If there is less of a difference in the peakiness, this could potentially reduce the overall MPE for a spectrally complex signal.

There may be another factor that affects the MPE observed in speech recognition, which relates to the modulation rates produced by the Schroeder maskers. Recent publications by Stone and colleagues have argued that masking release for speech is due to a release from modulation masking rather than a release from energetic masking (Stone et al., 2011, 2012; Stone & Moore, 2014b). Schroeder maskers would be expected to produce the same modulation rate at the level of an auditory filter, so any difference in speech recognition would not be due to differences in the rate of modulation produced by the two maskers. This may depress the difference in masking observed by Schroeder maskers (Green & Rosen, 2013; Summers & Leek, 1998) or sine-phase and random-phase maskers (Deroche et al., 2013).

The present study has two primary goals: 1) replicate the MPE using maskers that share characteristics with Schroeder maskers, and 2) determine if an MPE is observed for the detection of non-speech, spectrotemporally complex signals. For the first goal, if the frequency sweep rate (phase curvature) of the masker is a strong determinant in the amount of masking produced by two maskers with identical spectra, then the MPE should be replicable using maskers that are not constructed from the Schroeder formula but have
similar instantaneous frequency sweep rates. For the second goal, if the maskers poorly match the phase curvature of the basilar membrane at several points, then the MPE for the detection of spectrotemporally complex signals should be smaller than that of pure tones. Furthermore, if the reduced MPE observed by Green and Rosen (2013) in speech recognition was caused by the similar modulation spectra of the masker, the MPE for the detection of complex signals should be larger than was observed for speech signals by those authors. In experiment 1, repeated upward and downward frequency sweeps are used as maskers. Maskers with a constant sweep rate are compared to maskers with a non-constant sweep rate, as well as to Schroeder maskers. In experiment 2, four harmonic complexes designed for electrophysiological use ('chirps') are used as maskers and compared to Schroeder maskers. Chirps were chosen for their non-constant phase curvature that is based on electrophysiological data and models (Elberling, Don, Cebulla, & Sturzebecher, 2007). As the frequency dispersion in the cochlea is thought to result in the frequency glides that characterize the phase curvature of the basilar membrane (Shera, 2001a), we thought chirps might prove to be effective producers of the MPE. Both experiments used spectrotemporally varying signals, and experiment 2 also included pure tones as signals. As the human basilar membrane is thought to have a non-constant phase curvature over the range of frequencies used here (Oxenham & Dau, 2001b), both experiments include maskers that also exhibit this trait.
2.3 Experiment 1: Masking with frequency modulated tones

2.3.1 Methods

2.3.1.1 Listeners

Ten normal-hearing listeners (6 female) completed this experiment. Age range was 19-25, average 21 years. Normal hearing was determined by pure tone audiometry with thresholds ≤20 dB HL at octave frequencies from 250-8000 Hz (ANSI, 2004). The experiments were approved by the Ohio State University Institutional Review Board.

2.3.1.2 Stimuli

Three pairs of maskers (Figure 1) were tested in this experiment; Schroeder (schr+ and schr-) complexes, linear frequency sweeps (lin+ and lin-), and logarithmic frequency sweeps (log+ and log-). Maskers followed by a "+" had a downward sweeping instantaneous frequency, whereas maskers followed by a "-" had an upward sweeping instantaneous frequency. For schr+/-, an $f_0$ of 100 Hz was used, all harmonics from 200-5000 Hz were included, and the phase of each component was set according to Eq. 1. For lin+/-, and log+/-, the MATLAB chirp function was used to generate frequency sweeps from 200-5000 Hz over a period of 10 ms, which were then concatenated to create maskers 460 ms long. Lin+- was therefore set to have the same sweep rate as the schr+/-. Logarithmic maskers (log+/-) were used as they would place roughly equivalent amounts of energy in each ERB. Maskers were 460 ms long including 30 ms cos² ramps. Two signal types were used in this experiment: 1) spectrotemporal bursts up (SBU) consisting of a sequence of 10 ms tonebursts at 500, 700, 900, 1100, 1400, 1800, 2300, 3000, 3700, and 4600 Hz with 10 ms of silence between each tone, and 2) spectrotemporal bursts
down (SBD) consisting of a sequence of 10 ms tonebursts at 4600, 3700, 3000, 2300, 1800, 1400, 1100, 900, 700 and 500 Hz with 10 ms of silence between each tone. The SBU and SBD signals are therefore similar to those used in Hoglund and Feth (2009), but unlike their study the current experiments presented each 10 ms toneburst at the same dB SPL. All signals were positioned in the temporal center of the masker, and were tested at 60, 70, and 80 dB SPL. All stimuli were generated digitally in MATLAB at a sampling rate of 44.1 kHz and passed through RME Hammerfall DSP Multiface II audio interfaces before being presented monaurally over Sennheiser HD280 headphones. The frequency response of each headphone was measured using a KEMAR with G.R.A.S. ear simulators and a Larson-Davis model 824 sound level meter, and then appropriate correction factors were applied to create a flat frequency response (Russotti, Santoro, & Haskell, 1988). Listeners were seated in a single-walled sound attenuating booth.
Figure 1. Twenty millisecond sections of maskers used in experiment 1. Arrows mark a discontinuity caused by repeating signals generated in the temporal domain, which is referenced in the general discussion.
2.3.1.3 Procedure

Each subject was first briefed on the experimental procedures and asked to sign an informed consent form. After consent was given, the subject was tested for normal hearing as described above. No prior experience with psychoacoustic research was required for participation; however, introductory training (<two hours) was provided to assure familiarity with the procedure and signals and to assure stable thresholds. This training took the form of threshold estimates using the same procedures described below, and used stimuli that were not part of the experimental set. After the training, the subject provided data during sessions that did not exceed two hours, and included breaks at a minimum of every half hour. Subjects completed no more than one session per day. Most subjects completed all conditions within a two-week span. Subjects received a monetary incentive for their participation.

There were 36 conditions total (two signals at three presentation levels in six maskers). For each subject a random signal order was chosen, then a random masker order for each signal. All three signal presentation levels were completed for a combination of signal and masker before moving to the next combination. Each condition was completed within a single session, most sessions included multiple conditions. Each trial consisted of a single interval of the masker alone, or the masker and signal. The subject's task was to respond if the signal was present or not using a mouse to click the appropriate button in custom-scripted gui in MATLAB (Mathworks, 2012). The masker's level was initially set to the same level as the signal. A single-interval adaptive procedure (SIAM, Kaernbach, 1990) was used to vary the masker level in order to estimate the 75%
point on the psychometric function relating signal detection in percent correct to masker level. The initial step size was set to 8 dB, reduced to 4 dB after the first reversal, 2 dB after the third reversal, and 1 dB after the fifth reversal. Each run continued for seven reversals, the final four of which were averaged to estimate the threshold. The final reported threshold was determined by averaging threshold estimates. If three estimates from a single subject in a single condition spanned more than 5 dB, additional estimates were collected until three estimates were within a 5 dB span. The frequency with which this occurred varied across subjects. For one subject a repeated threshold estimate was necessary in almost 30% of the conditions. Across subjects, these repetitions appeared to be evenly distributed across maskers, signals, and signal levels. Results reported here include the mean of every threshold estimate, including those that fell outside the 5 dB span. Feedback was provided after every trial. After all conditions were completed, the MPE was calculated for each subject and each condition by subtracting the signal threshold in the downward sweeping masker (schr+, lin+, log+) from the signal threshold in upward sweeping masking (schr-, lin-, log-). The MPEs were then subjected to a repeated measures ANOVA with $\alpha=.05$ to assess potential statistical significance. Post hoc tests on significant interactions used Benjamini and Hochberg's (1995) approach to control familywise error.

2.3.2 Results

Results plotted in Figure 2 are data averaged across subjects and show the growth-of-masking for each signal, with maskers plotted as separate lines. As expected, rising frequency sweeps (schr-, lin-, log-) produced more masking than falling frequency
sweeps (schr+, lin+, log+). Growth-of-masking slopes are apparent in Figure 2 and have a mean of 1, ranging from 0.7 to 1.3. As the maskers here are designed in the spectral domain (schr+/-) and the time domain (lin+/-, log+/-), the overall masker level is displayed. Thresholds for SBU and SBD signals with the schr+/maskers were approximately 10 dB higher than those found by Summers and Leek (1998) using a pure tone at 4 kHz and presented at the same levels as the current experiment.
Figure 2. Growth of masking for experiment 1. The figure shows the average overall level of a masker at signal threshold for the SBU signal in the top panel and the SBD signal in the bottom panel. Filled symbols are maskers with downward sweeping instantaneous frequencies (schr+, lin+, log+), open symbols are maskers with upward sweeping instantaneous frequencies (schr-, lin-, log-). Lines are simple linear regression to show approximate slope. Error bars represent one standard deviation. Symbols have been shifted along the abscissa to add clarity.
A repeated measures ANOVA was performed on the MPE data, shown in Figure 3, with within-subject factors of masker type (schr, lin, log), signal type (SBU, SBD), and signal level (60, 70, 80 dB SPL). There were significant main effects of masker type ($F(2,18) = 26.50, p < .001$) and signal level ($F(2,18) = 8.69, p = .002$). All other effects failed to reach significance. Post hoc pairwise comparisons using the Benjamini & Hochberg (1995) FDR procedure to control for familywise errors showed that the schr masker produced a larger MPE ($M=12.26, SD=3.52$) than the lin ($M=7.87, SD=2.23$) masker ($t(9) = 5.55, p < .001$). The schr MPE was also larger than the log ($M=5.32, SD=1.63$) masker ($t(9) = 5.87, p < .001$). The lin MPE was larger than the log, ($t(9) = 2.91, p = .017$), though this was a smaller difference than was observed between either two maskers and the schr masker. MPEs generally decreased as the signal level increased, which is consistent with previous findings at these relatively high signal levels (Summers & Leek, 1998), and this effect was not pursued further statistically. The observed MPEs for schr maskers and these spectrotemporally varying signals were similar to those found in previous research using pure tone signals, ranging from ~8-15 dB (e.g., Carlyon & Datta, 1997a; Kohlrausch & Sander, 1995; Summers & Leek, 1998). The MPEs obtained here were substantially larger than those found recently for speech recognition, which averaged between 1.5-7 dB (Green & Rosen, 2013).
Figure 3. Mean MPEs for experiment 1. Error bars show standard deviation.
2.3.3 Discussion

The schr- and lin- showed similarly high levels of masking for both signals, with a difference between the masked thresholds for these two maskers of <1 dB at all three presentation levels. The lin+ maskers, however, were more effective maskers than the schr+ maskers, and this kept the lin MPE smaller. On average, log- produced 4.7 dB less masking than schr-, while log+ produced 1.4 dB more masking than schr+, keeping the log MPE smaller than both schr and lin. Following previous interpretations of the causes of the MPE (e.g., Kohlrausch & Sander, 1995; Lentz & Leek, 2001; Oxenham & Dau, 2001; Smith et al., 1986), it could be said that the schr+ produced the "peakiest" output waveforms for the auditory filters targeted by these signals, followed by log+ and then lin+. The maskers that were most effective at masking - producing what may be the least peaky output waveforms - would be the lin- and schr- maskers, followed by log-. That the masker constructed to have a similar sweep rate to the schr- masker should produce similar amounts of masking is consistent with the current interpretations of the underlying mechanism of the masker phase effect. Similarly, the differences between log and schr maskers may be explained as the result of differences in phase curvature matching by the two maskers. The log+ masker may have a phase curvature that is equal in magnitude but opposite in sign to some of the auditory filters targeted by the signals, whereas schr+ may match different filters more effectively. The ineffectiveness of the log- masker compared to the schr- masker may be explained in the same way. This interpretation does not offer a ready explanation for the dissimilarity between the thresholds for lin+ and schr+ maskers. Hypothetically, the lin+ and schr+ maskers should
produce equivalently peaky output waveforms and therefore equivalent amounts of masking, but they failed to do so in these subjects. Overall, the schr maskers were the most effective at producing an MPE for these signals and subjects.

One difference between the maskers that may influence thresholds stems from their construction; schr maskers were produced in the spectral domain, whereas the lin and log maskers were produced in the temporal domain as a series of repeating frequency sweeps. In short, there is discontinuity between the sweeps produced by this method (see Figure 1). As it is reasonable to assume that any effect produced by the stimulus construction methods would influence results for the + and - maskers equally, it seems unlikely that construction is the primary cause of this difference. Without further testing, however, that possibility cannot be ruled out.

The subsequent experiment constructed all maskers in the spectral domain to remove this possibility. Furthermore, rather than attempting to replicate the MPE found with schr maskers (i.e., lin) or attempting to expand it (i.e., log), physiologically-inspired signals (chirps) were used as maskers and compared to the Schroeder maskers. Shera (2001) has argued that the physiological manifestation of phase curvature - the frequency-glide - is a result of the frequency dispersive nature of the cochlea (see also Ramamoorthy, Zha, & Nuttall, 2010). Chirps have been designed by various researchers to attempt to compensate for the frequency-specific delay of the basilar membrane's traveling wave, so we chose to use established chirp formulae for the second experiment.
2.4 Experiment 2: Masking with harmonic complexes

2.4.1 Methods

2.4.1.1 Listeners

Ten normal-hearing listeners (7 female) completed this experiment. The age range was 19-34, average 23 years. One listener was unable to complete the 60 dB SPL signal levels, and is only included in the 50 dB SPL results and analysis. A separate listener participated in experiment 1, for that subject the threshold estimates were less than 1 dB different for conditions that were the same in each experiment (60 dB SPL presentation of SBU and SBD). Normal-hearing was determined by pure tone audiometry with thresholds ≤20 dB HL at octave frequencies from 250-8000 Hz (ANSI, 2004).

2.4.1.2 Stimuli

Five pairs of maskers were tested in this experiment. Schroeder maskers were included as in Experiment 1. Four of the chirps described by Elberling et al., (2007) were used as the basis of the other four masker pairs. Using Elberling et al.'s equations 5 and 8, their "Don," "Eggermont," "Neely" and "de Boer" chirps were created with an $f_0$ of 100 Hz and including all harmonics from 200-5000 Hz. The complementary downward sweeping chirps were created by reversing the sign of these formulas. Each of these chirps was based on the frequency-specific delay estimated from electrophysiology data or predicted by a model: the Don chirps (don+/-) were based on the observed wave V delay from narrow-band auditory brainstem responses, Eggermont (egg+/-) chirps were based the compound action potential from narrow-band electrocochleography, Neely (neely+/-) were based on frequency-specific auditory brainstem responses, and the de
Boer (deb+/-) chirps were based on de Boer's (1980) cochlear model using Greenwood's frequency-to-place mapping (1990). This chirp was therefore similar to the one proposed by Dau, Wegner, Mellert, and Kollmeier, (2000), but was constructed in the spectral rather than temporal domain. Each masker was 460 ms in duration including 30 ms raised cosine onset and offset ramps.

Five signals were tested in this experiment: pure tones at 1, 2, and 4 kHz, and the SBU and SBD stimuli from experiment 1. As in experiment 1, signals were located in the temporal center of the maskers. Pure tone signals were 260 ms long including 30 ms raised cosine onset and offset ramps, SBU and SBD signals had the same frequency and temporal characteristics as in experiment 1. As the chirps were constructed in part to capture the low-frequency response of the cochlea (Elberling et al., 2007), it was decided to test pure tones below 4 kHz. Different signal levels were piloted, and it was determined to use 50 and 60 dB SPL signals in order to allow some comparison with experiment 1 while keeping maskers at a comfortable listening level.

2.4.1.3 Procedure

The overall procedure was the same as experiment 1. As there were more conditions for this experiment (five signals at two levels in ten maskers), most subjects needed four or five weeks to complete all the conditions. Condition order was again determined by a random selection of a signal, then a random selection of a masker. Each combination was tested at both signal presentation levels before moving to the next combination.
2.4.2 Results

Results plotted in Figure 4 are data averaged across subjects and show the growth-of-masking for each signal. As expected, rising frequency sweeps generally produce more masking than falling frequency sweeps. Growth-of-masking slopes as determined by a simple linear regression are also plotted in Figure 4 and have a mean of 1.1 ranging from .8 to 1.5. In order to offer comparison with experiment 1, the overall masker level is displayed. In contrast to experiment 1, schr+ was not consistently the least effective masker across all signals. Schr- was on average the most effective masker for all signals except the 1 kHz tone. The results are consistent with previous simultaneous masking experiments using the Schroeder formula and similar parameters (Oxenham & Dau, 2001b; Shen & Lentz, 2009; Smith et al., 1986; Summers & Leek, 1998).
Figure 4. Growth of masking for experiment 2. Filled symbols are downward sweeping maskers ("+" maskers), open symbols are upward sweeping maskers ("-" maskers). Error bars are one standard deviation. Symbols have been shifted along the abscissa for clarity.
MPEs for the five masker pairs are plotted in Figure 5. Once again a repeated measures ANOVA was used to assess statistical significance. The Greenhouse-Geisser correction was applied when Mauchly's test indicated the assumption of sphericity had been violated. There were significant main effects of signal type ($F(1.87,14.94) = 11.028, p = .001$) and masker ($F(4,32) = 24.62, p < .001$), and significant interactions between signal type and masker ($F(16,128) = 7.56, p < .001$) and masker and signal level ($F(4,32) = 4.79, p = .004$). These interactions were investigated using the Benjamini and Hochberg (1995) FDR procedure to control the familywise error for the multiple pairwise comparisons, the results of which may be seen in full in the Appendix. Looking across signals, the masker and signal level interaction was driven primarily by the difference in the don MPEs relative to the other chirp maskers at 50 and 60 dB SPL presentation levels. At 50 dB SPL, the don masker did not produce MPEs that were significantly larger than the other chirp maskers, whereas at 60 dB SPL it did. The MPE from the egg masker at 50 dB SPL was also significantly larger than the neely masker, which was not true at 60 dB SPL. Averaged across signals, schr produced larger MPEs than each of the chirp maskers at both signal presentation levels.

The interaction between signal type and masker may be clearly observed in Figure 5. Schr maskers produce significantly larger MPEs than all of the chirp maskers for every signal except the 1 kHz signal, when the results were averaged over signal level. Smaller differences were also observed between the chirps for the SBU and SBD signals. At 1 kHz, the don masker produced a significantly larger MPE than the schr and egg maskers. The egg masker was particularly ineffective at producing an MPE, as both the deb and neely maskers produced MPEs that were significantly larger.
Figure 5. Mean MPEs for the signals in experiment 2. Error bars show standard deviation.
In light of the growth-of-masking data, the generally smaller MPEs for chirp maskers must be attributed to the upwards and downwards chirps being less effective at most of the signals tested. This was in contrast to experiment 1, where schr+-/- maskers bracketed the other maskers in effectiveness, producing a larger MPE. In experiment 2, the larger MPE is generated primarily by the effectiveness of schr- at masking all but the 1 kHz signal.

2.4.3 Discussion

For all signals except the 1 kHz tone, the observed MPEs were dominated by the relatively high effectiveness of the schr- masker, rather than by the lack of masking produced by the schr+ masker. An illustrative example may be seen for the 2 kHz signal; at both signal levels, downward sweeping chirp maskers produced thresholds +/-3dB of the downward sweeping schr+ masker. The upward sweeping chirp maskers, however, produced thresholds 6-9 dB higher than the schr- masker. This same pattern was observed for the 4 kHz, SBU, and SBD signals. Drawing inference about the phase response of the cochlea from these results, it could be said that the downward sweeping chirps produced auditory filter waveforms that were similarly peaky to those produced by schr+ masker. By contrast, the upward sweeping chirps produced auditory filter waveforms that were much peakier than those produced by the schr- masker for all but the lower-frequency filters. The detection of the SBU and SBD signals may have been dominated by the better "match" of the schr maskers to the higher frequency components, as in the first experiment. In this instance, however, that match is in the ability of the schr- masker to
produce more masking (a less peaky output from the relevant auditory filters) rather than by the ability of the schr+ masker to produce less masking. This has not been a common focus in studies researching the phase response of the cochlea.

2.5 General discussion

2.5.1 Frequency sweep rate

The two experiments above provided evidence that was generally consistent with previous interpretations of the MPE. Maskers with a downward sweeping instantaneous frequency were less effective as maskers than those with an upward sweeping instantaneous frequency, and it is reasonable to interpret those differences as being caused by variations of the peakiness of the output waveform from the relevant auditory filter(s). However, this explanation does not intuitively explain the greater masking produced by the schr- masker relative to most of the other upward sweeping maskers used. In order to explore this result, the log, lin, don, and schr maskers were input into a gammachirp filterbank (Irino & Patterson, 2001) along with a 10 ms, 4 kHz tone. The filterbank was set with the same parameters as those found in Irino and Patterson (2006). As shown in Figure 6, this offers qualitative support for the experimental results reported above. The maskers with downward sweeping instantaneous frequencies (schr+, lin+, log+, don+) generally reveal more of the signal than the maskers with upward sweeping instantaneous frequencies (schr-, lin-, log-, don-).

Examining the schr+ and lin+ signals, the difference in construction between the two (spectral and temporal, respectively) may have played a role in their relative effectiveness as maskers. It would appear from the gammachirp filterbank output that, if
anything, the lin+ masker should be less effective as a masker than the schr+, but that is not what was observed in experiment 1. It does appear that our choice of masker construction - concatenating repeated frequency sweeps - did result in a discontinuity that may have affected the results. The small peaks at 10 and 20 ms in Figure 6 (marked by arrows at the 10 ms point in Figure 1) are likely in response to that sharp discontinuity in the masker, as no such peaks are visible in the maskers constructed in the spectral domain. It is possible that these discontinuities affected the masking effectiveness in unpredicted ways.

The differences in the gammachirp filterbank responses to the maskers in the lower frequencies (not shown) did seem to reflect the differences observed in experiment 2; the don maskers had noticeably larger differences in energy distribution at lower frequencies than the schr maskers. Given that the chirps in experiment 2 were originally designed to synchronize basilar membrane activity across frequencies and in particular below 4 kHz, it is perhaps unsurprising that they would more closely match the phase curvature at these lower frequencies. Furthermore, as may be seen in Figure 6 the upward and downward don chirps produce substantially more peaky outputs than the Schroeder maskers at 4 kHz. This may explain the smaller MPEs observed for these maskers and also the effect observed above, that the schr- masker was more effective than the upward sweeping chirps for most of the signals and levels tested.

In summary, these experiments provide evidence that the instantaneous frequency sweep of the so-called Schroeder masker does play a substantial role in creating the MPE. When maskers were constructed that had similarly changing instantaneous frequencies
but without using the Schroeder formula (experiment 1), these maskers produced similar amounts of masking for the signals used here. Other maskers with non-constant phase curvatures also produced MPE for signal detection (experiments 1 and 2), though these MPE were often smaller than those produced by the Schroeder maskers.
Figure 6. Results from dynamic gammachirp filterbank for filter tuned to 4 kHz with an input of a 10 ms 4 kHz tone from 10-20 ms of each masker. The left column shows signals in the presence of maskers with downward sweeping instantaneous frequency (schr+, lin+, log+, don+). The right column shows signals in the presence of maskers with upward sweeping instantaneous frequency (schr-, lin-, log-, don-). Ellipses have been centered on the response to the 4 kHz pure tone.
2.5.2 Detection of spectrotemporally varying signals

The MPEs observed for the SBU and SBD signals in these experiments were closer to those found for pure tone detection than those found for speech recognition. Consistent with pure tone detection MPEs, detection of SBU and SBD signals decreased with increasing signal levels (Shen & Lentz, 2009; Summers & Leek, 1998). The size of the MPEs (8-18 dB for the schr maskers) was comparable to those found for pure tones at similar levels. A direct comparison was offered by experiment 2: for schr maskers the MPEs for the SBU and SBD signals ranged from 16.9-17.9 dB when presented at 50 or 60 dB SPL. The pure tone MPEs generated by the schr maskers in that experiment ranged from 9.0-16.9 dB. The larger spread of MPEs for the pure tones likely reflects the closer match of the schr+ masker used to the 2 kHz place on the basilar membrane; above and below that point the masker does not match the phase curvature as closely. Nevertheless, the size of these MPE for the spectrotemporally complex signals was as great or greater than those for pure tones with these subjects, contrary to our expectations. These data are not inconsistent with recent arguments put forward by researchers regarding masking release for speech in fluctuating maskers. The smaller MPE observed with speech by other researchers (Green & Rosen, 2013; Summers & Leek, 1998) likely has less to do with poor matches between the maskers and the relevant auditory filters, and more to do with the low relevance of the modulation rates induced by these maskers and the presence of spectral gaps between the resolved harmonics (Deroche et al., 2014). The 100 Hz modulation rate of these maskers is well above the modulation rates most relevant for
speech perception, believed to peak near 8-20 Hz (Füllgrabe et al., 2006; Gustafsson & Arlinger, 1994). Further experiments are planned to test this directly.

The larger MPEs for the SBU and SBD signals observed for the schr maskers compared to the other maskers used in these experiments was interesting. Were the MPEs for these signals determined by the most detectable portion of the signal, the chirp maskers should have produced MPEs more similar to those seen in the schr case. An average that accounts for the relative detectability of the included frequencies would seem to be a more likely explanation. As the signal was fixed at a particular level for each experiment, the lower frequency bursts of the SBU and SBD signals would have a lower sensation level than the bursts around 2-3 kHz.

In review, two experiments were presented that tested the relevance of the sweeping instantaneous frequency of maskers and the detection of spectrotemporally complex signals in these maskers. The results of these experiments were consistent with previous interpretations of the MPE; that the falling instantaneous frequency maskers were less effective than those with a rising instantaneous frequency. The maskers with a linearly upwards sweeping instantaneous frequency - the schr- and lin- maskers from the first experiment - provided equivalent amounts of masking at each signal level tested. The downward sweeping instantaneous frequency maskers showed similar patterns, but the schr+ masker was less effective as a masker than the lin+ masker, even though the two shared similar features. The experiments also provided evidence that the smaller MPE observed with speech in previous studies was not likely due to the spectrotemporally complex nature of the signals. The SBU and SBD signals provided
MPEs of equivalent size to the 2 kHz pure tone signals with the Schroeder maskers. These same signals provided smaller MPEs to the other maskers used in these experiments, and determining why the MPEs varied so greatly between signal and masker combinations will require further research.

2.6 Appendix

Pairwise comparisons from experiment 2 are shown in Table 1. Significance as shown by a paired-t test using Benjamini and Hochberg's (1995) FDR procedure is indicated by an asterisk.

<table>
<thead>
<tr>
<th>1 kHz signal:</th>
<th>Pair</th>
<th>Mean difference</th>
<th>s.d.</th>
<th>C.I.</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>don - deb</td>
<td>0.57</td>
<td>3.61</td>
<td>-2.01, 3.15</td>
<td>0.50</td>
<td>9</td>
<td>.629</td>
<td></td>
</tr>
<tr>
<td>don - egg</td>
<td>5.87</td>
<td>4.21</td>
<td>2.85, 8.88</td>
<td>4.41</td>
<td>9</td>
<td>.002*</td>
<td></td>
</tr>
<tr>
<td>don - neely</td>
<td>0.18</td>
<td>3.58</td>
<td>-2.39, 2.74</td>
<td>0.15</td>
<td>9</td>
<td>.881</td>
<td></td>
</tr>
<tr>
<td>egg - deb</td>
<td>-5.30</td>
<td>4.59</td>
<td>-5.58, -0.20</td>
<td>-3.65</td>
<td>9</td>
<td>.005*</td>
<td></td>
</tr>
<tr>
<td>egg - neely</td>
<td>-5.69</td>
<td>5.51</td>
<td>-9.63, -1.75</td>
<td>-3.27</td>
<td>9</td>
<td>.010*</td>
<td></td>
</tr>
<tr>
<td>neely - deb</td>
<td>0.40</td>
<td>2.34</td>
<td>-1.28, 2.07</td>
<td>0.53</td>
<td>9</td>
<td>.606</td>
<td></td>
</tr>
<tr>
<td>schr - deb</td>
<td>-2.89</td>
<td>4.23</td>
<td>-5.92, 0.13</td>
<td>-2.16</td>
<td>9</td>
<td>.059</td>
<td></td>
</tr>
<tr>
<td>schr - don</td>
<td>-3.46</td>
<td>2.68</td>
<td>-5.38, -1.55</td>
<td>-4.09</td>
<td>9</td>
<td>.003*</td>
<td></td>
</tr>
<tr>
<td>schr - egg</td>
<td>2.40</td>
<td>4.29</td>
<td>-0.66, 5.47</td>
<td>1.77</td>
<td>9</td>
<td>.110</td>
<td></td>
</tr>
<tr>
<td>schr - neely</td>
<td>-3.29</td>
<td>4.88</td>
<td>-6.78, 0.20</td>
<td>-2.13</td>
<td>9</td>
<td>.062</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2 kHz signal:</th>
<th>Pair</th>
<th>Mean difference</th>
<th>s.d.</th>
<th>C.I.</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>don - deb</td>
<td>-1.17</td>
<td>3.71</td>
<td>-3.83, 1.48</td>
<td>-1.00</td>
<td>9</td>
<td>.343</td>
<td></td>
</tr>
<tr>
<td>don - egg</td>
<td>-1.27</td>
<td>3.06</td>
<td>-3.46, 0.92</td>
<td>-1.31</td>
<td>9</td>
<td>.223</td>
<td></td>
</tr>
<tr>
<td>don - neely</td>
<td>2.23</td>
<td>4.29</td>
<td>-0.84, 5.30</td>
<td>1.64</td>
<td>9</td>
<td>.135</td>
<td></td>
</tr>
<tr>
<td>egg - deb</td>
<td>0.09</td>
<td>2.74</td>
<td>-1.87, 2.05</td>
<td>0.11</td>
<td>9</td>
<td>.916</td>
<td></td>
</tr>
<tr>
<td>egg - neely</td>
<td>3.50</td>
<td>5.47</td>
<td>-0.42, 7.41</td>
<td>2.02</td>
<td>9</td>
<td>.074</td>
<td></td>
</tr>
<tr>
<td>neely - deb</td>
<td>-3.40</td>
<td>4.11</td>
<td>-6.34, -0.46</td>
<td>-2.62</td>
<td>9</td>
<td>.028</td>
<td></td>
</tr>
<tr>
<td>schr - deb</td>
<td>6.89</td>
<td>2.48</td>
<td>5.12, 8.67</td>
<td>8.79</td>
<td>9</td>
<td>&lt;.001*</td>
<td></td>
</tr>
<tr>
<td>schr - don</td>
<td>8.07</td>
<td>3.82</td>
<td>5.34, 10.80</td>
<td>6.68</td>
<td>9</td>
<td>&lt;.001*</td>
<td></td>
</tr>
<tr>
<td>schr - egg</td>
<td>6.80</td>
<td>4.07</td>
<td>3.89, 9.71</td>
<td>5.28</td>
<td>9</td>
<td>.001*</td>
<td></td>
</tr>
<tr>
<td>schr - neely</td>
<td>10.30</td>
<td>4.26</td>
<td>7.25, 13.34</td>
<td>7.65</td>
<td>9</td>
<td>&lt;.001*</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Pairwise comparison results from experiment 2.
Table 1 continued

<table>
<thead>
<tr>
<th>Pair</th>
<th>Mean difference</th>
<th>s.d.</th>
<th>C.I.</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 kHz signal:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>don - deb</td>
<td>0.41</td>
<td>4.84</td>
<td>[-3.05, 3.87]</td>
<td>0.27</td>
<td>9</td>
<td>.796</td>
</tr>
<tr>
<td>don - egg</td>
<td>-2.11</td>
<td>4.87</td>
<td>[-5.60, 1.37]</td>
<td>-1.37</td>
<td>9</td>
<td>.203</td>
</tr>
<tr>
<td>don - neely</td>
<td>0.51</td>
<td>3.18</td>
<td>[-1.77, 2.79]</td>
<td>0.51</td>
<td>9</td>
<td>.625</td>
</tr>
<tr>
<td>egg - deb</td>
<td>2.52</td>
<td>3.10</td>
<td>[0.30, 4.74]</td>
<td>2.57</td>
<td>9</td>
<td>.030</td>
</tr>
<tr>
<td>egg - neely</td>
<td>2.62</td>
<td>3.56</td>
<td>[0.08, 5.17]</td>
<td>2.33</td>
<td>9</td>
<td>.045</td>
</tr>
<tr>
<td>neely - deb</td>
<td>-0.10</td>
<td>3.86</td>
<td>[-2.87, 2.66]</td>
<td>-0.08</td>
<td>9</td>
<td>.936</td>
</tr>
<tr>
<td>schr - deb</td>
<td>6.01</td>
<td>4.11</td>
<td>[3.07, 8.95]</td>
<td>4.63</td>
<td>9</td>
<td>.001*</td>
</tr>
<tr>
<td>schr - don</td>
<td>5.61</td>
<td>4.95</td>
<td>[2.07, 9.15]</td>
<td>3.58</td>
<td>9</td>
<td>.006*</td>
</tr>
<tr>
<td>schr - egg</td>
<td>3.49</td>
<td>3.03</td>
<td>[1.32, 5.66]</td>
<td>3.64</td>
<td>9</td>
<td>.005*</td>
</tr>
<tr>
<td>schr - neely</td>
<td>6.12</td>
<td>4.26</td>
<td>[3.07, 9.16]</td>
<td>4.54</td>
<td>9</td>
<td>.001*</td>
</tr>
</tbody>
</table>

SBD signal:                      |                 |      |            |      |    |      |
| don - deb  | -0.32           | 3.70 | [-2.97, 2.32] | -0.28 | 9  | .787 |
| don - egg  | -1.46           | 2.62 | [-3.33, 0.42] | -1.75 | 9  | .113 |
| don - neely | 1.43            | 4.09 | [-1.49, 4.36] | 1.11 | 9  | .296 |
| egg - deb  | 1.13            | 1.55 | [0.02, 2.24]  | 2.31 | 9  | .046 |
| egg - neely | 2.89            | 2.47 | [1.12, 4.66]  | 3.70 | 9  | .005*|
| neely - deb | -1.76           | 2.78 | [-3.75, 0.23] | -2.00 | 9  | .076 |
| schr - deb | 8.64            | 4.40 | [5.50, 11.79] | 6.22 | 9  | <.001*|
| schr - don | 8.97            | 4.28 | [5.91, 12.03] | 6.63 | 9  | <.001*|
| schr - egg | 7.51            | 4.23 | [4.49, 10.54] | 5.62 | 9  | <.001*|
| schr - neely | 10.40           | 5.76 | [6.28, 14.53] | 5.71 | 9  | <.001*|

SBU signal:                      |                 |      |            |      |    |      |
| don - deb  | 1.71            | 4.29 | [-1.36, 4.78] | 1.26 | 9  | .239 |
| don - egg  | -0.20           | 4.40 | [-3.35, 2.95] | -0.14 | 9  | .890 |
| don - neely | 2.27            | 2.20 | [0.70, 3.85]  | 3.26 | 9  | .010*|
| egg - deb  | 1.91            | 2.01 | [0.47, 3.35]  | 3.00 | 9  | .015*|
| egg - neely | 2.47            | 3.47 | [-0.01, 4.95] | 2.25 | 9  | .051 |
| neely - deb | -0.56           | 3.51 | [-3.08, 1.95] | -0.51 | 9  | .624 |
| schr - deb | 6.77            | 4.50 | [3.55, 9.99]  | 4.76 | 9  | .001*|
| schr - don | 5.06            | 4.16 | [2.09, 8.03]  | 3.85 | 9  | .004*|
| schr - egg | 4.86            | 3.45 | [2.39, 7.33]  | 4.45 | 9  | .002*|
| schr - neely | 7.33            | 3.93 | [4.52, 10.14] | 5.91 | 9  | <.001*|

Continued
Table 1 continued

<table>
<thead>
<tr>
<th>Pair</th>
<th>Mean difference</th>
<th>s.d.</th>
<th>C.I.</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>don - deb</td>
<td>-0.49</td>
<td>2.49</td>
<td>[-2.28, 1.29]</td>
<td>-0.62</td>
<td>9</td>
<td>.548</td>
</tr>
<tr>
<td>don - egg</td>
<td>-1.06</td>
<td>2.21</td>
<td>[-2.64, 0.52]</td>
<td>-1.52</td>
<td>9</td>
<td>.164</td>
</tr>
<tr>
<td>don - neely</td>
<td>1.34</td>
<td>2.16</td>
<td>[-0.20, 2.89]</td>
<td>1.96</td>
<td>9</td>
<td>.081</td>
</tr>
<tr>
<td>egg - deb</td>
<td>0.57</td>
<td>1.67</td>
<td>[-0.63, 1.76]</td>
<td>1.07</td>
<td>9</td>
<td>.313</td>
</tr>
<tr>
<td>egg - neely</td>
<td>2.40</td>
<td>2.08</td>
<td>[0.91, 3.89]</td>
<td>3.65</td>
<td>9</td>
<td>.005</td>
</tr>
<tr>
<td>neely - deb</td>
<td>-1.84</td>
<td>2.98</td>
<td>[-3.97, 0.29]</td>
<td>-1.95</td>
<td>9</td>
<td>.083</td>
</tr>
<tr>
<td>schr - deb</td>
<td>5.20</td>
<td>2.59</td>
<td>[3.35, 7.06]</td>
<td>6.34</td>
<td>9</td>
<td>.000</td>
</tr>
<tr>
<td>schr - don</td>
<td>5.70</td>
<td>2.59</td>
<td>[3.84, 7.55]</td>
<td>6.95</td>
<td>9</td>
<td>.000</td>
</tr>
<tr>
<td>schr - egg</td>
<td>4.64</td>
<td>3.06</td>
<td>[2.45, 6.83]</td>
<td>4.80</td>
<td>9</td>
<td>.001</td>
</tr>
</tbody>
</table>

**Signal level 50 dB SPL**

<table>
<thead>
<tr>
<th>Pair</th>
<th>Mean difference</th>
<th>s.d.</th>
<th>C.I.</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>don - deb</td>
<td>1.65</td>
<td>1.46</td>
<td>[0.53, 2.77]</td>
<td>3.39</td>
<td>8</td>
<td>.010</td>
</tr>
<tr>
<td>don - egg</td>
<td>2.03</td>
<td>1.74</td>
<td>[0.70, 3.37]</td>
<td>3.51</td>
<td>8</td>
<td>.008</td>
</tr>
<tr>
<td>don - neely</td>
<td>1.65</td>
<td>1.55</td>
<td>[0.46, 2.84]</td>
<td>3.19</td>
<td>8</td>
<td>.013</td>
</tr>
<tr>
<td>egg - deb</td>
<td>-0.38</td>
<td>1.38</td>
<td>[-1.45, 0.84]</td>
<td>-0.83</td>
<td>8</td>
<td>.431</td>
</tr>
<tr>
<td>egg - neely</td>
<td>-0.38</td>
<td>2.06</td>
<td>[-1.97, 1.20]</td>
<td>-0.56</td>
<td>8</td>
<td>.591</td>
</tr>
<tr>
<td>neely - deb</td>
<td>0.00</td>
<td>1.74</td>
<td>[-1.34, 1.34]</td>
<td>0.00</td>
<td>8</td>
<td>.999</td>
</tr>
<tr>
<td>schr - don</td>
<td>3.12</td>
<td>2.10</td>
<td>[1.51, 4.74]</td>
<td>4.46</td>
<td>8</td>
<td>.002</td>
</tr>
<tr>
<td>schr - egg</td>
<td>5.15</td>
<td>2.72</td>
<td>[3.06, 7.25]</td>
<td>5.68</td>
<td>8</td>
<td>.000</td>
</tr>
<tr>
<td>schr - neely</td>
<td>4.77</td>
<td>2.39</td>
<td>[2.93, 6.61]</td>
<td>5.99</td>
<td>8</td>
<td>.000</td>
</tr>
</tbody>
</table>
2.7 References


Chapter 3: Manuscript 2 - Masker Phase Effect in the Auditory Brainstem Response

3.1 Abstract

Objectives: The amount of masking produced by a harmonic complex can be substantially altered by changing the phase relationships of the components (Smith et al., 1986). For detection of pure tones the maskers with upward sweeping instantaneous frequencies can produce >20 dB more masking than the downward-sweeping maskers in behavioral tests. The proposed mechanism of this difference is based on animal and computer models, and posits that the interaction of the maskers' instantaneous frequency sweeps and the frequency glide on the basilar membrane result in differentially peaky output waveforms after cochlear transduction. The only extant physiological recording in humans examined cortical responses to the maskers, but if the difference in masking is produced at the cochlea, this difference should be observable at more peripheral levels. The purpose of this study was to determine if this difference in masking can be observed in subcortical physiological responses as measured by the auditory brainstem response (ABR).

Design: Twenty normal-hearing subjects participated in this study. ABRs were recorded using an ipsilateral montage and disposable silver-silver chloride electrodes connected to a Tucker Davis Technologies system 3 evoked potential system. A 4 kHz tone burst with a duration of 1.25 ms (0.5 ms cos² ramps) was presented alone and with
maskers over an ER-2 insert earphone at a rate of 17.1/second. Maskers were 49, equal-amplitude harmonics of 100 Hz from 200-5000 Hz with phases set in the normalization file according to the formula in Smith et al. (1986). The masker was 40 ms long (5 ms \( \cos^2 \) ramps) and was fixed at 60 dB SPL for the entire experiment. The tone bursts began at 80 dB pSPL and were decremented in 10-dB steps. The pure tones had an onset 26 ms after the onset of the masker, and two averages of 2000 sweeps were recorded for each condition. The recording window was 50 ms in duration with filter settings of 30-3000 Hz. Threshold was determined offline by the presence/absence of wave V.

Results: Maskers with an upward sweeping instantaneous frequency produced an average threshold of 57 dB pSPL, whereas maskers with a downward sweeping instantaneous frequency produced an average threshold of 47.5 dB pSPL. This difference was significant at the \( \alpha=0.05 \) level.

Conclusions: This study showed evidence of a difference in masking produced by phase-shifted harmonic complexes which can be observed at the level of wave V of the ABR. This is consistent with the explanation that the difference is produced at a peripheral site(s), most likely at the level of the cochlea. The smaller difference in thresholds observed here relative to other behavioral studies was likely due to the differential effects of overshoot on the two maskers.

3.2 Introduction

Over the past few decades, researchers have investigated the role of the phase relationships of harmonic complexes on cochlear processing. Research focused on human behavioral responses has used so-called Schroeder maskers to investigate masking of
pure tones (Carlyon & Datta, 1997a, 1997b; Hoglund et al., 2013; Kohlrausch & Sander, 1995; Lentz & Leek, 2001; Oxenham & Dau, 2001b, 2004; Smith et al., 1986; Summers, 2000; Summers & Leek, 1998; Wojtczak & Oxenham, 2009a) and speech (Green & Rosen, 2013; Summers & Leek, 1998). Large (>20 dB) differences in masking of pure tones and smaller differences in speech have been observed depending on the direction of the instantaneous frequency sweep of the masker: an upward-sweeping masker produces significantly more masking than a downward-sweeping masker. This has been explained as the result of the interaction between phase curvature of the masker and the phase curvature of the basilar membrane, which may be characterized as the frequency glide. When a masker with a downward-sweeping instantaneous frequency (Schroeder plus or schr+) interacts with a place on the basilar membrane that has an upwards-sweeping frequency glide, the supposed movement of the basilar membrane is peakier - sharper peaks and longer valleys in the envelope - than when an upward-sweeping masker (Schroeder minus or schr-) is used. Many of these behavioral studies have included auditory filter models that qualitatively support this interpretation (e.g., Kohlrausch & Sander, 1995). When the Schroeder maskers are passed through these filters, the upward-sweeping masker produces a noticeably flatter envelope.

As it is generally not possible to physically investigate the living human cochlea, animal models have also been employed to help understand its physiology. The frequency glides of different animal models at different points in the periphery have been described (de Boer & Nuttall, 1997; Carney, McDuffy, & Shekhter, 1999; Lin & Guinan, 2004; Recio & Rhode, 2000; Recio, Rich, Narayan, & Ruggero, 1998). They provide a
guide to what might be expected in the human periphery. Specifically, Schroeder phase complexes were presented to chinchillas (Recio & Rhode, 2000) and guinea pigs (Summers et al., 2003). Recio and Rhode (2000) used laser interferometry to observe the response of the basilar membrane of chinchillas to clicks, tones, and Schroeder maskers with and without an accompanying probe tone. The response to these waveforms was as predicted: schr+ produced “peakier” waveforms than schr- when observed qualitatively and when quantitatively measured as the amount of time the basilar membrane movement exceeded 20% of its maximal value. In addition to producing a peakier waveform, schr+ stimuli produced less suppression of a simultaneously presented pure tone than schr- did, by as much as 10 dB. Furthermore, the schr+ stimuli’s instantaneous frequency sweep was more accurately represented by the basilar membrane than the schr-. Summers et al. (2003) carried out a similar experiment on the guinea pig using laser velocimetry. Their results were qualitatively consistent with those of Recio and Rhode in that the responses recorded from schr+ signal were noticeably peakier than those elicited by the schr- signal. Unlike Recio and Rhode, however, the RMS values elicited by schr- and schr+ were nearly equivalent.

These data are consistent with the basic explanations for Schroeder masking, but there are numerous facets of human behavioral experiments that are not well understood. For instance, there appears to be an unexpected build-up of forward masking, forward masking of off-frequency tones depends on phase curvature (Wojtczak & Oxenham, 2009a), unexplained relationships between pure tone masking differences and speech recognition (Green & Rosen, 2013; Summers & Leek, 1998), and the relatively high
levels of masking produced by the schr- stimuli compared to other upward-sweeping maskers (Klyn et al., in preparation). As our understanding of the cochlear mechanisms does not currently account for these phenomena, it is reasonable to explore further how these stimuli are represented and processed at sites more central than the cochlea. To date, only one study that we are aware of has investigated physiological responses of the human auditory system beyond the cochlea, that of Rupp, Sieroka, Gutschalk, and Dau (2008). Rupp et al. examined the P1m component of the MEG in response to tones played in schr+ and schr- maskers, and found that this cortical response produced results qualitatively similar to the behavioral results.

One of the outstanding gaps in our understanding of the difference in masking produced by the Schroeder stimuli is the lack of physiological investigations in the subcortical human auditory system. It is unknown if the difference in masking observed behaviorally may be seen in subcortical measures of auditory function such as the auditory brainstem response (ABR). These early objective responses to auditory stimulation provide some evidence of the type of processing that occurs between the cochlea and the cortex, up to and including the inferior colliculus (Moore, 1987; Starr, 1976). The purpose of this study was therefore to test whether the difference in masking observed in behavioral experiments can be seen in the ABR.

3.3 Materials and methods

3.3.1 Participants

Twenty (20) adult subjects participated in this study (19 females; age range 19-28 years; mean age: 20.7 years). Only one ear per subject was tested for this study, and each
subject had pure tone behavioral thresholds equal to or less than 20 dB HL (ANSI, 2004) at octave frequencies from 250-8000 Hz for the test ear. Subjects were recruited from The Ohio State University community and were paid for their participation. The subjects were briefed on the study and then reviewed and signed an approved consent form with the opportunity to withdraw from the study at any time for any reason.

Twenty (20) subjects contributed to ABR datasets for the masked pure tone thresholds using schr+ and schr- maskers, whereas a subset (n=13) contributed to the datasets the schr+ and schr- stimuli alone. This was due to the relatively long test time required for the experiment and the resulting inability to obtain data for all conditions for all subjects; total testing including audiometric screening was kept under two hours for all subjects.

3.3.2 Stimuli

Stimuli were constructed in SigGenRP (Tucker Davis, 2001) and imported using BioSigRP (Tucker Davis, 2013) for presentation. Schroeder stimuli were constructed from equal amplitude harmonics from 200-5000 Hz using an f₀ of 100 Hz. Each component was set using a SigGenRP normalization file to 43.1 dB pSPL and with a phase set according to the formula used by Smith et al. (1986): \( \theta_n = \pm \pi n(n - 1)/N \) where \( n \) is the harmonic number, \( N \) is the total number of harmonics in the complex, and \( \theta_n \) is its phase. When the sign of the formula is set to (+) a schr+ stimulus is generated, whereas a (-) sign creates the schr- stimulus. For the masked threshold experiment, the Schroeder stimuli were set to a 40 ms duration including 5 ms raised cosine ramps. These relatively long ramps were chosen as previous research using noise bursts has shown that
longer ramp times reduce the evoked response amplitudes (e.g., Barth & Burkard, 1993), and the goal of this portion of the experiment was to examine the response to the pure tone presented with the Schroeder stimuli. Examples of these stimuli may be seen in Figure 7.

Schroeder stimuli (maskers) were presented alone and with a 4 kHz tone burst (signal). The tone burst was 1.25 ms in duration with 0.5 ms raised cosine ramps. The initial level of the tone burst was 80 dB pSPL, which was then decremented in six 10 dB steps. These levels were chosen based on previous behavioral tests and also pilot data.

Figure 7. Harmonic complexes ("Schroeder maskers") with phases set according to the formula proposed by Smith et al. (1986), an \( f_0 \) of 100 Hz, and components from 200-5000 Hz.
using 10 subjects not included in the present analysis. Previous research involving noise-masked thresholds for pure tone signals found that ABR thresholds were typically within 5 dB of behavioral thresholds for the same stimuli (Nousak & Stapells, 2005). While the subjects in this experiment did not participate in a behavioral task using these stimuli, the timing of the stimuli was chosen based on previous psychoacoustical research. Using long-duration Schroeder maskers and brief (5 ms) pure tone signals, Oh (2013) found a maximum difference in masking between the schr- and schr+ maskers with an f₀ of 100 Hz when the 4 kHz signal was presented 6 ms into a 10 ms period of each masker. Similar results have been reported with different signal frequencies and masker f₀’s (Carlyon & Datta, 1997b; Kohlrausch & Sander, 1995; Summers, 2000). In an effort to balance test time with minimizing overshoot effects (Summers, 2001a) or offset responses from the masker, the 4 kHz signal was presented 26 ms after the initial onset of the 40 ms masker. This did allow for the possible overlap of the wave V of the ABR to the tone burst and the Pa component of the middle latency response (MLR) to the Schroeder stimuli which will be discussed in the Results section below.

3.3.3 Procedures

All testing was performed in a small single-walled sound-attenuating booth. Calibration was performed monthly using a 2-cc coupler and a Larson-Davis model 824 sound level meter. The ER-2 insert earphone was tested at 1/3 octave frequencies from 100-5000 Hz, then these results were used to interpolate the amplitude response of the earphone at each component frequency using MATLAB’s spline function (Mathworks, 2012). The appropriate correction values were then used in the normalization file to
provide a flat spectrum over the frequency range of the stimuli and verified acoustically. Stimuli were presented monaurally to the test ear at a rate of 17.1/second.

ABR testing was performed using a TDT system 3 evoked potential system (Tucker-Davis Technologies, Gainesville, FL). Surface disposable silver electrodes were placed on the head at Fz, M1, and M2; the high forehead, ipsi- and contralateral mastoids. Electrode impedance was maintained at less than or equal to 5 kΩ as confirmed on the TDT equipment before and after each recording session. Recordings were amplified 100k, and responses exceeding ±50 µV were rejected from the averaged waveform through the automated artifact rejection. Each condition was recorded twice using 2000 sweeps and a 50 ms window bandpass filtered from 30-3000 Hz.

After signing the informed consent form, subjects' pure tone thresholds in air were estimated behaviorally (ANSI, 2004). Once the electrodes had been placed and the subjects were seated in the booth, subjects were instructed to remain as still as possible during the course of the testing. Prior to testing with the Schroeder stimuli, the tone bursts were presented in quiet in order to ensure recording quality and to find approximate individual amplitudes and latencies. All subjects provided clearly identifiable waveforms for the tone bursts in quiet. These data are not presented here. Subsequent to the tone burst in quiet, recordings were made with the schr+ and schr-maskers with the tone burst signal. If subjects produced large numbers of artifacts, the run was stopped, the subject was allowed a brief respite, and reinstructed to remain still during the testing. When the subject declared themselves ready to resume, the testing recommenced.
After data collection was complete, each pair of waveforms from the same condition were visually inspected for wave V, if present its latency and peak-to-trough amplitude was recorded. The threshold was estimated as the final signal level at which a repeated wave V could be identified. The threshold data were subjected to a paired t test to assess statistical significance. The experiment was approved by the Ohio State University Institutional Review Board.

3.4 Results

An example series of waveforms from ABRS to tone bursts presented with Schroeder stimuli may be seen in Figure 8. As can be seen in this example, the morphology of the response was substantially degraded relative to what may be seen for tone bursts in quiet. Thresholds were determined by visual inspection of wave V of the ABR, defined here as a reproducible positive peak preceding the most striking negative trough occurring 5-12 ms after the onset of the tone burst. There was the potential for an overlap between the wave V of the tone burst ABR and the Pa component of the masker MLR, as was noted above. However, we hoped to minimize the Pa by using the relatively long ramp (5 ms) for the masker. Furthermore, the masker was presented at a fixed level for all conditions; we can therefore assume that any changes observed were a product of the decrementing level of the signal. If a there was a large positive deflection in the waveform that did not change in amplitude or latency as the signal changed level then we would have to interpret it as a Pa rather than a wave V. As the results show, there was a consistent change that must be attributed to a signal response rather than a masker response.
Figure 8. Response waveforms from one subject to schr- (top) and schr+ (bottom) with a 4 kHz tone burst at 26 ms (vertical line). Arrows mark the wave V of the 80 dB condition.
3.4.1 Amplitude

As the signal level was lowered, the amplitude of the wave V response generally decreased. An individual regression line was calculated for the amplitude versus level function for each subject, and the slopes and intercepts of these lines were compared using paired $t$-tests. These tests fail to reject the null hypothesis that the slopes and intercepts of the signal in schr- and schr+ were different ($p > .05$). For a 4 kHz signal presented at 80 dB pSPL in the schr- masker, the average amplitude (peak to trough) was 0.273 $\mu$V (0.118 SD). As the level reduced, the minimum average amplitude observed was 0.157 $\mu$V (0.019 SD) at a signal level of 40 dB pSPL, though the final observed amplitude for the 30 dB pSPL signal was 0.258 $\mu$V. It is worth noting that, as the signal level decreased, fewer and fewer subjects contributed to the averaged data because the signal was below threshold for the majority of the subjects. Data from lower signal levels should therefore be interpreted with a great deal of caution. For the schr+ case, a statistically similar trend was seen, with an average of 0.269 $\mu$V (0.111 SD) at 80 dB pSPL decreasing to 0.178 $\mu$V (0.075 SD) at 50 dB pSPL followed by a slight increase as the number of subjects dropped. The mean amplitude as a function of signal level is plotted in Figure 9.
3.4.2 Latency

Latencies increased as the signal level decreased and the effective masking increased. An individual regression line was calculated for the latency versus level function for each subject, and the slopes and intercepts of these lines were compared using paired $t$-tests. These tests fail to reject the null hypothesis that the slopes and intercepts of the signal in schr- and schr+ were different ($p > .05$). Relative to the onset of the signal, the average latency was 7.79 ms (1.22 SD) for the 80 dB pSPL signal to 10.1 ms (2.2 SD) when presented at 40 dB pSPL signal when played with schr-. Again there
was a reversal of this general trend at the lowest signal level, with the average latency at 30 dB pSPL of 9.16 ms (2.41 SD). For the schr+ case the latencies increased from 9.47 ms (2.29 SD) to 10.27 ms (2.44 SD) from 80-30 dB pSPL. The latency as a function of signal level is plotted in Figure 10.

![Figure 10](image.png)

Figure 10. Average latency as a function of signal level with regression lines. Error bars show standard deviation. Symbols are horizontally offset for clarity.

3.4.3 Masked thresholds

Thresholds were determined as the lowest signal level at which a wave V was visually identifiable for each individual subject. Due to the noisiness of the recorded waveforms we recorded the schr- and schr+ without the signal present for a subset of
subjects (N=13). Example waveforms from one subject are reproduced in Figure 11. We then subtracted each subject's masker response waveforms from the signal and masker response waveforms in order to reduce the possible contributions of responses to the masker. While this should theoretically yield an ABR to the signal alone, for these recordings there was not an appreciable increase in the clarity of the response to the tone burst signals. As such, thresholds for the signal for all subjects (N=20) are reported here, and may be seen in Figure 12. The thresholds for the schr- condition (mean 57, s.d. 15.3) were higher than for the schr+ condition (mean 47.5, s.d. 16.2), a statistically significant difference as confirmed by a paired-t test ($t(19) = 2.08, p = .026, 95\% CI$ for the difference: -0.04-19.04).

Figure 11. Waveforms from one subject to schr- and schr+ maskers without a signal.
Figure 12. A box plot of estimated thresholds for all subjects (n=20) using the presence/absence of wave V of the ABR. Boxes show first thru third quartiles with the 10th and 90th percentiles shown as whiskers. Data outside these percentiles are shown as individual dots.

3.5 Discussion

3.5.1 Latency and amplitude

Two factors contributed to the increase in latency and the decrease in amplitude observed in both masker conditions. First, and most well-described, decreasing the signal level has a consistent and predictable effect on the wave V latency and amplitude. As the signal level is decreased the latency increases and the amplitude increases. For a 4 kHz tone burst, the functions that predict these changes have been in the area of -0.05 ms/dB
for latency (Gorga, Kaminski, Beauchaine, & Jesteadt, 1988) and 0.009 µV/dB for amplitude of wave V (Picton, Ouellette, Hamel, & Smith, 1979). For our subjects, the mean latency function for the 4 kHz tone bursts in quiet was -0.046 ms/dB, and the mean amplitude function was 0.005 µV/dB. Similarly, as the level of effective masking increases, the latency increases and the amplitude decreases (Burkard & Hecox, 1983a, 1983b; Hecox, 1974). Effective masking has been reported to affect the latency at ~0.025 ms/dB and the amplitude -0.003 µV/dB for wave V (Burkard & Hecox, 1983a). Previous research has used random noise maskers, with the result that the expected contribution of the noise to the ABR will be minimized in the final waveform through averaging - only the signal was time-locked to the recording window. For this reason it was logical for those researchers to keep the signal level constant and vary the noise level, as it allowed a clear view of the effect of masking on the signal-related ABR without confounding that change with different signal levels. Our study used maskers that were also time-locked to the acquisition window, so for this initial study we kept the masker at a constant level and varied the signal level. The result was that changes in the response waveform can be confidently attributed to changes in the signal rather than changes in the masker, but it does not allow us to divorce the contributions of reduced signal level from the contributions of the increased effective masking. The latency (-0.018 and -0.038 ms/dB) and amplitude (0.0006 and 0.0007 µV/dB) functions reported here very likely show the combined effects of the masking increasing as the signal level was decreased. Perhaps most relevantly, the reduced clarity of waveform morphology and the reduced number of subjects contributing to lower signal levels makes these functions less precise than would
be useful for further analytical purposes. The slopes of the average regression lines are
unduly influenced by the small number of subjects who contributed to the data at the
lowest signal levels.

The expectation would be that the slopes of the amplitude and latency functions
would be similar, but that there would be offset reflected in the intercept of the regression
lines. This was not what was observed, however, and the cause of this result is not clear.
It is possible that there was an interaction between the response to the maskers and the
changing response to the signal. In this scenario, as wave V increased in latency, it at
some point overlapped with an MLN response to the masker. This in turn could cause a
slight increase in amplitude or shift the peak of the signal-provoked wave V. Most likely
the unfortunately noisy recording situation resulted in a high level of variance in the
amplitude and latency data. Further experiments that vary the parameters of the recording
set up will allow us to investigate this phenomenon more thoroughly.

3.5.2 Morphology

As may be seen in Figure 8, the response waveform to the tone burst signal also
included noise from non-signal related sources. Some of this is due to experimental set-
up, such as the impedance of the electrodes or myogenic artifacts that do not meet the
threshold for automatic rejection. While this is always the case in far-field
electrophysiological recordings such as a scalp-recorded ABR, in this case it was
important to ensure that the interpretations of the waveform were not significantly
affected by the presentation of the masker that was also time-locked to the acquisition
window. As noted in the Results section, the subtraction of the masker-only ABR from
the signal+masker ABRs did not substantially alter the interpretation of the waveform for threshold purposes. As the Schroeder maskers are commonly compared to frequency-sweeps (i.e., "chirps"), this, too, was unexpected. Each period of the schr- masker contains an upward sweep of instantaneous frequency which is repeated four times over each presentation of our 40 ms masker. Similarly, the schr+ masker contains four downward instantaneous frequency sweeps. Research using chirps designed to compensate for the traveling wave delay in electrophysiological testing has shown that repeated presentations of chirps (i.e., "chirp trains") produce a repetition of the wave V component of the ABR for each chirp presented (e.g., Junius & Dau, 2005). Yet an examination of the waveforms in response to the schr- and schr+ maskers alone revealed no repeated wave V. As the Schroeder formula was not designed to compensate for the traveling wave delay but to create a flat amplitude envelope (Schroeder, 1970), it seems likely that responses to the schr- and schr+ stimuli’s individual periods created overlaps that reduced their contributions to the final recordings. In other words, a positive deflection in response to the first period of the stimulus may have interacted with a negative deflection to the second period of the stimulus, resulting in a negligible response at that time point. This possibility may be elucidated by further testing.

3.5.3 Masked thresholds

The primary hypothesis of this study was that the differences in masked thresholds with schr- and schr+ found in behavioral experiments will be observed in threshold estimates from the ABR. The difference in masking in this experiment (9.5 dB) was consistent with this basic hypothesis. The threshold values observed were slightly
different from those that have been obtained using behavioral methods. The most comparable paradigm in terms of stimulus parameters has been the masking period pattern (MPP) experiment, as these studies use a short duration signal presented simultaneously with a masker. Kohlrausch and Sander (1995) reported masked thresholds for different signal frequencies and masker combinations, the most comparable to the present experiment were MPPs for the 5 ms, 1100 Hz signal presented with schr- and schr+ maskers with $f_0$s at 25 and 100 Hz. With an $f_0$ of 100 Hz they observed differences larger than those reported here, ~8 dB at the minimum and ~20 dB at the maximum difference in MPP. The differences observed with an $f_0$ of 25 Hz were smaller, ranging from ~2.5 to 10 dB. Using an $f_0$ of 100 Hz and a signal frequency of 2 kHz, the differences observed were yet smaller, only a few dB over the range of the period.

Carlyon and Datta (1997) used an 1.1 kHz signal 5 ms in duration and found threshold differences of ~10 dB between the schr- and schr+ maskers at 6 and 8 ms into the 10 ms period of the maskers. Summers (2000) provided MPPs for 7 normal-hearing listeners detecting a 5 ms, 4 kHz tone burst in long duration schr+ and schr- maskers with an $f_0$ of 100 Hz and 49 components. The signal was fixed at either 60 or 80 dB SPL and the masker was varied adaptively. The maximum difference in masking for the 60 dB signal (~20 dB) occurred at 5.5 ms into the period of the maskers. The signal level relative to the masker component level at threshold was approximately 16.9 dB for schr- and -3.1 dB for schr+. For the 80 dB signal the maximum difference between thresholds (~10 dB) occurred 8 ms into the masker period. The thresholds were again ~16.9 dB for schr- but up to ~7 dB for the schr+ masker in this condition. Oh (2013) provided results that were
comparable to those reported by Summers, but adaptively varied the signal level as was done in the present experiment. Again a 5 ms, 4 kHz signal was presented in long duration schr- and schr+ maskers with an $f_0$ of 100 Hz and 49 components. For his three subjects, there was a clear maximum difference (>20 dB) at 6 ms into the period of the maskers. These thresholds were ~23 dB relative to the component level of the schr-stimulus and ~0 dB for the schr+.

The thresholds observed in the present ABR experiment were 13.9 dB and 4.4 dB above the level of the masker components for the schr- and schr+ maskers, respectively. These are qualitatively similar to, but bracketed by, the behavioral results reported previously, which typically exceed 10 dB. The behavioral results have reported higher thresholds for the schr- masker and lower thresholds for the schr+ masker than were observed in the present study. Several factors may have contributed to this difference. One explanation may be that the duration of the tone burst presented in this study affected its detectability. The tone burst was one-fourth the duration that has been used in behavioral studies, and this difference may have contributed to the reduced difference in masking. It has been well established that the duration of a tone affects its detectability (e.g., Garner & Miller, 1947), including down to the durations used here (Gerken, Bhat, & Hutchison-Clutter, 1990). The mechanism by which thresholds in schr+ and schr- would be differentially affected by the signal duration is unclear, but further behavioral or electrophysiological testing could provide evidence, or lack thereof, for this possibility.

A second, more likely, explanation is that the temporal proximity of the signal to the onset of the masker produced what is commonly called overshoot (Zwicker, 1965a,
This phenomenon has been consistently demonstrated by comparing the threshold of a short duration signal near the onset of a tonal or noise masker to one in the middle of the masker. The amount the threshold changes can be quite large depending on the parameters used, by 15-20 dB in some conditions. There are two frequently discussed mechanisms for this phenomenon that relate adaptation to the masker to signal detectability - one is adaptation at the auditory periphery, frequently assumed to occur at the level of the auditory nerve. At the onset of a stimulus the auditory nerve reaches its maximum firing rate quickly, then slowly adapts over a period of time (~150 ms) to a lower, steady firing rate (Smith & Zwislocki, 1975). The second is adaptation of the cochlear dynamic range by way of the efferent system, and particularly the medial olivocochlear reflex (MOCR) (Schmidt & Zwicker, 1991). This is believed to reduce the gain produced by the outer hair cells in response to the stimulus over a ~200 ms time course. This explanation has received support from several models (e.g., Jennings, Heinz, & Strickland, 2011; Strickland, 2001) which appear to explain a great deal of the overshoot data.

The cochlear processing of the schr+ and schr- stimuli results in output waveforms from the basilar membrane that are differentially peaky. The schr+ stimuli are, as a result, more affected by the compressive nonlinearity of the basilar membrane (Alcantara, Moore, Glasberg, Wilkinson, & Jorasz, 2003; Carlyon & Datta, 1997a, 1997b; Oxenham & Dau, 2004; Summers, 2000; Summers et al., 2003; Summers & Leek, 1998). This, in turn, results in differential overshoot when comparing the two maskers. Summers (2001) evaluated overshoot using schr+ and schr- and included a 5 ms, 4 kHz
pure tone stimulus presented at 60 dB and 80 dB SPL. There was a clear pattern of increased overshoot from the schr+ masker compared to the schr- masker, effectively reducing the difference in masking produced. As with many overshoot experiments, early signal placement was limited to less than 20 ms from the onset of the masker; less than was used in the current study. There is ample evidence that overshoot effects can be observed beyond the 26 ms delay used in this study, however (McFadden, Walsh, Pasanen, & Grenwelge, 2010; Zwicker, 1965b), so this is a plausible explanation for the difference between the physiological masking difference observed here and prior behavioral research.

3.6 Summary

This study demonstrated for the first time that differences in masking by schr+ and schr- stimuli may be observed in the tone burst ABR. A mean difference of 9.5 dB was observed for the 4 kHz pure tone embedded within the maskers, with schr+ producing less masking overall than schr-, consistent with published behavioral research. This provides a further link between physiological studies with animal models and behavioral studies with human subjects. The results here were consistent with the explanation that schr+ and schr- stimuli produce differentially peaky waveforms at the basilar membrane and that this results in different amounts of masking for pure tone stimuli. Further research will be necessary to explore the morphological features observed here, and to investigate human physiological responses of more peripheral origin than the wave V of the ABR.
3.7 References


Chapter 4: Manuscript 3 - An Adaptive Procedure for Psychophysically Estimating Phase Curvature

4.1 Abstract

A harmonic complex with phases set to affect a sweep in its instantaneous frequency can produce substantially different amounts of masking depending on the direction and rate of the sweep. The masker phase curvature that produces the least amount of masking for a particular frequency is thought to have the same magnitude but opposite sign as the phase curvature at the place on the basilar membrane stimulated by that frequency. By finding thresholds in different phase curvatures, previous research has found a minimum by way of fitting a curve to the averaged threshold data. This paper presents a procedure to estimate the phase curvature by adapting presentation of stimuli both in level and phase curvature in order to locate a minimum in the underlying curvature versus threshold function. Data from simulations and a psychoacoustical experiment support its effectiveness in estimating the phase curvature in fewer trials than the conventional procedure. Further refinements to improve reliability and speed are proposed.

4.2 Introduction

Over the past 30 years there has been substantial evidence that altering the phase relationship of the components of a masker can create large differences in masking of
simple signals, and smaller differences in masking complex signals like speech (Carlyon & Datta, 1997a, 1997b; Gifford et al., 2008; T. Green & Rosen, 2013; Hoglund et al., 2013; Lentz & Leek, 2001; Mehrgardt & Schroeder, 1983; Oxenham & Dau, 2001b, 2004; Rupp et al., 2008; Shen & Lentz, 2009; Smith et al., 1986; Summers, 2000, 2001b; Summers et al., 2003; Summers & Leek, 1998; Wojtczak, Beim, & Oxenham, 2014a; Wojtczak & Oxenham, 2009a, 2010). Since Smith et al. (1986), the explanation offered for the difference in masking known as the masker phase effect (MPE) rests on the interaction of the phase relationship of the components and the phase response of the basilar membrane. When a complex sound has its phase curvature set such that it is equal in magnitude, but opposite in sign, to the phase curvature of a point on the basilar membrane, the waveform of the membrane's response at that point will have a very high crest factor - it will be "peaky." Contrarily, if the complex sound has its phase curvature set with any other value, the output waveform of that same point on the basilar membrane will have a lower crest factor. The explanation continues that the peakier the response of the basilar membrane, the less masking is produced by the complex for a signal presented at that point on the basilar membrane. It follows that by finding the phase curvature that produces the least amount of masking for a pure tone signal, investigators can find the phase curvature of the point on the basilar membrane corresponding to that frequency.

Lentz and Leek (2001) proposed just such an experiment, where they modified the formula used by Smith et al. (1986) to include the scalar $C$ instead of simply changing the sign from positive to negative (formula 1).

$$\theta_n = C\pi(n - 1)/N$$  \hspace{1cm} (1)
Where $\theta_n$ represents the phase of the $n$th harmonic and $N$ is the number of harmonics in the complex. When the scalar, $C$, is set to -1 or +1, the formula is the original Schroeder equation used by Smith et al. (1986) and subsequent researchers. Lentz and Leek tested four pure tone signals at 1, 2, 3 and 4 kHz, and varied the scalar in ~10 steps between -1 and 2 for their 4 listeners. Their results were consistent with previous research and the explanation of the MPE discussed above, that a positive scalar value interacts with the negative phase curvature of the basilar membrane to produce a peaky internal response and a less effective masker. The $C$ values they chose were not equally spaced, but allowed Lentz and Leek to narrow down a possible band of $C$ values that produced the least amount of masking for their subjects, and thus a range of possible phase curvatures for those points on the basilar membrane.

Oxenham and Dau (2001) used the formula proposed by Lentz and Leek (2001) to systematically investigate the phase curvature of the basilar membrane. They tested signal frequencies at octaves between 250 and 8000 Hz, and also tested the effect of different masker levels. They furthermore varied the $f_0$ of the harmonic masker, which allowed for a second test of the effect of the phase curvature on the amount of masking produced. By fixing the bandwidth of the masker, raising and lowering the $f_0$ changes the number of components in the masker and thus the denominator of formula 1. Oxenham and Dau used nine values of the $C$ scalar from -1 to 1 in 0.25 steps for each signal frequency, masker level, and $f_0$ tested. Their data were consistent with the results reported by Lentz and Leek (2001), and showed variation in the amount of masking produced by the different scalars. When the masking threshold was plotted against the $C$ value, a clear
bowl-like shape was revealed in the data. Oxenham and Dau used two methods in interpreting their data; first, they found the $C$ value that provided the minimum amount of masking on average among their listeners, and second they fitted a sinusoidal function to the averaged threshold estimates and found the minimum of the fitted curve. There was some discrepancy between the results produced by these methods at the lower frequencies tested, but on the whole the results agreed well. The fitted minima approach - estimating thresholds for multiple $C$ values and finding the minimum of the function fit to the averaged thresholds - has been used consistently since Oxenham and Dau pioneered it (e.g., Oxenham & Dau, 2004; Oxenham & Ewert, 2005; Shen & Lentz, 2009; Wojtczak & Oxenham, 2009).

There were two facets of this procedure that prompted the current study. The first was the discrepancies between the results produced by the fitted functions and the found averaged minima. This does not seem, by itself, a difficult problem to solve, as further estimates could be obtained from different subjects to drive down the inevitable errors in measurement in threshold estimation. Furthermore, different $C$ values could be used to try to pin down which $C$ value provides the least masking for a given individual. This leads to the second facet; like many psychoacoustical experiments, psychophysically estimating the phase curvature of a subject for a particular frequency takes an extraordinarily long time. Each estimate of the phase curvature requires a threshold estimate, typically several, for each $C$ value tested. For Oxenham and Dau (2001) this means that for a single frequency at a particular level and with a particular $f_0$ for the masker, a minimum of 27 threshold estimates are needed, with a worst-case (and
unlikely) scenario requiring 54 threshold estimates. The typical choice for estimating a threshold in all of the studies cited above was the transformed up-down staircase method of Levitt (1971), which may not be the most efficient adaptive procedure available (i.e., Amitay, Irwin, Hawkey, Cowan, & Moore, 2006). Regardless of the adaptive procedure used, however, estimating thresholds for maskers with $C$ values that produce relatively large amounts of masking seems inefficient. Though these data points allow for a better fit of an assumed underlying function, if the goal of the research is to find the point of minimum masking then the description of the function is of secondary importance.

If researchers wish to continue to estimate the phase curvature of the basilar membrane it would be helpful if these estimates could be arrived at more efficiently. What is needed is a procedure that provides a psychophysical estimate of the phase curvature of the basilar membrane reliably and efficiently. One such procedure is proposed here. It makes use of small blocks of trials in order to find a point of minimal masking in the threshold vs $C$ value function. Like all previous published research on this phenomenon, this procedure assumes that the phase curvature of the basilar membrane for a particular frequency may be estimated as a constant. In addition, the assumption is that there is a single $C$ value which will produce a minimum amount of masking and that the amount of masking produced by other $C$ values raises monotonically as the value moves further away from the minimum.

4.3 Adaptive procedure

The proposed adaptive procedure tests multiple $C$ values (the "spread") at one presentation level. Once a certain number of trials have been completed for each $C$ value
(a "block"), the number of $C$ values for which a certain threshold has been reached is counted. If more than three $C$ values are at or above threshold, then the level of all of the maskers is reduced one step and the testing begins again. This is similar to the BUDTIF procedure described by Campbell and Lasky (1968). If fewer than two $C$ values are at or above threshold, then the level of all the maskers is increased one step and the testing begins again. If two or three $C$ values are at or above threshold, then the range of $C$ values is decreased to exclude maskers that are effective at that level, new $C$ values are chosen within that range, and the testing begins again. This process continues until a minimum value between two adjacent $C$ values is reached. A final round of trials is initiated, and the mean of the $C$ values that elicited the most correct answers in this final round is taken as the $C$ value providing the minimum amount of masking. A flowchart of the basic process may be seen in Figure 13.
Figure 13. Flow chart of proposed adaptive procedure with the threshold value of 0.75 and four trials per block. Percent correct is abbreviated p(c).

Certain rules were added to speed up the procedure. First, testing at any C value was stopped once the threshold value was reached. For instance, if the threshold value was
80% and there were five trials, two wrong answers meant the maximum percentage 
correct on that C value was 60% and testing at that C value would be stopped. Second, 
there would likewise be no use to continue testing after four correct answers with those 
parameters.

The level step-size followed conventions for the adaptive staircase procedure, 
halving in size after every reversal down to a predetermined size (Levitt, 1971). For the 
adaptive procedure in the simulations and experiments presented here, the initial step-size 
was 8 dB, and was halved at every reversal to minimum step-size of 1 dB. Likewise the 
procedure utilized two-alternative forced-choice (2AFC), but should be compatible with 
3AFC, 4AFC, etc.

4.4 Simulations

Simulations were conducted in MATLAB in order to test the validity of the 
proposed procedure and the effects of different parameters; stopping rule, inattention rate, 
threshold value, number of C values, and number of trials. For each simulation a random 
C value between 0 and 1 was chosen as the minimum. The C value of -1 was taken as the 
maximum, and a sine function was used to fit all remaining C values. A Weibull function 
was generated for each C value to imitate the psychometric function. Consistent with 
published data, the difference between the maximum and minimum was 20 dB (Oxenham 
& Dau, 2001b).

Ten-thousand simulations were run for each set of parameters and over 100 
combinations of parameters were tried. Stopping rules (the minimum difference between 
adjacent C values) ranged from 0.001 to 0.2, subject inattention rates ranged from 0-5%,
threshold values ranged from 0.75 to 1, number of $C$ values tested was either five or seven, and number of trials ranged from 4-20. For the sake of brevity, not all combinations are presented here, but only those relevant to showing general trends in parameter choices. As may be seen in Figure 14 and Figure 15, reducing the value of the stopping rule below 0.05 offered little improvement. As the procedure narrowed the tested $C$ values into the trough of the underlying sine function, the differences between the psychometric functions of the adjacent $C$ values became smaller and smaller and the procedure was less able to distinguish between them. Increasing the number of trials per block can overcome this for threshold values below 1, with the obvious trade-off that a larger amount of time will be necessary to create each estimate. When the threshold was 1 (Figure 14), increasing the number of trials per block slightly increased the average absolute error.
Figure 14. Average absolute error as a function of trials per block with error bars at one standard deviation, based on 10,000 simulations of the adaptive procedure using five C values, a 3% inattention rate, and a threshold value of 1 (see text). Symbols have been offset along the abscissa for clarity.
Subject inattention rates had a predictable effect on the performance of the procedure, with larger inattention rates producing larger average errors (Figure 16 and Figure 17). There was a substantial interaction between the inattention rates and the threshold values used. If one incorrect response was allowed (threshold of 0.75 for four trials, or 0.8 for five to nine trials, Figure 16) then a higher simulated inattention rate pushed the overall accuracy down. If no incorrect responses were allowed (threshold of 1, Figure 17) then the pattern that emerged depended on both the inattention rate and the
number of trials. For inattention rates of 2% there was little difference between five, six, or seven trials, and less than a 0.01 difference in average absolute error between four to eight trials. For inattention rates of 3%, five trials produced the smallest average absolute error and standard deviation, numbers of trials below and above five produced less accurate results. At 5% inattention rate, the more trials used, the less accurate and more disperse the results became. This was a logical consequence of introducing errors, allowing no margin for mistakes in the trials, and increasing the numbers of trials.

Figure 16. Average absolute error as a function of trials per block with error bars at one standard deviation, based on 10,000 simulations of the adaptive procedure using five C values and a threshold value of n-1/√n (see text). Symbols have been offset along the abscissa for clarity.
Figure 17. Average absolute error as a function of trials per block with error bars at one standard deviation, based on 10,000 simulations of the adaptive procedure using 5 C values and a threshold value of 1 (see text). Symbols have been offset along the abscissa for clarity.

With low inattention rates, such as might be expected to be found with practiced subjects, more trials generally increased the accuracy of the procedure. There was a clear tendency towards asymptote as the numbers of trials increased, however, and the increased precision may be offset by the increased time necessary to test for that many trials. The number of trials at the elbow of the function depended on the other parameters chosen, but with a relatively low subject inattention rate (3% or less), a small number of trials per block (four or five) produced absolute average errors that were smaller than the
step-size between adjacent $C$ values used in previous experiments. Looking at the number of trials required to produce an estimate (Figure 18), as expected, larger number of trials per block required larger numbers of trials for the final estimate.

Figure 18. Average number of trials needed to reach the stopping point for the adaptive procedure (ordinate) versus number of trials per block (abscissa) based on 10,000 simulations of the adaptive procedure using a 2% inattention rate (see text). Other inattention rates did not significantly change the outcome. Error bars are one standard deviation.

One unexpected result from the simulations was the effect of the number of $C$ values tested in each spread. Using a total of five $C$ values in each spread typically produced substantially lower absolute average errors than using the same parameters and seven $C$ values in each spread. The exception to this finding was with the highest
inattention rate tested (5%) where the larger spread mitigates the deleterious effect of a larger number of trials. At a threshold of 1 and inattention rate of 5%, having a spread of seven C values produces smaller error than a spread of seven C values for trials per blocks of five and greater. For every other combination of parameters tested, however, a five C value spread produced smaller average absolute errors than a seven C value spread. Contrarily, using five C values required larger numbers of trials in order to arrive at an estimate, regardless of the number of trials per block, threshold value, or inattention rate. This difference was typically around 20%. According to the simulations, then, there was a trade-off between speed and accuracy, as five C values produced slightly more accurate estimates slightly more slowly than seven C values.

4.5 Experiments

The adaptive procedure outlined above was tested experimentally with human listeners. In order to provide a comparison with more conventional methods, threshold estimates for multiple C values were found with the same subjects. These thresholds were then fit with a sine function to find an estimate of the C value that would produce a minimum amount of masking. This allowed for direct comparisons between the different parameters used in the new adaptive procedure and those obtained through the Levitt or other threshold-tracking procedures.

4.5.1 Methods

4.5.1.1 Listeners

Four normal-hearing listeners (3 female) participated in this study. The age range was 20-34 years, average 24 years. Normal hearing was determined by pure tone
audiometry with thresholds ≤20 dB HL at octave frequencies from 250-8000 Hz (ANSI, 2004). All subjects were well practiced with psychophysical testing before the beginning of data collection. The experiment was approved by the Ohio State University Institutional Review Board.

4.5.1.2 Stimuli

Three pure tone signals were tested; 1, 2, and 4 kHz. Each was 260 ms in duration including 30 ms raised cosine ramps, and was temporally centered in the masker. For the maskers, the Lentz and Leek (2001) modification of the Schroeder formula (equation 1) was used to set the phase relationship of the components. Equal amplitude harmonics of the fundamental from .4-1.6 of the signal frequency were used for each masker (Oxenham & Dau, 2001a). An f₀ of 50 Hz was used for all maskers. The maskers were 460 ms in duration including 30 ms raised cosine ramps. The overall amplitude of the masker was set to 75 dB SPL.

Rather than using C values from -1 to 1 for the conventional approach, C values were chosen to bracket the C value estimated to produce the minimum amount of masking by prior research (Lentz & Leek, 2001; Oxenham & Dau, 2001b; Shen & Lentz, 2009). A spread of nine C values was used with the conventional approach, which, combined with the C values chosen, meant the difference between C values used in this study was smaller than that used in prior research. For the 1 kHz and 2 kHz signals, C values spread from 0 to 1, and for 4 kHz, C values spread from -0.25 to 0.75. For the adaptive procedure, the C value spread began with endpoints at -1 and 1, with the C values chosen by the procedure as the experiment progressed.
4.5.1.3 Procedures

Conventional: The 1-up, 3-down adaptive staircase method of Levitt (1971) was used in a 2AFC task to estimate the 79.1% of the psychometric function at each $C$ value. The signal started at 75 dB SPL. The initial step-size was 8 dB and was halved at each reversal until a minimum step size of 1 dB was reached. Ten such reversals were collected, the first four were discarded, and the average of the final six was used as the threshold estimate. Three such estimates were taken for each condition, and if the thresholds varied by more than 5 dB, further estimates were taken until three were within 5 dB of each other. This occurred three times in total. All thresholds, including those outside the 5 dB range, were included.

Signals were chosen in a random order, and then $C$ values for each signal were randomly tested. In other words, all of the $C$ values for given signal frequency were tested prior to moving to a new signal frequency. Three $C$ value threshold estimates would be completed before moving on to the next $C$ value threshold estimate.

Adaptive: Based on the simulations reported above, four conditions were tested for each signal; threshold of 1 and four trials per block, threshold of 1 and five trials per block, threshold of 0.75 and four trials per block, and threshold of 0.8 and five trials per block. The test order for the 12 estimates (three signals and four parameters) for each subject was randomly assigned. We used a spread of five $C$ values, and the $C$ values ranged from -1 and 1, inclusive. The starting level of the signal was 75 dB SPL with an initial step-size of 8 dB which halved on each reversal to a minimum step size of 1 dB.
For both the conventional and adaptive testing, data were collected in sessions that lasted no more than two hours including a break at least every half hour. No more than one session was completed in a day, and subjects completed all conditions for the conventional and then the adaptive procedures over the course of several weeks. Each condition was completed within a single session, and most sessions collected data from multiple conditions. The stimuli were generated digitally in MATLAB at a 44.1 kHz sampling rate and were passed through RME Hammerfall DSP Multiface II audio interfaces. These were presented to the right ear of the subject via Sennheiser HD280 headphones while the subject was seated in a single-walled sound attenuating booth. The frequency response of the system was measured using a KEMAR with G.R.A.S. ear simulators and a Larson-Davis model 824 sound level meter; these measurements were then used to apply correction factors to create a flat frequency response without changing the phase relationships of the stimuli's components (Russotti et al., 1988).

4.5.2 Results

4.5.2.1 Conventional

The individual results for the different signal frequencies are plotted in Figure 19. The overall pattern for each subject is similar, but there were noticeable variations - particularly subject 1 (filled circles) at 1 kHz and subject 3 (filled squares) for each signal. The variability between subjects mirrored that observed by other researchers into Schroeder masking (e.g., Oxenham & Dau, 2001b; Shen & Lentz, 2009). Furthermore, Oxenham and Dau's Figure 1 and Shen and Lentz's Figure 1 also showed individuals with less consistently smooth threshold vs C value functions, similar to subject 3 in our
experiment. The source(s) of these differences is not clear, but if it is representative of a
difference in the underlying physiological process rather than a measurement error, it
may have substantial implications for the adaptive procedure as discussed below.

Figure 19. Individual average thresholds estimated with the conventional procedure.

The thresholds averaged across individuals are shown with standard deviations in
Figure 20, with a sinusoidal curve fit using four free parameters per signal ($y = y_0 +$
\(a \sin\left(\frac{2\pi x}{b}\right) + c\). From these results, the minimum point of masking for each signal was estimated by finding the minimum of the fitted curve. These \(C\) values do not, in themselves, provide much information about the maskers, as the phase curvature depends also on the fundamental and the number of components in the complex. The phase curvature is the second derivative of equation 1, i.e., \(\frac{d^2\theta}{df^2} = C \frac{2\pi}{Nf_0^2}\). The assumption underlying these procedures was that the phase curvature that produced the minimum amount of masking was equal in magnitude but opposite in sign to phase curvature of the stimulated auditory filter. Therefore, the auditory filter's phase curvature was estimated as \(-C_{\text{min}} \frac{2\pi}{Nf_0^2}\) using minima as described above. We then found the dimensionless phase curvature by taking these estimates and multiplying them by \(\frac{f_s^2}{2\pi}\), which allows for easier comparisons across the different signal frequencies (Shera, 2001a).

The minima of the curves in Figure 20 provide phase curvature estimates of -14, -9, and -6 for 4, 2, and 1 kHz signals, respectively. These were very similar to estimates provided in published research: for 4 kHz signals at similar levels, previous estimates were near -16 (Oxenham & Dau, 2001b; Shen & Lentz, 2009; Lentz & Leek, 2001 estimated ~25 but noted that their procedure likely overestimated the phase curvature at this frequency), for 2 kHz, the estimates were between -7 and -10 (Lentz & Leek, 2001; Oxenham & Dau, 2001b; Shen & Lentz, 2009), and for 1 kHz, between -3 and -8.3 (Lentz & Leek, 2001; Oxenham & Dau, 2001a, 2004). This similarity was impressive given that the current study used smaller steps between \(C\) values than previous studies.

Using the same curve fitting procedure for the individual data as was used for the average
produced phase curvature estimates between -8.2 and -24.5 for 4 kHz, between -3.3 and -10 for 2 kHz, and -3.7 and -8.8 for 1 kHz. As is clear from studying Figure 19, the use of a sine curve to fit the individual data required constraining the curve fit to have a single minimum over the range of $C$ values used. If a simple curve fitting procedure was followed using a sine function, subject 3 would have multiple minima over the range investigated. Previous studies that have investigated simultaneous masking using the Schroeder maskers have provided phase curvature estimates for averaged data only, so it is not possible to directly compare these results.

The average number of trials for these subjects to estimate a single threshold using the conventional procedure was 78 (range: 48-100). As nine $C$ values were collected in this and previous studies to estimate the phase curvature for a single frequency, the total average number of trials necessary using the conventional method was 702. This estimate excludes the repeated tests that were performed here and in previous research to ensure stable threshold estimates. This number of trials was therefore more comparable to a single estimate provided by the adaptive method.
4.5.2.2 Adaptive

The $C$ values that produced the minimum amount of masking for each subject at each frequency are presented in Figure 21, along with the individual's local minimum and fitted minimum from the conventional approach. It is unclear how to represent the uncertainty inherent in the conventional estimates, as these do not provide multiple estimates of the minimally masking $C$ value but only of threshold values at different $C$ values. As such, those estimates have been added to the figure as single points. Figure 21
shows that the adaptive procedure appeared to estimate a similar construct as the conventional thresholds and curve fitting procedure, though there are differences between these estimates at the level of an individual. There was substantial variability both within and between subjects, similar to what was seen in the conventional procedure (Figure 19).

Figure 21. Estimates of the $C$ value that produced the minimum amount of masking for four subjects. Larger symbols are from the adaptive procedure with error bars at a standard deviation. Smaller symbols are from the conventional procedure; filled are the local minima (the lowest threshold tested) and open are the minima from a curve fit to the individual's thresholds. Symbols have been offset along the ordinate for clarity.
Figure 22 contains box plots of the differences between the estimates from the conventional and adaptive methods. Figure 22A shows the differences between the adaptive procedure estimates and the local minima from the Levitt threshold estimates - the $C$ value that produced the lowest threshold for the individual. Figure 22B shows the differences between the adaptive procedure estimates and the minima from the fitted curve for each individual. These plots indicate that there were no systematic biases arising from the adaptive procedure relative to the conventional procedures for these subjects. While the five trials, n/n threshold condition did show a positive bias (the adaptive procedures generally estimated a higher $C$ value for minimum masking than either the local or fitted minimum), neither the other five trial condition nor the other n/n threshold condition showed evidence of a similar bias. It is possible there was some interaction between those two parameters, but there is no obvious cause of such an interaction and no interaction was observed in the simulations.
Figure 22. Differences between an individual's adaptive estimates and local minimum (A) and fitted curve minimum (B).
As there were no obvious systematic differences between the different parameters used in the adaptive procedure, we chose to collapse across these conditions to compare the averaged results with the conventional results. Since each individual provided three estimates at each C value in the conventional procedure, we randomly selected three of the adaptive estimates for each individual at each frequency before averaging. These averaged values were again transformed into dimensionless phase curvature estimates (section 4.5.2.1) in order to facilitate comparison. The averaged estimates were -10.5 for 4 kHz, -7.7 for 2 kHz, and -7.3 for 1 kHz. The estimates for the 4 kHz and 2 kHz were lower than the conventional, curve fit estimates (-14 and -9, respectively), whereas the 1 kHz was slightly higher (conventional: -6). As Oxenham and Dau (2001) noted, estimates within a factor of two may well be within the measurement error associated with the conventional measurement technique, but given the relatively consistent estimates noted above, the difference between our 4 kHz estimate and the conventional estimate (~3.5) was larger than hoped.

The number of trials needed to provide a single estimate was slightly lower than simulations predicted, averaging 484 trials over all of the testing conditions (range 199-602). For these subjects, as in the simulations, the higher threshold value (n/n) required fewer trials than the lower (n-1/n) by 13% on average. Four trials per block also required fewer trials than five trials per block by 17% on average, which was again broadly consistent with simulations.
4.6 Summary

This paper described a novel adaptive procedure for estimating the phase curvature of the auditory system using a spread of $C$ values and small numbers of trials at each $C$ value. The procedure then adapts both the level of the stimuli and the $C$ values used in order to estimate a $C$ value that provides the least amount of masking. The phase curvature of this masker is taken as the phase curvature of the basilar membrane at that frequency multiplied by -1. Simulations and an initial experiment provided evidence that the procedure can estimate the point of minimum masking with a reasonable degree of accuracy and in 30% fewer trials than the conventional method. The conventional procedure estimates thresholds at multiple phase curvatures and fits a curve to the estimates in order to locate a minimum. The data provided in the simulations and experiment indicate that this basic technique has promise in simplifying the estimation of the phase curvature but would benefit from further refinement and testing.

This study was, in many ways, a proof of concept. There are many parameters that have yet to be adjusted and modeled in order to further refine the procedure, such as level effects (i.e., Shen & Lentz, 2009), further number of $C$ values in a spread, adaptive numbers of trials, etc. Most importantly, test-retest reliability will need to be assessed using precisely the same conditions with the same subjects in order to assure the usefulness of the test. Perhaps most promising would be a reconsideration of the final phase curvature estimate. While the conventional procedure spends a considerable amount of effort in establishing thresholds at $C$ values that are of peripheral interest to the determination of the phase curvature at a particular frequency, the adaptive procedure
proposed here spends considerable effort in determining which $C$ values are not of interest and then does not use that information in its final estimate. The final estimates reported above only consider the most successful phase curvature values in the final spread of $C$ values, essentially discarding the information provided by the less successful values in that and previous spreads. Rather than discard this information, all of the trials could potentially be used in an adaptive curve fitting procedure similar to the maximum-likelihood adaptive procedures (e.g., Green, 1993; Pentland, 1980). The best manner in which to weigh the information provided by these trials will require further research.

Part of the future research will also necessarily need to investigate further reductions in the number of trials to arrive at the final estimate. The 30% reduction in the number of trials relative to the conventional approach is not insignificant, particularly in a 2AFC or 3AFC context, but is likely able to be matched through threshold estimation using alternative adaptive procedures (e.g., Kaernbach, 1990). Further refinement of the parameters in the basic procedure may allow for an improvement in this regard. For instance, in addition to a "keep" threshold, a "reject" threshold could be added to remove $C$ values that produce maskers that are too effective to require further testing. This could potentially improve accuracy of the final estimation, as the repeated testing at these $C$ values increases the likelihood that misleading positive responses will be offered.

It should be noted that this is not the only attempt to develop an improved method for estimating phase curvature. Recently, Rahmat and O'Beirne (in press) offered an adaptation of von Békésy's tracking procedure that rapidly estimates subject thresholds at different masker phase relationships in order to estimate the underlying phase curvature
of the auditory system. Their results were consistent with previously published research for the signal frequency tested (500 Hz) and the conventional method against which they compared it (Kaernbach, 1991). Their technique does not allow for a direct comparison on the number of trials. As each estimate is based on continuous sweeps from low to high (or the reverse) C values, it seems likely that it offers a larger gain in speed than the technique proposed here. Averaging up and down sweeps, as they propose, also reduces the potential hysterisis effects associated with the Békésy tracking procedure (Sęk, Alcántara, Moore, Kluk, & Wicher, 2005; Sęk & Moore, 2011). This is a promising approach, and warrants consideration by any researchers investigating phase curvature.

The results offered in this and previous research do provoke a question as to the cause of individual differences observed in the data from the conventional and adaptive techniques. As was seen in our Figure 19 and in other research that presented individual results (e.g., Gifford et al., 2008; Oxenham & Dau, 2001b; Shen & Lentz, 2009), the assumption that the masking pattern associated with masker phase curvature is well-represented by a sine curve is of questionable utility at the level of a single individual. Given the lack of an obvious mechanism that would cause the discontinuities in threshold versus phase curvature functions, it can be tempting to label these as products of measurement error. Certainly no threshold estimation procedure is perfect, and it is certainly possible that these individual differences primarily reflect that imperfection. If these differences are a result of measurement error then the basic gist of the adaptive procedure here retains validity. If, however, the individual threshold estimates that do not conform to the expected function are reflections of real differences, then the adaptive
procedure and the curve fitting conventional procedure will both fail to capture relevant individual differences. This is of particular importance for research questions directed at any possible changes in phase curvature. For example, Gifford et al. (2008) tested thirteen subjects pre and post cochlear implantation and used Schroeder masking as an assessment of changes to the cochlear nonlinearity. In addition to changes in the difference between the maximum and minimum amount of masking produced by Schroeder complexes (i.e., the MPE), their figures 5 and 6 show evidence that, for some subjects, the phase curvature that produced the minimum amount of masking changed after implantation. That the introduction of an electrode array into the cochlea could have some effect on its frequency dispersion should perhaps come as no surprise, but a literature search provided no evidence of research on this topic.

In summary, the present study presented a novel adaptive procedure for psychophysically estimating the cochlear phase curvature. Simulations and an experiment provided evidence that this procedure does find masker phase curvatures that produce minimum amounts of masking for a given pure tone signal. This phase curvature is assumed to be equal in magnitude but opposite in sign to the phase curvature of the basilar membrane at that frequency. Further refinement of the procedure would likely allow for yet more accurate and rapid assessment.
4.7 References


Chapter 5: Overall Discussion

5.1. Summary and discussion

The introductory chapter described the purpose of this document to be the contribution to scientific knowledge of the auditory system's phase response. The previous three chapters - written as separate manuscripts - have provided new data and methods to investigate the phase response and its role in masking in humans. As the purported mechanism of the MPE is a physiological response that is currently inaccessible to direct investigation in humans, other methods of exploration were required. In order to help connect the physiological explanation with the behavioral results, two chapters used psychoacoustical experiments (Chapters 2 & 4) and a third examined electrophysiological responses to phase-shifted harmonic complexes (Chapter 3). Each chapter approached the topic of the auditory system's phase response from a different angle, and hopefully together they can fill some portion of the gap between the physiological models and human behavior.

In Chapter 2, we investigated in two experiments whether the MPE could be elicited using maskers that shared superficial features with the Schroeder maskers. The primary interaction believed to cause the MPE is the interaction of the phase curvature of the basilar membrane - the frequency glide - and that of the stimulus - the sweep of the
instantaneous frequency. We compared Schroeder maskers to chirps with different sweep rates in order to test whether this feature was sufficient to induce an MPE. These chirps were from different sources; in Experiment 1, we used repeated chirps constructed in the temporal domain that swept across the frequencies 200-4000 Hz linearly or logarithmically. In Experiment 2, we used three chirps derived from electrophysiological studies and one from a cochlear model, all constructed in the spectral domain. We also introduced the detection of spectrotemporally complex stimuli as a possible intermediary between prior studies that used pure tones (Carlyon & Datta, 1997a, 1997b; Dau & Oxenham, 2001; Kohlrausch & Sander, 1995; Lentz & Leek, 2001; Oxenham & Dau, 2001b, 2004; Smith et al., 1986; Summers, 2000, 2001a; Wojtczak, Beim, & Oxenham, 2014b; Wojtczak & Oxenham, 2009a, 2010) and those that investigated speech recognition (Green & Rosen, 2013; Summers & Leek, 1998). Our results showed that a robust MPE can be elicited using these alternative maskers, though for most signals the MPEs for the chirps were smaller than those observed using Schroeder maskers. For all but the 1 kHz pure tone, the new maskers produced MPEs in the range of ~3-11 dB whereas the Schroeder maskers produced ~8-18 dB. At 1 kHz, three of the four chirps produced a larger MPE than the Schroeder maskers. These results provided partial support for the explanation of MPE offered above; the maskers with downward sweeping instantaneous frequency consistently produced less masking than those with upward sweeping instantaneous frequency. In the first experiment, the repeated upward linear chirp was as effective as the upward sweeping Schroeder in masking the spectrotemporally complex signals, but unexpectedly the downward sweeping linear
chirp was a more effective masker than the downward sweeping Schroeder stimuli. The interpretation of results was complicated by the stimuli construction method, as the repetition of the chirps introduced a break in the waveform at each repetition that may have influenced the amount of masking produced. This possible confound prompted us to use maskers constructed in the spectral domain for the second experiment. In that experiment, the smaller MPEs were clearly caused by the upward sweeping Schroeder stimulus being more effective maskers than the upward sweeping chirps.

The MPEs for the spectrotemporally complex signals used in these experiments were of a similar magnitude to those found for detection of pure tones, and much larger than those observed for speech recognition. One possible interpretation for this difference is that the masking release observed with modulated noise and speech is driven by a release from modulation masking rather than from energetic masking (Stone et al., 2011, 2012; Stone & Moore, 2014b). Upward and downward sweeping instantaneous frequency maskers produce modulations at the same rate in the auditory periphery. It is only the peakiness that is varied by changing the direction of the sweep. As such, we might expect a reduced MPE for speech recognition compared to signal detection, even if the signal is spectrotemporally complex like those used in Chapter 2.

In Chapter 3, we tested, for the first time, whether the MPE can be observed in an electrophysiological measure in humans. Short 4 kHz tone bursts were played with Schroeder maskers while ABRs were recorded from the scalp of subjects. The presence/absence of wave V of the ABR, generated by the lateral lemnisci and inferior colliculi, was used as a threshold marker. We found a statistically significant difference
(MPE) of 9.5 dB for the masked thresholds for our 20 subjects. This was the first time the MPE has been observed in an objective estimate of masked thresholds in the periphery, but the MPE was somewhat smaller than has been observed behaviorally. We were not able to find any research that tested MPE with a signal using the same parameters as those in our study, so it is possible that the reduced MPE was due to these parameters (i.e., ramp time, overall durations of signal and masker). There is evidence that the phenomenon of overshoot differentially affects Schroeder masking, reducing the MPE for signals presented near the onset of the maskers (Summers, 2001a), and that overshoot persists past the onset point of the signals used in our study (McFadden et al., 2010). Overshoot therefore offers another potential explanatory mechanism for the smaller MPE observed in the ABR.

In Chapter 4, we offered psychophysical estimates of phase curvature at three frequencies, using established methods but with greater resolution than other studies. These estimates were consistent with those in published research, a finding that supports the reliability of arriving at these estimates by estimating average thresholds at different phase curvatures and fitting a curve to find a minimum point of masking. The phase curvature that offers the smallest amount of masking can then be interpreted as having the same magnitude but opposite sign as the phase curvature of the basilar membrane at that frequency. The primary purpose of Chapter 4 was to offer a new adaptive procedure to psychophysically estimate the phase curvature, rather than to estimate thresholds. In computer simulations, this procedure produced reasonably consistent results in fewer trials than the conventional procedure. Different parameters and signals were then tested.
behaviorally, and the results from both the conventional and new adaptive technique were in broad agreement with previous estimates of the phase curvature that produces the minimum amount of masking for these complexes (Lentz & Leek, 2001; Oxenham & Dau, 2001b, 2004; Shen & Lentz, 2009). The results from the conventional procedure were closer to previous results also using the conventional procedure than were the adaptive procedure results for our subjects, but the differences were similar to others reported. The adaptive procedure produced phase curvature estimates in 30% fewer trials than the conventional procedure. We concluded that the adaptive procedure has promise in estimating the phase curvature, but might benefit from refinement we proposed in terms of both efficiency and accuracy.

Taken together, these three chapters contribute to our understanding of the phase response of the auditory system in the following ways:

1. They support the hypothesis that a relevant stimulus parameter for producing the MPE is the instantaneous frequency sweep (Chapters 2 and 4).
2. Show that replicating the rate of the instantaneous frequency sweep in non-Schroeder stimuli was not sufficient to reproduce MPEs of the same magnitude (Chapter 2).
3. Show that chirps designed for simultaneous stimulation along the basilar membrane can produce MPEs of the same size or larger than Schroeder maskers for a 1 kHz signal. Those same chirps produce smaller but still robust MPEs for higher frequency and spectrotemporally complex signals (Chapter 2).
4. They provide new estimates of the phase curvature at three frequencies (Chapter 4).
5. Provide evidence that the MPE may be observed in potentials generated in subcortical populations of neurons (Chapter 3).

6. Introduce a novel adaptive method for psychophysically estimating the phase curvature of the auditory system at a particular frequency (Chapter 4).

As with most research, however, the questions that these chapters provoke may be more compelling than the arguments they offer. There are numerous possibilities for future research based on these questions, but the following section will focus on only a few of them.

5.2. Future directions

5.2.1 Point of maximum masking

As noted briefly in Chapters 2 and 4, there is a curious result that has been shown repeatedly in the data but not fully explicated. While research using Schroeder maskers has been focused on the phase curvature that provides the least amount of masking, there is also a phase curvature value that produces a maximum point of masking. These exact values are not reported in the literature, but a survey of the results shows a curious pattern. A $C$ value of $-1$ has consistently produced the most amount of masking of all the $C$ values tested (Gifford et al., 2008; Lentz & Leek, 2001; Oxenham & Dau, 2001, 2004; Oxenham & Ewert, 2005; Rahmat & O’Beirne, in press; Shen & Lentz, 2009; Wojtczak & Oxenham, 2009). This is true regardless of the number of components in the complex or $f_0$ used (see, in particular, Oxenham & Dau, 2001), both of which shift the phase curvature of the harmonic complex. It is also true regardless of the signal frequency. It is possible that a $C$ value of $-1$ is simply the most positive phase curvature value used in
every experiment, and does not capture the point of maximum masking. The point of maximum masking would in that case be at various $C$ values less than -1 depending on the signal frequency and masker $f_0$ and number of components. In Chapter 4, however, the fitted curves for the three frequencies converged very close to -1, and this mirrors what can be seen in published results elsewhere.

Pursuing the phase curvature that produces the maximum amount of masking for different frequencies would provide new information into the auditory system's phase response. The common explanation for larger amounts of masking for negative $C$ values is that, when a stimulus is introduced into the cochlea, the interaction of an upwards sweeping signal with an upwards sweeping frequency glide results in a less modulated output waveform. As Shera (2001) argued, the frequency glide of the basilar membrane may be found by taking the functional inverse of the group delay. If stimuli designed to have the same group delay as the traveling wave are used as maskers, however, they do not produce as much masking as Schroeder maskers with a -1 $C$ value (Chapter 2). If the maximum point of masking does reflect the least modulated basilar membrane response, this would allow new stimulus control for testing the role of modulation in masking release.

### 5.2.2 Individual differences

The phase curvature that produces the minimum or maximum amount of masking is going to vary from individual to individual. The conventional procedure of estimating thresholds at different phase curvatures and fitting a sine curve to the averaged data appears to be questionable when individual data are considered (Chapter 4). It is not clear
if either the proposed adaptive or conventional procedures would allow for estimations accurate and consistent enough to adequately contrast individual differences or track small changes to phase curvature for an individual. Pursuing individual differences could help us understand the different physiological processes that drive the MPE, not just the phase curvature but also the role of the compressive nonlinearity associated with the release from masking and the potential role of efferent pathways. The amount of variability in our individual estimates makes correlations between MPE and other estimates of cochlear nonlinearity (e.g., forward masking or OAEs) and efferent activation (e.g., suppressive effects) quite difficult. In addition, as noted below (Section 5.2.4), the relationship between the phase curvature predicted from the MPE and the electrophysiological chirp stimuli needs further research. The ability to assess an individual's phase curvatures would further this goal.

5.2.3 Masking release

Until quite recently, the most prevalent explanation for the release from masking in speech when using interrupted noise was that listeners were able to "listen in the dips" of the modulated masker. In the past few years, however, researchers have argued much of the observed masking release springs from modulation masking release - noise that appears to have an acoustic amplitude envelope with little low-frequency modulation produces significant low-frequency amplitude modulations in the inner ear. The argument continues that when substantial acoustic modulations are introduced to the noise, they cause a release from modulation masking of the speech signal (Stone et al., 2011, 2012; Stone & Moore, 2014b). This is very similar to the mechanism believed to be driving the
MPE with Schroeder maskers, which also have flat acoustic envelopes and modulated envelopes at the basilar membrane. The different amounts of modulation that can be produced by varying the $C$ value from the Lentz and Leek (2001) variation of the Schroeder formula, and the different modulation rates produced by varying the $f_0$ allow a different way to systematically test whether the release is from energetic or modulation masking.

5.2.4 Electrophysiology and cochlear phase curvature

All of the chapters in this document contribute to our understanding of cochlear physiology in ways that may influence electrophysiological measures, particularly those arising in the auditory periphery. The results of Chapter 3 provided evidence of an MPE from structures more peripheral than any research yet published, but pushing further towards the cochlea (e.g., observing the MPE at the level of the compound action potential) would provide stronger evidence yet. These efforts may also yield data that helps explain extant gaps in our understanding of the physiological processing of complex signals like Schroeder stimuli, such as the dependence of off-frequency forward masking and its temporal build up with Schroeder maskers (Wojtczak et al., 2014b; Wojtczak & Oxenham, 2009a, 2010).

Perhaps most relevant for practical applications, the psychophysical and electrophysiological methods could cross-pollinate more productively. As noted above, the phase curvature of the basilar membrane is likely, at some frequencies at least, the functional inverse of the group delay as a function of frequency (Shera, 2001a). Chirps for electrophysiological research, such as those used in Chapter 2, have been designed
with the intent to compensate for the group delay as a function of frequency in order to simultaneously stimulate the entire length of the basilar membrane. The response to the most common stimulus used in electrophysiological threshold testing, the click, is dominated by the high frequency portions of the cochlea (≥ 4 kHz), so chirps are particularly designed to activate lower frequency portions of the basilar membrane and auditory nerve (e.g., Dau, Wegner, Mellert, & Kollmeier, 2000; Elberling & Don, 2008a, 2008b, 2010; Elberling, Don, Cebulla, & Sturzebecher, 2007; Fobel & Dau, 2004; Shore & Nuttall, 1985). These chirps have been based on models of the traveling wave, latencies from electrophysiological recordings, and estimates of traveling wave delay from otoacoustic emissions. To date there do not appear to have been any chirps derived from psychophysical estimates of the phase curvature, whether by traditional methods (Oxenham & Dau, 2001b) or by adaptive procedures (Chapter 4; Rahmat & O’Beirne, in press). This offers a new method of designing chirps for electrophysiological testing, but also offers a caveat for these stimuli. The psychophysical experiments in humans and physiological experiments in animal models support the notion that upward sweeping chirps produce a longer, less peaky response on the basilar membrane than downward sweeping chirps. This smeared pattern of activation at any particular point along the basilar membrane may result in desynchronized and smaller responses to the upward chirps at each point. Indeed, researchers using chirps do question whether the larger responses to chirp stimuli are the product of enhanced neural synchrony or simply increased activation (Petoe, Bradley, & Wilson, 2010a, 2010b). In addition, chirps that are derived from later components of the evoked electrophysiological response such as
wave V are perhaps more adequately described as compensating for cochlear-neural delay (Elberling, Callo, & Don, 2010), as they include delays introduced by cochlear and retrocochlear processes. Recent psychoacoustical experimentation has provided evidence that there is some mechanism for re-synchronizing signals that have been asynchronously transduced in the cochlea (Wojtczak, Beim, Micheyl, & Oxenham, 2012). As very little is known about this process, the responses used to estimate traveling wave latency at different frequencies may include some part of the process.

In addition to psychophysical experimentation providing new insight to electrophysiology, the reverse can also be a fruitful avenue of exploration. This was the original motivation for the second experiment in Chapter 2, as the chirps designed from models and electrophysiology had phase curvatures that were somewhat similar to those estimated psychophysically. As we showed, these chirps can be as effective as the Schroeder maskers at matching the basilar membrane phase curvature to produce a minimum amount of masking, especially at the lower frequencies tested. They failed to produce equally large amounts of masking as the upward sweeping Schroeder stimuli, however, as we discussed above. There have been many refinements to the chirp stimuli that were not included in this document. Some of these refinements may prove to more accurately represent the non-constant phase curvature of the basilar membrane along a larger frequency range. Iterative refinement using both electrophysiology and psychophysics may prove to be the most effective way to understand the phase response of the auditory system at the periphery.
5.3 References


157


References


Complexes in Several Avian and Mammalian Species. In *Hearing–From Sensory Processing to Perception* (pp. 1–10). Springer.


169


