SPECTRAL AND TEMPORAL INTEGRATION OF BRIEF TONES

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the Graduate

School of The Ohio State University

By

Evelyn M. Hoglund, M.A.

The Ohio State University
2007

Dissertation Committee

Professor Lawrence L. Feth, Advisor

Professor Robert A. Fox

Assistant Professor Christina M. Roup

Associate Professor John G. Neuhoff

Approved by

____

Advisor
Graduate Program in Speech and
Hearing Science

ABSTRACT

Spectral and temporal processing have an extensive history of research for the discrimination and integration of tones. The integration of both dimensions simultaneously, however, has received little attention in psychoacoustics. This dual integration is vital to our daily processing of sounds around us, and has also not been effectively addressed in the ecological acoustics research. For this reason, we still have essentially no understanding of how the auditory system processes sounds that are changing in both frequency and time domains at the same time.

This study was designed to begin the process of measuring the basic detection of signals that vary in both spectral and temporal dimensions. Baseline measures of detection for 10 msec pure tones were taken and the levels adjusted so that all the frequencies could be detected at the same level of attenuation. The thresholds were then obtained for spectral integration of the signals and for temporal integration, so that these results could be compared with prior research. The signals were then varied on both dimensions simultaneously in several ways: with equal spectral and temporal step sizes, different spectral and temporal step sizes, random presentation, and with doubled spectral or temporal information. The data were also analyzed along several differences: spectral step size, temporal step size, frequency range, direction, slope, and predictability.

The spectral and temporal integration conditions showed a good match with the results of prior research, showing that the current procedures and signals could be used to reliably compare to existing results. The spectrotemporal integration conditions showed the threshold for overall detection of the signals to be limited by the ability to integrate spectral information, while the temporal integration was much better. Additionally, very little influence could be seen by most of the differences in signals. Surprisingly, random presentation of frequencies did not negatively influence detection; rather, results indicated that differences in step size between the dimensions were a greater factor in the measured thresholds.

Dedicated to my father, Morris Crouse,

who was always excited about the prospect of my educational advancement.

I'm finally a Doctor.

ACKNOWLEDGMENTS

I wish to thank my advisor, Lawrence Feth, for his patience and persistence through all my delays in the process of reaching this goal. If it weren't for his encouragement, I may not have reached my dream. He has been a mentor to me, helping through difficult times, and allowing me the freedom to proceed as I was able. There were many times that other advisors would have given up, but he never did. In these final stages of the work, he has been especially supportive, pushing me to maintain my momentum, and offering many words, and hours, of encouragement. In addition to his role as mentor, he has been a friend, and provided support through a number of difficult times.

Thanks to the other members of my committee, Robert Fox, John Neuhoff, and Christina Roup, for their contributions to this project. To Dr. Fox for his help on the statistics, Dr. Neuhoff for the ecological perspective and for traveling to Columbus to participate in the meetings, and to Dr. Roup for jumping in at the last minute for the candidacy exam and staying with me through the dissertation. I additionally wish to thank David Huron for his contributions through the candidacy exam. All of these individuals have provided inspiration and support for my work.

I wish to thank my family and my husband's family for their support and encouragement through the many years it has taken to reach this point. They have also never wavered in their belief in me and my ability to reach this goal. Their excitement has carried me through many stresses and self-doubts.

Last but not least, I wish to thank my husband for his support and patience in my MANY long days through these years, and his willingness to move far away from family and friends so that I could return to school. He has kept the homefront going to allow me the freedom from some of the distractions that would have made this even more difficult to achieve.

This research was supported by a grant from the National Institutes of Health, NIH / NIDCD R01 – DC006879.

VITA

October 26, 1962…………….….………..….Born, Great Bend, KS

1984………………………….…….….……..B.A., Speech-Language Pathology, University of Kansas

1988…………………….……...……….M.A., Speech-Language Pathology, University of Kansas

1997 – present…………..…….….……….Graduate Research and Teaching Assistant, The Ohio State University

PUBLICATIONS

Research Publication

1. E. Hoglund, L. Feth, and A. Krishnamurthy, "Dynamic center of gravity effects: Matching virtual frequency glides (A)", J. Acoust. Soc. Am. 120(5), 3126 (2006).

2. N. Iyer, B. Aarden, E. Hoglund, and D. Huron, "Effect of intensity on sensory dissonance (A)", J. Acoust. Soc. Am. 106(4), p. 2208-2209, (1999).

3. A. Sheffield, A. Krishnamurthy, L. Feth, S. Davidson, E. Hoglund, and L. Roth, "Telephone speech enhancement for elderly hearing-impaired listeners (A)" J. Acoust. Soc. Am. 106(4), 2180 (1999).

4. N. Iyer, E. Hoglund, M. Jones, "The perception of three tone patterns by Indian and American listeners", J. Acoust. Soc. Am. 103(5), 2848 (1998).

FIELDS OF STUDY

Major Field: Speech and Hearing Science

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Our understanding of the environment around us is based primarily on our senses. How we process the information that reaches our eyes, nose, skin, tongue, and ears can determine how we respond to events and objects within that environment. This can relate to our safety, growth, socialization, and many more aspects of daily life. Sensory processing, in general, can be described in terms of both evolutionary and ecological significance. How our senses have developed must be closely intertwined with our need for them, as well our responses to the input. A long history of research can be found that has studied how human and animals use all their senses. Hearing is generally used throughout all aspects of daily activities, and the need to understand the mechanisms for auditory perception of sounds in our environment is vital to our understanding of how we interact with that environment. A large part of auditory research has attempted to measure the impact of the physical world on our beings, the perceptual limits of our sensory systems, and our interpretation of these signals. The research based on the physical world includes studies involving how sounds are produced by various materials and in different environments. Study of the perceptual limits has focused on the abilities of all portions of the ear, from the introduction of sound into the outer ear to the neural encoding of those sounds, both in the cochlea and into the brain. In addition to study of human perception, work has included a wide variety of animals, and comparison between species. Finally, much research has been done in order to study how these auditory signals are processed and used in a functional way, be it related to basic information to allow the organism to react appropriately, or in the context of spoken language. This has also included work with animals, as most species use sound for communication.

Prior research has included much work to measure the basic capacity of our auditory system to perceive specific sounds, both simple and complex, produced in laboratory settings. This research allows the specific manipulation of portions of a sound in order to attempt to measure the influence of each aspect on auditory processing. The manipulation of physical attributes allows for precise control of the signals, and very close measures can be made for detection and discrimination of sound components. The signals are approached in a primarily unidimensional fashion, with descriptions and study based on such features as intensity, frequency, and duration. Studies may even analyze such aspects as loudness and pitch. These features are psychological perceptions, but can be related closely with the acoustic components of intensity and frequency. The level of complexity of the signals used, however, seldom exceeds the combination of pure tones, as can be seen in a vast array of psychoacoustic literature.

On the other hand, other research has addressed our responses to real sounds within our environment, primarily focusing on sounds such as music and naturally occurring environmental sounds. This end of the auditory research "continuum" has used existing sounds and environments, studied on a more contextual level. The sounds involved in this work are much more complex, and cannot be considered simply on the basis of such dimensions as loudness/intensity or pitch/frequency. A majority of this work is more descriptive in nature, which limits the possibility of manipulating the sounds, making it very difficult to isolate the components that are instrumental in our perception. Some work has been done to control some of the acoustic factors in research related to these phenomena, but it has been difficult to limit the scope of the scales and features. The perceptions being studied in these studies are also psychological, but cannot be "mapped" onto physical correlates due to the complexity, both of the sounds and the responses. In fact, these perceptions may even follow the continuum from acoustic "gestures", or individual events, to "textures", or a general background (Shafer, 1994).

Corresponding ecological studies have addressed perceptions such as size, hardness, nasality, wetness, friendliness, dissonance, and percussiveness (ecological perception). While parameters related to spectrum, amplitude, or timing may be varied in these studies, the changes in perception measured are mostly inferred. These perceptions may be characterized as *interpretations* of the incoming sound more so than loudness or pitch. These phenomena may imply the nature of the sound source based on the acoustic characteristics in the signal. For example, inanimate objects may tend to have a general vibration with a broad frequency spectrum, such as with motors, resulting in a more noisy sound; or impact sounds, such as hammers or falling objects. Organisms may produce sounds with a system of resonating cavities, such as an airway and an oral cavity, which would result in a more tonal and harmonic sound. But, conversely, organisms may also produce impact sounds while machines may produce tonal "whines". While the acoustics are important in differentiating the sound producer, there is much overlap in source characteristics, requiring that the listener use familiarity to assist in the perception of the variety of sounds. Within the context of an assortment of environmental sounds, many factors can affect their identification, both in terms of accuracy and response time (Ballas, 1993). After the acoustic characteristics are processed, the perception can then be altered on this basis to focus on phenomena specific to the type of source.

These perceptions are also more multidimensional, based on many aspects of the sound that may or may not be separable. As such, they cannot be paired with the acoustical characteristics directly, and measuring the responses becomes complex, apparently including many factors beyond the auditory input. While loudness and pitch are multidimensional to an extent, these appear to be more of a primary level of perception, more closely paired with the acoustical characteristics, and are less vulnerable to interpretation. Ecologically based phenomena are also much more related to learned and cultural impact. The connotations of such perceptions as "wetness" or "friendliness" may, while deriving from the same physical characteristics, have vastly different meaning in different parts of the world. In fact, even in the context of a single culture, many different meanings may be drawn from such perceptions. The categorical boundaries between a "wetness" that may be pleasant or desirable and one that may be threatening can vary between individuals. The perception of "wetness" may vary from a slight moistness to a flood, and the border between pleasant and unpleasant may be much more variable than the border between, for example, loudness that is pleasant or unpleasant. Additionally, an individual's background experiences will likely play a role in the location of this boundary; the pleasantness of "wetness" is likely to be different for a survivor of a hurricane than for a person living in a drought-ridden region. While experience can affect the perceptions of such phenomena as loudness and pitch, as well, these are less likely to be quite as vulnerable to change. If one were to rate loudness on the basis of pleasantness, the attribute of "annoyance" would become relevant, with variability in this rating more likely related to experiential or cultural differences, while the rating of loudness, per se, may be more consistent.

Unfortunately these two extremes in the research have required very different methods and interpretations. In fact, they have ultimately measured different aspects of the same processing mechanism, with much of the laboratory work more generally measuring capability and the ecological studies measuring proclivities (Watson, 2004). While the laboratory research measures specific aspects of sound, the ecological studies are more descriptive. The psychoacoustical perceptions may be viewed in terms of the "details" of processing, while the ecological perceptions are more of a "big picture" of auditory processing.

The two lines of research have their bases in very different origins, too, with very different purposes for the work. The laboratory studies originated with work by engineers with a focus on audition as a channel for communication. As such, the important aspects of sound were those that related to the intelligibility of speech signals. The inclusion of nonspeech sounds was developed primarily in the context of studying the effect of specific dimensions on intelligibility. These dimensions were to be manipulated and analyzed for the purpose of developing communications technology. On the other hand, ecological research grew from an interest in how real sounds impacted our responses, and studies our perceptions rather than the signals. This places the focus on the ear for perception of the world, and communication is secondary. Communicative value is based on our interpretation of existing sounds and our learning of what actions are connected to these sounds.

As a result, the conclusions from these different lines of studies are very difficult, if not impossible, to compare. At this point, it is becoming increasingly important to interpolate between these two lines of study, filling in the middle portion of the "continuum", so that the laboratory research may become more relevant to daily processing, and the ecological studies may become more tenable to manipulation and replication. Some of this attempt to bridge the gap has included speech signals in comparison with environmental sounds, since speech is one of the acoustic signals extensively studied in laboratory experiments. It also is a complex signal, unlike most of the sounds used in psychoacoustic research.

Much of this work has been done in the context of comparing auditory processing of sounds versus speech. Results of much of this work indicate that speech is not processed as a separate type of signal, but that environmental sounds are processed in much the same way as speech. They suggest that perception of sounds differs more on the basis of complexity or frequency range.

The studies comparing speech with other sounds are a beginning in the attempt to bring the two extremes of auditory research together. This study is an effort to add to this beginning, in order to build an experimental structure that can complete the continuum connecting these two major areas. The experiments contained here are psychoacoustic studies using laboratory sounds consisting of specific frequencies in specific time windows, with only one frequency presented at a time, or multiple tones presented in a single time interval. These are produced with consideration of the potential to create sounds more likely to occur in the natural environment. In fact, during the course of the work, some listeners noted the similarity of a few of the signals to naturally occurring sounds, even though the sounds used are very simplified. The goal was to build on the strong tradition of using simple sounds in a laboratory setting to determine the capabilities of the human auditory system, and using increasingly complex signals for measuring the basic limits. By increasing the complexity of the signals in a stepwise fashion, the various acoustic dimensions can be controlled so that clear measures can be obtained. In contrast, the speech signals used in other laboratory studies are more complex, and as such still are not measuring the basic capability of the auditory system. The signals in this study are designed to measure factors related to the integration of sounds along the auditory frequency spectrum, and at the same time determine the influence of the temporal character of basic sounds. The "duality" of time and frequency dictates that when one of these domains is changed, then it by necessity affects the other. When one dimension is varied while the other is held constant, we cannot be sure that we are getting an accurate measure of auditory processing. It is important to quantify the effect the two have on each other. Is one dimension more salient than the other? Do we respond to both in the same way? Or does changing both of them impair our ability to detect those changes? The interaction of these two dimensions is intrinsic to the processing and recognition of the sounds we hear every day, including speech. In our everyday environment, these dimensions cannot be taken separately, and are constantly "working together" to make sounds identifiable. There is always reciprocity in time and frequency, with frequency specificity lost with decreasing duration, and temporal specificity lost in order to gain spectral acuity. The influence of frequency change on threshold for different durations has not successfully been quantified in even the most basic level of signals, nor has the influence of duration change on thresholds of different frequencies. Very few studies have attempted to address the duality of time and frequency in our basic sensory capability, and much work is needed to provide an understanding of our time-frequency responses. We must have an understanding of our sensory capability in order to better understand our responses to the everyday sounds around us. An understanding of our ability to detect signals varied along both dimensions can allow us to expand our study to suprathreshold measures of the same interactions. We already know that the effect of increasing duration is decreased detection thresholds, and that this corresponds to a similar increase in the perception of loudness when the duration is increased with a more intense signal. As we learn about the thresholds for signals varied along both time and frequency, we can study whether a similar correspondence occurs with more intense signals. Then, as we begin to measure responses to simpler forms of potential environmental sounds in the laboratory, we can aspire to gain a better understanding of how we use the sounds surrounding us outside the laboratory.

By working from the background of psychoacoustics, a better understanding of the acoustic world can be formed by understanding our auditory capabilities as they relate to real world sounds. We can form this understanding by learning how to effectively manipulate the different aspects of sounds, and analyzing what effect those manipulations have on our detection, discrimination, and identification of those sounds. This will allow researchers to more effectively control the relevant dimensions of sound, and to continually improve our ability to imitate them. As we understand more of our acoustic world and how we process it, further progress can be made in understanding the effects of hearing loss, and in advancing technology to compensate for hearing loss.

Additionally, further research should be advanced from the ecological acoustics realm, in order to provide improved understanding of the naturally occurring sounds and our perception of those signals. From that perspective, our ability to identify and discriminate different classes of sounds and their sources can be measured with an increasing degree of accuracy and specificity. We can learn more about what components of specific classes of sounds cause us to be able to identify them, rather than just labeling a sound as "sharp" or "brassy" or "percussive", for example. We can determine whether duration, or periodicity, or modulation may cause us to distinguish between sounds within a classification. We can even learn what acoustical characteristics are relevant to the determination of what a class of sounds should be. This can further be used to improve our ability to emulate the many nuances of our acoustic environment. In time, the two progressions could complete the span in our knowledge, resulting in a much improved understanding of how we process all levels of sound.

CHAPTER 2

BACKGROUND RESEARCH

The attempt to understand the nature of human auditory perception has provided a wide variety of directions for study. Two of the important aspects of sound that have been of great interest include the temporal processing of sounds, and processing within the frequency domain, or spectral processing. For both of these perspectives on auditory perception, there is the dilemma related to specificity versus generality. In other words, how do we process both the acuity (resolution), or "small picture", and at the same time the integration (summation) of the "big picture"? Most of the research has been conducted on one of these dimensions while keeping the other constant. But the real world of sounds requires processing of dynamic sounds in both dimensions. Do we process both the "small picture" and the "big picture" for both dimensions at the same time? If so, how do we do it? How can we measure the processing in both dimensions at the same time?

Temporal processing:

In their work on temporal integration, Viemeister and Wakefield (1991) introduced the idea of a "multiple looks" model for temporal integration which allows for an explanation for both good temporal summation and resolution. Prior to this model, integration was studied in terms of either a long integration "window" (hundreds of milliseconds) which provided a good explanation for the summation, or a short integration window (3-5milliseconds), which provided a good explanation for the resolution. Unfortunately, neither type of model could offer a satisfactory explanation for the opposite extreme: long windows could only explain resolution via "leaky" integration, and short windows offered poor prediction for summation.

The multiple looks model proposed that the auditory system actually uses short time constant windows or "looks" at the acoustic input. Thus, one look could detect a short duration signal, consistent with thresholds measured on resolution tasks. As the duration of the signal increases, the auditory system uses an increasing number of these "looks" consecutively, and the information from these windows is accumulated for detection, accounting for improved thresholds for longer duration signals. By measuring detection thresholds for one or two pulse signals, Viemeister and Wakefield were able to show that listeners are able to utilize "intelligent" sampling to detect sounds. In quiet, detection thresholds for 200 microsecond pulse pairs exhibited a 4 dB improvement relative to single pulses at a 1 msec separation, and this threshold increased until the separation was 5 msec, with no further change for longer separations. This suggests integration in a single window at 1 msec, partial integration up to five msec, then independent processing with longer separations. They then introduced a noise, which they either incremented by 6 dB, decremented by 6 dB, or left unchanged for the middle 50 msec of a 100 msec separation. The alteration of the noise would change the total energy in the signal being processed, and based on a long "window" would affect the detection threshold. Increasing the noise level should increase the threshold for the pulse pair, while decreasing the noise level should decrease the threshold relative to the unchanged condition. When this intervening noise was introduced, the detection thresholds for two pulses averaged 2.5 decibels lower than for single pulses, regardless of the level of the noise. The consistent improvement in threshold despite the noise, they argued, supports the use of “intelligent” processing of sounds, in that the looks are stored in memory and selectively used for processing, rather than summation of the entire duration of the signal, as would be assumed for a long integration window.

Further research by Buus (1999) considered the weighting of the pulses based on their temporal location in the signal. There was some inconsistency between the predicted and observed rate of improvement in threshold for signals with durations under 200 msec. If it is assumed that the contribution of individual looks increases with increased time during these shorter durations, then the predicted improvement more closely matches the observed thresholds. He found that the temporal location of any component had only a small effect on its detectability, and therefore, the weights of the pulses are approximately independent, regardless of the noise masker used. Additionally, Buus found that the improvement in detection grows less quickly for single-band (50 Hz wide) and incoherent maskers than for coherent maskers. Incoherent maskers utilize sinusoidal components with random amplitudes and phases, while coherent maskers utilize matched amplitudes and phases. His results support the predictions of the multiple looks theory for the single-band and incoherent masker conditions, but the coherent masker condition is inconsistent with the theory. But then, the result from that condition is inconsistent with any current temporal integration theory.

Later research by Kidd, Mason, and Richards (2003) studied the multiple looks theory in relation to informational masking rather than energetic masking. Generally, informational maskers are those sounds that raise the threshold of a signal by means other than energy, perhaps by interference in attention, or by segregation of different aspects of the acoustic input. The type of informational masker used by Kidd, et al., is a sequence of multitone bursts composed of a small number of pure tones, with components drawn randomly from a wide range of frequencies, but excluding the range around the signal tone in order to minimize the energetic masking (the masking due to noise energy, generally within the critical band containing the signal). They used their multiple bursts different (MBD) masker with 1, 2, 4, and 8 sixty msec bursts and five different interburst intervals (IBI). The IBIs used ranged from 0-400 msec. The MBD masker consisted of a series of bursts with 8 frequency components randomly drawn from an equally logarithmically distributed range from 200 to 5000 Hz. The components never included frequencies between 851-1175 Hz, the critical band around the signal frequency of 1000 Hz. Thus, the only tone that was consistent in all bursts was the 1000 Hz signal. Subjects showed individual differences in the amount of masking produced by the MBD signals, consistent with their previous research on informational masking. However, for all subjects, they found that for increasing number of bursts, the threshold decreased, consistent with the multiple looks theory. The theory would predict that the increased number of looks would result in more sensitive detection. On the other hand, the results of their study showed an increase in threshold for increasing interburst intervals. This is NOT predicted by the multiple looks theory. The theory would predict that increasing the interval would not result in a change in detectability beyond the 5 msec window, as the looks would become independent at that point. (The only IBI used in this study that was less than 5 msec was the 0 msec condition.) They alternatively offer two potential explanations for the results. The first is a loss of information due to memory "noise", that is, some of the signal cues are lost for integration as a result of loss from the memory stores needed for "intelligent sampling". However, the amount of change in threshold cannot be readily explained by loss of memory trace for the interburst intervals used here. The other explanation is related to auditory stream coherence. They consider this more likely due to the "time frame involved", however, caution must be used for either interpretation in the context of their specific experiment. The multiple looks theory also predicts similar performance across the conditions (interburst interval x number of bursts), but their results showed different amounts of change in threshold. The thresholds showed the greatest improvement for 0 IBI with increasing number of bursts and progressively less improvement with increasing IBI, with the least change in the 400 IBI condition. The overall conclusion from this study was that multiple looks were not adequate to explain the changes in threshold observed with the multiple burst different masker used. The coherence of the auditory stream made up of the signal tone imbedded in the masker was the most likely explanation for the thresholds. However, a variety of studies in the area of informational masking have shown consistent differences between this form of masking and energetic masking. The difference in results relative to the multiple looks model may be confounded by basic differences between the types of masker.

In general, the experiments based on multiple looks for temporal integration support the idea of using selected samples of the auditory signal for resolution of short duration signals as well as summation for long duration signals. While it appears there may be more to be learned about the use of these "looks", the evidence thus far supports the idea that the auditory system may be able to use the same mechanism for both extremes of temporal signal processing. Of course, in light of the fact that some of the data has been inconsistent with the predictions of the theory, modifications may be discovered that encompass different signals, or alternative processing strategies may be needed to account for some types of signals, such as informational masking.

Spectral processing:

In terms of the dimension of spectral integration, the measurement has been related to the critical bands in the auditory system. These bands are defined as the width of the individual frequency filters for processing sounds, and are generally regarded as the limit for spectral resolution. More recently these have also been described in terms of equivalent rectangular bandwidth (ERB) units, or the measure of a theoretical rectangular filter representing a critical band. Most psychoacoustic research has related to the processing of sounds within the context of an individual critical band, and thus the ability to resolve acoustic signals, but some has also extended to the processing of sounds across many critical bands.

Early work involving critical bands included a major study by French and Steinberg (1947) using a variety of high pass and low pass filters to measure the articulation index for use with telephone systems. They were able to reliably measure the width of critical bands within the ear, and thus the resolution for speech sounds, resulting in their articulation index. This study of critical bands was not the first accomplished, but demonstrated an effective measure for the purpose of analyzing spectral resolution, or acuity. Other research into spectral integration has been able to use these measures of critical bands as the limit or "look" for spectral integration.

In 1979, Spiegel studied the critical band and spectral integration by using two experimental methods. The first experiment was designed to determine the maximum limit of integration and the second to determine the critical bandwidth for the same listeners. These two measures had not previously been done in the same listeners and with the same procedures. Of importance to the idea of spectral integration, he found that thresholds for his noise signal increased at 1.5 dB per octave above a bandwidth of 50 Hz. This is consistent with the energy detector model, which predicts that for signals beyond the critical band, several bands will be combined to process the signal. By combining the bands, an increase in signal energy is required to compensate for the increased bandwidth (Green, 1958; Green, 1960) as more total energy (signal band plus wideband masker) is passed through the filter. Based on his use of a masker noise with a bandwidth of 100-3000 Hz, he concluded that spectral integration could potentially occur through the entire range of audibility. Of additional interest, he noted a comparison between the flexibility of the auditory system for temporal integration and the conclusion that the same flexibility may exist for spectral integration.

Another approach to spectral integration has been the development of profile analysis (Green, 1983). A full description of this procedure will not be undertaken here, as this is relevant primarily in terms of the basic paradigm. In profile analysis, a two interval forced choice format is used. A standard signal is presented which includes a complex of 3-20 sinusoidal components, logarithmically spaced and equal in level.

Another stimulus, the signal, is also presented, which is identical to the standard stimulus except that the level of one (or more) of its components is incremented. The subject is to listen to the two signals and detect which is the stimulus signal. The important aspect of this paradigm is the simultaneous comparison across the spectrum of a signal in order to make a categorical decision about the presence or absence of a stimulus. This simultaneous comparison is useful for consideration of a spectral "look" for integration. Data generally show an improvement in detection of a change in the spectrum with increasing number of components, consistent with the idea of "multiple looks" as presented in the temporal dimension.

An example of profile analysis in terms of multiple "looks" for spectral integration can be found in work done by Bernstein and Green (1987). They studied the ability to detect an increment in the level of one or more tones in a 21-component complex ranging logarithmically from 200-5000 Hz. Their initial experiment measured the threshold for an increment in a single component with a flat standard complex. Previous research had demonstrated a frequency dependence for detectability of an increment. This experiment showed greatest sensitivity to changes in the middle region of the spectrum, which was consistent with the previous research. Using this as a basis, they then introduced more complex changes in the 21-component stimulus, while keeping the flat spectrum standard the same. They introduced "step-up" and "step-down" spectra, where the frequencies above some critical frequency were increased (or decreased) by a specified amount while below that frequency, the amplitudes were decreased (or increased) by the same amount. The critical frequency was left unaltered. Next they introduced "tilt-up" and "tilt-down" spectra, which consisted of changes that caused the amplitude of successive components to increase (or decrease) linearly with component number, pivoting around the 1000 Hz tone. Thresholds for these signals were measuring the degree of tilt that was just detectable. For their third experiment, they used an alternating spectrum, in which successive components were incremented and decremented. In their final experiment, they modified the signal such that the signal had a "sinusoidally rippled" spectrum. With the strictly limited restrictions on the signals as used in this series of experiments, the thresholds can be predicted on the basis of using only two channels (or "looks") for comparing the standard with the stimuli. Unfortunately, the two-channel predictor is inconsistent with evidence that detection improves with increases in the number of components. If only two channels were used for optimum detection regardless of number of components, a two-component complex should be equally as detectable as a 21-component complex. Evidence does not support this. The paradox that results from this is that the entire spectrum appears to contribute to the estimate of a flat spectrum, but only two channels contribute to the detection of the signal.

Further research into spectral integration has expanded on the use of multiple sine waves, such as those used in profile analysis. Buus, et al., (1986) examined spectral integration by measuring thresholds of 3 pure tones and an 18 tone complex consisting of harmonics separated by one critical band and presented at an equal level. They found that the threshold for the complex was lower than for the individual tones, at a level such that the level of each individual component was 6 dB lower than in isolation. Using these results, they rejected the no summation (single channel) model, since this would not predict any improvement in detectability. They could not determine whether the results were more consistent with either the independent thresholds or multiband energy detector model, so they presented the same stimuli with a randomized presentation order. This randomization resulted in a small increase in thresholds, which is not predicted by the independent channels model, in which the probability of the signal occurring in any particular channel is unrelated to other trials. It *is* consistent with the multiband energy detector model, in which a listener tunes to bands with the highest likelihood of containing a signal, so that the uncertainty increases the number of channels to be monitored. However, they suggested a modification to the model, suggesting that the listeners were able to "turn on" three of the 18 channels, rather than all of the channels, due to the increased probability of a signal in these three, since most of the signals presented were the pure tones. This modification provided for a good match to their observed data. This also could be interpreted as the use of "intelligent" processing of sounds by use of multiple "looks" at intervals yielding maximum information. A later study by Hicks and Buus (2000) showed that spectral integration for short duration signals (4.7 cycles) was more efficient than for long duration signals (150 cycles), and that further, the integration was influenced by the phase relationship of the components in the short duration tones. This short duration integration was actually greater than that predicted by the energy detector model.

Later work by Grose and Hall (1997) found that thresholds for tone complexes improved with increasing number of tones and noise bands. Their first experiment looked at detection of pure tones in 1, 2, 4, and 8 tone complexes with each tone presented in a 20 Hz wide band of noise centered around it. The frequencies were selected on the basis of spacing with one nonoverlapping ERB between each tone.

Detection thresholds improved at a rate of $\sqrt{\mathbf{N}}$, consistent with predictions based on the energy detection model (Green & Swets, 1966). The energy detection model predicts improvement in thresholds for signals with multiple frequency components at a rate of $\sqrt{\mathbf{N}}$. They then studied the thresholds for increments in the narrow band noise signals used in the first experiment. They found a parallel improvement (still $\sqrt{\mathbf{N}}$), but with thresholds 2 dB higher than the pure tones. This difference was noted to be consistent with other research showing that detection of pure tones is superior to discrimination of increments in fluctuating noise bands. They further studied decrements in the noise signal as a comparison to the increment detection. Using the same base signals, they decremented the level, and found the improvement in detection to be less than the $\sqrt{\mathbf{N}}$ observed for the previous experiments. In a final experiment, they carried the principle of decremented signals to its extreme, performing a gap detection experiment with the same narrow band noise complexes. For this experiment, they found that the improvement in detection was greater than the predicted $\sqrt{\mathbf{N}}$. A notable difference between these two decrement experiments relates to possible temporal integration differences. The decrement study used a 200 msec signal, but the gap detection study involved detection of gaps between 20-100 msec.

Bacon et al. (2002) did additional studies in spectral integration for tones in narrow bands of noise. They considered the influence of the type of masking noise used, since prior research had used noise that included inherent modulation even though the basic assumption was that it was unmodulated. Additionally, they critiqued previous work on the basis that the tones were presented at equal physical levels, so they were not necessarily equally detectable. In fact, the authors express doubt that the tones were equally detectable based on some of the thresholds obtained in this study. As a result, they measured pure tone thresholds for each of their component tones, combined them at equally detectable levels, then measured the threshold for the complex resulting from this combination. All of these measures were taken within the context of a 100 Hz narrow band noise masker centered at each frequency. The stimulus frequencies used were not spaced on the basis of ERBs, but rather had four different spatial relationships, all but one of which were harmonic. They were all spaced with at least one ERB between frequencies. Their first experiment measured thresholds for the individual tones and the three tone complexes in coherently modulated and unmodulated noise. They found an improvement in threshold of 11 dB for all frequencies and the three-tone complex when a coherently modulated noise was used as a masker. Additionally, they generally found a decreased threshold for the three-tone complex compared with the individual tones, demonstrating a spectral integration for the frequencies used. The integration for unmodulated noise corresponded well to predictions from the multiband energy detector model, while the modulated noise condition was dependent on the frequency relation of the three tones. One inharmonic combination showed no integration, otherwise the modulated noise resulted in improved thresholds, or improved spectral integration, relative to the unmodulated noise. The inharmonic relation did not show any effect on integration for unmodulated noise, thus casting questions about why this difference should be present. When incoherently modulated noise was used, the results were very similar to the results using unmodulated noise, demonstrating integration as predicted by the multiband energy detection model. However, the amount of integration appeared to decrease with increased spacing between the components, possibly exhibiting a limit in spectral integration. Closer spacing shows better than expected integration in this case, possibly related to influence of comodulation masking release (CMR). CMR is the release from masking caused by the difference in one or more attributes of on-frequency and flanking band envelopes caused by the addition of the signal to the on-frequency masking band. Bacon et al.'s experiments added a signal to all three bands, so there should be less masking release, but there still may be a difference between the bands due to the differences in the presentation levels for each component in most cases. This difference in integration based on CMR cues may also be important to consider in terms of processing of information in everyday sounds.

The research regarding spectral integration has demonstrated the ability for listeners to attend to very wide frequency ranges and select the regions, or channels, which contain the most useful information. This has been demonstrated with both noise and pure tone signals. While the type of noise used as a masker can make a difference in the amount of integration observed, in all cases the presentation of multiple signals in different, *independent*, critical bands, or ERBs, demonstrates an improvement. This use of different channels can be compared with the use of multiple "looks" in the temporal integration domain. The listener is "looking" at the different frequencies in much the same way as the different time windows. Also similar to the multiple looks for temporal integration, most of the data and models reported reflect improvement in detection based on increased opportunities for the "looks".

Spectrotemporal processing:

Since there is evidence to support the idea of multiple looks in both the time and the frequency dimensions of sound integration, the question remains about how these two processes may relate to each other. There has been little work published relating to both dimensions simultaneously. Brief references can be found to the interactions between the two dimensions in the context of some of the work already cited here. Spiegel made reference to the flexibility of the auditory system in processing both the short duration signals for the resolution of sounds and the ability to sum sounds over a much longer duration for improved detection. His data supported the idea that the spectral processing of sounds must be similarly flexible. However, he did not publish any work to relate the two areas quantitatively.

In their study of spectral integration, Grose and Hall noted that the differences between their experiments on level decrement detection and gap detection may have been due in part to the differences in the temporal integration of the signals, since the decremented signals were much longer than the gaps (200 msec vs. 20-100 msec). Again, however, no quantitative analysis of this suggestion is available.

Dai and Green (1993) published a series of experiments comparing thresholds for multitone complexes with varying durations. They used complexes with 3 and 21 tones in a profile analysis paradigm. Their first experiment compared thresholds for the two types of profile spectra with signal durations ranging from 10-1000 msec, but an additional condition was added for 5 msec with the 3-component complex. They found a much lower threshold for the 3-component complex at 10 msec, but the 21-component complex showed more rapid improvement in detectability. To determine if the difference in threshold curves was related to the number of components or the frequency spacing between components, they ran a follow up experiment using another 3-component complex, but with the same frequency spacing as that in the 21-component complex. The results indicated that the improvement in thresholds was related to the frequency spacing of the tones, rather than the total number of components. They invoked the multiple looks model for explanation of the results for the different duration conditions. However, this model would predict improved thresholds at the rate of $\sqrt{\mathbf{N}}$ regardless of duration and frequency spacing. This kind of improvement is only apparent in the data for durations under 100 msec. To account for the steady portion of the thresholds they found in durations between 100-1000 msec, the looks would have to only accumulate information during the first 100 msec of the signal. As a modification to the multiple looks model, Dai and Green suggested a "single look" model, in which the sensitivity is improved as the look is delayed from the onset of the signal. Their model assumes that the filters are initially wide, and become increasingly narrow, becoming more tuned after onset of the signal, with a maximum tuning being reached by 100 msec. This maximum tuning would also account for the steady portion of the thresholds. As further examination of the idea of dynamic filters, they measured the thresholds for additional 3-component complexes with different frequency spacing ratios. The data from this experiment provided further support to the narrowing of filters during the signal. Narrower frequency separations showed greater reduction in threshold with increasing signal duration than wider separations, as the tuning of the filters during longer signals allowed more independent processing of the individual components. These signals with narrower spectral range also required longer durations to reach their lowest thresholds. The complex with the narrowest frequency spacing did not reach the steady portion of the threshold curve until 100 msec signal duration, supporting the suggestion that the filters reach maximum tuning at that point. Wider frequency separations reach the steady portion more quickly as the components become separated into different filters at shorter durations, allowing independent processing of the components. A further experiment using stepped and alternating spectra provided additional evidence that the filters narrow with increasing duration, as the alternating spectrum showed greater improvement in thresholds with increasing duration than the stepped spectrum. The narrowing of the filters would be expected to have greater effect on the narrow differences in level with this alternating signal than the wider differences of the stepped spectrum. Improvement in the stepped signals should not be related to the narrowing filters, but may rather be due to an increasing number of filters to compare, as each becomes narrower, thus providing more components available for comparison, and thus, more looks in the spectral domain.

While a number of researchers have addressed the two dimensions of time and frequency integration, including the multiple looks hypothesis, separately, only a limited amount of work has been done in the area of the integration of both in the same signals. The relationship between the two dimensions has been recognized in the context of the work that has been done, but not much study has been done to quantify this correlation. To that end, this study was designed to address the detection of sounds that varied in both domains. Basic quiet thresholds for these sounds can provide valuable information on which to build further study of integration. This interaction of both temporal and spectral dimensions can begin to be quantified, allowing for the systematic understanding of sounds. Even this basic level of understanding has not been accomplished to date. Only after this is accomplished can we measure the relevant factors in the sounds in the world around us.

Based on the current study, further research can be conducted to expand our understanding of a variety of sounds varying in frequency across time. It is not possible to measure integration in the highly complex environmental and musical sounds around us without a foundation using simple laboratory sounds. This early research must be completed in order to allow an adequate understanding of the auditory system's responses, so that we can then understand what is perceived in those complex sounds.

A number of questions are addressed in light of the paucity of information about spectrotemporal integration. Is the integration of one dimension more salient than the other? Does overall integration remain consistent when both dimensions are varied, or is it decreased, or even increased? Do spectral or temporal limits affect the integration of particular combinations of tones? Does the multiple looks hypothesis apply to the spectral domain, or to spectrotemporal integration? For this study, the hypothesis was that sounds varying in both dimensions would be integrated in the same way as sounds varying in only one of the dimensions, consistent with the multiple looks hypothesis for temporal integration.

These questions are merely the beginning in terms of understanding our integration of these time and frequency varying signals. Additionally, ongoing research along this continuum may help to provide further clarification of the number of looks that may be useful in the integration of sounds, and what limits may exist in the "intelligent" selection of the location for these looks. This is clearly a realm of psychoacoustic research that deserves a closer "look".

CHAPTER 3

METHODS

In order to further the research into the spectrotemporal integration of sounds, it is useful to first replicate some of the previous research in the temporal and spectral integration dimensions individually. In this way, the basic reference points can be established for a new group of listeners and new stimulus production and procedures. Also, the measurement of thresholds for spectral integration and for temporal integration using the same procedures as used for the spectrotemporal conditions can allow for comparison of the data with the prior studies. For this series of experiments, the basic stimulus parameters and experimental procedures were modified from those that most inspired it (Viemeister and Wakefield, 1991, and Grose and Hall, 1997) in order to provide consistency with the additional experiments utilizing the combination of the two dimensions.

Subjects:

Six normal hearing young adult listeners (ages 18-38 years) were included in the present study. An audiological evaluation was done on each subject prior to inclusion in the study, including pure tone thresholds (both air and bone) and otoscopy. Inclusion criteria included pure tone thresholds at 20 dB or less at all audiometric frequencies in the ear used for the study, with air-bone gap of less than 10 dB, and normal otoscopic findings. Five female subjects and one male subject were included, the youngest was 20 years old and the oldest was 37 years old. No prior experience with psychoacoustics research was required for participation; however, introductory training was provided to assure familiarity with the procedures and signals, and to assure stable quiet threshold measurements. The same subjects participated in all experiments. Completion of all experiments required 4 to 5 weeks.

General procedures:

All experiments were conducted in a sound attenuated booth. Signals were generated digitally with Matlab (version 2006b), processed through Digital Audio Labs CardDeluxe sound cards, attenuated with a TDT PA4 programmable attenuator, and presented monaurally over Sennheiser SD580 headphones. All subjects used the right ear, with the exception of subject #6, who used the left ear due to a slight increase in pure tone threshold in the right ear at one frequency. Subjects saw a response screen on a standard computer monitor, which included two rows of lighted squares. The top row of squares included a warning light that indicated that the experimental run was about to begin, a light to indicate when a response could be accepted, and a light to show that the response was accepted. The bottom row included four squares, of which only the middle two were used. These middle squares indicated the first and second experimental intervals. Subjects selected the interval in which they heard the signal with a left mouse click for the first interval and a right mouse click for the second interval. They also had the option of using a left or right click on a computer touch pad due to occasional confusion related to pointing the mouse at the indicator light and using a left click for all responses.

The experiments were conducted using a basic 1 up 3 down 2IFC procedure to target a threshold at the 84% level (Levitt, 1971). This procedure was selected to maximize the reliability of the threshold measure by targeting a point high on the psychometric curve. This allows for a higher accuracy in the threshold estimate. The step size for signal level was initially set at 5 dB, and was reduced to 2 dB after the first reversal in level direction, then reduced again to 1 dB after the next reversal. The starting level for each run was at an attenuation level of 10 dB on the attenuator. Each run consisted of 50 trials, and included between 5 and 10 reversals. The first four reversals were discarded in the calculation of the threshold estimate. The subjects' thresholds were calculated as the average of three separate runs for each condition. As a result, each threshold estimate was based on a total of 150 trials, and each condition represented in the data points in the chapter 4 figures are based on a total of 900 trials. Occasionally, subjects ran more than three runs for a condition, if one or more run yielded an inconsistent result for either the threshold or the standard deviation. The standard deviation was used as a measure of stability only, in addition to a graph that was drawn at the end of the fifty trials, in determining the need for an additional run. These inconsistencies were generally a result of one or more errors early in the run, causing the step size to decrease from 5 dB to 2 dB, causing the presentation level to change too slowly over the course of the trials. An additional cause for inconsistency in the threshold for the run was a wide variation in levels for the retained reversals, causing a particularly wide range around the eventual calculated threshold.

Stimuli:

Eight sine wave tones were used in all the experiments, with frequencies at 356, 494, 663, 870, 1125, 1442, 1838, and 2338 Hz. These frequencies were selected based on work by Grose and Hall (1997), with each tone being in alternating critical bands, to assure that each tone is presented with limited risk of interaction in the cochlea, and to allow for the independent processing of each tone. All tones had a duration of 10 msec, consisting of a 5 msec onset and offset. Due to the brevity of the tones, there was no steady state component. The duration of the stimuli was based on the duration of the tones used in the temporal integration work of Viemeister and Wakefield (1991). Overall duration of signals and silent intervals were specified to insure that all temporal integration be contained within a 150 msec maximum window. The nominal stimulus duration was 160 msec for all the experimental conditions, but the final 10 msec interval was always silent. A 300 msec interstimulus interval was used. All signals were presented in quiet. Figure 3.1 represents the matrix of time and frequency cells used in the experimental signals. Rows in the figure represent the temporal domain, while columns represent the spectral domain. All spectrotemporal conditions were produced by combining tones along a diagonal within these cells. The diagonal to be used was determined by the conditions within the context of each particular experiment. Diagrams demonstrating the specific signal conditions for each experiment are represented in Appendix A.

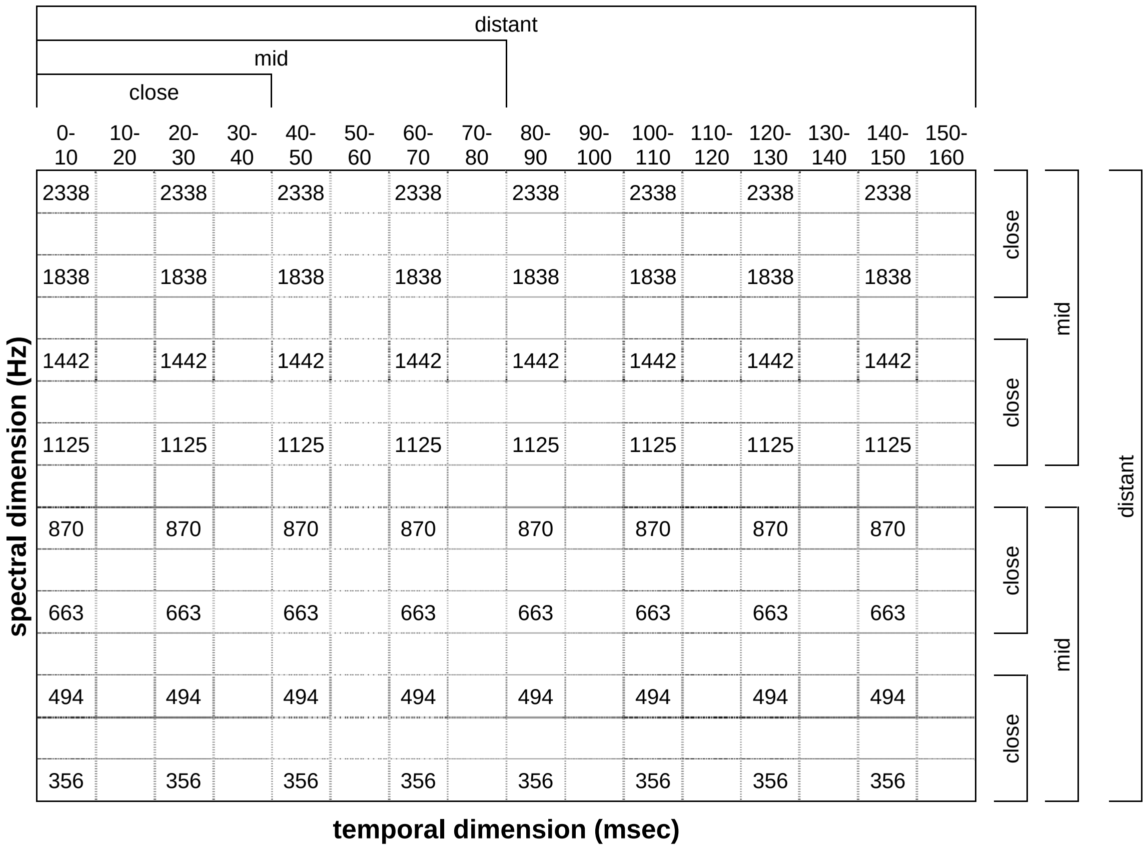


Figure 3.1: Matrix of frequencies presented and time windows. Blank cells in the vertical dimension are ERBs in which no tones were presented. Blank cells in the horizontal dimension are time windows in which no tones were presented. The times indicated above the matrix are the starting and ending time for each cell in milliseconds. Rows represent the temporal domain while columns represent the spectral domain. Brackets along the top and right side reflect the spacing for two tone signals, four tone signals vary from these as indicated in the experimental descriptions. All spectrotemporal integration is represented along a diagonal, with the specific conditions determining what diagonal is involved.

Baseline:

Procedure:

The baseline experiment was conducted to determine the initial thresholds for the individual sine wave tones. After these thresholds were obtained, the amplitudes of each of the tones were adjusted for the purpose of equalizing the detectability of the tones across the frequencies. In this way the signals could be presented for all other conditions at equally detectable levels. This step was included as a result of the critique of Grose and Hall offered by Bacon et al. (2002), in which they asserted that detection could have been based on only the most detectable components of the multiple frequency tones. This level adjustment ensured that the detection of the overall signal was not based solely on the signal(s) with the lowest individual threshold(s), while the tones that were more difficult to detect would reach subthreshold levels. Thresholds were obtained for each frequency at an initial voltage of 0.1V peak value at the sound card output. This voltage was selected to avoid the risk of overdriving the hardware, and introducing distortion. After these thresholds were obtained, the voltages were adjusted by calculating the difference between these thresholds and an attenuation level on the programmable attenuators of 35 dB. The input voltage was then modified by applying this difference to the original voltage (.1) and this result was then used for the subsequent generation of each tone. These conditions were run and adjusted repeatedly, and in a cumulative fashion, by modifying each voltage in the manner just described until all the measured thresholds were within a range of less than 2.5 dB, and the mean was between 34.75 and 35.25 dB of attenuation.

Stimuli:

The eight sine wave tones described above, and illustrated in Figure 3.1, were presented individually in quiet. Time-frequency representations (TFRs) of these signals can be found in Figure A1.

Experiment 1:

Procedure:

The first experiment was conducted to establish a reference for spectral and temporal integration within the context of the frequencies and time window to be used in this study. This reference is an extension of the work by Viemeister and Wakefield in temporal integration, and Grose and Hall in spectral integration, with modifications designed to be consistent with experiments 2 through 5, so that the results can be reasonably compared.

Using the adjusted presentation levels for the individual frequencies obtained in the baseline experiment to assure equal detectability, thresholds were determined for spectral integration for combinations of 2, 4, and 8 signals. Two tone combinations were presented at 1838 and 2338 Hz for the close-high condition, at 870 and 1125 Hz for the close-mid condition, and at 356 and 494 Hz for the close-low condition. These conditions use tones next to each other within the experimental matrix, and are separated by only one silent ERB. The configuration of these signals can be seen in the TFR shown in Figure 3.1. Additionally, the "high", "mid", and "low" terms were used to describe the location within the frequencies included throughout the study. For example, the "high" condition included the frequencies at the upper end of the spectrum used in the study. The tones used for the mid condition were 663 and 1442 Hz, and are separated by 5 silent ERBs. The distant condition consisted of 356 and 2338 Hz, using the most disparate frequencies in the matrix, and separated by 13 ERBs with no signal. The four tone combinations consisted of 663, 870, 1125, and 1442 Hz for the mid condition, again including adjacent tones and having only 1 silent ERBs between components, and 356, 663, 1125, and 1838 Hz for the distant condition, with three silent ERBs between components. The eight tone combination included all the identified frequencies. All temporal and spectral integration conditions were presented randomly, in order to eliminate the risk of order effects.

Temporal integration was measured for 2, 4, and 8 signals, again using the adjusted presentation levels obtained in the baseline experiment. The signals were presented at 356, 1125, and 2338 Hz. These frequencies were selected to represent the high, middle, and low frequency ranges in the study. Again, the configuration of the signals can be seen in the Figure 3.1. Two tone signals were presented with an intervening silent interval of 10 msec for the close condition, 50 msec for the mid condition, and 130 msec for the distant condition. Four tone signals were presented with a 10 msec inter-tone interval in the mid condition and a 30 msec inter-tone interval for the distant condition. The temporal integration for eight tones was measured with 10 msec intervals between signals.

Stimuli:

The eight sine wave tones were presented in quiet in the combinations described above for 2, 4, and 8 tone combinations. The terms "close", "mid", and "distant" were used for both dimensions throughout the study to simplify the descriptions of the conditions. These were selected to reflect the locations of components within the context of the matrix, given above, to facilitate comparisons across the dimensions. While not set to be equivalent, both the spectral and temporal dimensions were designed with a "silent" interval between potential components, to assure independent observations. Visual (TFR) representations of these signals can be found in Figure A2.

Experiment 2:

Procedure:

The second experiment consisted of measuring thresholds for signals differing in both the temporal and spectral domains simultaneously. Signals included complex tones of two, four, and eight sine wave tones presented within the spectrotemporal matrix identified in Figure 3.1. In this experiment the two dimensions were altered in a "symmetric" fashion, that is, they were altered in equally spaced "steps" in both the spectral dimension and the temporal dimension. The term "symmetric" reflects that the dimensional changes were the same, resulting in signals that followed the major and minor diagonals in the TFR. For example, the two tone stimuli were presented in conditions defined as close-close (adjacent frequencies and adjacent time intervals), mid-mid (steps that spanned approximately one half the total possible range in both frequencies and temporal intervals), and distant-distant (endpoint values for frequency and time). The close-close conditions were presented in both the high frequency range, that is, with 1838 and 2338 Hz tones (upper left corner of Figure 3.1), and the low frequency range, with 356 and 494 Hz tones (lower left corner of Figure 3.1). The mid-mid conditions also used the high frequency range (1125 and 2338 Hz) and the low frequency range (356 and 870 Hz). The close conditions presented tones in the first and third time windows, at 0 and 20 msec starting times (left side of Figure 3.1), and the mid conditions presented the tones in the first and fourth time windows, with starting times of

0 and 60 msec. In similar fashion, four tone stimuli were presented in “symmetric” signals, with mid conditions in consecutive frequencies and time windows, and presented in the high and low frequency ranges. The high frequency conditions included 1125, 1442, 1838, and 2338 Hz tones, presented in time windows with starting times of 0, 20, 40, and 60 msec. The four tone distant signals included alternate frequencies starting with the endpoint value. For example, the “up” condition included the 356, 663, 1125, and 1838 Hz tones presented in the time windows with 0, 40, 80, and 120 msec starting times. The eight tone stimuli were all presented in “close” intervals, since this was the only possibility within the spectral and temporal constraints of the experiment. All of the signals were presented in both an “up” and “down” condition, creating both upward and downward glide-like perception. In this way, it was hoped that if any directional differences exist, they might be seen in this context.

Stimuli:

Signals consisted of tones from the same TFR (Figure 3.1), but the signals were presented with different frequencies in different temporal intervals. The signals varied in either a rising or falling slope based on the steps used in each specified condition, with the slope falling on either the major or minor diagonal. A total of 10 two tone conditions, 6 four tone conditions, and 2 eight tone conditions were used. These conditions were described in the procedures above. Visual (TFR) representations of these signals can be found in Figure A3.

Experiment 3:

Procedure:

The third experiment measured the thresholds for signals varying in both the temporal and spectral domains simultaneously, as in experiment 2. In this experiment, the signals were varied in an asymmetric fashion, that is, the step sizes in each dimension were different. For example, the steps in the spectral domain may be distant, while the temporal steps may be mid or close. The alternative difference was also presented, with steps in the spectral domain being smaller than those in the temporal domain. In this way, the perceptual slope for the signals may be either steeper than 1 (as in the first example) or shallower than 1 (as in the second example), depending on the specific condition. Because the signals required a difference in step size between the two dimensions, no eight tone signals could be used, as the constraints of the experimental design required that these signals were always at a "close" step size for both dimensions. As in experiment 2 the signals were presented in both upward and downward slopes.

Stimuli:

Signals consisted of tones from the same frequency and time matrix as used in previous experiments, but followed an asymmetric pattern through the representation. The signals were again presented in both a rising and falling slope, to measure the perceptual thresholds for both directional variations. A total of 8 two tone conditions and 4 four tone conditions were used. These conditions were designed to include close, mid, and distant spectral and temporal steps, and were centered around the middle frequencies in the matrix. Visual (TFR) representations of these signals can be found in Figure A4.

Experiment 4:

Procedure:

The fourth experiment was conducted to measure the thresholds for the complex signals when the component tones were randomly selected. The signals thus had no sequential order between the tones, which precluded comparisons on the basis of frequency slope or direction within the time-frequency matrix. Additionally, the signals were randomly generated without replacement for each trial, so the frequency contour of the signal was different for each presentation through the run, with the frequencies always being different between the time windows. The tones could be at any of the eight frequencies, in any order, with only the temporal and spectral step size constraints limiting the selection. Thus, the two tone close condition consisted of two neighboring frequencies presented in consecutive temporal windows, but the two tones were randomly selected from among the eight possible. Similarly, the two tone mid condition consisted of two frequencies spaced approximately half the distance of the frequency range, and in the first and middle temporal windows, with the specific frequencies randomly selected. The conditions used in this experiment included 3 two tone, 2 four tone combinations, and an eight tone condition. The different temporal distances (close, mid, and distant) were represented in the conditions.

Stimuli:

The signals consisted of the same tones as previous experiments, but in experiment 4 the signals were generated by randomly selecting frequencies without replacement for each trial. In this way, every presentation of the experimental signal was different. All eight frequencies were equally likely to appear in the temporal intervals for the signal, so that within the same run, the high, medium, and low ranges of the frequency spectrum were all equally likely to be represented. Samples of signal complexes can be found in Figure A.5.

Experiment 5:

Procedure:

After completion of experiment 4, and preliminary analysis of the data, a fifth experiment was designed and run with three of the original six subjects. This experiment was designed to follow up on differences seen between the results of experiments 2 and 3. The effect of the slope of the spectrotemporal glide on the threshold was unexpected and suggested a need for immediate further investigation. Additional input was provided by increasing either temporal duration or spectral input. In this way, the difference noted in the "asymmetrical" data could be studied by changing the information presented within the signals. In this experiment, the conditions consisted of eight tone complexes that differed in the use of the frequencies and time windows. Long signals were generated with each frequency presented twice before a change. For example, the downward sloping signal using the high frequency range was generated with two 10 msec tones at 2338 Hz, followed by two tones at 1838 Hz, two tones at 1442 Hz, and two tones at 1125 Hz. The result of this configuration was a set of signals with slopes of less than 1. Short signals were generated with two frequencies in each time window. One of these signals was presented with 2338 and 1838 Hz tones at 0 msec, 1442 and 1125 Hz at 20 msec, 870 and 663 at 40 msec, and 494 and 356 Hz at 60 msec. The signals with this configuration are represented by slopes of greater than 1.

Stimuli:

The same eight frequencies were used within the eight time windows as represented in Figure 3.1, but the durations and complexity were modified as indicated above. The signals were again presented in both a rising and falling slope, to measure the perceptual thresholds for both directional variations. A total of 6 "long" conditions and 2 "short" conditions were used. The long conditions were presented in the high, middle, and low frequency ranges in the matrix. Visual (TFR) representations of these signals can be found in Figure A6.

CHAPTER 4

RESULTS

Quiet threshold measures were obtained for signals consisting of one to eight tones in up to eight time windows. Subjects were able to complete all experiments within approximately 4 weeks, with the greatest amount of time required for the baseline. The experiments were completed consecutively, in sessions of ½ to two hours duration. At times, subjects would complete one experiment and begin the next within the same session, if time allowed. After completion of the first four experiments, three of the subjects were asked to return for completion of experiment 5 as a follow up to observed data.

For all figures, a dashed line is included at the obtained threshold for the individual tones used in the baseline measure. This is included to provide a reference against which to measure the integration for the multitone complexes. With the exception of Figures 4.1, 4.6 and 4.18, the abscissa for all figures represents the number of component tones in the signal. In Figure 4.1, the abscissa identifies the frequencies used in the study, to illustrate the thresholds for the final baseline measures. In Figure 4.6, the abscissa identifies the frequency of the signals used in the temporal integration conditions. In Figure 4.18, the abscissa identifies the frequency range for the long duration tones. For all figures, the ordinate represents the number of dB of attenuation required to reach threshold for the signals. Also, with the exception of Figure 4.1 all figures include a line or lines representing the theoretical improvement in threshold predicted by the energy detection model. This predicted threshold is also consistent with the improvement expected in the context of the multiple looks hypothesis, as this implies the accumulation of energy through the "looks" used by the auditory system.

Throughout this chapter, efforts have been taken to maintain consistency in discussion of all the experiments in order to facilitate comparisons of the results. In all cases, the spectral dimension is described first, with the temporal domain second. Following the first experiment this specific organization can be noted primarily in stimulus description, as all the signals themselves are changing in both dimensions. However, this organization is also maintained in description of signals that are changed differently in the two dimensions (experiments 3 and 5). Additionally, for all experiments, the results are first presented as an overview, averaging all conditions and plotted according to number of tones for the initial discussion. After this discussion, the results are described according to a variety of variables to provide more information about the integration involved in the particular experiment. Finally, at the end of the reviews for each experiment individually, a brief comparison is provided for further analysis across the experiments.

<u>Baseline:</u>

Subjects repeated the single frequency experimental conditions for 356, 494, 663, 870, 1125, 1442, 1838, and 2338 Hz until the measured thresholds were approximately 35 dB of attenuation. This level was selected as the target to allow for enough adjustment from the starting level of 0 dB of attenuation to accommodate the potential range of thresholds. This allowed more readily detected signals to reach the threshold level early enough in the run to allow enough reversals to calculate, but also not cause the less detectable signals to reach the threshold so early in the run as to be discouraging to subjects unfamiliar with the task. A minimum of three repetitions of all the blocks of conditions were required for the thresholds to be adjusted to the 35 dB level, however, the number of repetitions was not limited. Subjects repeated entire blocks until the thresholds were equalized. In some cases, if most frequencies were satisfactorily set to 35 dB, the levels for the remaining few frequencies were adjusted an additional time and a final threshold measure was obtained for just those conditions before continuing to experiment 1. The mean threshold for all frequencies ranged from 34.56 dB to 35.64 dB, with an overall mean of 35.14 dB. A repeated measures ANOVA indicated no significant difference between frequency thresholds, $F(1,7) = .891$, $p = .484$, $\eta^2 = .151$. Variance in the frequency thresholds for individual subjects ranged from 2.18 to 2.41 dB. The thresholds for individual frequencies were averaged across subjects, and then the average of these values was calculated. Variance for these averaged thresholds was 1.08 dB. These results support the assertion that the signals were presented at equally detectable levels, and the experimental signals were detected on the basis of the energy from all presented frequencies.

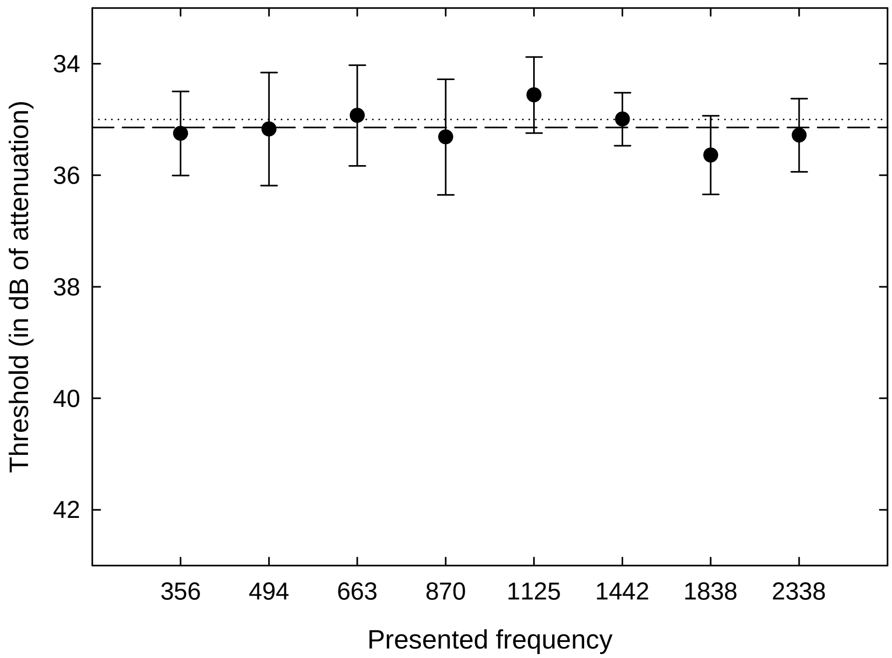


Figure 4.1: Thresholds for baseline frequencies averaged across subjects. Error bars represent 1 standard deviation around the mean. The dotted line is set at 35 dB of attenuation (the target threshold), and the dashed line is set at 35.14 (the obtained average for all eight frequencies). The abscissa is the frequency for which the threshold was obtained, and the ordinate is the dB of attenuation for that threshold.

Differences between specific thresholds were likely related to listener variability and measurement error, and the signals were assumed to be equally detectable with the signals presented at these levels. Just noticeable differences for signals with a duration of 15 ms are between 3 and 4 dB (Oxenham & Buus, 2000), when presented at a level of 65 dB. It was assumed in the current study that the just noticeable differences (JNDs) at absolute threshold would not be substantially different than those at higher presentation levels. For this reason, the variance found within individual subjects' thresholds was believed to be acceptable for the assumption of equal detectability. To check the accuracy of this assumption, the calculated improvement in thresholds based on the obtained data was compared with the improvement predicted by the energy detection model. The model predicts an improvement of 9.03 dB over all eight of the frequencies. When the average thresholds were submitted to the same calculation, the improvement in threshold was 8.9 dB, with the most disparate overall anticipated threshold being 8.76 dB of improvement (for subject 4).

Experiment 1:

In the first experiment, spectral integration and temporal integration thresholds were obtained individually. Results obtained are generally consistent with those of prior studies using slightly different experimental conditions. All figures included in this document for multicomponent signals include a dashed line to represent the obtained average threshold for single tones. This allows for easy comparison of any signals to the baseline measures. With the exceptions noted earlier, all figures show the number of tones along the abscissa and the dB of attenuation along the ordinate.

Spectral integration:

The spectral integration measures show an improvement in threshold between approximately 1 and 2.5 dB for each doubling of the number of tones. The overall improvement from 1 to 2 tones is 1.64 dB, from 2 to 4 tones is 1.02 dB, and from 4 to 8 tones is 2.44 dB. The cumulative improvement in threshold from 1 to 8 tones is 5.17 dB. In Figure 4.2, this can be seen with the filled symbols and solid lines representing the average of the thresholds of all conditions with each total number of tones. This graph includes all conditions plotted by number of tones. A repeated measures ANOVA indicated a significant difference between the observed spectral integration and the prediction from the energy detection model, $F(1, 5) = 26.765$, $p = .004$, $\eta^2 = .843$, over the one to eight tone signals. Additionally, the thresholds with increasing numbers of tones are significantly different, $F(1, 3) = 187.40$, $p = .000$, $\eta^2 = .974$. The interaction between the number of tones and condition was also significant, $F(3, 3) = 15.496$, $p = .001$, $\eta^2 = .756$.

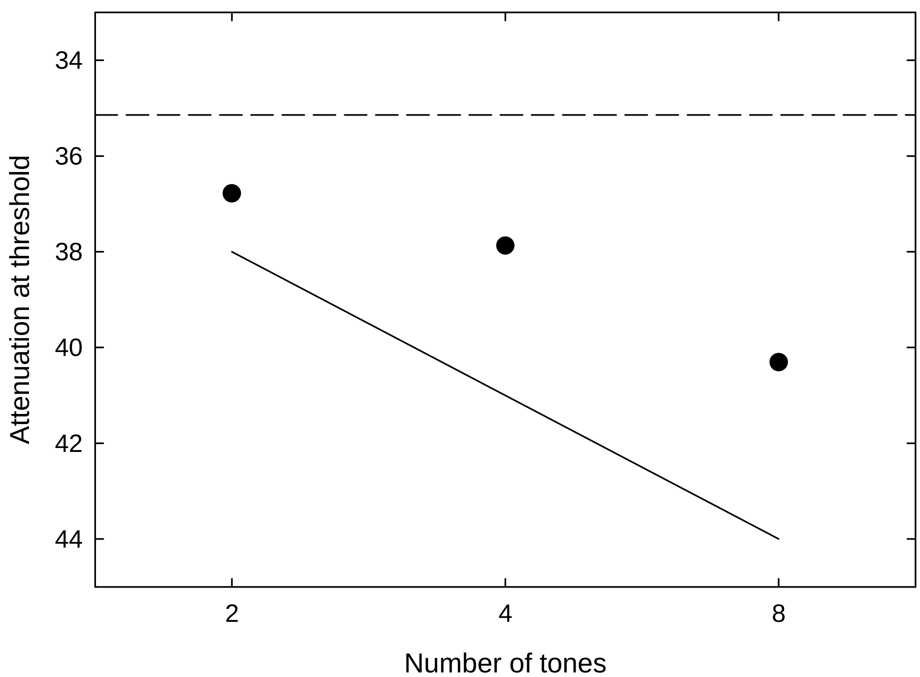


Figure 4.2: Overall spectral integration of multicomponent tones averaged across subjects. Filled circles are the thresholds for 5 two tone conditions, 2 four tone conditions, and 1 eight tone condition. The dashed line is the reference for integration of single tones, and the solid line is the theoretical integration threshold of 3 dB per doubling.

Variability can be seen in the spectral integration based on the frequency range of the close conditions, specifically the high frequency condition (1838 and 2338 Hz tones) showed better integration than the other two frequency ranges. The low frequency condition (356 and 494 Hz) showed the least integration. These frequency ranges were included for conditions with closely spaced tones in order to determine if there may be a difference in integration related to the location in the auditory spectrum, at least for the frequencies used here. These results do show greater integration for two tone signals in the high frequency range, with 37.48 dB of attenuation, and less improvement in quiet threshold seen as the frequency range decreases, with 37.25 dB for the middle frequencies and 36.55 dB for the low frequencies. In fact, the threshold for the two tone signals in the high frequency range was nearly the same as that for the four tone signals with the close frequency separation, which were presented in the mid frequency range (663, 870, 1125, and 1442 Hz). The improvement in threshold from one tone to two tones for the high frequencies is 2.34 dB, for the middle frequencies is 2.11 dB, and for the low frequencies is 1.41 dB. This can be observed in the circles on Figure 4.3. The open circle represents the high frequency range, the filled circle represents the middle range, and the half filled circle represents the low range. The differences in average thresholds are .70 dB and .23 dB between low and mid, and between mid and high ranges, respectively. These differences, however, are not significantly different, $F(2, 5) = .532$, $p = .594$, $\eta^2 = .096$, so it is unlikely that further differentiation of the frequency range would be revealing.

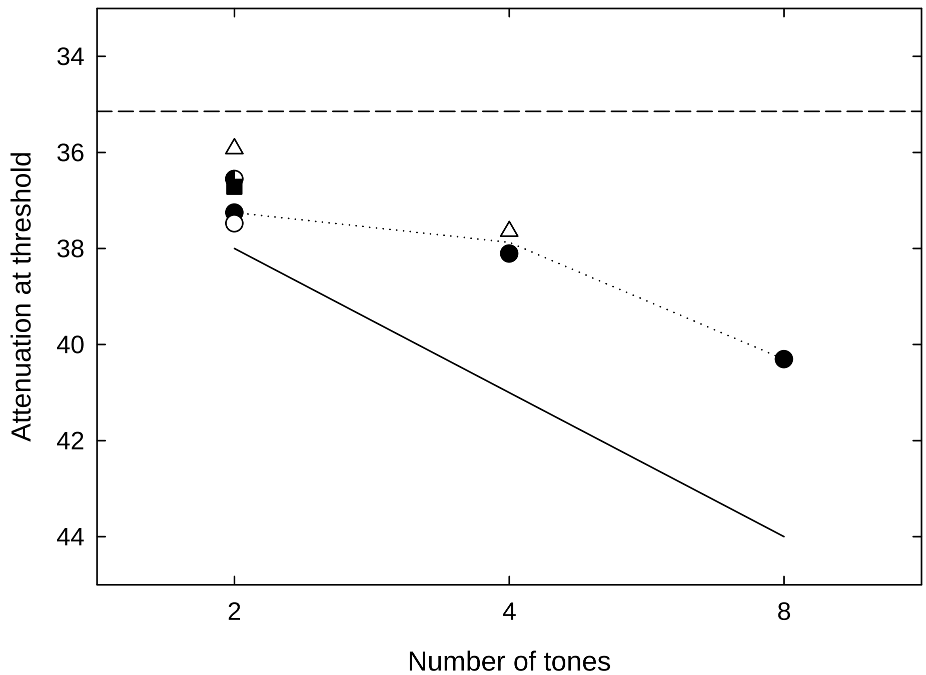


Figure 4.3: Spectral integration by spectral distance and frequency range. Circles are signals with close spectral spacing, open triangles are mid spectral spacing, and the filled square is distant spectral spacing. For the two tone signals, the open circle is the high frequency range, filled circle is the middle frequency range, and the half filled circle is the low frequency range. The dashed line is the reference for integration of single tones, and the solid line is the theoretical integration threshold of 3 dB per doubling. The dotted line shows the average thresholds for all conditions centered on the middle frequency range.

In addition, it can be seen that for the two component tones, the spectral spacing between the components affects the amount of integration. Close spectral spacing resulted in the lowest thresholds, and thus the most integration. When comparing the middle frequencies for two tone signals, the close condition yielded the lowest threshold with 37.25 dB of attenuation, while the middle spectral spacing resulted in 35.92 dB and the distant spectral spacing resulted in 36.72 dB of attenuation. Interestingly, this shows a lower threshold for distant spacing than mid spacing, although all of these differences are quite small. It can also be noted that the difference in thresholds for close and mid separations in the two tone signals is much greater than that for the four component tones. For the four tone signals, the threshold for signals with the close spacing is 38.11 dB of attenuation, while it is 37.64 dB for the middle spectral spacing. The difference between two tone thresholds for close and mid spectral spacing is 1.33 dB, while the comparable difference for four tone thresholds is 0.47 dB. Again, the difference in thresholds based on the spectral spacing was not significant, $F(1, 5) = 4.133$, $p = .098$, $\eta^2 = .453$, although judgment is suspended on whether this may be significant upon further investigation. The interaction between number of tones and spectral spacing was also not significant, $F(1, 5) = .495$, $p = .513$, $\eta^2 = .090$. The spectral gap between tones for the mid conditions of the two tone signals included 5 ERBs with no signal, whereas the spectral gap for the comparable four tone conditions was three ERBs. It is possible that this difference in spacing, while still not significant, could have influenced this small difference in threshold.

Only the two tone signals included conditions in the high and low frequency ranges, so all comparisons across number of components are based only on the middle frequency range. When considering only the middle frequency range signals, the improvement from 1 to 2 tones is 1.57 dB, from 2 to 4 tones is 1.17 dB. The improvement from 4 to 8 tones and the overall improvement remain 2.44 dB and 5.17 dB, respectively. The improvement with increasing number of tones was significant, $F(1, 5) = 24.012$, $p = .000$, $\eta^2 = .828$.

In all conditions, the integration is less than the energy detection model predicts. That model predicts 3 dB of improvement for every doubling of the number of tones, as the total energy in the signal doubles. As can be seen in Figure 4.3, the decrease in thresholds roughly follows the same slope as the theoretical prediction, but the amount of integration is less. When only the signals that are centered in the same frequency range are considered, the improvement is more consistent, although there remains a greater change between 4 and 8 tones than any other tone comparison. This change in threshold is represented in Figure 4.3 as a dotted line.

Temporal integration:

Temporal integration exhibits a more straightforward improvement in threshold with each doubling of number of tones. The thresholds decrease by 2.47 dB when the signal changes from 1 to 2 tones, by 2.39 dB when the signal changes from 2 to 4 tones and by 1.98 dB with a change from 4 to 8 tones. The cumulative improvement in threshold from 1 to 8 tones is 6.84 dB. In Figure 4.4, this can be seen with the filled symbols. Each data point is an average of the thresholds for all two tone conditions (including close, mid, and distant temporal spacing, and 356, 1125, and 2338 Hz tones), all four tone conditions (including close and mid temporal spacing, and 356, 1125, and 2338 Hz tones), and all eight tone conditions (356, 1125, and 2338 Hz tones). Analysis revealed that temporal integration was not significantly different than the energy detector prediction, $F(1, 5) = 3.636$, $p = .115$, $\eta^2 = .421$. The improvement with increasing number of tones was highly significant, $F(1, 3) = 287.417$, $p = .000$, $\eta^2 = .983$. The interaction between number of tones and condition was also significant, $F(3, 3) = 5.059$, $p = .04$, $\eta^2 = .503$, although this was not a strong effect.

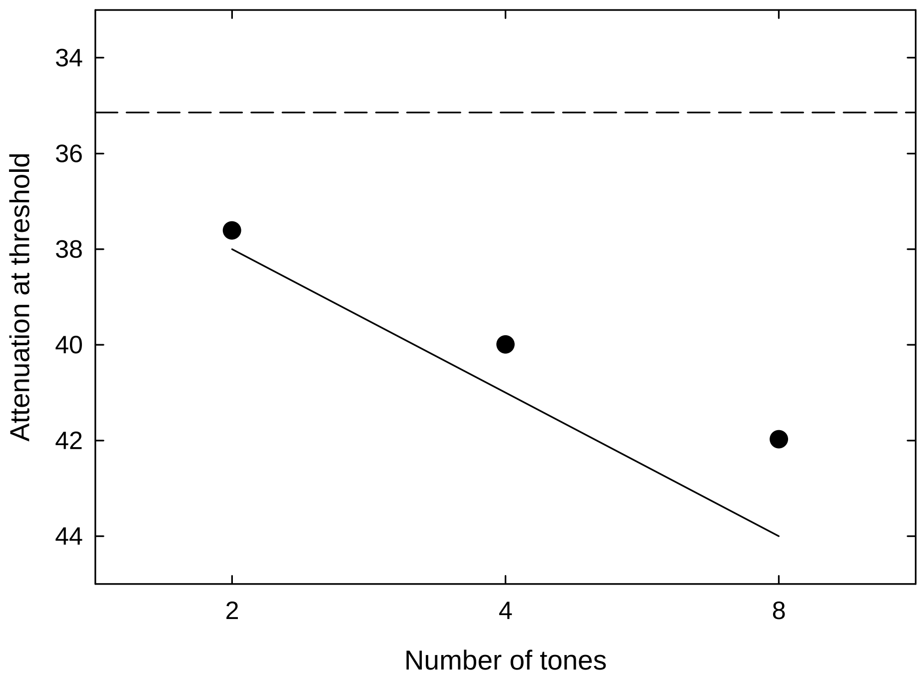


Figure 4.4: Overall temporal integration of multicomponent tones averaged across subjects. Filled circles are the thresholds for 9 two tone conditions, 6 four tone conditions, and 3 eight tone conditions. The dashed line is the reference for integration of single tones, and the solid line is the theoretical integration threshold of 3 dB per doubling.

The temporal integration conditions were run at three frequencies to sample potential differences on the basis of the spectral range. The threshold for all two tone signals presented at 356 Hz is 37.40 dB of attenuation, at 1125 Hz is 37.63 dB, and at 2338 Hz is 37.81 dB. For the four tone signals, the corresponding thresholds are 39.76 dB for the 356 Hz signals, 40.35 dB for the 1125 Hz signals, and 39.89 dB for the 2338 Hz signals. The threshold for the eight tone signals are 42.04 dB, 41.56 dB, and 42.34

dB for 356 Hz, 1125 Hz, and 2338 Hz, respectively. These values can be seen in Figure 4.5 with circles representing 356 Hz, triangles representing 1125 Hz, and squares representing 2338 Hz. The change in detection threshold on the basis of the frequency is not significant, $F(2, 3) = .119$, $p = .826$, $\eta^2 = .023$. A relatively wide range can be seen between thresholds for each number of tones, but no frequency shows a consistently higher or lower threshold. This result confirms that the remaining conditions can readily be centered around any of the included frequencies without concern about the influence of any spectral differences confounding the perception of these signals.

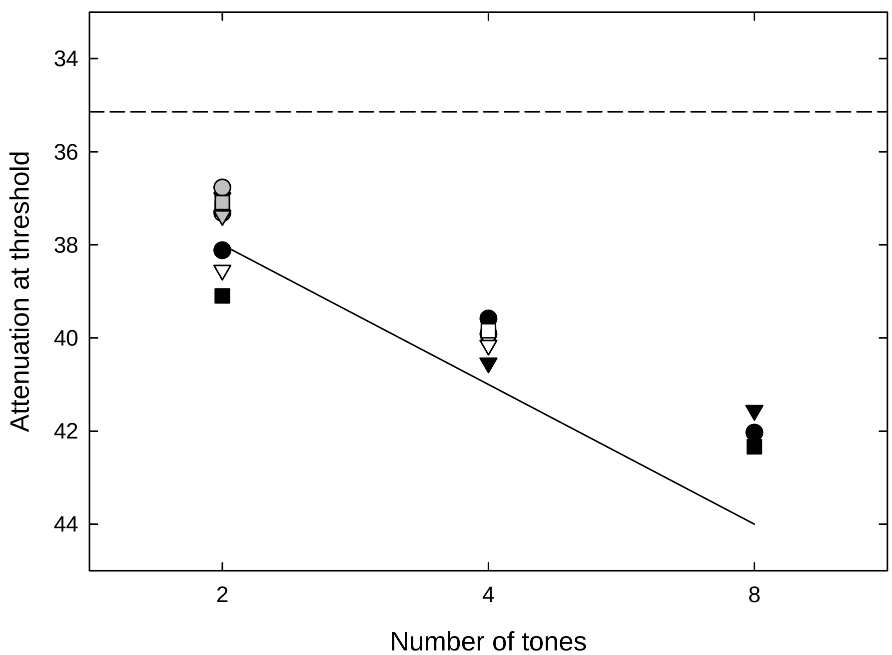


Figure 4.5: Temporal integration by number of component tones. Signals with closely spaced tones are represented by filled symbols, mid spaced tones by open symbols, and distant spaced tones by gray symbols. Circles represent 356 Hz signals, triangles represent 1125 Hz signals, and squares represent 2338 Hz signals. The dashed line is the reference for integration of single tones, and the solid line is the theoretical integration threshold of 3 dB per doubling.

Alternatively, in Figure 4.6, the thresholds are plotted against frequency, with the number of tones as the parameter. In this way, one can see a greater range in the measured thresholds for two tone signals, particularly for the close and mid temporal distance conditions. The distant spacing for the two tone signals and the four tone and eight tone conditions all show a more consistent threshold across all frequencies. The two tone distant, four tone distant, and eight tone conditions all have the commonality of extending across all (or nearly all) eight time windows, while the two tone mid and four tone mid conditions span only the first half of the potential time windows. The two tone close condition uses only the first two time windows, making it the most unique of the signals in the overall duration. This difference, however, does not show a clear relationship to the variability.

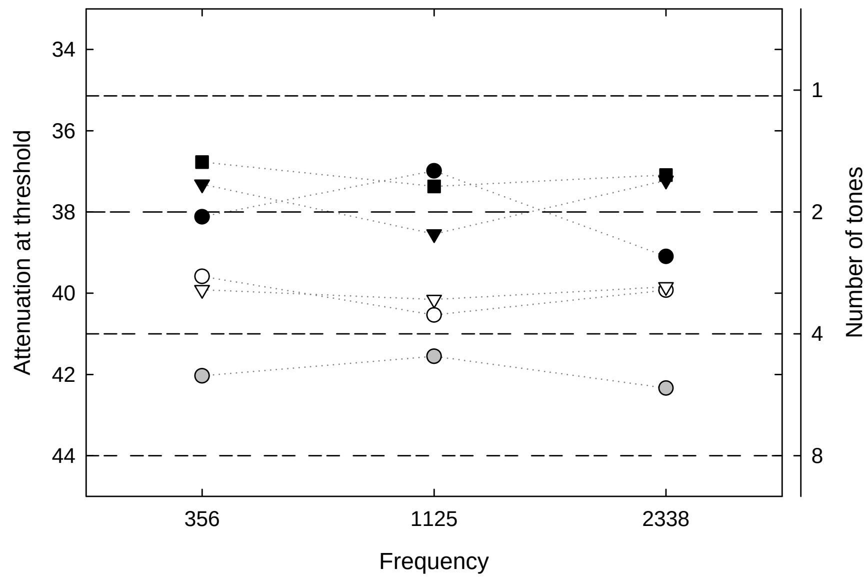


Figure 4.6: Temporal integration by frequency. Two tone signals are represented by filled symbols, four tone signals by open symbols, and eight tone signals by gray symbols. Close temporal spacing is represented with circles, mid spacing with triangles, and distant with squares. Dotted lines show the change in thresholds across frequencies within the same temporal spacing condition. The dashed line is the reference for integration of single tones, and the lines with grouped dashes are the theoretical integration thresholds for 3 dB per doubling, at 2, 4, and 8 tones.

Figure 4.7 replots the temporal integration data shown in Figure 4.5, but with the data averaged across frequencies. Thus, each circle represents closely temporally spaced signals at the given number of components, each triangle represents the mid temporally spaced signals, and the square represents the distant temporally spaced signals. This demonstrates very clearly that, for these short duration signals, when collapsed across frequencies, the thresholds are not affected by the differences in duration of the silence between tones. Thus the integration appears to be the same throughout the overall time window used in all the signals, as would be expected on the basis of prior work in temporal integration. A large amount of prior research has established a temporal window of 150 to 200 msec for integration, including the work supporting the multiple looks hypothesis. The current data support that, within the context of these signals ranging from 30 to 150 msec in total duration, the difference in integration is based on the energy in the signal, that is, the number of components, rather than the duration.

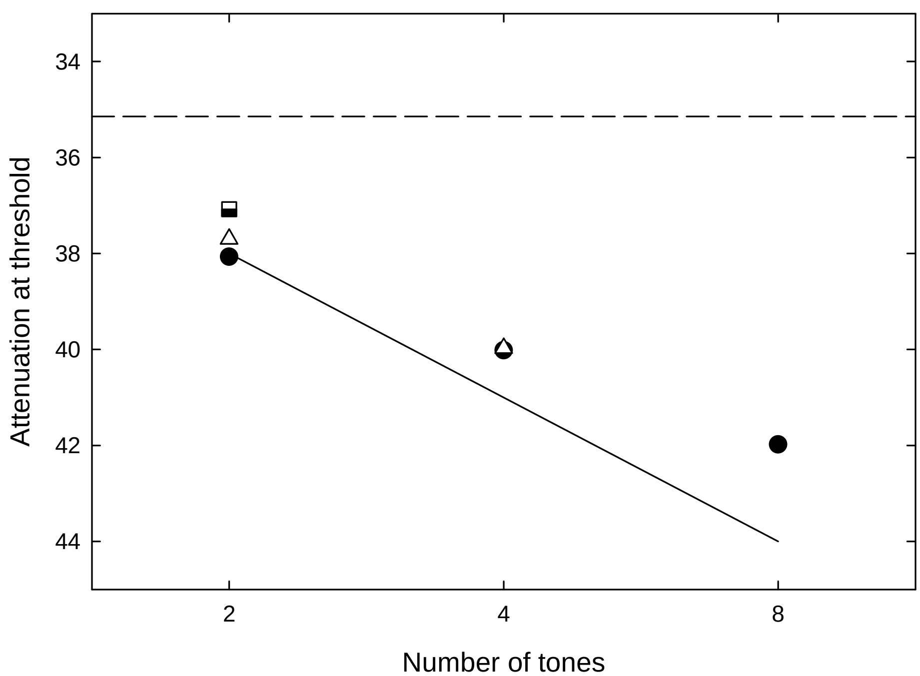


Figure 4.7: Temporal integration by distance. The filled circles represent the closely temporally spaced signals, open triangles represent the mid spaced signals, and the half filled square represents the distant spaced signals. The dashed line is the reference for integration of single tones, and the solid line is the theoretical integration threshold of 3 dB per doubling.

Experiment 2:

For the spectrotemporal signals, the integration appears to be similar to that of the temporal integration alone in the consistency of the improvement, but with higher thresholds across all numbers of tones. The decrease in threshold from one tone to two is 1.30 dB, from two tones to four tones is 1.50 dB, and from four tones to eight is 1.06 dB. The cumulative improvement in threshold from 1 to 8 tones is 3.86 dB. These improvements in thresholds are just over half those exhibited for temporal integration alone, with a greater discrepancy between the two data points as the number of tones increases. Repeated measures ANOVA revealed that the thresholds are significantly different than those of the ideal listener, $F(1, 5) = 70.271$, $p = .000$, $\eta^2 = .934$, as well as for temporal integration alone, $F(1, 5) = 43.184$, $p = .001$, $\eta^2 = .896$. These data are also more consistent than the thresholds for spectral integration alone, in the actual amount of integration overall, however, the spectral integration shows a more evident elbow in the curve at 4 tones. Figure 4.8 shows the overall thresholds for these symmetric signals plotted by number of tones. As before, the dashed line represents the average threshold for single tones. Additional reference lines are added for comparison to the spectral integration data, the temporal integration data, and the theoretical energy detection prediction. This shows the reduced effect of increasing number of components, as can be seen by the shallower slope when compared with the temporal integration. The data also show a greater stability in the integration of increased tonal components when compared with the spectral integration as seen by the consistent slope. Statistical analysis of these thresholds show that the differences between spectrotemporal and spectral integration were not significant, $F(1,5) = 1.667$, $p =.253$, $\eta^2 = .250$.

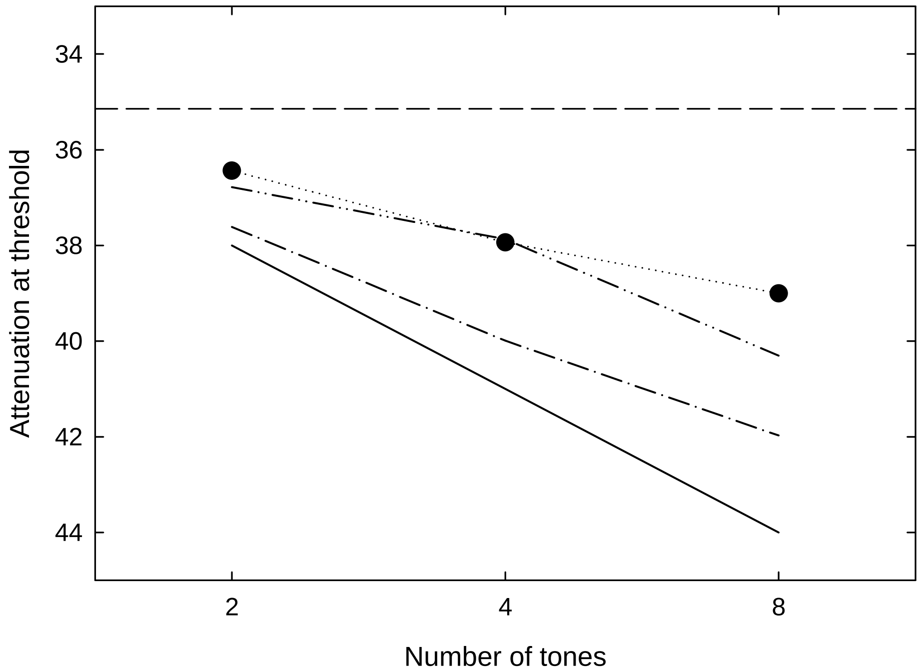


Figure 4.8: Spectrotemporal integration of multicomponent tones with symmetric signals. Filled circles are the thresholds for 10 two tone conditions, 6 four tone conditions, and 2 eight tone conditions. The dashed line is the reference for integration of single tones. For further reference, the dot-dot-dashed line reflects the obtained thresholds for the spectral dimension alone, and the dot-dashed line reflects the thresholds for temporal integration alone. The solid line illustrates the theoretical threshold of 3 dB per doubling.

The data also were considered on the basis of spectral and temporal spacing. Because of the design of this experiment, both dimensions are altered the same amount in each condition. The improvement from one to two tones was nearly identical in all spacing conditions. The two tone signals were detected at 36.25 dB of attenuation for the closely spaced condition, at 36.54 dB for mid spacing, and at 36.60 dB for distant spacing. Four tone signals had a slightly greater difference on the basis of spacing, with threshold at 38.15 dB for close spacing, and 37.53 dB for mid spacing. The differences were not significant in either of these comparisons. Pairwise comparisons between two tone signal conditions yielded t values greater than .060, with $p > .5$ for all comparisons.

A t-test comparing the four tone conditions was not significant at the $p = .1$ level. It is of interest that the detectability of the signals based on the spectral and temporal distance is reversed between the number of tones in the signals. That is, the closely spaced signals are the most detectable for the two tone signals, with the thresholds higher for signals with greater spacing. On the other hand, the four tone signals are more detectable in the conditions with greater spectral and temporal spacing. The improvement from two to four tones, and from four to eight tones for the closely spaced signals was 36.25 dB for two tones to 38.15 dB for four tones, and to 39.00 dB for eight tones. These differences were 1.90 dB and 0.85 dB, respectively. The improvement in threshold from two mid spaced signals to four was 36.54 dB to 37.53 dB, a difference of 0.99 dB. This difference in order, while potentially interesting, must be considered in the context of the nonsignificant difference.

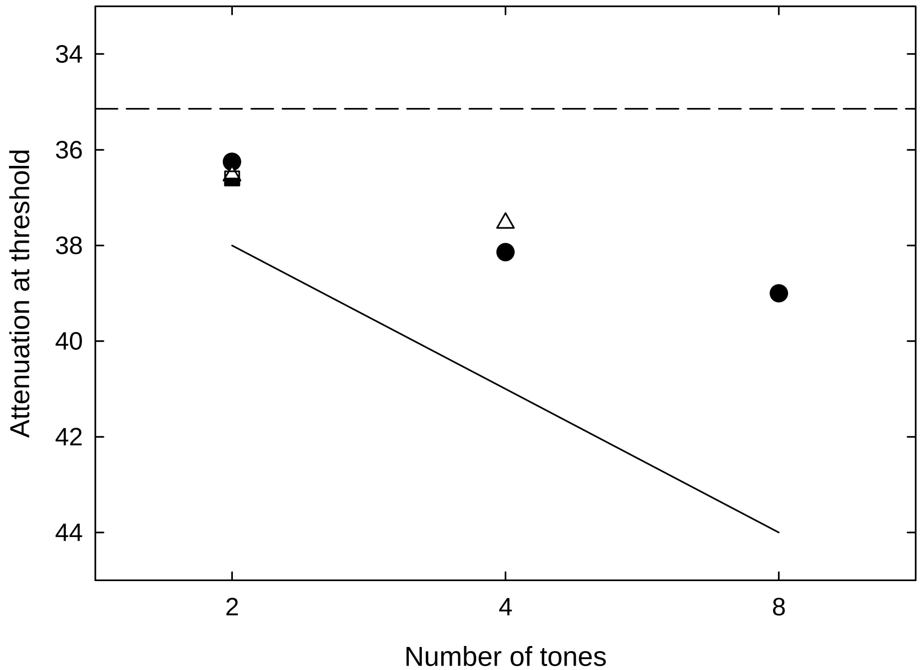


Figure 4.9: Spectrotemporal integration by spectral and temporal distance. Filled circles are closely spaced signals, open triangles are mid spaced signals, and the half filled square is the distant spaced signals. Mid spaced two tone signals and closely spaced signals with two and four tones include 4 conditions in each data point; all other data points represent 2 conditions. The dashed line is the reference for integration of single tones, and the solid line is the theoretical integration threshold of 3 dB per doubling.

Since conditions were included to address integration of signals with both upward and downward slopes, the data were also analyzed on the basis of direction of the change. The improvement in integration was parallel for signals in both directional conditions, with the upward slopes resulting in somewhat lower thresholds. The downward changing signals have thresholds of 36.26 dB, 37.86 dB, and 38.73 dB of attenuation for the 2, 4, and 8 tone signals, respectively. The upward changing signals have thresholds of 36.62

dB, 38.02 dB, and 39.28 dB of attenuation for the same tone conditions. Again, these results were not significant, $F(1, 5) = 1.757$, $p = .242$, $\eta^2 = .260$.

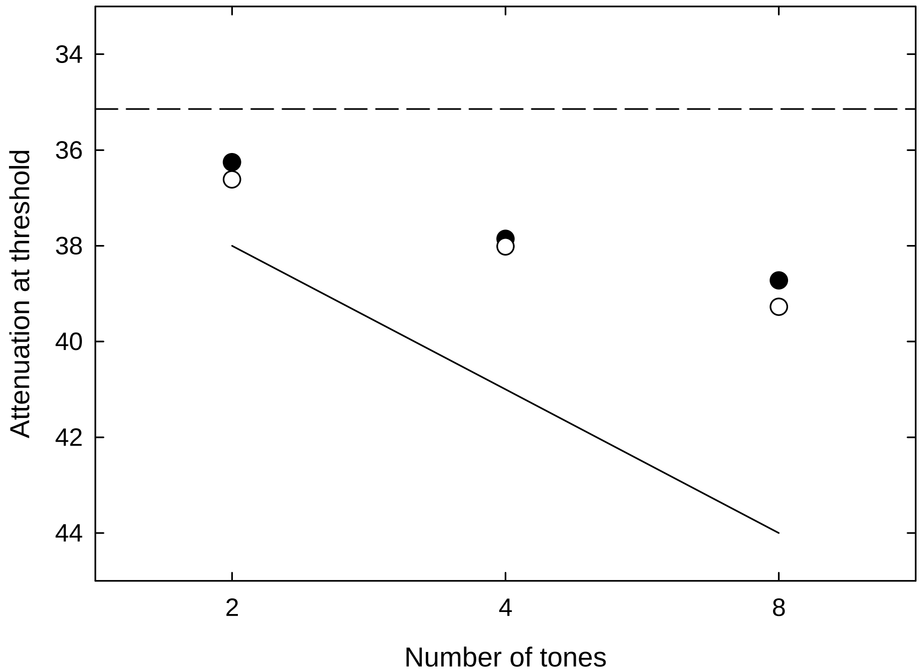


Figure 4.10: Spectrotemporal integration by direction of slope. The filled circles are the downward sloped signals, the open circles are the upward sloped signals. The dashed line is the reference for integration of single tones, and the solid line is the theoretical integration threshold of 3 dB per doubling.

Experiment 3:

For this experiment, the signals changed in frequency with either a faster or slower rate than in experiment 2. They are represented in the TFR with a slope of either greater or less than 1. These conditions were included in order to allow for the possibility of determining a difference in the amount of influence on the threshold due to either the spectral or temporal dimension. If one of these dimensions has a greater influence on the overall integration than the other, then the thresholds in this configuration should show an influence of slope. The integration for these signals considered all together is less than the comparable conditions in experiment 2. The threshold for asymmetric signals improved by 0.63 dB from 1 to 2 tones, and by 1.11 dB from 2 to 4 tones, with an overall improvement from 1 to 4 tones at 1.74 dB. This is different from the improvement in threshold for the symmetric signals, which showed an improvement of 2.80 dB for the same change from 1 to 4 tones. The ANOVA results indicate that the difference is nonsignificant, $F(1, 5) = 2.486$, $p = .176$, $\eta^2 = .332$, however this difference is greater than that for the comparison between symmetric spectrotemporal thresholds and spectral integration thresholds. Figure 4.11 shows a large discrepancy between this integration and the theoretical amount predicted by the energy detection model. Recall that there were no conditions that included 8 tones, since the constraints of the TFR would not allow a difference in the step size for the temporal and spectral domains, while still remaining within the matrix.

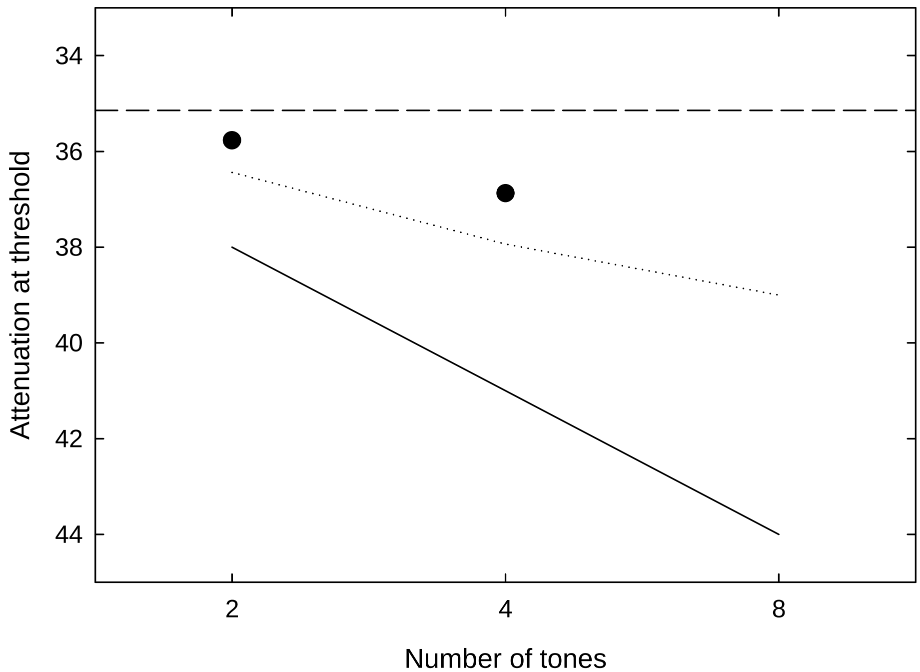


Figure 4.11: Spectrotemporal integration with asymmetric signals. Circles represent thresholds for 8 two tone conditions and 4 four tone conditions. The dashed line is the reference for integration of single tones, the dotted line is the threshold curve for symmetric signals, and the solid line is the theoretical integration threshold of 3 dB per doubling.

In further analysis of these conditions, the signals were first compared based on the spectral distance between the tones. The signals with close spectral spacing were considered together, regardless of the temporal distance or direction of the change in the signal. There were 2 two tone signals and 2 four tone signals in the closely spaced data, 2 two tone and 2 four tone signals in the mid spaced data, and 4 two tone signals in the distant spacing data. The closely spaced signals showed improvement from 35.60 dB for two tones to 37.32 for four tones, a difference of 1.72 dB. The mid spaced signals showed improvement from 35.32 dB to 36.44 dB for the same increase in number of tones, for a difference of 1.12 dB. Again, a t test was used to compare the spacing, with all comparisons nonsignificant, $p > .1$ for all. This change in threshold shows almost no improvement in integration for signals with a spacing of 3 to 5 silent ERBs between them. When the two tone signal conditions are compared to each other on the basis of the spectral distance, the differences appear small, with 35.60 dB for close, 35.32 dB for mid, and 36.09 dB for distant spacing. The total difference between the most and least integration for this comparison is only 0.77 dB, with the distant spaced signals being the most easily detected. The corresponding comparison across two tone signals for experiment 2 is an overall difference of 0.35 dB, showing less effect of spectral spacing when the signals are symmetric, although the difference is quite small in either experimental condition. In both experiments, the thresholds became slightly lower with the distant spectral spacing. The four tone signals included 2 conditions with close spacing and 2 conditions with mid spacing. The difference in integration in this comparison is also small, with a threshold of 37.32 dB for the closely spaced signals and 36.44 for the mid spaced signals, for a total difference of 0.88 dB. In fact, the amount of integration is less with the mid spacing than the close spacing for both the four tone signals and the two tone signals, as well as the four tone signals in experiment 2. The relationship is reversed for the two tone conditions when the signals are symmetric, that is, the threshold for the mid spaced condition is lower. The corresponding comparison with the symmetric four tone signals shows a difference of 0.62 dB, again with the mid spaced signals more easily detected. As reported above, the difference was nonsignificant, with $p > .1$.

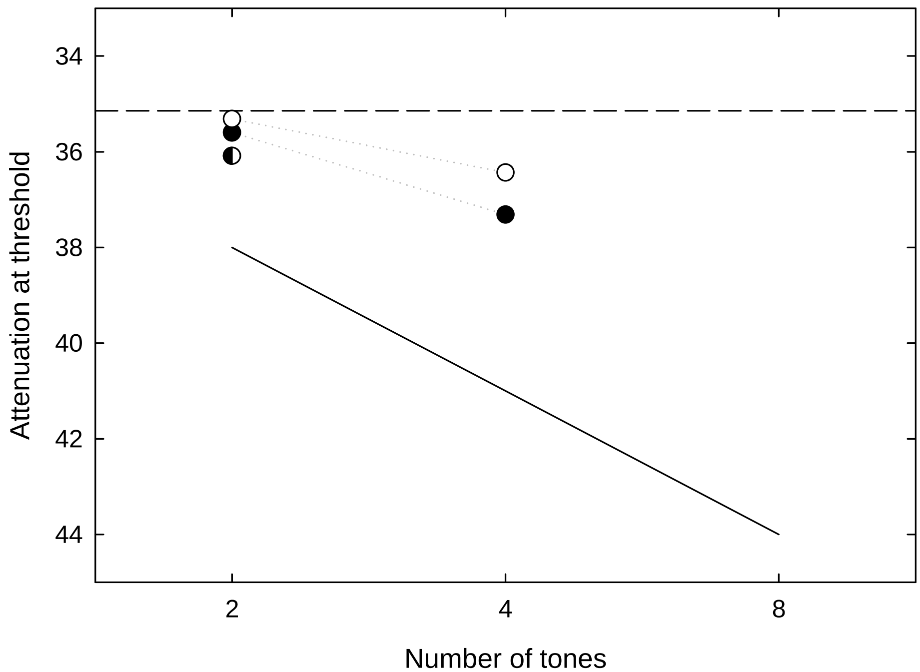


Figure 4.12: Spectrotemporal integration with asymmetric signals by spectral distance. Filled circles are the closely spaced signals, open circles are the mid spaced signals, and the half filled circle is the distant spaced signals. The dashed line is the reference for integration of single tones and the solid line is the theoretical integration threshold of 3 dB per doubling.

Next, the signals were compared on the basis of the temporal distance between the tones. The signals with close temporal spacing were considered together, regardless of the spectral distance or direction of the change in the signal. As for the spectral distance comparison, there were 2 two tone signals and 2 four tone signals in the closely spaced data, 2 two tone and 2 four tone signals in the mid spaced data, and 4 two tone signals in the distant spacing data. The closely spaced signals showed no change between two tones and four tones, with a threshold of 36.44 dB for both. The mid spaced signals showed improvement from 35.73 dB to 37.32 dB for the same increase in number of tones, for a difference of 1.59 dB. The change in threshold for close temporal spacing actually shows no change in integration when given additional tones, but the mid temporal spacing shows improvement for signals with a spacing of 30 to 50 msec of silence between them, more consistent with other measures of integration in the temporal domain. When the two tone signal conditions are compared to each other on the basis of the temporal distance, the differences appear small, with 36.44 dB for close, 35.73 dB for mid, and 35.46 dB for distant spacing. The total difference between the most and least integration for this comparison is only 0.98 dB, with the closely spaced signals being the most easily detected. The corresponding comparison across two tone signals for experiment 2 is an overall difference of 0.35 dB, showing less effect of temporal spacing when the signals are symmetric, as with the spectral distance comparison, although again the difference is quite small in either experimental condition. The four tone signals included 2 conditions with close spacing and 2 conditions with mid spacing. The difference in integration in this comparison is also small, with a threshold of 36.44 dB for the closely spaced signals and 37.32 for the mid spaced signals, for a total difference of 0.88 dB. In fact, the amount of integration is reversed relative to the two tone signals, with the close spacing being more easily detected than the mid spacing when two tones are presented. As with the comparison based on spectral distance, the corresponding comparison with the symmetric signals shows a difference of 0.62 dB, again with the closely spaced signals more easily detected. These comparisons were also not significant with a pairwise comparison t test, $p > .1$ in all cases.

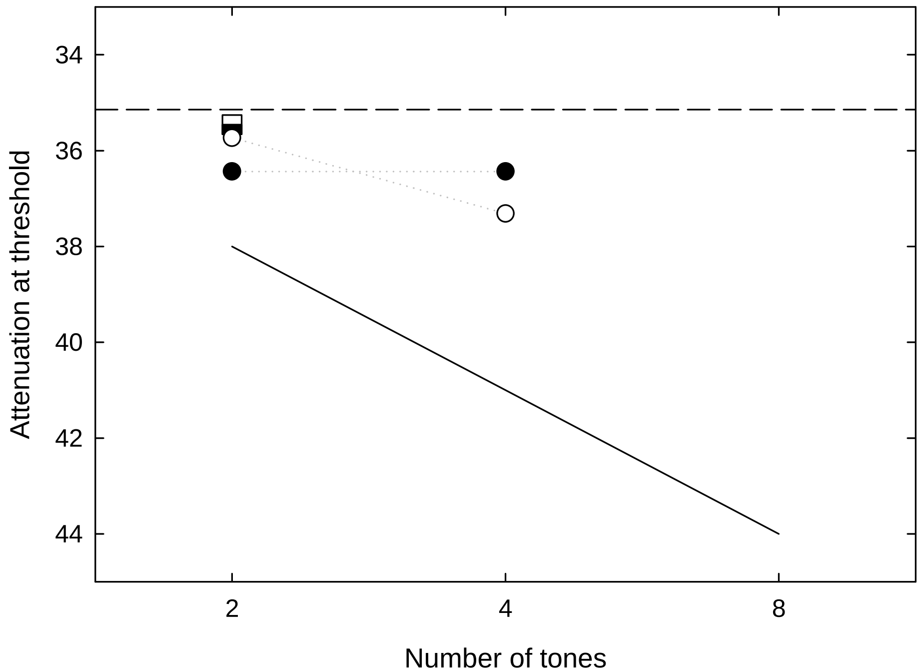


Figure 4.13: Spectrotemporal integration with asymmetric signals by temporal distance. Filled circles are the closely spaced signals, open circles are the mid spaced signals, and the large half filled square is the distant spaced signals. The dashed line is the reference for integration of single tones and the solid line is the theoretical integration threshold of 3 dB per doubling.

As in experiment 2, the data were also analyzed on the basis of direction of the slope. In contrast to the results with symmetric signals, the improvement in integration was different for signals in the two directional conditions, with the upward slopes showing less change in threshold with the addition of two more tones, with a threshold of 35.52 dB for two tones, and 36.31 dB for four tones, and less overall integration. The downward sloping signals showed integration more consistent with other comparisons undertaken here. The two tone thresholds for these are 36.03 dB and the four tone threshold are 37.45 dB, for an improvement of 1.42 dB. The differences were not significant for direction, $F(1, 5) = 1.789$, $p = .239$, $\eta^2 = .263$, however.

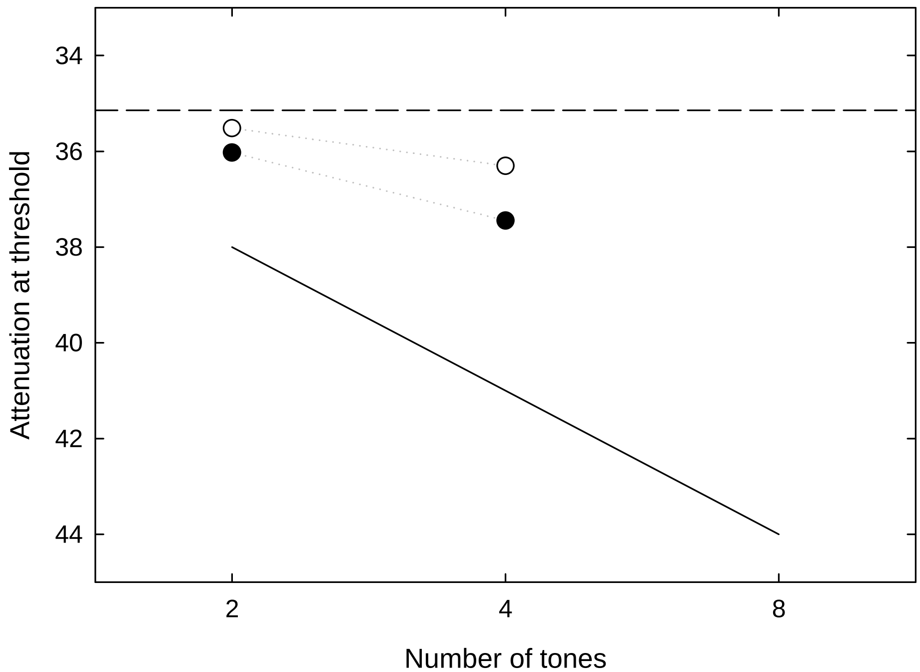


Figure 4.14: Spectrotemporal integration with asymmetric signals by direction. Filled circles are downward sloping signals, open circles are upward sloping signals. The dashed line is the reference for integration of single tones and the solid line is the theoretical integration threshold of 3 dB per doubling.

Additionally, the data were compared with those of experiment 2 on the basis of the slope of the TFR of the signals. The data in experiment 3 were all represented by slopes of greater than or less than 1, while all the experiment 2 signals had a slope equal to 1. The signals with slope of less than 1, that is, the conditions with closer spectral spacing and more distant temporal spacing, showed an improvement in threshold that paralleled that of the signals with a slope equal to 1. The signals with the shallower slope showed a change in threshold of 1.86 dB, with a threshold of 35.46 dB for the two tone signals and 37.32 dB for the four tone signals. The signals with a slope of 1 showed a change of 1.50 dB from two to four tones, with thresholds of 36.44 dB and 37.94 dB respectively. The eight tone signals with a slope of 1 showed further improvement in threshold of 1.06 dB, with a threshold of 39.00 dB. In contrast to these threshold improvements, the signals with slopes greater than 1 showed little improvement with the increase from two to four tones. The threshold for two tone signals was 36.09 dB and the threshold for four tone signals was 36.44 dB, for a change of only 0.35 dB. Surprisingly, the results did not show a significant difference between the slopes, $F(2, 10) = 1.742$, $p = .231$, $\eta^2 = .258$, despite the apparent differences on visual analysis. However, the thresholds between subjects showed greater variability for these conditions than for the previous experiments. Judgment should be suspended until further data may be gathered related to these individual differences.

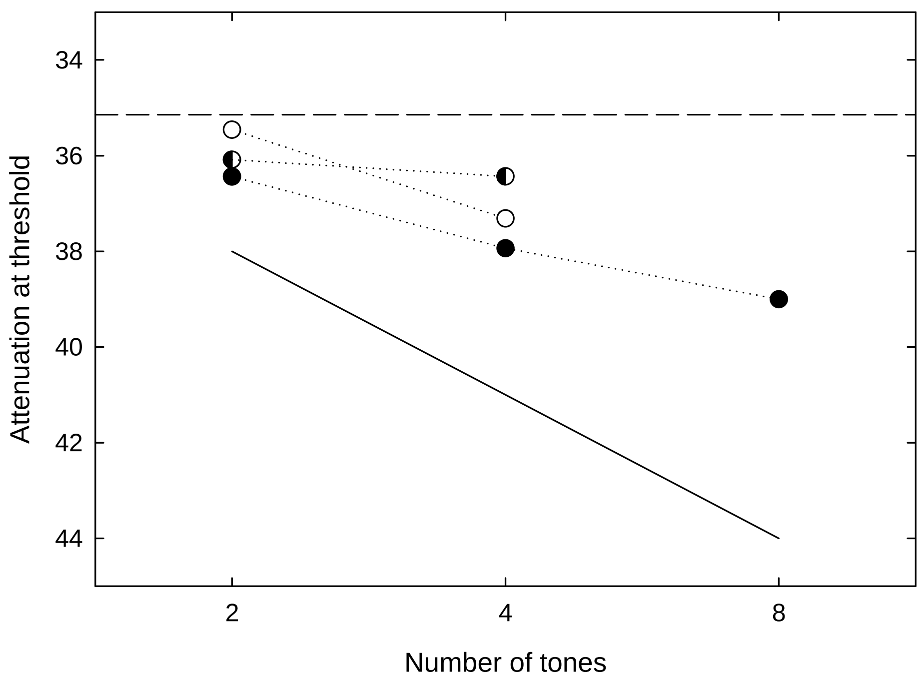


Figure 4.15: Spectrotemporal integration with asymmetric signals by slope. Filled circles are signals with slope = 1 (taken from Experiment 2), open circles are signals with slope <1, half filled circles are signals with slope >1. The dashed line is the reference for integration of single tones and the solid line is the theoretical integration threshold of 3 dB per doubling.

Experiment 4:

The signals for experiment 4 were random presentations of two, four, and eight tone signals, with limits set so that the spectral and temporal distances were matched within the signals. As a result of the established constraints, the data were analyzed on the basis of the number of tones and the spectral and temporal distance only. The overall improvement in threshold showed a very close match to the thresholds for the signals in experiment 2. The thresholds for the random signals were 36.01 dB for the two tone signals, 37.42 dB for the four tone signals, and 38.50 dB for the eight tone signals. These can be compared with 36.44 dB, 37.94 dB, and 39.00 dB for the corresponding signals from the symmetric conditions. Thus, there appears to be little difference in the spectrotemporal integration for signals when there is no ability to predict the frequency range in which the tones will occur. The difference was not significant compared with the symmetric thresholds, $F(1, 5) = 0.834$, $p = .403$, $\eta^2 = .143$, nor with the asymmetric thresholds, $F(1, 5) = 0.319$, $p = ..597$, $\eta^2 = .060$.

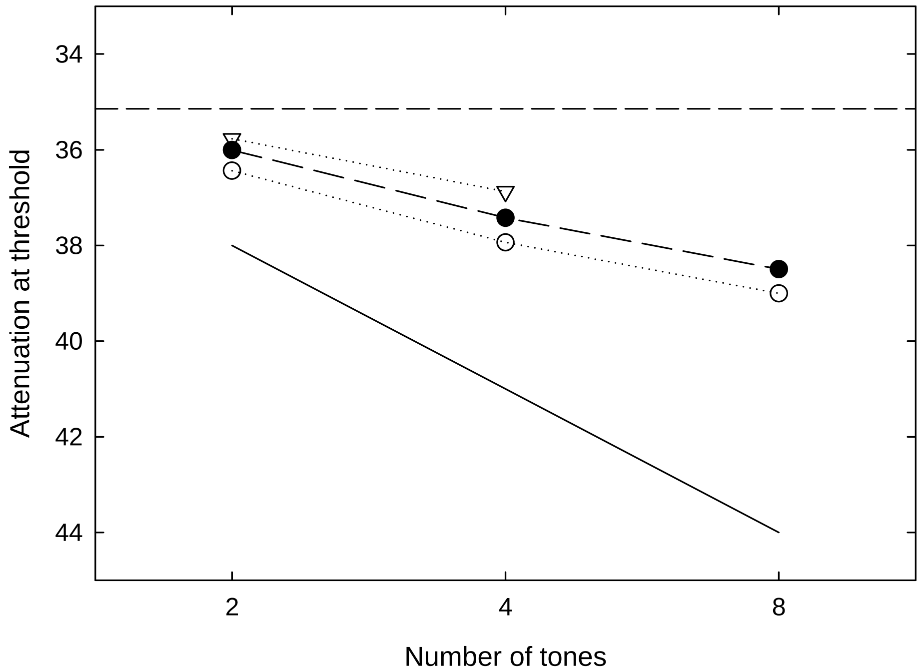


Figure 4.16: Spectrotemporal integration for experiments 2, 3, and 4. Filled circles are the random signals, open circles are the symmetric signals, and open triangles are asymmetric signals. The dashed line is the reference for integration of single tones and the solid line is the theoretical integration threshold of 3 dB per doubling.

When the random signal conditions were analyzed on the basis of the spectral and temporal spacing, a larger difference can be seen in the thresholds for mid spaced signals between two tones and four tones than for any other comparison, with a change from 34.98 dB to 36.88 dB (1.90 dB). In fact, the threshold for two mid spaced random signals is somewhat poorer than for single tones. The thresholds for the closely spaced random signals show an improvement more consistent with other conditions. The spread seen between the signals with different spacing is greater with the random signals than for any other conditions. The comparison between spacing shows the only significant difference between the two tone conditions with close and mid spacing, p = .05. All other comparisons were nonsignificant.

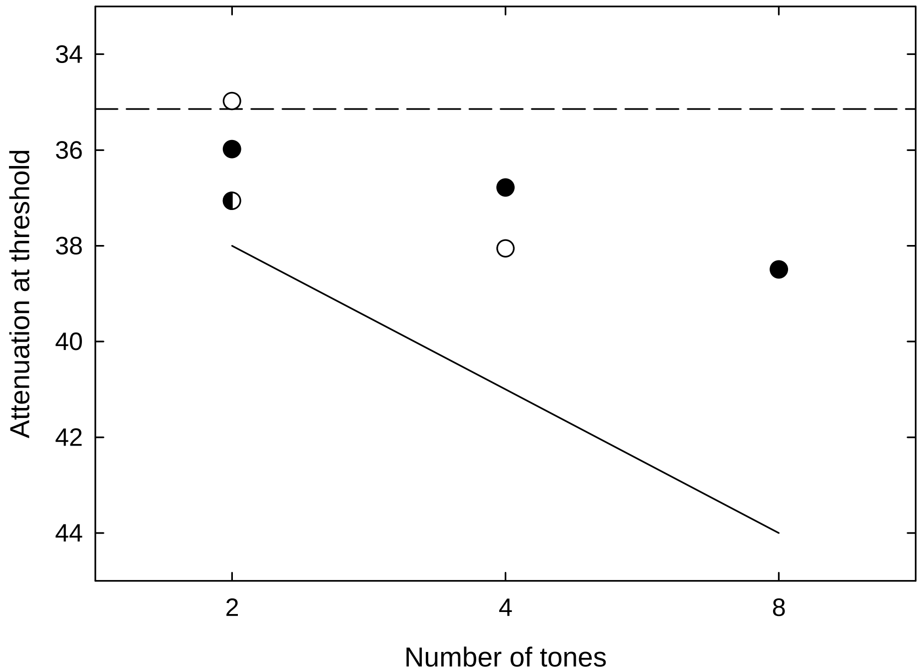


Figure 4.17: Spectrotemporal integration of random signals by spectral and temporal spacing. Filled circles are the closely spaced signals, open circles are the mid spaced signals, and the half filled circle is the distant spaced signals. The dashed line is the reference for integration of single tones and the solid line is the theoretical integration threshold of 3 dB per doubling.

Experiment 5:

This experiment was used to provide further information related to the differences between the slopes of the spectrotemporal signals. The integration of the signals presented in experiment 3 appeared to be significantly different than other conditions during data collection, inspiring further investigation. Upon statistical analysis, the difference did not reach the level of significance, however, the observed differences remained of interest. The results of the signals with the tones doubled either in frequency within time windows, or in number of time windows at a given frequency were analyzed for further information. Data were gathered at three frequency ranges, represented on the abscissa in Figure 4.18. The average thresholds for the mid and high frequency ranges were consistent across these conditions for the long duration tones (doubled 10 msec components), with average thresholds of 44.06 dB for the low frequency range, 42.64 dB for the mid range, and 42.99 dB for the high range. The signals with the complex spectral information (doubled tones in four time windows) spanned the entire frequency range, so are represented in the mid range in Figure 4.18. The threshold for these signals is slightly higher than those for the conditions with doubled temporal information, with 42.06 dB. Pairwise comparisons reveal no significant difference between the thresholds for the different frequency ranges, $p > .05$, although the difference between the high frequency range and low frequency range approached the level of significance, at $p = .081$. The difference between the signals with doubled temporal information and those with doubled spectral information were not significantly different from each other, $F(1, 2) = 2.497$, $p = .255$, $\eta^2 = .555$.

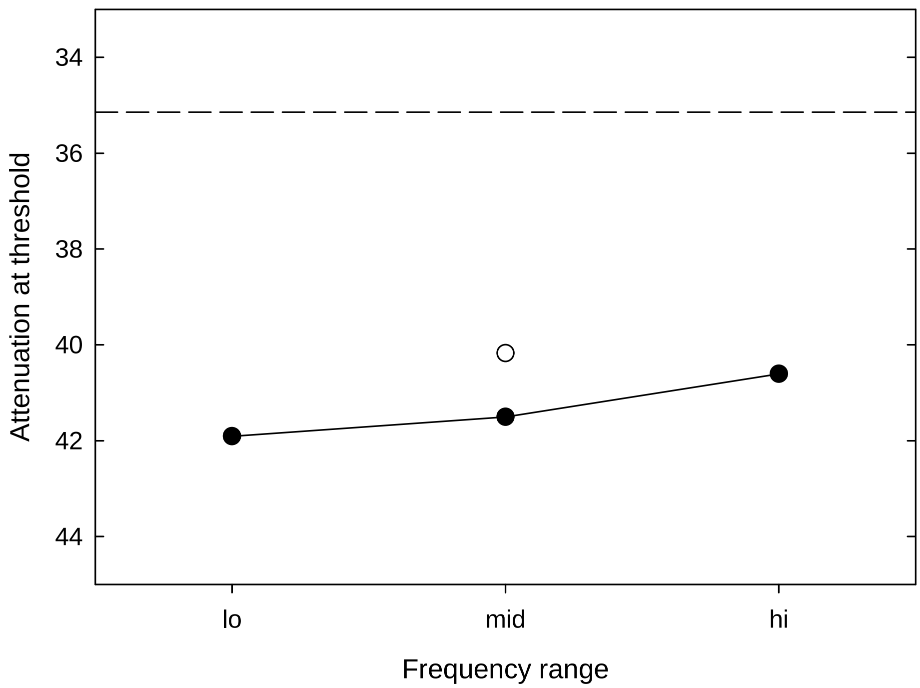


Figure 4.18: Spectrotemporal integration for signals with increased spectral or temporal information. The filled circles are the signals with doubled temporal input, the open circle is signals with doubled spectral input. The dashed line is the reference for integration of single tones.

These signals were further analyzed by direction of change and by slope. The data for direction were plotted in Figure 4.19 with prior data from experiments 2 and 3 for comparison, with those prior data shown in gray. The circles represent asymmetric signals, thus the two and four tone data are taken from experiment 3, while the eight tone data are from this experiment. The triangles are the symmetric signals, and are taken from experiment 2. Filled symbols are downward moving signals, and the extended signals of this experiment resulted in an average 41.87 dB threshold. The open symbols are upward moving signals, and these results showed an average 41.96 dB threshold.

These are nearly identical, with a difference of only 0.09 dB and appear more consistent with the symmetric signals from experiment 2 than the earlier asymmetric signals. If a line were to be fit between these points and the experiment 2 data points, it would be parallel with the predicted thresholds based on energy detection. The thresholds for direction are not significantly different than those for the previous asymmetric conditions, $F(1, 2) = 5.553$, $p = .143$, $\eta^2 = .735$, but again, thus these thresholds appear to follow the same curve as those initial asymmetric signals.

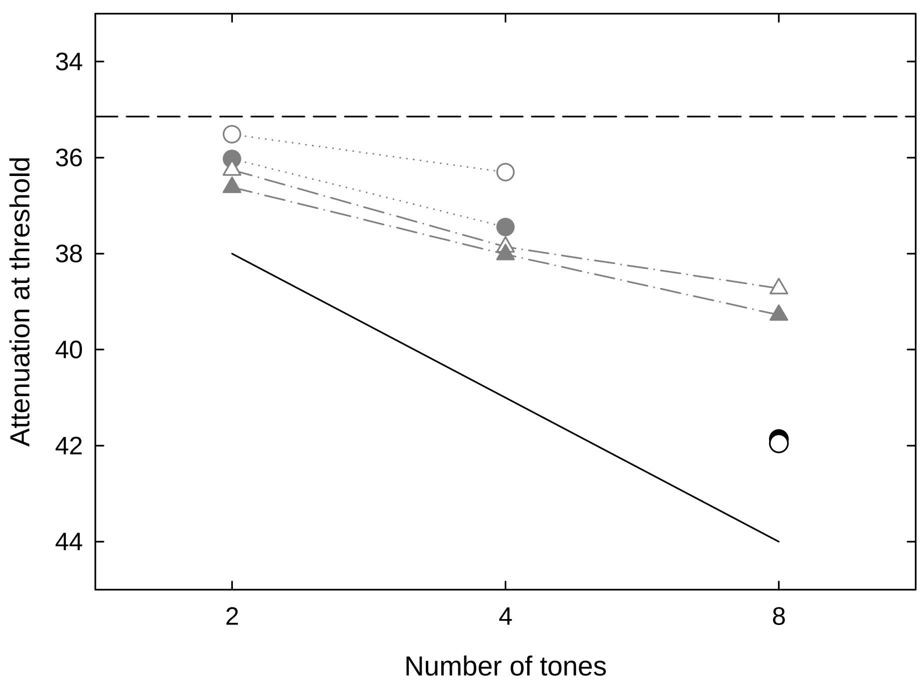


Figure 4.19: Spectrotemporal integration with symmetric and asymmetric signals by direction. Signals from experiment 5 were added for eight tone conditions. Filled circles and triangles are downward sloping signals, open circles and triangles are upward sloping signals. Circles are asymmetric signals and triangles are symmetric signals. Data from experiments 2 and 3 are shown in gray. The dashed line is the reference for integration of single tones and the solid line is the theoretical integration threshold of 3 dB per doubling.

When analyzed by slope, similar comparisons can be seen. The integration for the more complex signals of experiment 5 appear to be more in line with the earlier symmetric signals than the asymmetric signals. The relationship between the slopes remain comparable within the asymmetric signals between the four tone and eight tone conditions here. For these numbers of tones, the slopes of less than 1 showed more integration than those with slopes greater than 1. For the eight tone signals the threshold for less than 1 were 43.23 dB, while for greater than 1 they were 42.06 dB, for a difference of 1.17 dB. These differences were not significant, $F(1, 2) = 5.019$, $p = .154$, $\eta^2 = .715$.

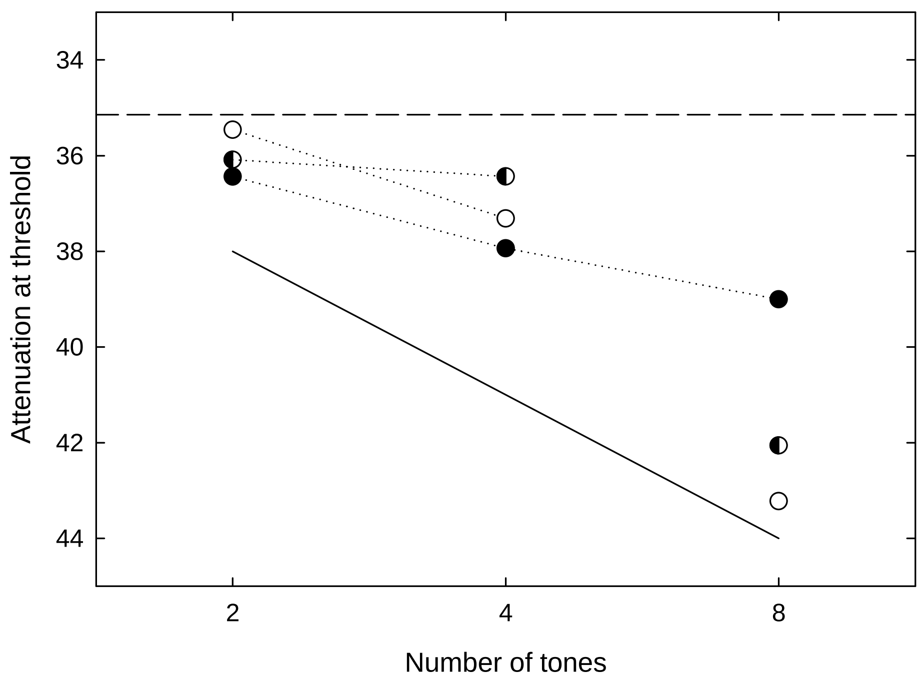


Figure 4.20: Spectrotemporal integration with asymmetric signals by slope. Signals from experiment 5 added for eight tone conditions. Filled circles are signals with slope = 1 (taken from Experiment 2), open circles are signals with slope <1, half filled circles are signals with slope >1. The dashed line is the reference for integration of single tones and the solid line is the theoretical integration threshold of 3 dB per doubling.

General results:

Integration was first measured for the spectral domain and the temporal domain individually. A significant difference can be seen between the two measures, with the spectral integration being significantly less than the theoretical ideal listener, while temporal integration was not. Spectrotemporal integration was not significantly different than the spectral dimension alone. The signals throughout the study were analyzed on the basis of a number of different features. Features considered were spectral distance between tones, temporal distance between tones, direction of slope representing the signal change, and steepness of that slope. Spectral and temporal differences were considered in terms of having same or different step sizes within the signals. The signals were even randomized on the basis of the matched distances in the two dimensions. Across these different features, the only condition that demonstrated a distinct difference was the combination of different spectral and temporal spacing, forming asymmetric signals. While still statistically nonsignificant, the results were near the level of significance, and appeared to be different on the basis of visual inspection. This discrepancy from the other conditions was overcome by providing additional information within the tones, either by doubled frequencies, or by doubled time windows at a single tone.

CHAPTER 5

DISCUSSION

The study of auditory processing of individual sounds in the laboratory has addressed many aspects of sounds in both the spectral and the temporal domains in many contexts. However, they have held one dimension constant, while manipulating the other. This has allowed the understanding of the auditory system that we have at this time. The number of studies that have addressed both dimensions simultaneously, though, is very limited. At this time in auditory research, the study of both spectral and temporal integration in the same signals is due. The ability to manipulate fine details of both at the same time is available, and the foundations have been laid.

The study of both dimensions is needed to help determine whether listeners are able to integrate one dimension more effectively than the other when both are challenged, if both are handled equally, whether integration is better or worse in both dimensions when compared with performance when one dimension is kept stable, or even if integration is completed only in one dimension at a time.

In the current study, the study of the individual dimensions laid the groundwork for measuring the detection of the spectrotemporal signals. Initial thresholds were obtained for single pure tones, then these tones were adjusted for equal detectability. This was accomplished with a degree of accuracy that allowed for the assurance that no frequency would provide greater information than others when they were combined for the more complex signals. In this way, the thresholds could be relied upon to reflect the detection of the overall signals. This was confirmed by calculating the expected improvement in threshold based on the energy detection model, which resulted in expected improvement of 9 dB from 1 tone to 8 tones. Each subject's equalized thresholds were also summed in the same way, resulting in expected improvements very close to those expected by energy detection. Thus, the results could reasonably be attributed to the integration of the signals, without being confounded by differences in the presentation level of the various frequencies. With this established, the primary aim of this study could be accomplished; that is, the tones could be combined to measure the integration in the time and frequency dimensions.

Spectral integration:

The spectral integration conditions of the first experiment were conducted to replicate the work by Grose and Hall (1997), with minor adjustments to the method to allow for comparison with the remainder of the experiments within this study. Additionally, the same subjects were used across all the experiments, so that the integration could be compared within the same individuals. In Grose and Hall's study, they found an improvement of 1.5 to 2 dB per doubling of tones, using the same frequencies as those included in this study. Their signals were 400 msec in duration, compared with the 10 msec here. The current study showed improvements of 1.02 to 2.44 dB per doubling of tones. These improvements were a good match for their results, despite the different durations, as well as the adjustment to the signal levels to make them equally detectable. The wider variation in the threshold improvements may be related to the brief duration of the current tones. Little data is available related to the stability of threshold measures for such brief tones, so this cannot be ruled out as the cause for the wider range of improvements. Regardless of this potential, the thresholds and improvements were equivalent to those reported in that 1997 work. While the previous research used only signals centered around the 1125 Hz signal, this study also included two tone signals in the high and low frequencies in order to consider the potential for differential integration on the basis of frequency range. This showed that the signals consisting of 1838 Hz and 2338 Hz to have the greatest integration. The low frequencies, 356 Hz and 494 Hz, showed the least integration. While the differences aren't significant in these data, it suggests the possibility that if longer duration tones were used, they could be integrated more effectively in these higher frequencies. It is uncertain at this time why that may be, although the higher frequency tones are within the range more common to speech signals.

Additional differences include a reduction in integration relative to the theoretical improvement expected according to the energy detection model. Some difference could be expected on the basis of central noise. According to Durlach, Braida, and Ito (1986) this could be related to the comparison of signals across frequency. Whether this degradation increases with greater spectral distance is inconsistent across studies. Green, et al., did not see the degradation based on distance, and the current study is in agreement with Green's results, however, this central noise may contribute to the reduced integration relative to predictions for the "ideal" listener. The central noise may increase the uncertainty in the threshold, thus limiting the amount of integration possible. Some further reduction is likely due to the basic variability of the testing procedure. The thresholds were based on the average of three runs, but this would not eliminate all the procedural noise. As a result, it would not be possible for real listeners to attain the theoretical level of integration, even under ideal circumstances. The threshold levels measured here are consistent with previous results, indicating that despite the very brief signals used here, the integration across the represented frequency range is possible. Since the tones were carefully adjusted to be equally detectable, the thresholds are unlikely related to the detection of only some tones. All frequencies were presented so that they reached threshold at the same time. As a result, differences in integration can be attributed to auditory processing rather than signal detection artifacts. The additional step of calculating theoretical integration on the basis of individual subjects' obtained thresholds was completed in order to assure the reliability of the assumption of equal detectability. Since the poorest match was 8.76 dB of improvement, the differences in the thresholds can be assumed to be due to the spectral integration.

To consider the integration across the spectrum, the concept of multiple looks could be considered in the frequency domain. The auditory system may be combining the signals across independent filters in a spectral "multiple look" in much the same way that Viemeister and Wakefield (1991) proposed for the temporal domain. The auditory system may be monitoring a wide range of critical bands, and selectively responding to the bands that include energy. The distance between component tones did not affect the integration of the signals in this experiment. This result was in contrast to the results reported by Bacon et al., (2002) who showed poorer integration with wider frequency spacing. The duration of their signals was 500 msec, much different than the 10 msec used here. Their conclusion was that the difference could reflect a spatial limit to spectral integration. The current results, however, suggest that this may not be the case, at least for all tasks. In fact, even predictability of the location of component tones may not be required for the integration, as some results obtained in the spectrotemporal conditions later in this study suggest. The multiple looks hypothesis would suggest that the auditory system can select the intervals containing information for integration. It appears with the current results that this may be the case in the spectral domain. Additionally, the independent channels model can be used to consider the combination of tones for these thresholds. The critical bands can be combined in independent observations, much like the observations in the temporal domain, and the information combined for improved integration. The specific location of the observations will not affect the detection threshold, as all bands are monitored independently. These independent channels can thus be applied in both domains, and presumably, for signals varying in both simultaneously.

Differences in thresholds for the close and mid spectral spacing can be seen when compared across the two tone conditions and the four tone conditions. The reduction in integration observed with the four tone signals is less than for the two tones when separated more in frequency. One important difference in the signals may be related to this difference. The mid spacing for the two tone signals included 5 silent ERBs, while the equivalent four tone signals included only 3 silent ERBs. It is possible that there is incomplete independence of the observations in the spectral domain until the spacing exceeds the 3 ERB separation. Since the differences were not significant, though, this difference is not a cause for questioning the independence of the observations for this study.

Temporal integration:

The temporal integration conditions of the first experiment were also conducted to replicate earlier research. In their work, Viemeister and Wakefield (1991) demonstrated an improvement in threshold of 2.5 dB from 1 tone to 2 tones using 200 microsecond rectangular tone pulses. The current study showed improvements of 1.98 to 2.47 dB per doubling of tones, using 10 millisecond pure tones, with an extension beyond the work of Viemeister and Wakefield by including up to 8 tones. The greatest improvement seen here (2.47 dB) occurred when the signal doubled from 1 to 2 tones, for an exact match to their result. No significant difference can be seen for different temporal spacings, from 10 msec of silence between tones to 130 msec of silence. All of these distances are greater than the 5 msec gap required for independence of the observations. Thus, the multiple looks hypothesis for temporal integration is supported with these signals. The same result can be seen at all the frequencies tested. A surprising result seen here is greater integration for some two tone signals than predicted by energy detection. This is not seen when all frequencies are averaged together. The explanation for this result is most likely related to the nonsignificant statistical difference between the data and the predicted integration. The differences may be due to chance variation in the measurements.

The similarity in the detection of the signals presented at the three frequencies (356 Hz, 1125 Hz, and 2338 Hz) additionally confirms that the temporal integration task does not vary with the spectral range being used. As a result further study can reliably be pursued in any of the frequency range included here, with good assurance that the results will be the same. Not only is the integration across the different temporal spacings comparable between the frequencies, the detection thresholds are also essentially the same.

<u>Spectrotemporal integration:</u>

The combination of the two integration domains yields additional information about the function of the auditory system that has previously not been explored. How do the two dimensions of acoustic signals interact in our detection? The results found in this study indicate that the limits of the spectrotemporal integration are very much due to the limits in the ability to integrate spectral information. The detection thresholds found for the signals with different frequencies in different time windows were very similar to those for the spectral integration conditions. This was especially true when the step sizes for the two dimensions were the same. The signals did not show significant differences on the basis of size of spectral and temporal steps or direction. This, again, supports the multiple looks hypothesis, interpreted in the spectrotemporal signals, as well. The distant spacing was not integrated differently than the closely spaced signals, and neither were the mid spaced signals. Thus, the auditory system appears to be capable of monitoring the filters throughout the entire range of frequencies included here, and detect signals that occur in any of the time windows. The information contained in those tones is accumulated over the course of the longer, overall, time window for use in detection.

While the signals described as symmetric were processed in a way that was very similar to those varying in only the spectral dimension, the asymmetric signals were different. This came as a surprise after the data collection was initiated. Prior expectation was that there should be no difference between the two types of spectrotemporal signals. After all, the frequencies should be integrated in similar fashion regardless of whether the spectral and temporal steps were the same or different. One would expect that this prediction would be further supported based on the nonsignificant differences related to the step sizes in the symmetric tone complexes. While the statistical analysis showed the difference between these two types of signals not to be significant, either, inspection of the data showed a difference in the average thresholds. The nonsignificance is likely due to the greater variability in the data. The question of the importance and amount of the difference between the integration for these signals will require further investigation to arrive at a satisfactory answer. It is possible that the differences seen in the data may be representative of the lower limit for processing of the rate of change in temporally and spectrally changing signals at threshold. Figure 4.15 showed that the thresholds for signals with slopes greater than 1, that is, those that changed over greater spectral than temporal range, showed essentially no improvement with the increase from two tones to four. Also, signals with close temporal spacing showed no improvement with the same increase in number of tones. These signals all necessarily had slopes greater than 1 by definition of the asymmetric conditions. It is possible that the auditory system is not able to track changing signals at this rate, particularly over the duration of 30 msec involved in these signals, with spectral changes from 7 to 15 ERBs. Further study with different signal durations, different spectral ranges, and suprathreshold signals may provide additional insight into this uncertainty. It is also for this reason that experiment 5 was included in this current study.

Random:

Random signals were included in the study to consider the question of predictability. If a pattern such as auditory streaming is involved in the detection of these signals, then random presentation should have a negative effect on the integration of the tones. Alternatively, if selective monitoring of critical bands is important for this detection, then again, the random presentation of frequencies should have a negative effect on integration. Rather, the results showed that the detection of the randomly generated signals was closer to that of the symmetric signals than was detection of the asymmetric signals. This supports the multiple looks hypothesis yet again, in that clearly the listeners are not simply establishing which critical bands contain the component tones and then monitoring those in order to detect the total signals. Rather, they are monitoring all the critical bands that may be included and detecting the tones wherever they may be present. Auditory streaming also does not appear to be a factor in the detection, since for the random signals, the closely spaced condition actually showed the poorest detection, which would contradict the influence of streaming. In considering the results with the symmetric signals in the predictable conditions, the thresholds showed no difference on the basis of spectral spacing, again contradicting streaming as a factor. However, since the differences in the thresholds are quite small overall, and the signals are brief, the analysis of the potential influence of streaming in spectrotemporal integration cannot be fully addressed here.

Doubled information signals:

The addition of doubled spectral or temporal information to the spectrotemporal signals resulted in conditions that echoed the asymmetric signals of experiment 3. The slopes of the corresponding TFRs were either greater or less than 1. However, the thresholds echoed more closely those of experiment 2 with symmetric signals. That is, the thresholds are closer to those predicted by the energy detection model than the thresholds obtained for the earlier asymmetric signals. The relationship of the slopes was maintained from the four tone conditions for the asymmetric signals, so that the steeper slope showed poorer integration than the shallower slope. Again, the differences failed to reach statistical significance, but the variance in the data, particularly for the asymmetric conditions limited this potential. In Figures 4.19 and 4.20, these signals were represented as eight tone signals, despite having 4 changes throughout the duration. This was done because of the total overall energy in the signals, rather than the number of changes, which could have been used as the defining difference. These results provided further support for the need for additional study to determine the relationships between spectral and temporal information in the detection of many types of auditory signals. Rather than the match to the spectral integration data noted for the other spectrotemporal conditions, the detection of these signals was improved to the extent that the thresholds were closer to the temporal integration data.

General discussion:

Many of the relationships considered in the context of this study were shown not to be statistically significant. However, on visual inspection of the results, some differences can be seen. Future work with this research may benefit from additional subjects or additional trials used in the calculation of the thresholds, in order to allow for decreased variability in the results. Then, if the apparent differences remain nonsignificant, they can more easily be disregarded. Some of the current results that are near significance, and visually appear different, could prove to be of further interest.

Throughout all of the discussion, the references to the multiple looks hypothesis and the energy detection model have been used, in some cases, interchangeably. This is not to imply that the two concepts are, in fact, interchangeable. The independent channels model may also be included in this discussion. The current data show a small increase in threshold for random signals, which is more consistent with the energy detector model, but this is not significant, so that it cannot rule out independent channels. Rather, since the current thresholds were measured in quiet, the influence of the three hypotheses cannot be separated. In future work, further measures should include signals with noise, and further modification to identify significant differences, in order to facilitate differentiation between these three ideas more specifically. If the thresholds are unchanged despite inclusion of noise between components, then the multiple looks hypothesis would be further supported, while the original energy detector would be contradicted.

All of the results presented here are consistent with the multiple looks hypothesis, with some limits in the spectral domain. It is unclear why spectral integration is poorer than temporal integration, but the analysis of the predicted threshold improvement based on the actual baseline measures indicates that it is not due to differences on the detection based on frequency thresholds. The results are also consistent with prior, separate studies of these dimensions. The suggestion offered by Kidd, Mason, and Richards (2003) that increase in detection threshold for signals with wider temporal separation due to a memory "noise", while considered with caution in their study, is not supported by the current results, since there was no significant difference between conditions with wider or narrower temporal or spectral spacing. They also suggested an auditory stream coherence explanation for their results. This also cannot be accepted directly on the basis of the current results, since the random presentation conditions were not significantly different than the predictable conditions. In fact, the asymmetric changes were more different. While it seems reasonable that auditory stream coherence may play a role in the detection of these signals, further study is required, particularly related to the potential limit on detection related to rate of change seen here.

Future research:

The results of this study lead to many more questions about the limits of spectrotemporal integration. A primary question is the influence of different durations on the detection of these signals. The durations of the individual tones could be altered, as well as the proportion of the total signal. The current signals were 10 msec tones separated by 10 msec silent intervals. The tones and silences could be altered differently to study the psychometric curve for the detection of these signals. If the total composite signal exceeds the 150 msec time window, will the detection be affected? Another question that relates closely to the extension of duration is what will happen to the integration if the spectral range is widened. Will the spectral integration be changed for conditions spanning more than 15 ERB? The audibility curve will begin to be a greater factor in the upper and lower limits of a widened range, making the equal detectability requirement more important than it is for the current frequencies.

Further investigation should address the difference seen in the asymmetric conditions. This may be readily achieved by studying different signal durations and spectral ranges. This could reveal information relative to the question of limits in the processing of rate of change in the signals. If the same slope is maintained, but the rate of change of the signal is reduced, will the integration improve? Perhaps the signals presented here are at the limit of that capacity. This cannot be determined without additional information.

The signals here primarily consisted of pure tones presented in consecutive time intervals for the spectrotemporal conditions. What changes in integration will occur if the signals are more complex? Other adjustments to the stimuli in the spectral domain could involve use of different frequencies and frequency ratios. Tones could be used with different relationships, including harmonics. More components can be included, possibly up to 21, as used by Dai and Green. The levels of the combined tones could later be altered to form sounds more like environmental sounds, as our understanding of the basic sounds improved.

Additionally, further study should include use of noise during the intertone intervals, as in the Viemeister and Wakefield work leading to the multiple looks hypothesis. This could help to differentiate between this view and energy summation.

Other work could use noise to facilitate the study of integration at suprathreshold levels to determine if the same relationships apply in that context. It could also be used as a replacement for the pure tones used here to extend the study for more complex signals, much like the combination of additional tones. Other suprathreshold research could include loudness summation measurements, to expand on other work in loudness perception.

<u>Summary:</u>

The paradox of integration versus acuity has long been considered for sounds in either the domain of time or frequency. However, the two dimensions have not been well studied together in the same sounds. But signals that vary along both of these dimensions are constantly present in real world sounds around us. A better understanding of these dimensions in the controlled environment of psychoacoustic research could provide valuable insight into the way humans process these important components, and the knowledge gained could be carried out of the laboratory and used for a multitude of purposes in daily life.

Once more is understood about the use of multiple looks in both time and frequency for static signals, further research could also be expanded for a better understanding of the integration of dynamic signals. This would then help to span the gap between psychoacoustics and the study of naturally occurring sounds in the environment. Without knowing the limits of our auditory system's ability to process sounds, it is impossible to effectively mimic that processing for use in assistive devices, such as hearing aids or cochlear implants. The ability to program integration into technology could allow individuals with hearing loss to use environmental sounds effectively. This would allow those sounds to regain the importance in daily functioning observed in the ecological studies being conducted currently and in the past. It would also be beneficial to be able to understand what about those sounds contributes to our responses, both in the sources and our perception of the acoustics.

While this study used laboratory generated sounds, it is a start to addressing the issue of translation between the two different ends of the research continuum. Some of the sounds used here were described by at times in terms of environmental sounds. Descriptions of bullfrogs, crickets, and raindrops were used. This inspires the expectation that the ability to simulate environmental sounds in the laboratory with strictly controlled signals may not be too difficult to achieve with persistence and strategy, and by building on the work that has already been done to this time.

APPENDIX A: TIME FREQUENCY REPRESENTATIONS FOR EXPERIMENTAL SIGNALS

For Figures A.1 through A.7: abscissa is time domain, ordinate is frequency domain.

Figure A.1: Baseline experimental arrays

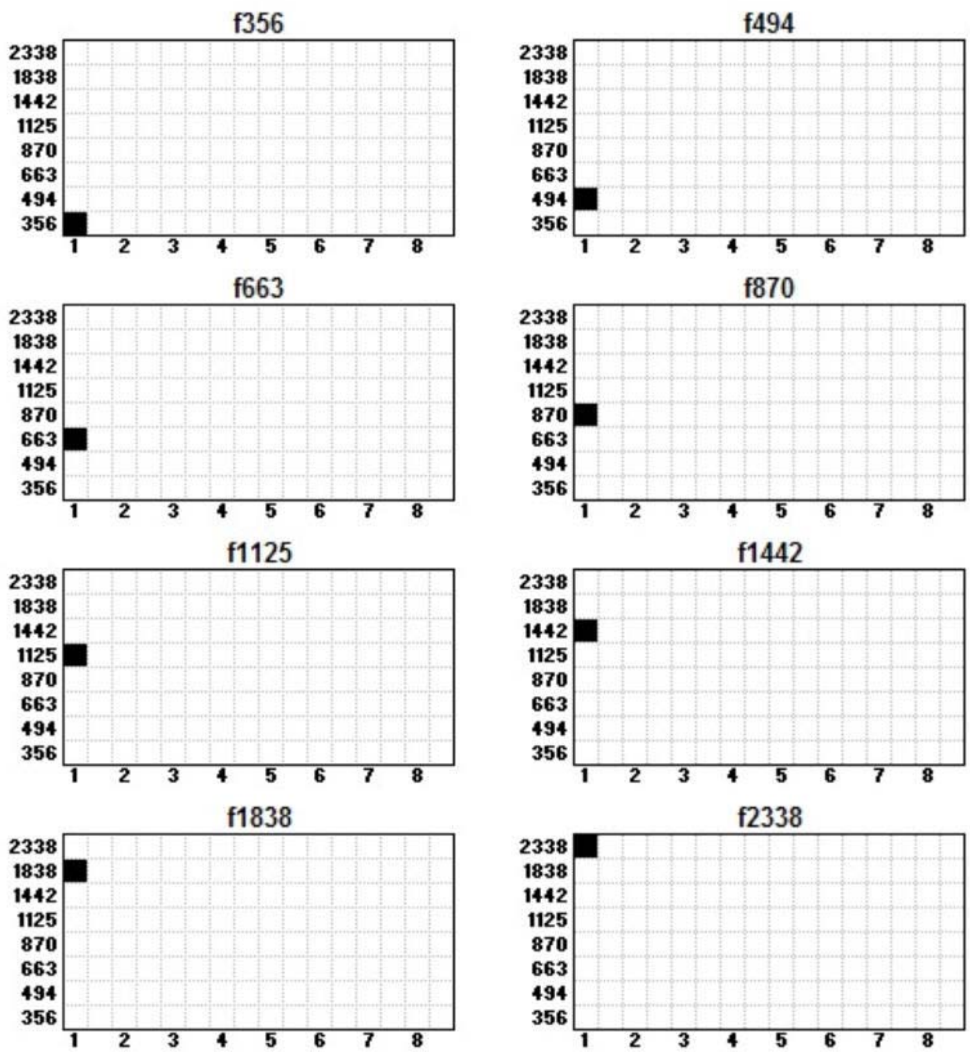


Figure A.2: Experiment 1 spectral arrays

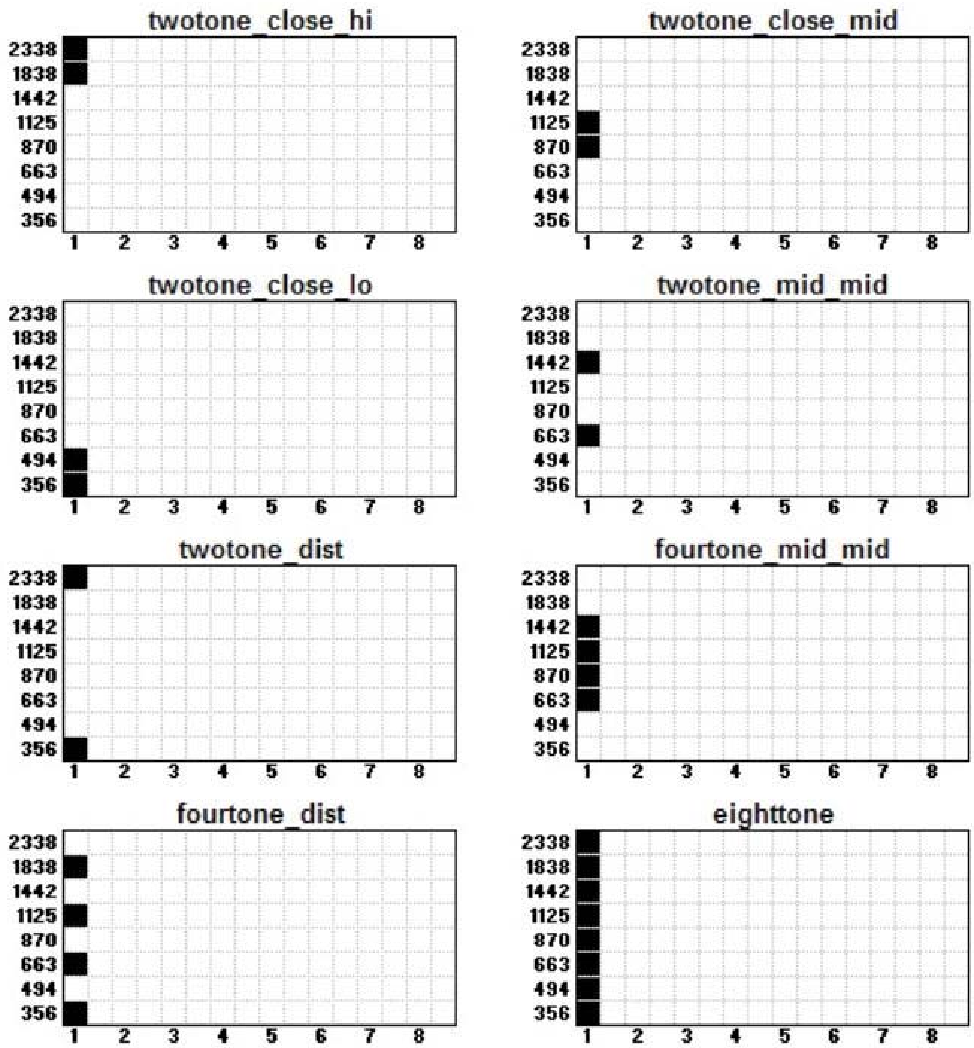
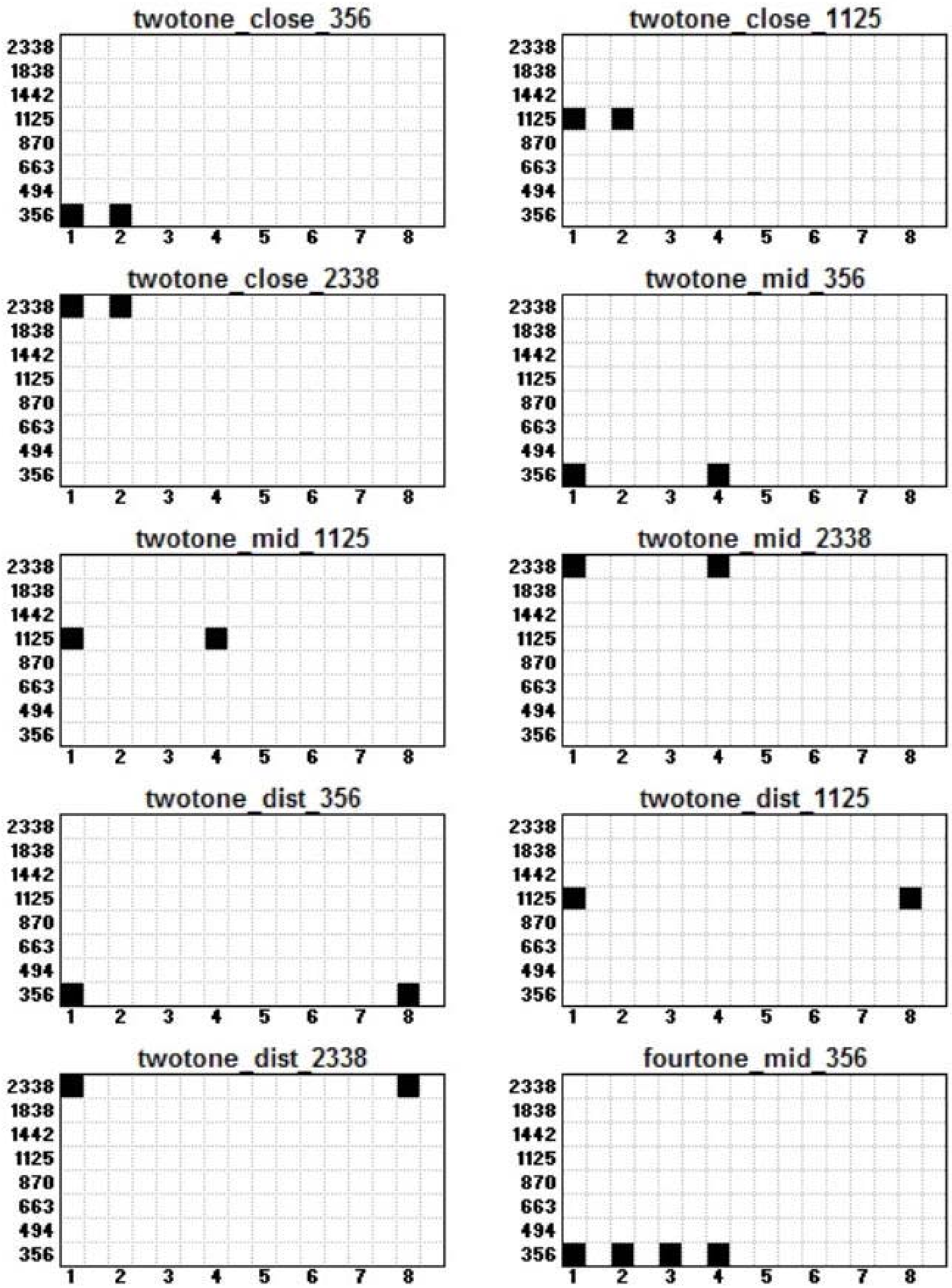


Figure A.3: Experiment 1 temporal arrays



Continued

Figure A.3: Experiment 1 temporal arrays, continued

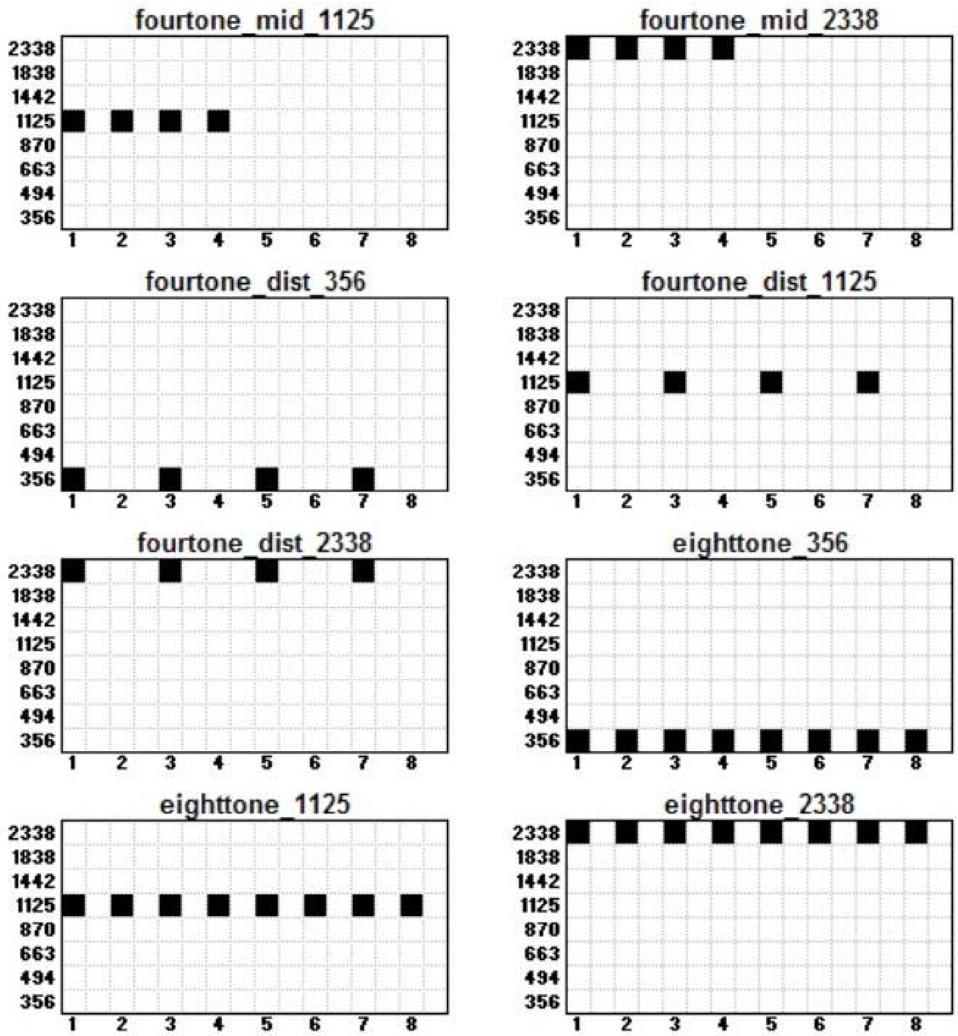
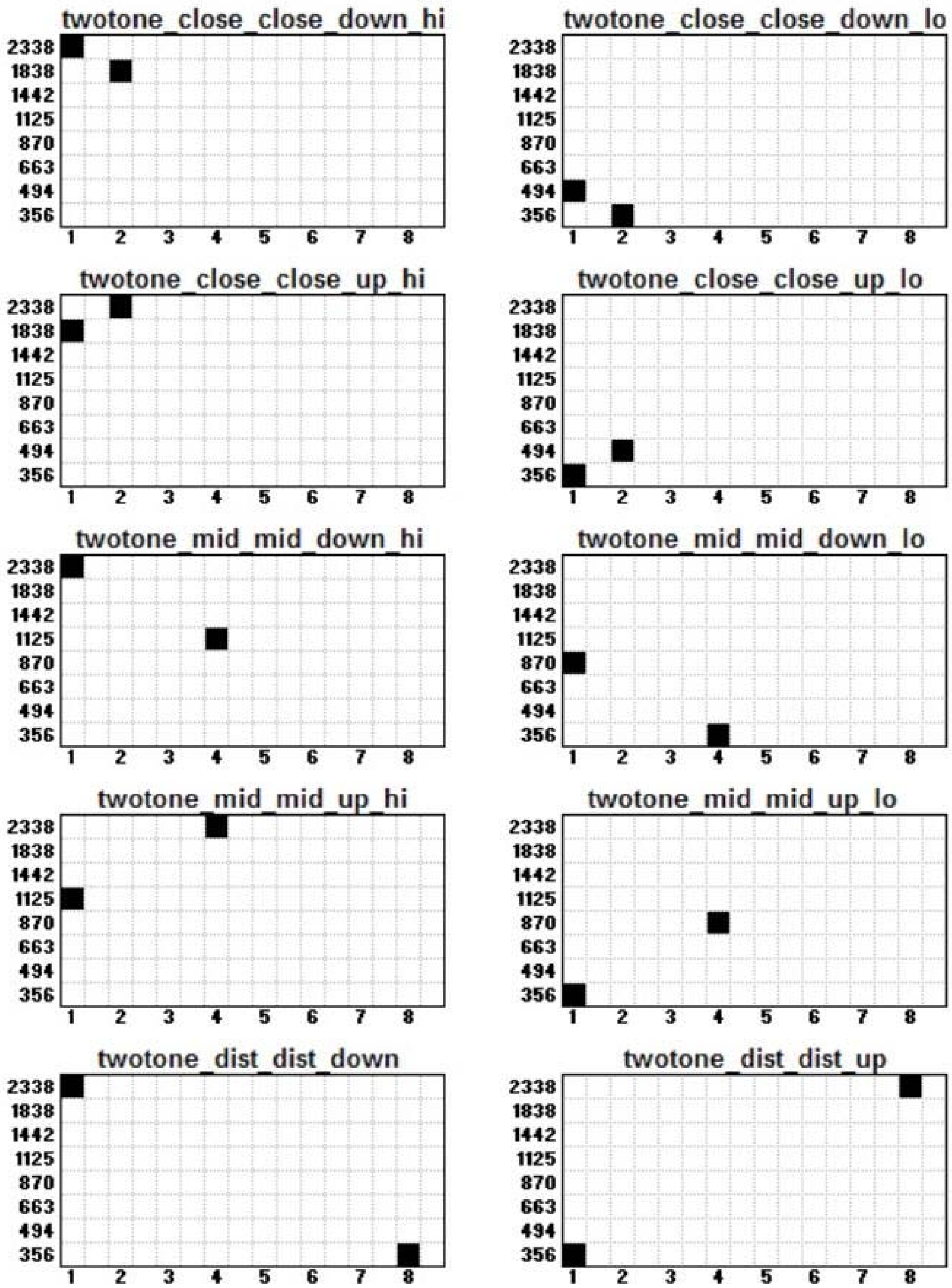


Figure A.4: Experiment 2 arrays



Continued

Figure A.4: Experiment 2 arrays, continued

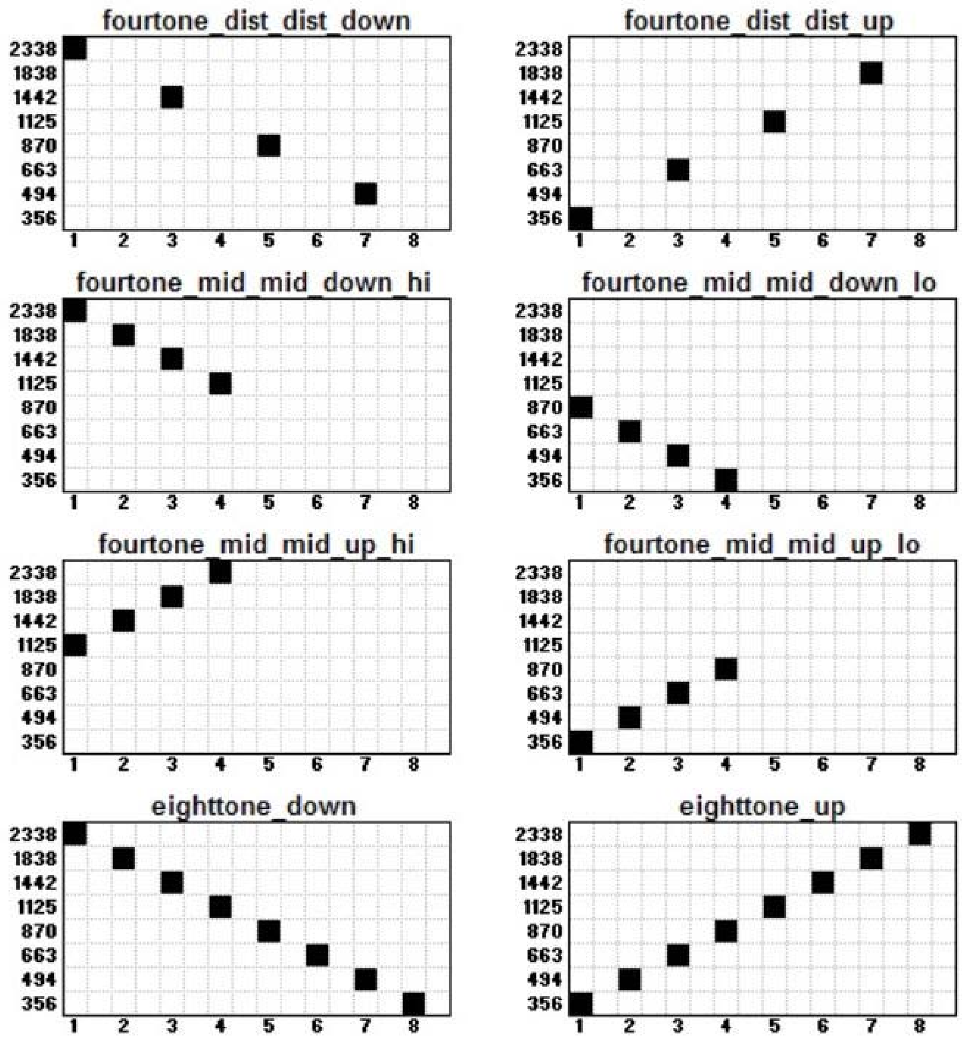
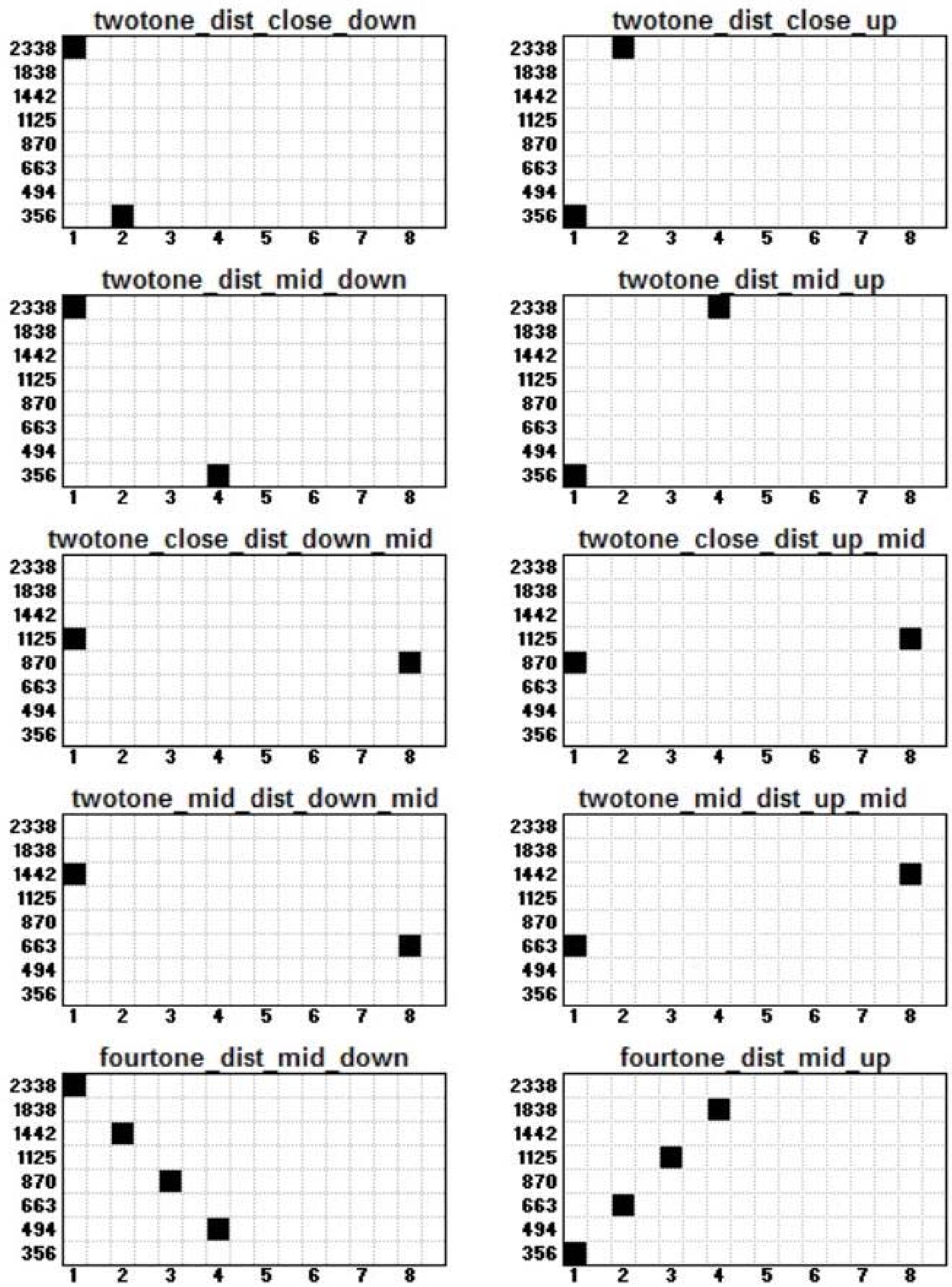


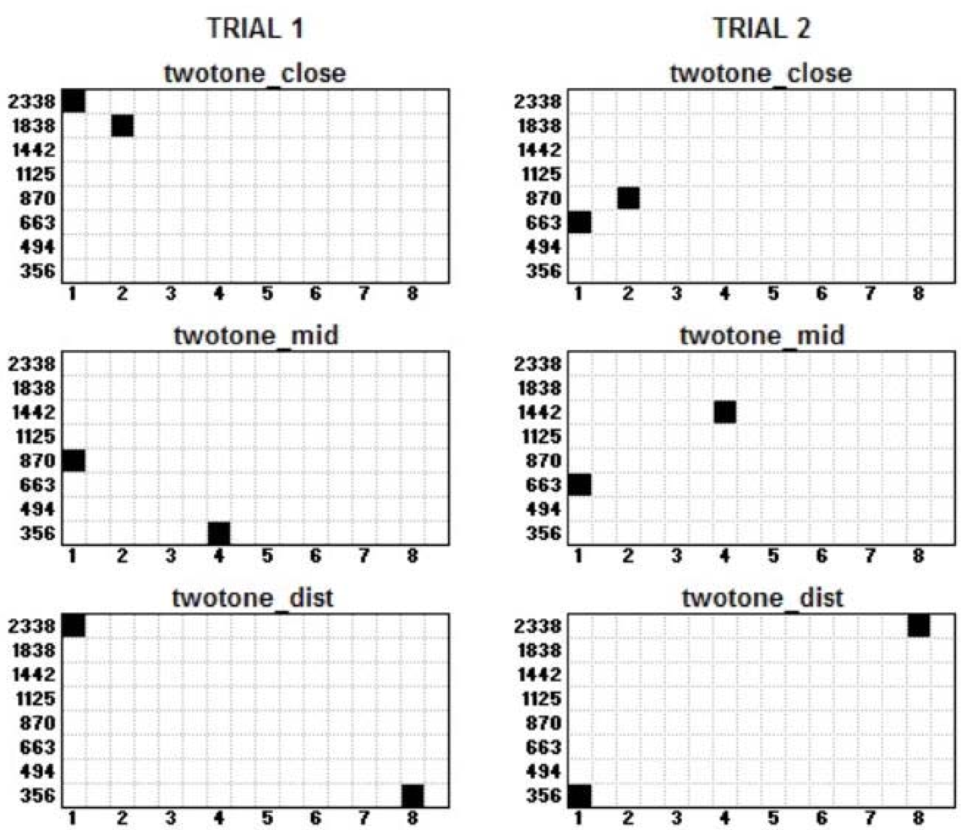
Figure A.5: Experiment 3 arrays



Continued

Figure A.5: Experiment 3 arrays, continued

Figure A.6: Experiment 4 array examples



Continued

Figure A.6: Experiment 4 array examples, continued

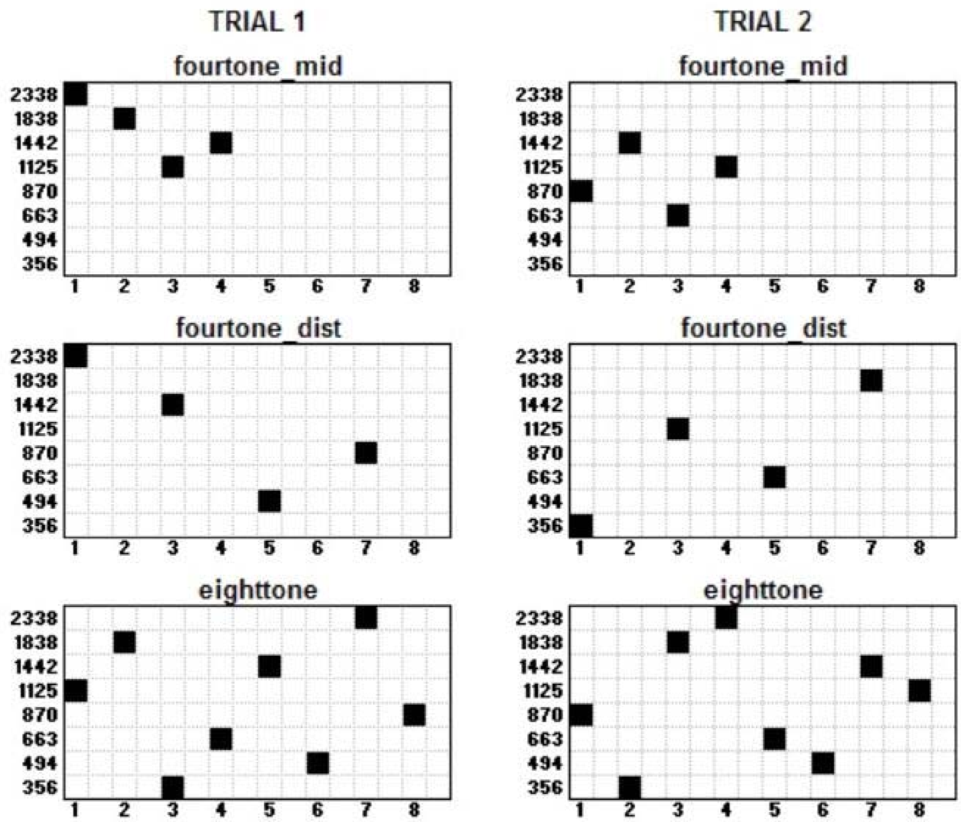
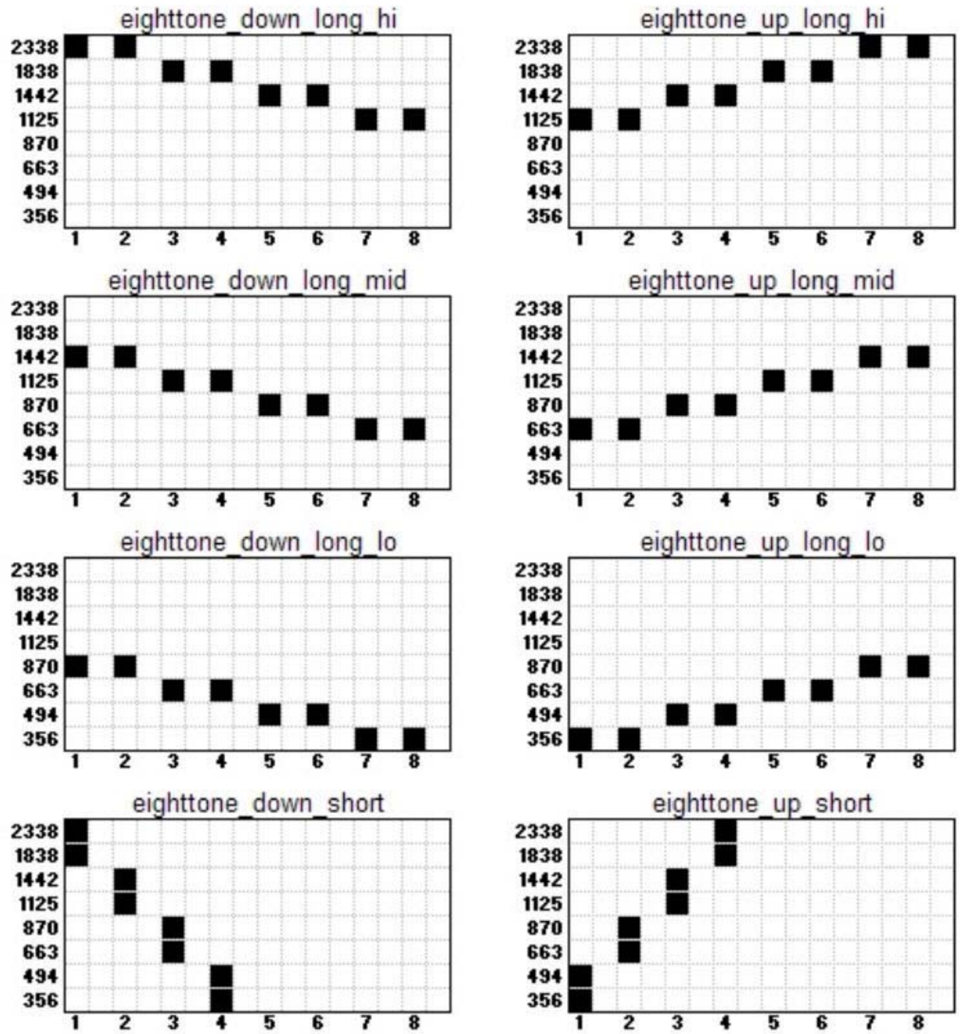


Figure A.7: Experiment 5 arrays



APPENDIX B: AMPLITUDE VECTOR CALCULATION

Decibel conversion:

$$\text{dB } (35 - \Theta) = 20 \log (V_2/V_1)$$

target: dB difference = 0

Initial adjustment to amplitude of individual frequencies:

Θ_1 = obtained threshold with signal level at 0.1 V
V_1 = adjusted voltage

$$\frac{\text{Antilog } (35 - \Theta_1)/20}{10} = V_1$$

Additional adjustments to amplitude of individual frequencies:

Θ_2 = obtained threshold with signal level as adjusted in previous step

V_{adj} = voltage used to obtain factor for adjusting voltage toward 35 dB threshold

$$\frac{\text{Antilog } (35 - \Theta_2)/20}{10} = V_{adj}$$

$V_{adj}*10$ = multiplier

multiplier* $V_1 = V_2$

Example:

frequency	threshold	equal detectability conversion			
2338	40.34	-5.340	-0.267	0.541	**0.05408**
1838	40.98	-5.977	-0.299	0.503	**0.05025**
1442	40.05	-5.047	-0.252	0.559	**0.05593**
1125	41.68	-6.677	-0.334	0.464	**0.04636**
870	38.03	-3.030	-0.152	0.706	**0.07055**
663	39.26	-4.260	-0.213	0.612	**0.06124**
494	35.70	-0.697	-0.035	0.923	**0.09229**
356	34.11	0.890	0.045	1.108	**0.11079**

thresholds with .1 V for each frequency
column F is the voltage to be applied for equally detectable thresholds set to 35 dB

frequency	threshold	equal detectability conversion				
2338	34.51	-0.230	-0.011	0.974	0.09739	**0.973868**
1838	32.70	1.580	0.079	1.199	0.11995	**1.199499**
1442	35.73	-1.450	-0.072	0.846	0.08463	**0.846253**
1125	33.84	0.440	0.022	1.052	0.10520	**1.051962**
870	34.34	-0.060	-0.003	0.993	0.09931	**0.993116**
663	35.00	-0.720	-0.036	0.920	0.09204	**0.92045**
494	33.80	0.480	0.024	1.057	0.10568	**1.056818**
356	34.32	-0.040	-0.002	0.995	0.09954	**0.995405**

column B is measured thresholds after adjusting to be equally detectable
column G is the multiplier to be applied to the threshold for further tweaking
toward the mean

frequency	threshold	calc. V	multiplier	
2338	34.51	0.05408	0.9739	**0.0527**
1838	32.70	0.05025	1.1995	**0.0603**
1442	35.73	0.05593	0.8463	**0.0473**
1125	33.84	0.04636	1.0520	**0.0488**
870	34.34	0.07055	0.9931	**0.0701**
663	35.00	0.06124	0.9204	**0.0564**
494	33.80	0.09229	1.0568	**0.0975**
356	34.32	0.11079	0.9954	**0.1103**

column C is taken from column F in first grid
column E is final voltage value to be applied to ampVector array for the subject

Subject 1

frequency	ampVector
2338	0.0227
1838	0.052
1442	0.0258
1125	0.0362
870	0.0453
663	0.0645
494	0.0668
356	0.1063

Subject 2

frequency	ampVector
2338	0.0318
1838	0.0639
1442	0.0519
1125	0.0308
870	0.046
663	0.0706
494	0.1533
356	0.1786

Subject 3

frequency	ampVector
2338	0.0373
1838	0.026
1442	0.0351
1125	0.1114
870	0.1025
663	0.1448
494	0.1575
356	0.2191

Subject 4

frequency	ampVector
2338	0.0406
1838	0.0656
1442	0.0805
1125	0.0633
870	0.0593
663	0.0851
494	0.1171
356	0.1465

Subject 5

frequency	ampVector
2338	0.0294
1838	0.06
1442	0.0363
1125	0.0456
870	0.0417
663	0.0507
494	0.0903
356	0.1081

Subject 6

frequency	ampVector
2338	0.0401
1838	0.1015
1442	0.0535
1125	0.0388
870	0.056
663	0.0516
494	0.0808
356	0.1111

BIBLIOGRAPHY

Bacon, S. P., Grimault, N., & Lee, J. (2002). Spectral integration in bands of modulated or unmodulated noise. *Journal of the Acoustical Society of America,* 112(1), 219-226.

Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception and Performance,* 19(2), 250-267.

Bernstein, L. R., & Green, D.M. (1987). Detection of simple and complex changes of spectral shape. *Journal of the Acoustical Society of America,* 82(5), 1587-1592.

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

Buus, S. (1999). Temporal integration and multiple looks, revisited: Weights as a function of time. *Journal of the Acoustical Society of America,* 105(4), 2466-2475.

Buus, S., Schorer, E., Florentine, M. & Zwicker, E. (1986). Decision rules in detection of simple and complex tones. *Journal of the Acoustical Society of America,* 80(6), 1646-1657.

Dai, H., & Green, D.M. (1993). Discrimination of spectral shape as a function of stimulus duration. *Journal of the Acoustical Society of America,* 93(2), 957-965.

Durlach, N.I., Braida, L.D., & Ito, Y. (1986). Towards a model for discrimination of broadband signals. *Journal of the Acoustical Society of America*, 80(1), 63-72.

Fowler, C. A., & Rosenblum, L.D. (1990). Duplex perception: a comparison of monosyllables and slamming doors. *Journal of Experimental Psychology: Human Perception and Performance,* 16(4), 742-754.

Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *Journal of the Acoustical Society of America,* 87(1), 311-322.

French, N. R., & Steinberg, J.C. (1947). Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America,* 19(1), 90-119.

Green, D. M. (1958). Detection of Multiple Component Signals in Noise. *Journal of the Acoustical Society of America,* 30(10), 904-911.

Green, D. M. (1960). Auditory Detection of a Noise Signal. *Journal of the Acoustical Society of America,* 32(1), 121-131.

Green, D. M. (1983). Profile analysis: A different view of auditory intensity discrimination. *American Psychologist*, 133-142.

Green, D. M., and Swets, J.A. (1966). *Signal Detection Theory and Psychophysics*. Los Altos, California: Peninsula Publishing.

Grose, J. H., & Hall, J.W. (1997). Multiband detection of energy fluctuations. *Journal of the Acoustical Society of America,* 102(2 pt. 1), 1088-1096.

Gygi, B., Kidd, G.R., & Watson, C.S. (2004). Spectral-temporal factors in the identification of environmental sounds. *Journal of the Acoustical Society of America,* 115(3), 1252-1265.

Handel, S. (1989). *Listening: An Introduction to the Perception of Auditory Events*. Cambridge, MA: The MIT Press.

Hicks, M.L. & Buus, S. (2000). Efficient across-frequency integration: Evidence from psychometric functions. *Journal of the Acoustical Society of America,* 107(6), 3333-3341.

Kidd, G., Mason, C.R., & Richards, V.M. (2003). Multiple bursts, multiple looks, and stream coherence in the release from informational masking. *Journal of the Acoustical Society of America,* 114(5), 2835-2845.

Li, X. F., Logan, R.J., & Pastore, R.E. (1991). Perception of acoustic source characteristics: Walking sounds. *Journal of the Acoustical Society of America,* 90(6), 3036-3049.

McAdams, S., Beauchamp, J.W., & Meneguzzi, S. (1999). Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. *Journal of the Acoustical Society of America,* 105(2), 882-897.

Moore, B. C. J. (1995). *Hearing*. San Diego: Academic Press, Inc.

Oxenham, A.J., and Buus, S. (2000). Level discrimination of sinusoids as a function of duration and level for fixed-level, roving-level, and across-frequency conditions. *Journal of the Acoustical Society of America,* 107(3), 1605-1614.

Schafer, R. M. (1994). *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester, VT: Destiny Books.

Spiegel, M. F. (1979). The range of spectral integration. *Journal of the Acoustical Society of America,* 66(5), 1356-1363.

Viemeister, N. F. (1996). Auditory temporal integration: What is being accumulated? *Current Directions in Psychological Science,* 5(1), 28-32.

Viemeister, N. F., & Wakefield, G.H. (1991). Temporal integration and multiple looks. *Journal of the Acoustical Society of America,* 90(2 pt. 1), 858-865.

Warren, H., & Verbrugge, R.R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology, Human Perception and Performance,* 10, 704-712.

Watson, C. S. (2004). Notes on response proclivities and sensory capabilities.