# Feature Assignment in Perception of Auditory Figure and Ground

A Dissertation Presented

by

Melissa Kay Gregg

to

The Graduate School

In Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Experimental Psychology

Stony Brook University

May 2010

Stony Brook University

The Graduate School

Melissa Kay Gregg

We, the dissertation committee for the above candidate for the

Doctor of Philosophy, hereby recommend

acceptance of this dissertation.

Arthur Samuel – Dissertation Advisor

Professor of Psychology

Nancy Franklin – Chairperson of Defense

Associate Professor of Psychology

Nancy Squires

Professor of Psychology

Margaret Schedel

Assistant Professor of Music

This dissertation is accepted by the Graduate School

Lawrence Martin

Dean of the Graduate School

Abstract of the Dissertation

Feature Assignment in Perception of Auditory Figure and Ground

by

Melissa Kay Gregg

Doctor of Philosophy

in

Experimental Psychology

Stony Brook University

2010

Many critical skills rely on the ability to successfully sort out the auditory sensory information in an auditory scene. For example, in order to communicate successfully, listeners must be able to segregate incoming speech from other co-occurring sounds, such as a telephone ringing, background music, car horns, and speech coming from a nearby source. The task of segregating an auditory object of interest from other co-occurring sounds is one of identifying an organized auditory figure against the unattended auditory ground. In this project, I conducted a series of experiments intended to contribute to a basic understanding of how auditory figure and ground analysis is accomplished. The first part of this project utilized behavioral methods to determine how the relationship between the object in figure and the objects in ground affects the auditory system's assignment of a feature to those objects. One finding that occurred across experimental manipulations was that the auditory system is more likely to allocate a feature to an object that needs the feature in order to be a meaningful object. This finding did not occur as often in more complicated perceptual scenarios, presumably because of increased competition for the feature.

The second part of this project was designed to address the fate of features in auditory ground. Behavioral and physiological methods were used to determine if features are assigned to objects in perceptual ground or to perceptual groups in ground. The results indicated that features are assigned to objects in the background, but not to perceptual groups. The results suggest that the nature of perceptual ground is not a free-floating perceptual limbo. Rather, objects appear to be well constructed in ground.

Table of Contents

List of Figures

Acknowledgements

First, I thank my advisor, Arty Samuel, for the guidance on this project, as well as for the excellent advice over the last four and a half years (which I'm sure I will continue to seek throughout my career). I also thank Arty for teaching me how to organize my thoughts and put them to use to do good science. Second, I extend my gratitude to the woman who can make a computer do anything, Donna Kat, for creating many wacky programs for me and for being patient with my programming ignorance. Many thanks to my very helpful committee members, Nancy Franklin, Nancy Squires, and Meg Schedel for the advice on and enthusiasm for this project. I also thank two wonderful research assistants, Carolyn Creary and Kavita Patel, for helping me collect data. And, many thanks to two fellow graduate students, Jiwon Hwang and Vera Hau, for devoting their time to help me learn how to conduct and make sense out of an ERP experiment. Thank you to my supportive husband for moving 3000 miles across the country with me so I could work with some guy on Long Island, for asking what my projects are about and sincerely wanting to know, and for being the first pilot subject in every experiment I conduct, including every experiment in this project. And, last but not least, thank you to my 10 month old daughter, Ella, for keeping me smiling and lighthearted throughout this project. My guess is that "dissertation" will be a part of her early vocabulary.

Feature Assignment in Perception of Auditory Figure and Ground

Listeners must deal with auditory sensory information from many sources in any given auditory scene. Vast numbers of auditory features, including frequency components and their intensities, must be sorted out so that the features that belong to one object are correctly grouped together and features that belong to other objects are correctly excluded from objects of interest. Though auditory research has provided some important guiding principles for how feature analysis is accomplished (see Bregman, 1990), the process of auditory feature analysis is complicated by recent findings of non-veridical representation of an auditory scene. The auditory system has been shown to prefer efficient processing over detailed processing of all of the information available in a scene (e.g., Shinn-Cunningham, Lee, & Oxenham, 2007). If we do not perceive all of the information in a scene, what guides the part that we do perceive? The purpose of this project is to shed light on this issue by addressing how auditory features are combined to form auditory objects of attention and to examine what happens to features that remain in the unattended background.

*The Problem of Auditory Scene Analysis*

The human listener is faced with a substantial amount of sensory information from many different sound sources in the environment, most of which overlap in time and in space. Successful navigation and communication rely on the ability to sort out this information. Take for example, two co-workers walking from their office building on Wall Street in Manhattan to the train station at the end of the day. As the two co-workers are talking, each must be able to filter out the acoustic information from other sound sources to achieve successful interpersonal communication, so that the phrase, “I hear that Omar is getting fired,” is not incorrectly heard as “I hear that you are getting fired.” Each worker must also be able to identify potentially dangerous objects, such as a car approaching while crossing a street, and must be able to identify important objects, such as a train change announcement at the station. These are just a few of the auditory perception feats the two workers must accomplish in the face of overlapping sounds from other sources, such as people talking, people yelling, traffic, emergency sirens, jackhammering, etc.

With so much overlapping sensory information in any given listening situation, the task of disentangling the relevant sensory components from the irrelevant components can be quite challenging. In order to make sense of the incoming information, the auditory system has to be able to parse information into potential objects in the environment. An auditory scene can contain a huge number of features, so the system must also be able to select and bind together the features that belong together in order to perceive distinct objects in any detail. How are these parsing and binding processes accomplished by the system?

The answer to this question is the focus of Bregman’s theory of auditory scene analysis (ASA: see Bregman, 1990 for an extensive review). The goal of the theory is to explain how we determine what parts of the auditory input belong to the same auditory object, or source. Bregman’s theory treats both parsing and binding as being the result of an analysis process that combines the physical components that belong together to make up one stream of sounds, such as a clock ticking, for all sources of sound in a scene. The analysis yields multiple co-occurring auditory streams, such as a voice, a telephone, and a clock ticking. Whether the auditory features in any given listening situation are deemed to belong together is determined by two factors: low-level Gestalt grouping principles and higher-level knowledge (Bregman, 1990).

The grouping of auditory features into objects, or streams, by Gestalt grouping is based on the physical properties of sound components, such as frequency and harmonic structure. The Gestalt grouping principles were established by the Gestalt psychologists (Koffka, 1935; Köhler, 1947; and Wertheimer, 1923), who maintained that the world was organized into units on the basis of many guiding principles. For example, their principles of similarity and proximity state that auditory features that are physically similar to each other and that occur in close temporal and spatial proximity to each other tend to be parts of the same object; thus, they will typically be integrated into one object.

Higher-level knowledge-based guidance for auditory feature grouping, called schema-based processing by Bregman (1990), is based on attention, memory, motivation, and pre-existing knowledge of auditory objects. Higher-level processes can be a strong influence on how auditory features are integrated to form objects (see Carlyon, Cusack, Foxton, & Robertson, 2001; Shinn-Cunninham, Lee, & Oxenham, 2007; Sussman, Horvath, Winkler, & Orr, 2007). Listener knowledge can often dominate perception so much that acoustic features are integrated together despite physical evidence that they do not belong to the same object. For example, when listeners are presented with distorted versions of common folk tunes, such that each tone is randomly moved up or down one octave, most people cannot recognize the tunes. However, once listeners are told what the tune is, they can easily hear the familiar melody (Deutsch, 1972).

One of the key constraints on the process of auditory scene analysis is the principle of exclusive allocation (Bregman, 1990). According to this principle, once a feature has been assigned to an object, it cannot be concurrently assigned to another, separate object. The act of auditory object formation involves an ongoing evidence assessing process (Bregman, 1990). The auditory system continually assesses the probability that each piece of physical evidence belongs to an object and updates the grouping of information accordingly. Thus, the process of grouping physical input into an auditory stream is cumulative, i.e., the tendency to group things into one stream builds up over time (Anstis & Saida, 1985; Bregman, 1978).

While Bregman’s (1990) theory of auditory scene analysis has been foundational in establishing properties that guide auditory feature analysis, ASA theory does not explain how auditory figure and ground are established. Perceptual figure and ground are terms that refer to the organizational tendency of the perceptual system to perceive the attended portion of a scene as an organized object (i.e., figure) against the unattended and mostly unorganized background (Rubin, 1921). There are several problems that make ASA theory incapable of consistently predicting the percept that emerges from a scene. One problem is that counter-evidence has been presented for the claim that there are multiple co-occuring streams in any given listening situation. Several studies have shown that only one stream may be perceived in any detail at a time; i.e., detail is perceived in a stream that is the perceptual figure but not in streams that are in perceptual ground (e.g., Botte, Drake, Brochard, & McAdams, 1997). The theory is undercut by exceptions to the principle of exclusive allocation: It has been extensively demonstrated that an auditory feature can contribute to two objects at the same time (e.g., Rand, 1974).

Perhaps one of the most critical problems for Bregman's (1990) theory, however, is recent evidence against veridical perceptual representation of the environment. Several lines of research have shown that auditory perception of a scene may be determined by efficiency, rather than including all of the detail within the scene (e.g., Gregg & Samuel, 2008; 2009; Lee & Shinn-Cunningham, 2008). This poses a problem for auditory scene analysis because it suggests that the grouping principles established by ASA paradigms do not reliably guide the percept that emerges as auditory figure or ground. Collectively, these problems reflect an incomplete picture of how auditory features are allocated to auditory figure and ground because neither low-level Gestalt grouping principles nor higher-level knowledge based processes can consistently predict the percept that emerges from a scene.

*The Problem of Non-Veridical Representation: Is Hearing Believing?*

Most people would assert that their experience of the world is rich and detailed. Confident assertions such as, "I will believe it when I see it," "I know what I saw," or "I know what I heard" illustrate this common assumption. But, there is a growing body of research that casts doubt on these subjective impressions. For example, it has been found that auditory features do not follow the rules of energy trading that should hold if perception is veridical. Physically, if one auditory component contributes to two objects simultaneously, then the total energy in that component should be split between the two objects so that the sum of the amount of energy the component contributes to each object equals the total amount of energy in the component. A recent study has shown that this intuitive rule for energy trading does not apply when two objects are competing for the same feature (Shinn-Cunningham et al., 2007).

Other evidence of non-veridical perceptual representation is the remarkable inability of people to detect large changes to visual and auditory scenes, even though the changes seem blatantly obvious after the fact.For example, in the visual domain, studies have shown that observers miss changes to many different visual scenarios (see Simons & Rensink, 2005 for a review). Such change blindness can occur during a visual interruption, such as a flicker (e.g., Rensink, O'Regan, & Clark, 1997) or a blink (e.g., O'Regan, Deubel, Clark, & Rensink, 2000). For example, people fail to notice a huge airplane engine disappearing in a flickering picture display (Rensink et al., 1997). And, around two-thirds of participants fail to notice a change in the identity of a real-life conversational partner when the conversation is interrupted by a brief visual occlusion (Simons & Levin, 1998). Change blindness also can occur when attention is focused away from the changing object; when instructed to attend to a group of basketball players, viewers rarely notice a person dressed up in a gorilla suit walking behind the players (Simons & Chabris, 1999).

In the auditory domain, changes to an unattended stream of auditory input (such as a reversal of speech, change of language, or change of speaker) are often missed while shadowing a spoken message presented to an attended stream of auditory input (Cherry, 1953; Treisman, 1960; Vitevitch, 2003). Listeners are often unaware of changes to attended information as well, such as changes to scenes consisting of common environmental sounds (Eramudugolla et al., 2005). Change deafness has been shown to be resistant to familiarity training with the sounds, and although change detection improves when the object that changes is relatively far on acoustic dimensions from the sound it replaced, performance is still rather poor (Gregg & Samuel, 2008). Other change deafness work has shown that listeners encode the semantic gist of an auditory scene, rather than the acoustic detail; i.e., listeners miss changes to objects that are within the same semantic category (such as a small song bird chirp changing to a seagull squawk) significantly more than changes to objects from different categories (Gregg & Samuel, 2009).

In addition to change deafness demonstrations, there are even cases of deafness to auditory features that are not changing. A recent study has demonstrated that when two objects are both competing for the same feature, the feature can completely disappear from the percept (Shinn-Cunningham et al., 2007). In this study, two objects were presented: A vowel and a tone sequence. A target feature was used that could potentially group with each object. If the feature were incorporated into the vowel, the vowel would sound like /ɛ/ as in "pet;" if not, the vowel would sound like /I/ as in "pit." Similarly, the presence/absence of the feature tone would affect the rhythmic properties of the tone sequence. If the feature were perceived as part of the tone sequence, the sequence would have a constant rhythm; otherwise it would be perceived as "galloping." The spatial relationship between the feature and the two objects was manipulated: The feature could be presented at the same spatial location as the vowel (or tone sequence) to encourage it to group with the vowel (or tone sequence); or, to reduce the likelihood of the feature grouping with one of the objects, the feature could be presented at a different spatial location.

When only a single object was presented with the target feature, Shinn-Cunningham et al. (2007) found that the object (either the vowel or the tone sequence) grouped with the feature, even when the feature was presented at a different spatial location than the object. The tone sequence grouped with the feature to produce the even rhythm percept, and the vowel grouped with the feature to produce the /ɛ/ percept.

The results were quite different when two objects were presented. In the two-object condition, the vowel and the tone sequence objects were presented simultaneously in a repeating sequence. In one block of trials the listeners were instructed to attend to the vowel; in another, listeners were instructed to attend to the tones. The listeners' task was to label the object they were attending to, i.e., report "even" or "galloping" when attending to the tone and "/ɛ/" as in "pet" or "/I/" as in "pit" when attending to the vowel. Shinn-Cunningham et al. (2007) found that in the attend-tone block, listeners heard the feature as contributing to the tone sequence when it was presented at the same spatial location as the sequence and when it was presented in a spatial location that did not match the sequence or the vowel. But, the feature did not contribute to the perception of the tone sequence when it was presented at the same spatial location as the vowel. So, the spatial cues were not strong enough to prevent the feature and the object from binding, except in the most extreme spatial configuration in which the feature was presented at the same spatial location as the object in the background. Oddly, in the exact same configuration in the attend-vowel block, i.e., when the feature was presented at the same spatial location as the vowel, the feature did not group with the vowel, i.e., the percept of the vowel was /I/. Because the feature did not contribute to the percept of either the tone sequence or the vowel, it as if the feature disappeared from the mixture, resulting in a case of feature nonallocation.

Collectively, the findings of feature nonallocation along with the energy trading (Shinn-Cunningham et al., 2007) and change deafness work (e.g., Eramudugolla et al., 2005) call into question the extent of detail with which we perceive and represent our environment. These findings suggest that perceptual processing may favor efficiency over detailed processing. If we do not hear everything there is to hear, what guides the subset of the environment that we do perceive? How are the numerous features in the environment assigned to objects in perceptual figure and what happens to features in perceptual ground? Is it possible to establish some rules for which features get assigned to objects in figure and which ones do not? These questions are the motivation for this project. The first part of the project is concerned with feature assignment in figure and the second with feature assignment in perceptual ground.

### *Feature Assignment in Auditory Figure: Experiments 1-4*

The recent finding that when two objects are competing for the same feature, an object in perceptual figure will not group with a feature despite perfect spatial alignment between that object and the feature (Shinn-Cunningham et al., 2007) is surprising. It is in direct contrast with an extensive line of research demonstrating that an object in figure will group with a feature despite evidence that the feature belongs to another source (e.g., Cutting, 1976; Darwin & Hukin, 1998; Rand, 1974; Shinn-Cunningham & Wang, 2008).

For example, demonstrations of duplex perception have shown that an object will group with a feature despite evidence based on spatial location that the feature does not belong to the object. Duplex perception was first demonstrated by Rand (1974). In this study, presentation of the formants in a consonant-vowel syllable was split across the two ears. A formant is a frequency band of increased intensity in a sound; there may be multiple formants in any given sound, the frequency of which can determine the identification of vowels (Handel, 1989). A formant transition is the change in frequency from a consonant to a vowel produced by the place of articulation of both (Handel, 1989). The frequency of the formant transition is a critical component for identification of consonants. It is possible to synthesize speech in which the transition of the third formant is the sole determinant of whether a syllable is heard as “da” or as “ga.”

In the original demonstration of duplex perception (Rand, 1974), the second and third formant transitions from a syllable, e.g., “da,” were presented to one ear while the rest of the syllable (i.e., the first formant and the remaining second and third formants) was presented to the other ear. This stimulus generated two simultaneous percepts (hence the term duplex perception): Listeners reported hearing a fully intact syllable in one ear and a nonspeech chirp-like sound in the other ear. The identity (“da” vs. “ga”) of the syllable was determined by the third formant transition. So, even though the critical feature for identification of the syllable, i.e., the third formant transition, was presented at a separate spatial location from the rest of the syllable, the feature was integrated with the other components to create a coherent, identifiable percept (while at the same time creating the separate percept of a chirp).

Duplex perception has been found to be surprisingly resistant to a variety of other manipulations of the third formant transition, such as onset time asynchrony (e.g., Bentin & Mann, 1990; Nygaard, 1993; Nygaard & Eimas, 1990), amplitude differences (Bentin & Mann, 1990; Cutting, 1976; Whalen & Liberman, 1987), fundamental frequency (Cutting, 1976), and periodicity differences (Repp & Bentin, 1984). The effect is so strong that it has even been found to occur when the isolated formant transition is not necessary to form a coherent percept: When the entire intact syllable is presented to one ear and a redundant transition is presented to the opposite ear, listeners report hearing a clearer percept of the syllable and a chirp located in the center of the head (see Nygaard & Eimas, 1990).

The duplex perception phenomenon is not limited to speech objects; it has been demonstrated with musical stimuli as well (e.g., Hall & Pastore, 1992; Pastore, Schmuckler, Rosenblum, & Szczesiul, 1983). For example, when two simultaneous piano notes are presented to one ear while a single note is presented simultaneously to the other ear, the resulting percept is of both the single tone and a fused chord (Pastore et al., 1983). Duplex perception also has been demonstrated with environmental sounds (see Fowler & Rosenblum, 1990 for a demonstration of duplex perception with slamming doors).

Given that there are extensive demonstrations of duplex perception showing that an object in figure will group with a feature despite cues that the feature belongs to a separate source (e.g., Whalen & Liberman, 1987), why then is there a recent case in which an object in figure will not group with a feature despite the fact that the spatial cues are perfectly aligned (Shinn-Cunningham et al., 2007)? One possibility is that the object does not group with the feature because it is already an identifiable object, i.e., a vowel, and incorporating the feature into the object would change the category of that object. Perhaps then, one rule for feature assignment to perceptual figure in multiple object contexts is that an object will not group with a feature if it does not need it. If this rule is true, then it could potentially explain one way in which the perceptual system maintains object stability in an unstable, constantly changing auditory world. Experiment 1 is designed to address this possibility. Experiment 1 first attempts to replicate the main finding of Shinn-Cunningham et al. (2007) and then attempts to determine if an object in perceptual figure will only group with a feature, which could belong to something in the background, if it needs it to form an identifiable object.

## Experiment 1a: Replication of Shinn-Cunningham et al. (2007)

### *Method*

#### *Participants*

Twenty four undergraduates from Stony Brook University (SBU) with no reported hearing deficiencies participated in this experiment. The mean age was 22.7 years. All listeners received course credit as compensation for their participation.

#### *Stimuli*

Stimuli consisted of two objects -- a vowel and a tone sequence -- and a target feature that could potentially contribute to the percept of each object. If the feature is incorporated into the vowel, the extra energy the feature creates at the vowel's first formant pushes the percept from /I/ (as in "pit") to /ε/ (as in "pet"). The presence/absence of the feature also affects the rhythmic properties of the tone sequence. If the feature is perceived as part of the tone sequence, the sequence has an "even" rhythm; otherwise it is perceived as "galloping."

Both objects and the feature were created by first generating a 60 msec vowel. The vowel was synthesized at a sampling rate of 44.1 kHz with a source filter synthesis algorithm using Praat (Boersma & Weenik, 2007). The fundamental frequency of the vowel was set at 125 Hz. The first three formants were 490, 2100, and 2900 Hz, and their corresponding bandwidths were 90, 110, and 170 Hz. The fourth harmonic, 500 Hz, was filtered out of the vowel (see Figure 1a). This 500 Hz component made up the feature and the individual tones of the tone sequence. The amplitude of the vowel and the tone was equated.

The vowel and the 500 Hz tone were combined in different temporal combinations to create 3 second long repeating sequences. As can be seen in Figure 1b, there were 10 repetitions of a 300 ms sequence consisting of two 60 msec tones and a 60 msec vowel. There was a 40 msec silent interval after each object in the sequence. In the target present condition, the target feature was played at the same time as the vowel.

Spatial cues were manipulated by presenting stimuli over 2 speakers: one arranged straight ahead of the listener and one $45^0$ to the right of the listener. Four different spatial configurations were used (see Figure 1c, bottom-left panel). The amplitude of the stimuli was roved over a 14 dB range to prevent listeners from using amplitude as a labeling cue. All stimuli were presented at a comfortable listening level, never exceeding 80 dB.

*Conditions.* Figure 1c depicts the conditions used in this experiment. As can be seen in the figure, the paradigm included both a single- and a two-object condition (Shinn-Cunningham et al., 2007). In the single-object condition, the vowel and the tone sequence were each presented either with the feature at the same spatial location (target present), or without the target feature (target absent). The single-object prototypes served as controls to ensure that the feature grouped with the object in the target present condition to produce the even rhythm percept, or with the vowel to produce the /ɛ/ percept. Both objects also were presented with the feature at a different spatial location (target location doesn't match).

In the two-object condition, both the vowel and the tone sequence were presented without the target feature (control, no target) and with the target feature (target present condition). As can be seen in Figure 1c, for the target present condition, across the four versions, the location of the target matched one object, both objects, or neither object.

*Procedure*

The procedure consisted of three sequential phases: familiarization, training, and test. During the familiarization phase, target present and target absent versions of both objects were presented while the correct label was displayed on the computer screen. For example, the vowel /I/ played while the label "/I/ as in pit" was displayed. The vowel and the tone sequence were presented in separate blocks, and within each block there was a 1 second ISI. There were 12 familiarization trials total (3 repetitions of target present/absent versions of both objects). Listeners were given the opportunity to repeat the familiarization phase before moving on with the experiment.

Listeners then completed a training phase that required them to label each object. Listeners indicated their labeling response (in this phase and in the test phase) by pressing one of two buttons on a button board labeled "i" or "e" for the vowel training and "galloping" or "even" for the tone sequence training. The labels "i as in pit" or "e as in pet" for the vowel or "galloping" or even" were also displayed on the computer screen. Listeners received feedback during training. The computer screen displayed “Correct” or “Wrong” after each response. The vowels and the tone sequences were presented in separate blocks, and there was a 500 msec inter-trial interval. There were at least 24 training trials total (6 repetitions of each object). Listeners were required to repeat the training phase until they could achieve a 90% accuracy rate. Most listeners required one repetition of the vowel training block and no repetitions of the tone training block.

Next, listeners completed the test phase. There were two blocks: an attend tones block and an attend-vowel block. In the attend tones block, listeners were instructed to listen to the tone sequences while ignoring any other sound that was presented. They were told to listen to the sequence on each trial and to label the sequence as “galloping” or as “even.” In the attend-vowel block, listeners were instructed to listen only to the vowel, while ignoring any other sounds. They were instructed to label the vowel they heard on each trial as “i as in pit” or as “e as in pet.”

In each block, the single- and two-object conditions were randomly intermixed. There were 8 conditions within each block. There were three single-object conditions: target absent, target present, and target present at a different location (see Figure 1c). And, there were five two-object conditions: target absent, target present at same location as object in figure, target present at same location as object in ground, target present at same location as both objects, and target present at a location matching neither object. There were 20 repetitions of each of the 8 conditions, resulting in 160 trials per block. Brief rest breaks were provided half-way through each block and in between blocks. The computer launched into the next trial 500 ms after each response, timing out after 5 seconds.

Order of blocks in all three phases (familiarization, training, and test), attend-tone or attend-vowel, was counterbalanced across listeners. The experiment was conducted in a sound attenuated chamber and took about 40 minutes to complete.

### *Results and Discussion*

As in the work by Shinn-Cunningham et al. (2007), an a priori criterion was set (here and in Experiments 2-4) excluding listeners who could not reliably distinguish between the single-object target present and target absent prototypes. Thus, a d' score of at least 0.7 between the target present and target absent prototypes was required in order for listeners to be included in the data analysis. All listeners met the criterion with the tone stimuli. Eleven listeners were excluded for failing to meet the criterion with the vowel stimuli. While a subject loss of nearly half (11 out of 24) may seem high, this rate is essentially the same as the attrition rate in Shinn-Cunningham et al. In their study, half of the subjects were excluded for failing to meet the vowel criterion. In the current study, one additional listener was excluded for failing to respond to 85% of the test trials. The following analyses were conducted on the remaining twelve listeners.

Accuracy (percent of correct labeling) was computed for each listener in each condition. As in the work by Shinn-Cunningham et al. (2007), signal detection theory was used (here and in Experiments 2-4) to analyze the accuracy scores. This analysis provided a measure of perceptual distance between each stimulus and the single-object prototypes.

The measure used to calculate the perceptual distance between the target-present and target-absent prototypes was $d'_{present:absent}$. This measure was calculated using the formula: $d'_{present:absent} = \Phi^{-1}$ [Pr ("target present" / target present)] - $\Phi^{-1}$ [Pr ("target present" / target absent)], where $\Phi^{-1}$ represents the inverse of the cumulative Gaussian distribution and [Pr ("target present" / stimulus)] represents the probability that the listener responded that the target was present, given the particular stimulus.

The perceptual distance between each stimulus and the single-object target-absent conditions was then calculated for each listener using the formula: $d'_{stimulus:absent} = \Phi^{-1}$ [Pr ("target present" / stimulus)] - $\Phi^{-1}$ [Pr ("target present" / target absent)]. To calculate the perceptual distance between a particular stimulus and the target present and target absent prototypes, the following formula was used: $\delta_{stimulus:absent} = d'_{stimulus:absent} / d'_{present:absent}$. A $\delta_{stimulus:absent}$ value of $< 0.5$ indicates that the stimulus was perceived as more like the target absent version of the object, while a $\delta_{stimulus:absent} > 0.5$ indicates that it was perceived more like the target present version of the object.

The results of the decision theory analysis are depicted in Figure 2. The Single-object condition (right-most point in both graphs) replicated Shinn-Cunningham et al. (2007). Both the vowel and the tone sequence grouped with the target feature, despite being presented at separate spatial locations (mean δ score was 0.81 for the tones and 0.63 for the vowel). This finding is consistent with other studies that also have found feature integration despite evidence that the feature does not belong to the object (e.g., Rand, 1974). Thus, it appears that the auditory perceptual system is quite tolerant of physical inconsistencies between features and objects when there are no other objects in the background that the feature could be attributed to.

The Two-object condition also replicated Shinn-Cunningham et al. (2007). Recall that the critical result in their study was that the target feature did not contribute to either object when it was presented at the same spatial location as the vowel (referred to as the critical condition here and in Experiments 2-3). The circled markers in Figure 2 represent the critical condition. As can be seen in the figure, the target feature did not contribute to the perception of either object. The feature grouped with the tone sequence at every spatial location except for the critical condition (mean $\delta = 0.41$). And, as shown in the bottom panel, the target feature did not contribute to the vowel in the critical condition (mean $\delta = 0.35$). Because the target feature did not contribute to the percept of the tone sequence or the vowel in this condition, it is as if the feature disappeared from perception.

## Experiment 1b: The Feature Reappears to Perception

Shinn-Cunningham et al. (2007) and Experiment 1a demonstrated that there are situations when two objects are competing for the same feature where neither object will take the feature. And, an object in perceptual figure (the vowel) will fail to group with the feature despite perfect spatial alignment between the two. The purpose of Experiment 1b was to determine if this finding occurred because the object in perceptual figure was already an identifiable object without the feature. This experiment used a modification of the Shinn-Cunningham et al. paradigm to examine whether an object in figure will group with the target feature if the feature is needed to form an identifiable percept. Critically, in this version, the vowel was only an identifiable object when it grouped with the target feature.

## *Method*

### *Participants*

Sixteen SBU undergraduates with no reported hearing deficiencies served as listeners for course credit. The mean age was 19 years. All listeners received course credit as compensation for their participation.

### *Stimuli*

As in Experiment 1a, a tone sequence, a vowel, and a target feature were used. The tone sequence was perceived as rhythmically even with the target feature, and as galloping without the target feature. However, instead of using a vowel stimulus that is perceived as /I/ without the feature and as /ɛ/ with the feature, a stimulus was used that is perceived as a non-identifiable speech sound without the target feature and as /I/ with the feature. Figure 3 depicts the vowel stimulus that was used. The non-identifiable speech sound was created by synthesizing a vowel using the same parameters as described in Experiment 1a, but without the first formant (the critical formant for vowel identification). Thus, the non-identifiable object had a fundamental frequency of 125 Hz and formants at 2100 and 2900 Hz with corresponding bandwidths of 110 and 170 Hz. As can be seen in Figure 3, the target in this experiment was the first formant of the vowel used in Experiment 1a, which peaked at 490 Hz (recall that the target in

Experiment 1a was the 500 Hz harmonic of the vowel). The target was created by synthesizing a vowel with only the first formant (490 Hz with a bandwidth of 90 Hz) on a fundamental frequency of 125 Hz. As in Experiment 1a, the target was used as the individual members of the tone sequence.

As in Experiment 1a, the target feature was combined with the objects in different temporal combinations to create 3 second long repeating sequences (see Figure 1a). There were 10 repetitions of a 300 ms sequence consisting of two 60 msec tones and a 60 msec non-identifiable speech sound. There was a 40 msec silent interval after each object in the sequence. In the target present condition, the target feature was played at the same time as the non-identifiable speech sound.

All other aspects of stimulus creation and presentation were as described in Experiment 1a.

*Conditions*

The conditions were the same as described in Experiment 1a (refer to Figure 1c).

*Procedure*

All aspects of the procedure were the same as described in Experiment 1a, except for the response button labels in the attend-vowel block. In this experiment, listeners were required to label stimuli in this block as “not a vowel” or as “/I/ as in pit.”

### *Results and Discussion*

All listeners met the a priori criterion with the tone stimuli. One listener was excluded for failing to meet this criterion with the vowel stimuli. In addition, one listener was excluded for failing to respond to 75% of the test trials, and two listeners were excluded because the experimenter failed to save their responses. The following analyses were conducted on the remaining twelve listeners.

Two ANOVAs, one for the attend-tone block and one for the attend-vowel block, were conducted to compare the δ scores from each condition of this experiment with the corresponding conditions in Experiment 1a (refer to Figure 1c for the 6 conditions in each block). Both ANOVAs produced a significant interaction between Experiment and Condition, attend-tones $F(5, 55) = 14.93$, $p < .01$, $\eta^2 = .58$, attend-vowel $F(5,55) = 9.14$, $p < .01$, $\eta^2 = .45$.

*Attend-Vowel.* The Single-object condition in the attend-vowel block was similar to Experiment 1a: The vowel grouped with the target feature when the two were presented at different spatial locations (mean δ = 1.22). In fact, the δ scores in this condition were significantly higher, i.e., more like the target-present prototype, than in Experiment 1a, $p < .01$.

The central question this experiment intended to address was whether the target feature would contribute to the vowel in the Two-object condition in which the feature had previously disappeared. As can be seen in the circled marker in the bottom panel of Figure 4, the target feature re-appeared to perception. The feature contributed to the perception of the vowel in the critical condition (mean δ = 1.21) significantly more than in Experiment 1a (mean δ = 0.35), $p < .01$. In fact, the feature contributed to the perception of the vowel in every Two-object target present condition, except for when the feature location matched the tone sequence (left-most point in Figure 4).

In summary, the use of a stimulus that needed the feature in order to be perceived as an object caused the feature to reappear to perception, except in the most extreme case in which the feature was presented at the same location as the object in the background. This result contrasts with the findings of Shinn-Cunningham et al. (2007) and Experiment 1a. Perhaps the finding of a disappearing feature in both of those studies occurred because the object in figure simply did not need the feature to form an identifiable object. When an object *does* need the feature to form a coherent percept, it may take it from any point in space that is temporally concurrent, unless there is evidence that the feature belongs to an object in the background. This finding supports Shinn-Cunningham et al. claim that more evidence is needed to pull a feature into perceptual figure than to reject it to the perceptual ground.

*Attend-Tones.* The Single-object condition in the attend-tone block was different than Experiment 1a. The tone sequence did not group with the target feature when the two were presented at different spatial locations (mean $\delta$ = 0.36). Planned comparisons confirmed that the $\delta$ scores in this condition were significantly lower than in Experiment 1a, $p < .01$.

Though this condition is not of particular interest, a finding opposite of the results of Shinn-Cunningham et al. (2007) and Experiment 1a warrants a potential explanation. The tone sequence stimuli used in the present experiment were different than the tone sequence stimuli used in Shinn-Cunningham et al.

and Experiment 1a. The stimuli used in this experiment were the frequency components making up the first formant of a vowel, rather than a 500 Hz tone. Recall that the individual tone sequence components and the target feature are the same stimulus. In all three experiments, two different percepts can occur when the tone sequence and the feature are presented at separate locations. The two can fuse into one rhythm of tones that repeats every 60 msec (as it did in Shinn-Cunningham et al. and in Experiment 1a). Or, the two can segregate into two different rhythms: a sequence consisting of two tones that repeat every 140 msec and a one-tone sequence that repeats every 240 msec. The limited frequency range used in Experiment 1a made the fused percept more likely to occur, and the additional frequency components used in this experiment provided cues that made the segregated percept more likely to occur.

The central finding in the Two-object condition in the attend-tone block was again replicated. The target feature did not contribute to the perception of the tone sequence in the critical condition (see circled marker in the top panel of Figure 4). In fact, the feature contributed less in this condition than in did in Experiment 1, $\delta = 0.07$ vs. $\delta = 0.41$, $p < .01$.

The results of the other Two-Object conditions were slightly different than Experiment 1a. Recall that in Experiment 1a, the target feature contributed to the perception of the tones in every target present condition, except for the critical condition. In this experiment, the target feature only contributed to the perception of the tone sequence when both the tone sequence and the feature were presented at the same spatial location in this experiment (even when the other object was also presented at this location). Planned comparisons confirmed that δ scores when the tone sequence and the target were presented at different locations were lower than in Experiment 1a, *p's* < .05. This finding is similar to the Single-object condition and is probably attributable to the nature of the tone components making segregation more likely.

## Experiment 2: The Effect of Feature Variation on Perceptual Coherence

Experiment 1a demonstrated that a feature can disappear from perception, and Experiment 1b revealed that the feature will reappear to perception when it is needed to form an identifiable object. Experiment 2 examines the fate of the feature when the feature is altered relative to the objects in figure and ground. Natural listening situations are often complicated by many overlapping sounds, so having sound components that are convoluted by other sound sources is a common problem for the auditory system. One way to investigate how the auditory object construction process solves this problem is to measure the effect of a modified feature on that feature's contribution to the perception of the object in figure.

Thus, the purpose of Experiment 2 was to determine how a modified feature affects perceptual coherence between the feature and an object. Experiment 2a tested the possibility that a modification that occurs in many natural listening situations, noise masking, will affect feature and object binding. Experiment 2b tested the possibility that a modification that does not occur in natural listening situations, mistuning of sound components, will affect the likelihood of a feature binding with an object. The central question this experiment addressed was how variation of the target feature affects the probability that the feature will group with an object in figure. This question was addressed by examining how the level and type of variation of the feature affects perceptual coherence between the feature and object.

## *Method*

### *Participants*

Twelve SBU undergraduates (mean age = 19.2) with no reported hearing deficiencies participated in this experiment for course credit.

### *Stimuli*

The target feature, tone sequence, and vowel stimuli from Experiment 1b were used in this experiment. The tone sequence was perceived as rhythmically even with the target feature, and as galloping without the target feature. The vowel was perceived as a non-identifiable speech sound without the target feature and as /I/ with the feature.

Stimuli for the noise manipulation were created by synthesizing a 60 msec segment of white noise using a preset function in Cool Edit Pro (version 2.1). Three different levels of noise were generated: noise 10dB lower than the target feature, noise equal in amplitude to the feature, and noise 10dB higher than the feature.

Stimuli for the mistuning manipulation were created by shifting the pitch of the target feature both upward (+) and downward (-) using Cool Edit Pro. Previous research has demonstrated, with single-object conditions, that mistuning a feature by 8% is sufficient to remove the feature from an object, i.e., to cause the feature to no longer contribute to the identity of the object (see Darwin & Gardner, 1986). The pitch of the target feature was shifted by +/- 8, 16, and 32% in this experiment.

*Conditions*

The conditions from Experiment 1a were modified to include the noise and the mistuning manipulation. The conditions are depicted in Figure 5. As can be seen in Figure 5a, both the Single- and Two-object conditions were modified to include three levels of noise masking of the target feature: low (10dB lower than feature), medium (same level as feature), and high (10dB higher than feature). Figure 5b depicts the mistuning manipulation. The conditions were modified to include three levels of mistuning: low (8%), medium (16%), and high (32%). These conditions test how altering the target feature affects the probability of the feature binding with each of the objects. The Single-object target present/absent conditions from Experiment 1b (with no noise or mistuning manipulation) also were used to allow a comparison of how each modified condition was perceived relative to the Single-object prototypes.

*Procedure*

Each manipulation, noise and mistuning, and each block, attend-vowel and attend-tones, consisted of 20 conditions (the 18 conditions depicted in Figure 5 for each object plus the two Single-object target present/absent conditions). There were 20 repetitions of each condition. The result was 1600 trials: 400 trials in the attend-tones noise manipulation block, 400 in the attend-vowel noise manipulation block, 400 in the attend-tones mistuning manipulation, and 400 in the attend-vowel mistuning manipulation block, The 1600 trials were presented over a series of 4 days. Each listener completed 400 trials a day for four consecutive days. On each day, listeners completed one of the manipulations (noise or mistuning) and one of the blocks (attend-tones or attend-vowel). A balanced Latin square was used to determine the order of manipulation and block for each participant. Rest breaks were given after every 100 trials. All other procedures were the same as described in Experiment 1b.

*Results and Discussion*

One listener failed to meet the a priori criterion for the attend-vowel block and was excluded from the data analyses. An additional listener was excluded for failing to respond to any two-object condition. The following results are reported for the remaining ten listeners.

*How does the level of variation affect perceptual coherence?* Previous research has demonstrated that the auditory system often makes use of inconsistent information for perceptual coherence (e.g., Rand, 1974; Shinn-Cunningham et al., 2007). Thus, it was predicted that the system will be quite tolerant of the feature variation at the lower levels of noise masking and mistuning, but will become less tolerant as both manipulations become more extreme. Tolerance of feature variation is reflected in the δ scores. High δ scores indicate that the system was quite tolerant, given that the feature contributes to the perception of the object when δ scores are greater than 0.5, and low δ scores indicate that the system was not very tolerant, given that δ scores less than 0.5 indicate that the feature does not contribute to the perception of the object. To test the prediction, ANOVA's were conducted with Object (tone sequence or vowel), Condition and Manipulation Level (low, medium, or high) as factors on the δ scores from both the mistuning and the noise manipulations.

The prediction was supported by both analyses. The mistuning manipulation revealed a significant effect of Level, $F(2, 18) = 33.56$, $p < .01$, $\eta 2 = .79$. Planned comparisons indicated that the system tolerated the low and medium levels (mean δ scores were .51 and .51) of mistuning better than the high level of mistuning (mean δ = .31), *p's* < .05. The noise manipulation also revealed a significant effect of Level, $F(2, 18) = 8.89$, $p < .01$, $\eta 2 = .50$. Planned comparisons indicated that the system tolerated the low (mean δ= .59) and medium (mean δ =.37) levels of noise better than the high (mean δ = .04) level of noise, *p's* < .01.

The finding that both objects grouped with the modified feature (at low and moderate levels) suggests that the auditory system can be flexible in interpreting evidence that two things may belong together. In this experiment, the system made use of incomplete (feature masked by noise) and inconsistent (feature with a different pitch) information to accomplish complete object perception.

*How does the type of variation affect perceptual coherence?* It was predicted that the auditory system would be more tolerant of the feature manipulation that could occur in a natural listening situation, the noise masking, than of the feature manipulation that is not likely to occur in a natural listening situation, the mistuning. The hypothesis was tested with ANOVAs conducted on the δ scores from both the tone sequence and the vowel with Manipulation Type (noise or mistuning), Condition and Level (low, medium, or high) as factors.

Figure 6 depicts the results collapsing across the level factor. The "x" markers in the figure represent the δ scores from the attend-tones block, and the diamond markers represent the δ scores from the attend-vowel block. As can be seen in the figure, the answer to the question of which type of variation is tolerated better by the system depends on whether the object in figure was the tone sequence or the vowel.

When the object in figure was the tone sequence, the system tolerated the noise manipulation better than the mistuning manipulation (δ = .48 vs. .35). This trend can be seen in Figure 6: The δ scores from the noise manipulation were higher (x markers in Figure 6a) than the δ scores from the mistuning manipulation (x markers in Figure 6b). This difference contributed to a significant effect of Manipulation Type, $F(1,9) = 16.52$, $p < .01$, $\eta 2 = .65$ and a significant interaction between Manipulation Type and Condition, $F(5, 45) = 10.98$, $p < .01$, $\eta^2 = .55$. Planned comparisons revealed that noise was tolerated better than mistuning in 4 of the 6 conditions, *p's* < .05. The δ scores from the condition of particular interest in this experiment, i.e., the critical condition, were similar to the δ scores from the corresponding condition in Experiments 1a and 1b. As in those experiments, the feature did not contribute to the perception of the tone sequence; this occurred both when the feature was modified by noise and by mistuning.

When the object in figure was the vowel, however, the noise manipulation resulted in the feature disappearing from perception in the critical condition (see circled diamond marker in Figure 6a; mean $\delta = .16$). When the feature was modified by mistuning, however, the feature did contribute to the perception of the vowel in the critical condition (see circled diamond marker in Figure 6b; mean $\delta = .77$). In addition to the critical condition, the system tolerated the mistuning manipulation better than the noise manipulation in the majority of the conditions when the vowel was in perceptual figure ($\delta = .54$ vs. .18). This difference contributed to a significant interaction between Manipulation Type and Condition, $F(5,45) = 7.94$, $p < .01$, $\eta 2 = .47$. The system tolerated the mistuning manipulation better in 5 of the 6 conditions, *p's* < .05.

In summary, the answer to the question of how the type of feature variation affects perceptual coherence is: It depends on which object is in perceptual figure. When the tone sequence was in figure the system tolerated the noise better, and the vowel was in figure the system tolerated the mistuning better. Why would there be differences in the two objects? Recall that the vowel needs the feature in order to be an identifiable object. So, one potential reason for this finding could be that the vowel's need of the feature caused the system to be more tolerant of any kind of feature variation, natural or not, unless the feature is masked so much that it cannot be perceived (which may have been the case in the noise manipulation).

## Experiment 3: Properties of Perceptual Ground that Affect Perceptual Coherence

Experiments 1 and 2 established that one influence on feature assignment to objects in perceptual figure is that assignment depends on whether features are needed, but there still remains a discrepancy in prior research. A single object will group with a feature regardless of whether the feature is needed or not, but the same object will not group with the same feature when there is another object present in the background (Shinn-Cunningham et al., 2007). The discrepancy between one- and two- object situations suggests that there may be a caveat to the rule that features are assigned when they are needed. The caveat may be that the rule only applies if there is an object in the background that the feature could possibly be attributed to if it is not grouped with the object in figure.

There is evidence to suggest that an object in figure is less likely to group with a feature if there are other objects in the background that the feature could potentially belong to. Several studies have found that when an already identifiable object is presented alone with a target feature, the two will be integrated even though the integration pushes the perception of the object to a different category (e.g., Darwin, 1983; Darwin & Sutherland, 1984; Shinn-Cunningham et al., 2007). However, an object in figure is less likely to group with a target feature that could change the category of the object if there is a perceptual group in the background that the feature could be a member of. For example, if several tones that are the same frequency as the feature are added before and after the rest of the auditory display, the sequence will capture the feature, causing it to contribute less or not at all to identification of the object (Ciocca & Bregman, 1989; Darwin, Pattison, & Gardner, 1989; Darwin & Sutherland, 1984).

Experiment 3 was conducted to further explore this potential exception to the rule. The central question this experiment addresses is: Does the probability that a feature will contribute to the object in figure vary as a function of whether the object in figure needs the feature and whether the object in ground needs the feature? Experiment 3 was a modification of Experiment 1b with four new conditions that vary in feature compatibility with the object in figure and the object in ground.

## *Method*

### *Participants*

Twenty SBU undergraduates (mean age = 19.2 years) with no self-reported hearing deficiencies served as listeners for this experiment. All listeners received course credit for their participation.

### *Stimuli*

The target feature, tone sequence, and vowel from Experiment 1b were used. Two different vowels, one that needed the feature and one that did not need the feature and two different tone sequences, one that needed the feature and one that did not need the feature, were created from these stimuli. The vowel that needed the feature was the same target-absent stimulus used in Experiment 1b. This vowel was perceived as a non-identifiable sound without the target feature. The target feature was added to this sound in the target present version of this vowel, which resulted in the percept of the vowel /I/. The vowel that did *not* need the feature was the vowel /I/ in the target absent condition. In the target present version of this vowel, the feature was added to the vowel /I/. The result was the percept of the vowel /ɛ/. Thus, the feature was a redundant feature (see Figure 7a) with twice as much energy than in the corresponding condition in the vowel that needed the feature.

The tone sequence stimuli were created in a similar manner. The tone sequence that needed the feature resulted in a “galloping” percept in the target absent condition. The feature was added to this sequence in the target present condition to create the percept of an “even” rhythm. The tone sequence that did *not* need the feature was the same “even” rhythm in the target absent condition. The target present version was created by adding the feature to the “even” sequence. The resulting percept was still “even.” As with the vowel, the added feature was a redundant feature with twice as much energy as the target present feature in the tone sequence that needed the feature (see Figure 7a).

*Conditions*

The Two-object conditions from Experiment 1b were modified to include four different relationships between the figure and ground: figure needs feature/ground does not need feature, figure needs feature/ground needs feature, figure does not need feature/ground does not need feature, and figure does not need feature/ground needs feature. As depicted in Figure 7b, the two-object conditions from Experiment 1b were repeated four times, once for each figure/ground configuration. As Figure 7a shows, the Single-object conditions from Experiment 1b were repeated twice for each object to incorporate both types of figure objects: need and no need.

The vowel that needed the target feature in order to be perceived as an identifiable object was perceived as a non-identifiable speech sound when the target feature was absent and as the vowel /I/ when the target feature was present. The vowel that did not need the feature was perceived as /I/ without the feature and as /ε/ with the feature. The tone sequence that needed the feature was perceived as galloping without the feature and as even with the feature. The tone sequence that did not need the feature was perceived as even with and without the feature. The feature made no contribution to the category of the tone sequence because it was a redundant feature (see Figure 7a).

*Procedure*

During the attend-vowel block, listeners were required to choose one of three responses, "/ε/ as in pet," "/I/ as in pit," or "not a vowel." There were 20 repetitions of the 23 conditions in each block. Thus, there were 460 trials in the attend-tone block and 460 trials in the attend-vowel block. The experiment was split up over two consecutive days of 460 trials each. Order of blocks, attend-tone or attend-vowel, was counterbalanced across subjects. Rest breaks were given after every 115 trials. All other procedures were as described in Experiment 1b.

### *Results and Discussion*

Twelve listeners failed to meet the a priori criterion for the vowel stimuli, and thus were excluded from all data analyses. The following results are for the remaining eight listeners.

Previous research has provided evidence that an object in figure is less likely to group with a feature if there are other objects in the background that could potentially take the feature (e.g., Darwin, 1983). Thus, it was predicted that the object in figure would be more likely to group with the target feature when it is not likely that the feature belongs to the object in perceptual ground. To test this prediction, an ANOVA with Condition (5 conditions) and Figure/Ground Relationship (figure needs feature/ground does not need feature, figure needs feature/ground needs feature, figure does not need feature/ground does not need feature, and figure does not need feature/ground does need feature) was conducted on the attend-vowel and attend-tones blocks.

In the attend-vowel block, the analysis revealed a significant interaction, $F$ (12, 84) = 6.79, $p < .01$, $\eta 2 = .49$. Overall, the feature was more likely to contribute to the perception of the vowel that needed the feature than to the perception of the vowel that did not need the feature, p's < .05. The contribution of the feature to each vowel as a function of whether the feature could contribute to the object in ground was slightly different (see Figure 8a).

*Vowel that Needed the Feature*. The feature contributed to the perception of the vowel in the critical condition. As the circled markers in the top panel of Figure 8a show, this finding occurred when the feature could contribute to the object in ground *and* when the feature could not contribute to the object in ground, *n.s*. As can be seen in the markers to the left of the critical condition, the feature also contributed to the perception of the vowel in the other condition in which the feature was presented at the same location as the vowel (i.e., when the feature was presented at the same location as both objects) regardless of whether the feature could contribute to the object in ground, *n.s.*. Thus, the results of the conditions in which the feature location matched the vowel did not support the hypothesis. The feature contributed to the perception of the vowel when the two were presented at the same location regardless of whether the feature could contribute to the object in ground.

The hypothesis was supported, however, in the two conditions in which the location of the feature did not match the vowel. When the location of the feature matched the object in ground or neither of the objects, the feature was more likely to contribute to the perception of the object in figure when the feature could not contribute to the object in ground. The differences between the x and diamond markers in the first two conditions in the top panel of Figure 8a were significant, $p's < .05$. Overall, when the vowel needed the feature, spatial location cues were sufficient for the system to group the feature and object together, even when there was an object in the background that could potentially "take" the feature. When spatial cues between the feature and the object in figure were not consistent, the auditory system seemed to consider the structure of the object in the background, allocating the feature to the figure less when the feature could contribute to the object in ground.

*Vowel That Did Not Need the Feature.* The feature contributed to the perception of the vowel in the critical condition. As the circled markers in the bottom panel of Figure 8a show, this finding occurred only when the feature could also contribute to the object in ground. This finding is in the opposite direction of the hypothesis, and is opposite of the results of Shinn-Cunningham et al. (2007) and Experiment 1a, where the feature had disappeared from perception. This unexpected finding may be a result of stimuli differences between the experiments. As described above (see Stimuli), the target present version of the vowel had twice as much energy at the target feature as the original vowel used by Shinn-Cunningham et al. (and in Experiment 1a). The extra energy in this experiment may have made the category of the vowel more salient.

The hypothesis was supported in the other two-object conditions. The feature was more likely to contribute to the perception of the vowel when it could not contribute to the object in ground (see Figure 8a). Planned comparisons revealed that the δ scores in these conditions were higher when the feature could not contribute to the object in ground than when the feature could contribute to the object in ground, *p's* < .05.

The results from both vowels suggest that the feature's potential contribution to an object in ground affects that feature's contribution to the perception of the object in figure when spatial cues are ambiguous, i.e., when the feature is presented at a location that does not match the object in figure. When the feature and object in figure are presented at the same location, the object in figure "takes" the feature regardless of whether it can contribute to the object in ground.

In the attend-tones block, there was a significant interaction between Condition and Figure/Ground Relationship, $F\,(15, 135) = 20.35$, $p < .01$, $\eta 2 = .69$. Though there was a significant interaction, the results of the attend-tone block were generally in the opposite direction than the results of the attend-vowel block. This finding was not surprising, however, because the object "needing" the feature actually applies only to the vowel in this paradigm. The vowel needs the feature in order to be a meaningful object. The tone sequence, however, is a sequence with or without the target feature. The only thing that changes between target present and absent conditions is the rhythm of the sequence.

*Tones that Needed the Feature*. As in Experiments 1-2, the feature did not contribute to the perception of the tones in the critical condition, i.e., when the feature was presented at the same location as the vowel. As the circled markers in the top panel of Figure 8b show, this finding occurred when the feature could contribute to the object in ground *and* when the feature could not contribute to the object in ground, *n.s*. The feature's potential contribution to the object in ground did affect the results of the other two-object conditions. When the feature matched the location of the tones and when it matched the location of both objects, the results were in the opposite of the predicted direction. The feature was more likely to contribute to the tone sequence when the feature could also contribute to the object in ground (see x markers versus diamond markers in the top of Figure 8b). When the location of the feature matched neither object, the feature was more likely to contribute to the perception of the tone sequence when the feature could not contribute to the perception of the object in ground, $p < .01$.

*Tones that Did Not Need the Feature.* In contrast to Experiments 1-2, the feature did contribute to the perception of the tones in the critical condition. As the circled markers in the bottom panel of Figure 8b show, this finding occurred when the feature could contribute to the object in ground *and* when the feature could not contribute to the object in ground, *n.s*. Recall that when the vowel did not need the feature, the feature also contributed to the perception of the vowel in the critical condition (when the feature could also contribute to the object in ground). Thus, for the first time, there is a case of feature sharing between the two objects in this paradigm. There were no differences in the δ scores when the feature could contribute to the ground and when the feature could not contribute to the ground in all other two-object conditions.

Experiment 4: Multiple Objects Competing for the Same Feature

Experiments 1 – 3 are helpful in establishing the rules for feature binding in perceptual figure, but the information that they can provide is limited to two potential uses for a feature. Natural listening situations rarely involve such a simple scenario. Experiment 4 provides a test of the generality of the rules for feature and object binding: Will the same pattern of results found in Experiment 1b hold when there are three uses for the same feature? Experiment 4 examines the contribution of a target feature to an object when there are three potential uses for the feature.

*Method*

*Participants*

Twenty-six SBU undergraduates (mean age = 20.2 years) served as listeners in this experiment. No listener reported any hearing deficiency. All listeners received course credit for their participation.

*Stimuli*

The target feature from Experiment 1b was used. Recall that this feature was the first formant of the vowel /I/, which peaked at 490 Hz. In addition, three objects were used as stimuli: a tone sequence and two vowels. The tone sequence and one of the vowels were the same as in Experiment 1b. The tone sequence is perceived as “even” if it groups with the target feature and as “galloping” if it does not. The vowel, called Vowel 1 in this experiment, created the percept of “/I/” if it groups with the feature and a non-speech like sound, “not a vowel” if it does not group with the feature. The second vowel was the same vowel used in Experiment 1a. This vowel, called Vowel 2, creates an “/ε/” percept if it groups with the feature and “/I/” if it does not group with the feature.

*Conditions*

Figure 9 depicts the conditions that were used in this experiment. As can be seen in the figure, the Single-object condition consisted of each of the three objects presented alone with and without the target feature. On target present trials, the feature was presented at the same spatial location as the object (to promote grouping) or at one of the three different spatial locations (to discourage grouping). All three objects were presented simultaneously in the Three-object condition. As can be seen in Figure 9, the three objects and the target feature were presented at various spatial locations. The target feature was presented at the same spatial location as the attended object, at a spatial location matching one of the other objects, or at a spatial location that did not match any of the three objects.

*Procedure*

There were three separate blocks. In each block, the task was to attend to one of the objects: the tone sequence, Vowel 1, or Vowel 2. The attended object was always presented at a speaker located directly in front of the listener. In the attend-tone block, the task was to label the tone sequence as "even" or as "galloping" by pressing the corresponding key on a button board. In the attend-Vowel 1 block, the task was to label the vowel as "/I/ as in pit" or as "not a vowel." In the attend Vowel 2 block, the task was to label the vowel as "/ε/ as in pet" or as "/I/ as in pit." There were four possible spatial locations: the center (where the attended object was always presented), $45^0$ to the left, $90^0$ to the right, and $90^0$ to the left.

There were 7 conditions in each block: Single-object target present, Single-object target absent, Single-object target location does not match, Three-object target location matches object in figure, Three-object target location matches one the first object in ground, Three-object target location matches the second object in ground, and Three-object target location matches no object. There were 24 repetitions of each condition in each block, except for the two conditions in which the target location matched one of the objects in ground. This condition was split so that there were 12 repetitions for each object in ground, e.g., in the Attend-Vowel 1 block, there were 12 repetitions of the feature being presented at the same location as Vowel 2 and 12 repetitions of the feature being presented at the same location as the Tones. There was a total of 120 trials per block. Order of blocks was counterbalanced across listeners. Listeners received a rest break after each block. All other procedures were as described in Experiment 1.

*Results and Discussion*

Thirteen listeners failed to meet the a priori criterion for Vowel 2 and one listener failed to meet the criterion for Vowel 1. The following analyses are reported for the remaining twelve listeners.

The contribution of the feature to each of the objects was tested by conducting an ANOVA with Condition as a factor on the δ scores for each block. Recall from the description of the signal-detection theory analysis in Experiment 1a that the Single-object target present/absent conditions are used to calculate the δ scores in the other conditions. As a result, there are no δ scores for these two conditions and they are not included in the ANOVA. The remaining five conditions were submitted to the ANOVA. The results are depicted in Figure 10.

*Attend-Tones.* As can be seen in the top of Figure 10, the feature only contributed to the perception of the tone sequence when the two were presented at the same spatial location (see circled marker in figure). This trend contributed to a significant effect of Condition, $F$ (4, 44) = 89.16, $p < .01$, $\eta 2 = .89$. Planned comparisons confirmed that the δ scores in the Three-object condition in which the target feature matched the location of the tones was significantly higher than the δ scores in all other conditions, *p's* $< .01$. This finding is quite different from the results of Experiment 1b. Recall that in Experiment 1b the target feature contributed to the tone sequence when the feature matched the location of the tones and when it matched the location of both objects: the tones in figure and the vowel in ground. Thus, it appears that when three objects are competing for the same feature, assignment of the feature to the tone sequence is more selective than when two objects are competing for the feature.

*Attend-Vowel 1.* Similarly, the target feature only contributed to the perception of Vowel 1 when the two were presented at the same spatial location (see circled marker in middle of Figure 10). The ANOVA conducted on the attend-Vowel 1 block revealed a significant effect of Condition, $F\ (4, 44) = 9.53$, $p < .01$, $\eta 2 = .46$. Planned comparisons confirmed that the δ scores in the Three-object condition in which the target feature location matched Vowel 1 were significantly higher than the δ scores in all other conditions, *p's* $< .05$. Recall that in Experiment 1b, the feature contributed to the perception of the vowel in 5 of the 6 conditions. As with the tone sequence, the auditory system's assignment of the target feature to Vowel 1 was not as flexible when three objects were competing for the feature. The feature only contributed to the object in figure when there was evidence based on spatial location that the feature belonged only to the object.

*Attend-Vowel 2.* As can be seen in the bottom of Figure 10, the feature never contributed to the perception of the vowel. There were no differences in the δ scores across the Conditions, $F(4, 44) = 1.09$, $p > .05$. This finding is not surprising: The target feature never contributed to this vowel in Shinn-Cunningham et al. (2007) experiment with two objects and only contributed to one of six conditions in Experiment 1a. This vowel was already an identifiable object (the vowel /I/) without the feature, which is presumably why the feature was not used by it.

In summary, the target feature only contributed to the perception of two of the objects, the tone sequence and Vowel 1, and only when there was the best evidence that the two belonged together. The target feature only contributed to the object in figure when the feature was presented at the same location as the object in figure *and* when neither of the ground objects was presented at that same location. This finding is similar to Experiments 1 – 3 in that the feature is more likely to contribute to an object in figure if the feature and the object are presented at the same location. However, spatial consistency was not necessary for feature and object grouping in Experiments 1-3 as it was in the present experiment. The results of this experiment suggest that the rules for feature assignment to auditory figure become more stringent when three objects are competing for the same feature, as opposed to two.

### *Feature Assignment in Auditory Ground: Experiments 5-7*

William James's (1890) famous quote, "a baby's impression of the world is one great blooming, buzzing confusion" illustrates the complexity involved in our ability to integrate a select set of features from the “blooming, buzzing confusion” in any particular environmental situation. James’s quote hinted that sensory features are at some point unorganized and ill-defined. The notion that features in the environment are free-floating developed into an important theoretical claim in Treisman’s feature integration theory (e.g., Treisman & Gelade, 1980). According to feature integration theory, there are two steps involved in visual object perception: An initial, automatic, preattentive stage that processes free-floating features in parallel, and a second, serial stage where features are put together to form objects via the allocation of focal attention. Demonstrations of illusory conjunctions (e.g., Treisman & Schmidt, 1982) have been taken as evidence for the free-floating nature of features in perceptual ground. Illusory conjunctions involve the incorrect combination of two features, such as a color and a shape, and presumably occur because two or more features from the free-floating field may be incorrectly combined. Though many assumptions of feature integration theory have been challenged (e.g., Duncan & Humphreys, 1989; Wolfe, Cave, & Franzel, 1989), it has been one of the most prominent and influential theories of object perception, and is still the foundation for many current theories of object perception.

The idea of an unorganized feature space in perceptual ground was recently invoked by Shinn-Cunningham et al. (2007), who have also suggested that the fate of features in perceptual ground is a free-floating perceptual limbo. The primary motivation for Experiments 5-7 is to explore the notion of a perceptual limbo. Shinn-Cunningham et al. found that features in perceptual ground seem to disappear. The experiments in this section attempt to determine where such features go. If perception is not veridical, then features that are not used in perceptual figure really may disappear into a perceptual limbo. Experiment 5 will address this possibility.

## Experiment 5: Assignment of Features to Objects in Perceptual Ground

The purpose of Experiment 5 is to examine the nature of perceptual ground. The finding that a feature disappears from perception has been taken as evidence for a perceptual limbo for features that do not contribute to objects in perceptual figure (Shinn-Cunningham et al., 2007). This intriguing notion may very well be true, but there has been no direct test of the existence of a limbo because perception of objects in the background has not been measured. The approach in Experiment 5 is to try to determine if there is actually a perceptual limbo by examining whether or not features are assigned to objects in perceptual ground. If features are actually assigned to objects in the background, then the notion of a limbo in its extreme form is not likely to be correct. However, if features are not assigned to objects in the background, then there may be some validity to the idea of a floating feature field.

The paradigm used to determine if features are assigned to objects in perceptual ground included multiple phases. The two phases of particular interest were an initial labeling task in which each member of a vowel continuum was presented to listeners and a second labeling task in which the same stimuli were presented simultaneously with one of five backgrounds presented at a separate spatial location. The backgrounds consisted of target present and target absent conditions. The logic of the design is as follows: If the target feature is assigned to an object in the vowel background, the category of the percept that integration of the feature with the background object creates should affect the primary task when the two are consistent. For example, if the target feature is integrated into the vowel in the background, the percept that results from that integration should affect the labeling responses, relative to the labeling responses from the primary task without a background. For example, if the background forms the vowel object /ɛ/, judgments of "pit" or "pet" may become biased toward "pet," increasing the percentage of /ɛ/ responses.

## *Method*

### *Participants*

Twenty-eight SBU undergraduates (mean age = 19.9 years) with no self-reported hearing deficiencies served as listeners for this experiment. All listeners received course credit for their participation.

### *Stimuli*

Figure 11 depicts examples of the stimuli that were used. For the primary task, seven stimuli were used; each stimulus was a 60 msec member of a vowel continuum that changed in equal steps along F1 (i.e., the first formant). All other formants being equal, the average center frequency of F1 determines whether a vowel is heard as /I/ as in "pit," or /ɛ/ as in "pet." The continuum stimuli were created using a source-filter algorithm in Praat (Boersma & Weenik, 2007). Each stimulus was synthesized on a fundamental frequency of 125 Hz at a sampling rate of 44.1 kHz. The first member of the continuum, and the most /I/ like, had formants at 375, 2100, and 2900 Hz. The corresponding bandwidths were 90, 110, and 170 Hz. The second and third formants and the bandwidth values were constant for all members of the continuum. Only the first formant value changed in 21 Hz steps. The last member of the continuum, and the most /ε/ like, had formants at 501, 2100, and 2900 Hz.

The background stimuli consisted of target present and target absent conditions (see Figure 11). All background stimuli were 40 msec in duration. The first background (vowel background 1) consisted of a spectral structure that results in the percept of the vowel /I/. If the target feature is present and it groups with the vowel, the vowel percept will change to /ε/. Because there is a possibility that the feature may not group with the vowel if the vowel is already an identifiable object, there was a second version of the vowel background (vowel background 2). In this version, if the target feature is present and it groups with the vowel, the percept will be /I/; otherwise the percept will be a non-speech like sound. Both vowels and the target feature were the same as those used in Experiment 4.

The control background stimuli were created by band-pass filtering each of the following frequency components from the vowel background 1: 115, 200, and 175 Hz. The result of this combination of frequencies was a buzz-like sound. There were target present (using the same target as the experimental backgrounds) and target absent control backgrounds.

Five additional 40 msec stimuli were created for a surprise memory test (see details of the test below). All three were synthesized with a fundamental frequency of 125 Hz using a source-filter synthesis algorithm in Praat (Boersma & Weenik, 2007). Two vowels that were categorically distinct from the experimental stimuli were selected to serve as foils: /a/ was selected as a foil for the vowel /ε/ and /u/ was selected as a foil for the vowel /I/. The vowel /a/ was the foil for the target present stimulus in *vowel background 1*. The vowel was synthesized with formants values at 730, 1090, and 2440 Hz (Peterson & Barney, 1952), and corresponding bandwidths of 90, 110, and 170 Hz. The vowel /u/ was the foil for the target absent stimulus in *vowel background 1* and for the target present stimulus in *vowel background 2*. The vowel was synthesized with formant values at 300, 870, and 2240 Hz (Peterson & Barney, 1952) and corresponding bandwidths of 90, 110, and 170 Hz. The foil for the target absent version of *vowel background 2,* i.e., the "not a vowel" sound, was also a non-speech sound created by band-stop filtering the first formant, which peaked at 300 Hz, out of the vowel /u/. The foil stimuli for the control backgrounds were frequency sequences that were distinct from the control backgrounds. The target present control foil was created by band-pass filtering each of the following frequency components from the vowel background 1: 85, 125, 150, and 165 Hz. The target absent foil consisted of the frequencies: 85, 150, and 165 Hz.

*Procedure*

Listeners completed the following phases (listed in order) within one hour long testing session. In all tasks, stimuli were presented over headphones at a comfortable listening level of approximately 80 dB. Stimuli were presented to either the left ear or to the right ear. Listeners were instructed to attend to either the right or the left ear throughout the entire experiment. The attended-ear was counterbalanced across listeners. There was a 500 msec inter-trial interval in all phases.

*Familiarization:* Listeners were first familiarized with each member of the continuum. Familiarization began with the lowest F1 end of the continuum and proceeded in order, ending with the highest F1 end of the continuum. The computer screen displayed “/I/ as in pit” when the first stimulus was played and “/ε/ as in pet” when the last stimulus played. Nothing was displayed on the monitor while the other stimuli played. Listeners were instructed to simply listen to the continuum and told that it begins /I/ like and ends /ε/ like.

*Practice:* Listeners were presented with one randomization of the seven members of the continuum. They were instructed to label each sound as “/I/ as in pit” or as “/ε/ as in pet” by pushing the corresponding key on a button board. The labels “/I/ as in pit” or as “/ε/ as in pet” were also displayed on the computer screen. In this and in the following tasks, there was a 500 ms ITI that began after the listener's response.

*Labeling:* Each member of the continuum was randomly presented to listeners in a labeling task (without a background). The listeners’ task was to label each sound as “/I/ as in pit” or as “/ε/ as in pet.” There were 10 repetitions of the continuum stimuli (70 trials total).

*Labeling and Ignore:* Each member of the continuum was presented simultaneously with one of five backgrounds. As can be seen in Figure 11, there were actually six backgrounds because the /I/ background served two purposes:

the target absent version of background 1 and the target present version of background 2. To keep the number trials for each background equivalent, the /I/ background was used once.

The continuum stimuli were presented to the attended ear, while the background stimuli were presented to the ignored ear. The background stimuli were 20 msec shorter and 6dB lower than the primary task continuum stimuli to prevent the two from fusing into one percept.

Listeners were instructed to ignore the background and to attend to and label the stimulus in the attended ear as "/I/ as in pit" or as "/ɛ/ as in pet." The seven continuum stimuli combined with the five backgrounds resulted in 35 different combinations. There were 10 random repetitions of the combinations (350 trials total).

*Surprise Memory Test:* There is a potential limit to this design (and to the previous research on the organization of the auditory background; e.g., Sussman et al., 2007). The current design does not eliminate the possibility that listeners could switch attention from the primary task to the background. So, there is no direct assessment of whether the background is really a non-attended stimulus in this experiment. Incorporating such an assessment in this experiment (and in Experiments 6 and 7) is a difficult task, especially since most direct measures would cause attention to be focused on the background at some point. Although there is no direct measure of whether the background really is unattended in this experiment, there was a measure of whether the backgrounds were encoded. A surprise memory test was used for this purpose, as has been done in the visual attention literature (Mack, Tang, Tuma, Kahn, & Rock, 1992).

Listeners were presented with five probes for the backgrounds that had been presented. The probes consisted of two different background choices presented sequentially. One pattern had been presented (the target), and one had not been presented (the foil). The target-foil test pairs were: /ɛ/-/a/, /I/ - /u/, non-speech sound derived from /I/-non-speech sound derived from /u/, control target present-control target present foil, and control target absent-control target absent foil. The two patterns were presented 500 msec apart, with the order randomly selected. Listeners were required to select the background that was presented to the ignored ear during the experiment by pressing a key on the button board labeled “Background 1” (the first member of the pair) or “Background 2” (the second member). The words “Background 1 or Background 2” were also displayed on the computer screen during each trial. The memory questions were only administered at the end of the label-and-ignore trials, rather than after each trial, to prevent participants from intentionally directing their attention to the backgrounds during the experiment.

*Identification and Discrimination Tasks.* A failure to find an effect of the backgrounds in the Labeling and Ignore phase could potentially be due to listeners' inability to distinguish between the /I/ and /ɛ vowels (especially given the high exclusion rates in Experiments 1a, 3, and 4). Two additional tasks were added to tease apart this explanation for a null effect from a true null priming effect. The first task was an identification task; the second task was a discrimination task. This task was added because a failure to label appropriately may not necessarily reflect an inability to distinguish the vowels.

*Identification 1:* Listeners were presented with the three experimental backgrounds (/I/, /ɛ/, and the non-vowel; now, as the only stimuli, not as backgrounds) and were instructed to label them as "/ɛ/ as in pet", "not a vowel", or "/I/ as in pit" by pressing the corresponding key on a button board. The labels were also displayed on the computer screen during each trial. There were 10 random repetitions of the three stimuli (30 trials total).

*Discrimination 1:* Pairs of stimuli were presented with a 250 msec ISI. There were three types of Same trials (each background presented with itself). There were 10 repetitions of each Same pair. There were also three types of Different trials (/I/ paired with /ɛ/, /I/ paired with not a vowel, and /ɛ/ paired with not a vowel). The Different trials consisted of 5 repetitions of each pair in both orders, 10 repetitions of each pair total. There were 60 trials in this phase. The listeners task was to report whether the two stimuli presented on each trial sounded the Same or Different by pressing a corresponding key on a button board.

Most listeners had to be trained to accurately distinguish between the two vowels in Experiments 1-4, so it is likely that listeners may not be able to perform well on the Identification and Discrimination tasks without training. A training task was implemented, followed by another identification and discrimination task to assess any learning.

*Training:* First, listeners were familiarized with the correct labels for the background vowels. Each vowel was played while the computer screen displayed the correct label, either "/I/ as in pit"," /ɛ/ as in pet", or "not a vowel". There were 4 non-random repetitions of the stimuli. Listeners were instructed to simply listen while noting the correct labels. Next, listeners were presented with 4 random repetitions of the stimuli and were instructed to label them as "/I/ as in pit," "/ɛ/ as in pet," or "not a vowel" by pressing the corresponding response key. After each response, listeners received feedback indicating whether their response was correct or incorrect. The words "Correct" or "Wrong" were displayed on the computer screen after each response. Listeners were required to repeat the feedback training until they reached 90% accuracy.

*Identification 2:* This task was the same as Identification 1.

*Discrimination 2:* This task was the same as Discrimination 1.

### *Results and Discussion*

Examination of the data revealed three different groups: listeners who could label and discriminate the background stimuli before training (achieved a d' of at least 0.7 in both the Identification 1 and Discrimination 1 tasks, n = 13), listeners who could label and discriminate the background stimuli after training (achieved a d' of at least 0.7 in both the Identification 2 and Discrimination 2 tasks, n = 9), and listeners who could not label and discriminate the background stimuli before or after training (the d' scores were less than 0.7 in the Identification 1 and 2 and Discrimination 1 and 2 tasks, n = 4). Performance from one listener was substantially worse after training. This listener was excluded from the data analyses. One additional listener was excluded for pressing the same button throughout the entire experiment. The following results are reported for the remaining twenty-six listeners.

The percentage of /ε/ responses in the Labeling task was compared to the percentage of /ε/ responses in the Labeling and Ignore task for all three groups of listeners. An ANOVA was conducted on each group with Background (no background, /ε/, /I/, non-vowel, control target present, and control target absent) and Point Along the Continuum (F1 peaks at 375, 396, 417, 438, 459, 480, and 501 Hz) as factors.

*Before Training Group.* Figure 12 shows the labeling performance for the listeners who could accurately distinguish the vowels without training. The top panel of the figure allows a comparison of labeling performance when there was no background with performance on each of the experimental backgrounds, /ε/, /I/, and non-vowel. The bottom panel of the figure depicts labeling performance when there was no background and labeling performance on the two control backgrounds. As can be seen in the top panel of the figure, the experimental backgrounds affected the primary Labeling task, $F\,(5, 60) = 13.93$, $p < .01$, $\eta 2 = .53$. Planned comparisons revealed that the percentage of /ε/ responses was higher when the background was /ε/ (46.9%) than when the background was /I/ (27.9%) or the non-vowel (41.9%), *p's* < .05. Also, there were fewer /ε/ responses (and thus more /I/ responses) when the background was /I/ (27.9%) than when there was no background (48.5%), the background was /ε/ (46.9%), and both control backgrounds (target absent = 43% and target present = 46.2%), *p's* < .05.

Responses to the two control backgrounds did not differ. The results of the Surprise Memory Test revealed that listeners were only accurate at choosing which backgrounds had been presented 37% of the time. This level was actually significantly lower than chance performance, $t\,(12) = -2.38$, $p < .05$.

In summary, the backgrounds affected labeling performance in the group of listeners who could distinguish between the background vowels without any training. The /ɛ/ background increased the percentage of /ɛ/ responses relative to the other two vowel backgrounds, and the /I/ background increased the percentage of /I/ responses relative to the /ɛ/ background. The backgrounds affected labeling performance even though listeners were unaware of the acoustic details of the backgrounds. Thus, for this group of listeners it appears that the target feature was assigned to the objects in the background.

*After Training Group and Not Before or After Group.* The backgrounds did not affect labeling performance in the group of listeners who could distinguish between the vowel backgrounds only after training, $F\,(5, 40) = 1.43$, $p > .05$, and in the group of listeners who could not distinguish between the background vowels before or after training, $F < 1$. Performance on the Surprise Memory Test was slightly lower than, and not significantly different from chance for both groups (48.9% for the After Training Group and 45% correct for the Not Before or After Group).

This experiment was designed to determine if features are assigned to objects in perceptual ground, as opposed to floating freely in a perceptual limbo. The results suggest that features are assigned to objects in the background, but only when those objects are meaningful to the listener. In some circumstances, including situations in which the background objects are not meaningful to the listener, the idea of a perceptual limbo may be valid: The background objects did not affect labeling performance for the listeners for whom the backgrounds did not include “good” objects. This scenario is consistent with a perceptual limbo where features in perceptual ground remain unbound.

## Experiment 6: ERP Assessment of Assignment of Features to Objects in Perceptual Ground

Experiment 5 provided behavioral evidence that features are sometimes assigned to objects in the background. Experiment 6 was conducted to determine if there is physiological evidence that features are assigned to objects in the background. One method that has been shown to be sensitive to changes in the auditory background is the MMN (mismatch negativity) component of ERPs (event-related brain potentials). The ERP methodology is an electrophysiological technique that records neural responses to external stimuli, providing an indication of when perceptual or cognitive processing of the stimuli has occurred (Luck, 2005). The MMN is an ERP component that has the virtue of being sensitive to neural responses to unattended sounds. The MMN is elicited by deviations that occur within an otherwise regular auditory pattern. As noted, the MMN can be found regardless of whether the auditory input is being attended to or not. As a result, the MMN can be used as a measure of when the grouping of auditory features has occurred by providing an indication of when the brain has detected that an auditory object in the background has changed.

Several studies have shown that an MMN is elicited when one auditory sequence in the background changes to two sequences (e.g., Ritter, Sussman, Molholm, 2000; Sussman et al., 2007). Listeners in these tasks are typically required to complete a primary task while ignoring the tones in the background (see Sussman et al., 2007). This paradigm has offered a way to explore auditory organization outside the focus of attention. Experiment 6 uses a similar paradigm to determine if auditory features are assigned to objects in the background.

Two to-be-ignored backgrounds were used. Both backgrounds were composed of two rhythmic patterns, each of which is regular enough to not elicit an MMN. The temporal point where the two patterns meet created the percept of a third, new object in the experimental background, but no new object in the control background. If features are assigned to objects in the background and there is some neural sensitivity to this, then a larger MMN should be elicited in the experimental background at the point where the two sequences meet (where a new object is created) than in the control background at the same point (where no new object is created).

*Method*

*Participants*

Eighteen right-handed listeners participated in this experiment (mean age = 20.2). Sixteen listeners were SBU undergraduates who received course credit for their participation and two were SBU graduate students. No listener reported any hearing deficiencies.

*Stimuli*

Stimuli consisted of an experimental sequence and a control sequence. Figure 13 illustrates both stimuli. As can be seen in the figure, the sequences were composed of two frequency components that decreased and increased in temporal separation. Each component consisted of three segments that were 60 msec in duration and spaced 40 msec apart.

To create the sequences, a vowel was synthesized according to the same parameters used in Experiment 1a to create the vowel /I/ (see Experiment 1a). The vowel was synthesized on a fundamental frequency of 125 Hz at a sampling rate of 44.1kHz using a source-filter synthesis algorithm in Praat (Boersma & Weenik, 2007). The formant values were 490, 2100, and 2900 Hz, and the corresponding bandwidth values were 90, 110, and 170 Hz.

The first two segments of the experimental sequence were identical. These segments were created by band-pass filtering the 125 Hz component out of the vowel. The third segment of the experimental sequence was created by band-pass filtering the frequencies between 125 and 1800 Hz out of the vowel. The fourth segment was created by band-pass filtering the frequencies between 1800 and 3125 Hz out of the vowel. The fifth and sixth segments were identical and were created by band-pass filtering 1800 Hz out of the vowel.

The first two and last two segments of the control sequence were identical and were created by band-pass filtering 125 Hz out of the original, intact vowel. The third and fourth segments were also identical and were created by band-pass filtering the frequencies between 125 and 1800 Hz out of the vowel.

The two frequency complexes making up the experimental and control sequences were separated by a 250 msec gap that gradually decreased in equal temporal steps of 50 msec until the two sequences met (see Figure 13). Then, the sequences gradually increased in 50 msec steps back to the 250 msec temporal gap. The point at which the two sequences met resulted in a subset of the tones being temporally aligned. This configuration resulted in the percept of the vowel /I/ in the experimental sequence. In the control sequence, the point at which the tone components met did not result in the percept of a separate object. Rather, this point resulted in a non-identifiable tonal sound that matched what was present in each sequence already.

Two additional sequences were created for a surprise memory test: one foil for the experimental sequence and one foil for the control sequence. The memory test stimuli were created by first synthesizing the vowel /a/. The vowel was synthesized on a fundamental frequency of 125 Hz at a sampling rate of 44.1kHz using a source-filter synthesis algorithm in Praat (Boersma & Weenik, 2007). The formant values were 730, 1090, and 2440 Hz (Peterson & Barney, 1952), and the corresponding bandwidth values were 90, 110, and 170 Hz. The 500 Hz frequency component was band-pass filtered out of the vowel and used as the first two and last two segments of both foil sequences. In the experimental sequence foil, the third segment was created by band-pass filtering the frequencies between 0 and 1000 Hz out of the vowel, and the fourth segment was created by band-pass filtering the frequencies between 1001 and 5000 Hz out of the vowel. In the control sequence, the third and fourth segments were both created by band-pass filtering the frequencies between 0 and 1000 Hz out of the vowel. The timing of the frequency components within each foil sequence was the same as described above for the experimental and control background sequences.

*Procedure*

The experimental and control sequences were presented to listeners over headphones while EEG responses were recorded. Listeners were instructed to watch a silent movie [1] with closed captioning while ignoring any sounds presented over the headphones. Listeners were instructed to be as still as possible and to minimize eye blinks. The experiment was conducted in a sound-attenuating chamber, with participants seated comfortably in a reclining chair.

The experimental and control sequences were presented in separate blocks. In each block, the sequence went through a complete cycle 100 times. That is, the sequences began with the two frequency components 250 msec apart. The temporal separation of the frequency components decreased until they met. And, the temporal separation increased until the frequency components again were 250 msec apart. There were four blocks for each sequence, yielding 8 blocks and 800 trials total. Blocks were approximately 5 minutes long. Brief rest breaks were provided after each block.

At the end of the experiment, a surprise forced-choice memory task was presented to listeners to measure awareness of the content of the backgrounds. The test consisted of two probes: one for each sequence. Both probes presented

[1] The movie played during the experiment was the computer animation “Wallace and Gromit: The Curse of the Were-Rabbit.

the old sequence that had been present during the experiment and a new sequence that had not been presented. The two sequences were presented sequentially with a 500 msec ISI. The order of the background presented first, old or new, was randomly selected. Listeners were asked to select the one they thought had been presented during the experiment.

*Electrophysiological Recording.* The EEG was recorded continuously using a 64-channel electrode cap (Neuroscan Inc., Sterling USA). Recordings were obtained using a fronto-central electrode as ground and mastoid electrodes as reference. The horizontal EOG was monitored from electrodes at the outer canthi of the eyes, and the vertical EOG was monitored from electrodes above and below the left eye. Impedances for all electrodes were kept below 10 KO. The EEG and EOG signals were digitized at 1000 Hz and were amplified with a gain of 500.

### *Results and Discussion*

EEG data were low-pass filtered at 20 Hz. To eliminate EOG artifact, trials with EEG voltages exceeding 75 mV were rejected from the data analyses. Rejection of half, or more, of the total trials resulted in exclusion from the data analyses. Nine listeners were excluded because of this criterion. Out of the remaining nine listeners, an average of 26% of the total trials was excluded because of artifacts. ERP epochs began 200 msec prior to stimulus onset and continued for 900 msec. ERP averages were baseline corrected at 100 msec.

The MMN response was assessed by averaging the ERP elicited by the experimental and control backgrounds separately. For both backgrounds, the average ERP at each temporal point in the sequences, the deviant (0 msec), and the 5 standards (50, 100, 150, 200, and 250 msec), was calculated. The grand mean of the deviant – each standard was calculated over the time window from 300 to 800 msec. Differences in the size of the MMN response elicited by the two backgrounds were tested with an ANOVA conducted on the difference scores with Background (experimental, control) and Temporal Separation (50, 100, 150, 200, and 250 msec) as factors.

The ANOVA was conducted on several different electrode sites. One ANOVA was conducted on the F3, F4, P3 and P4 electrodes using Background, Temporal Separation, Scalp Position (anterior or posterior: F vs. P) and Hemisphere (left or right, 3 vs. 4) as factors. The analysis revealed a significant effect of Background, $F(1, 8) = 31.55$, $p < .05$, $\eta^2 = .80$. The size of the MMN was larger in the experimental background than in the control background. There also was a significant interaction between Background and Scalp Position, $F(1, 8) = 7.34$, $p < .05$, $\eta^2 = .48$. Planned comparisons indicated that the MMN elicited by the experimental background was larger in the anterior Scalp Position than in the posterior Scalp Position, *p's* < .05. The ANOVA also revealed a significant effect of Temporal Separation, $F(4, 32) = 4.89$, $p < .05$, $\eta^2 = .38$. Planned comparisons indicated that the MMN elicited at the 200 msec temporal separation was larger than the MMN elicited at 150, 100, and 50 msec, p's < .05.

A second analysis was conducted to examine other electrode sites. It has been suggested that there are two different MMN's: one frontal and one temporal (e.g., Deouell, Bentin, & Giard, 1998). An ANOVA was conducted on the F3, F4, P7, and P8 electrodes to test for any differences between frontal and temporal scalp positions. The ANOVA used Background, Temporal Separation, Scalp Position (frontal or temporal: F vs. P) and Hemisphere (left or right: 3,7 vs. 4,8) as factors. The ANOVA revealed a significant effect of Background, $F(1, 8) = 23.52$, $p < .05$, $\eta^2 = .75$. Again, the MMN elicited by the experimental background was larger than the MMN elicited by the control background. The interaction between Background and Scalp Position was significant, $F(1, 8) = 8.64$, $p < .05$, $\eta^2 = .52$. Planned comparisons indicated that the MMN elicited from the experimental background was larger at the frontal scalp positions than at the temporal positions, p's < .05. There was a significant effect of Temporal Separation, $F(4, 32) = 4.68$, $p < .05$, $\eta^2 = .37$. Again, planned comparisons indicated that the MMN elicited at the 200 msec temporal separation was larger than the MMN elicited at 150, 100, and 50 msec.

A final analysis was conducted on the FZ electrode alone. This electrode has been claimed to be the area of maximal signal-to-noise ratio for MMN (e.g., Sussman et al., 2007; Ylinen, Shestakova, Huotilainen, Alku, & Näätänen, 2006). The ANOVA was conducted on the difference scores using Background and Temporal Separation as factors. The results revealed that neural responses were sensitive to an object in the background. The experimental background resulted in a larger MMN than the control background, $F(1, 8) = 17.1$, $p < .05$, $\eta^2 = .68$. The results of this analysis are depicted in Figure 14. As can be seen in the figure, the difference between the deviant and each standard (black and red lines in each panel) was larger in the experimental background than in the control background. Temporal Separation was significant, $F(4, 32) = 4.61$, $p < .05$, $\eta^2 = .37$. Planned comparisons indicated that the MMN elicited at the 200 msec temporal separation was larger than the 150, 100, and 50 msec separations, *p's* $< .05$.

The results of the surprise memory test indicated that listeners selected the experimental background with 89% accuracy and the control background with 100% accuracy. This finding may seem surprising, given the poor accuracy on the surprise memory test in Experiment 5, but the two paradigms were actually quite different. The background stimuli were more salient in this experiment than in Experiment 5. Recall that sounds were presented to both ears in Experiment 5, and the listeners' task was to attend to one ear while ignoring the opposite ear. In this experiment, listeners were instructed to ignore the only sounds presented during the experiment, i.e., the backgrounds, while attending to a silent movie.

Although listeners in this experiment could accurately identify the sounds they were instructed to ignore, this finding is not necessarily a problem for the claim that listeners are sensitive to objects in the background. Typical studies examining the MMN component do not test for memory of the to-be-ignored sounds because the MMN is a physiological response that is evoked when the sounds are unattended *and* when they are attended (Näätänen & Winkler, 1999). Thus, although listeners were more aware of the sounds in this experiment than in Experiment 5, the difference in neural responding to the experimental and control backgrounds suggests that listeners are more sensitive to objects in the background (experimental sequence) than to other background changes (control sequence). Overall, the results of Experiment 6 seem to provide support for the finding in Experiment 5 that the auditory background is organized into objects, to some extent.

## Experiment 7: Assignment of Features to Groups in Perceptual Ground

Experiments 5 and 6 indicated that features are assigned to objects in the background, but only for some listeners. It may be the case for the remaining listeners that features are at least assigned to perceptual groups in the background. There is support in the literature for this notion. For example, the finding that a perceptual group can capture a feature out of an object (e.g., Darwin & Sutherland, 1984; Darwin, Pattison, & Gardner, 1989) suggests a potentially significant role for perceptual groups in the background. The purpose of Experiment 7 was to explore this possibility using a paradigm similar to Experiment 5.

The difference between an object and a group (at least in this experiment) is that an object forms one distinct, identifiable unit, whereas a perceptual group is made up of a sequence of units that follow the same pattern, e.g., the even/galloping rhythms used in Experiments 1-4. Though the rhythms were called “objects” in those experiments (following the convention of Shinn-Cunningham et al., 2007), one could argue that they are actually perceptual groups.

Previous research has shown that either unchanging or descending patterns, or perceptual groups, can be successful at pulling a feature out of an object (e.g., Darwin, Pattison, & Gardner, 1989; Darwin & Sutherland, 1984).

These findings suggest that there might be an important role for groups in perceptual ground. Experiment 7 addresses this possibility using a paradigm similar to Experiment 5. Stimuli from a continuum were presented in a labeling task alone and again in a labeling task with a to-be-ignored background. If the target is integrated with the perceptual group in the background, then consistent patterns in the figure and ground should affect labeling responses. For example, an even pattern in the background may result in a higher percentage of "even" responses than a galloping background.

### *Method*

#### *Participants*

Twenty-seven undergraduates (mean age = 19.9 years) from Stony Brook University served as listeners for this experiment. No listener reported any hearing deficiencies and all received course credit for their participation.

#### *Stimuli*

Figure 15 depicts examples of the stimuli that were used. The stimuli for the primary task were members of a rhythm continuum. The continuum stimuli were created using the target present tone sequence from Experiment 1a. The tone sequence was 1.5 seconds long: five repetitions of three 500 Hz tones that were 60 msec in duration and spaced 40 msec apart. This sequence resulted in an "even" rhythm percept. A continuum was created from this sequence by shifting the amplitude of every third tone by equal steps of 2 dB. The "even" end of the continuum was created by shifting the amplitude by 0 dB, and the "galloping" end of the continuum was created by shifting the amplitude by -10 dB.

The background stimuli consisted of target present and target absent conditions (see Figure 15). The experimental background stimuli were the same tone sequences that were used in Experiment 1a. Recall that in Experiment 1a, the individual tones of the sequences were obtained by band-pass filtering the 500 Hz component out of a vowel /ɛ/. The target present background consisted of a tone sequence that results in the percept of an “even” rhythm. The target absent background consisted of a tone sequence that results in the percept of a “galloping” rhythm. The background stimuli had the same timbre as the foreground continuum stimuli.

As with the experimental stimuli, the control background stimuli were created by filtering out frequency components from the vowel /ɛ/. The first two tones of both the target present and target absent control sequences were 104 and 90 Hz. The target absent sequence was a repetition of these two tones (with a 120 msec interval in between repetitions). The target present sequence was a repetition of 104, 90, and 500 Hz (the same target as the experimental stimuli). The result of both control stimuli was the percept of a bouncing rhythm. All background stimuli were 20 msec shorter and 6 dB lower than the continuum stimuli to prevent the two from fusing into one percept.

Additional foil stimuli were created for the surprise memory test. As with the experimental and control stimuli, these stimuli were created by filtering out frequencies from the vowel /ɛ/. The frequencies were selected so that the rhythms would be quite different from the experimental and control stimuli. The experimental background foil stimuli consisted of a repeating ascending rhythm sequence of 90, 95, and 500 Hz for the target present version and 90 and 95 Hz for the target absent version. The control background foil stimuli consisted of a repeating ascending sequence of 90, 104, and 500 Hz in the target present version and 90 and 104 Hz in the target absent version.

*Procedure*

Experiment 7 consisted of the same phases as Experiment 5. The only difference between the two experiments is that listeners were instructed to label the attended stimuli as “even” or as “galloping” in this experiment. All other aspects of the procedure were as described in Experiment 5.

### *Results and Discussion*

One listener was excluded from the data analyses for failing to push any buttons during the experiment, and one listener was excluded for pushing unlabeled buttons that were irrelevant to the experiment. The following results are for the remaining twenty-five listeners.

Examination of the data indicated two groups of listeners: listeners who could accurately label and discriminate between the even and galloping rhythms before training (achieved a d' of at least 0.7 on the Identification 1 and Discrimination 1 tasks, $n = 21$), and listeners who could accurately label and discriminate the even and galloping rhythms after training (achieved a d' of at least 0.7 on the Identification 2 and Discrimination 2 tasks, $n = 4$).

Previous research has shown that perceptual groups in the background can capture a feature out of perceptual figure (e.g., Darwin & Sutherland, 1984). Thus, it was expected that the target feature would be assigned to the perceptual groups in perceptual ground in this experiment. This finding would be indicated by the experimental backgrounds affecting labeling responses to the tone sequences in figure.

*Before Training Group.* Figure 16 depicts the labeling functions from the group of listeners who could distinguish between the tone sequences without any training. As can be seen in the figure, the background sequences did not affect labeling performance, $F(4, 80) = 1.14$, $p > .05$.

*After Training Group.* The background sequences also did not affect labeling performance in the group of listeners who could distinguish between the sequences with training, $F(4, 12) = 1.45$, $p > .05$.

The finding that the backgrounds did not affect labeling performance in both groups of listeners suggests that the target feature was not assigned to the perceptual groups in the background. The contrast between feature allocation to objects in perceptual ground (found in Experiment 5) and feature non-allocation to perceptual groups in ground could potentially occur because perceptual groups may be easier to keep separate from an object in figure. Perhaps the global configurations making up perceptual groups are generally simpler stimuli that are less likely to compete for features than complex objects. Given that Experiment 5 found that features are sometimes assigned to *objects* in the background, the present results suggest that the nature of a perceptual limbo is not completely free-floating. Rather, it may be a field that is defined by certain rules. For example, features may bind to some objects, but not to similar features making up a perceptual group.

*General Discussion*

The auditory environment consists of numerous features, sets of which must be combined at some point to accomplish successful auditory object identification. Though there are established principles to guide how features may be grouped together, such as physical similarity and prior knowledge (Bregman, 1990), the general applicability of these principles is threatened by the claim that perception is not veridical, i.e., listeners perceive only a small portion of a given scene rather than all of the detail available within the scene (e.g., Shinn-Cunningham et al., 2007). The purpose of this project was to establish some principles for how the portion of the auditory environment that is perceived is analyzed to create the perception of auditory objects.

*Summary of Results.* The first part of this project aimed to establish principles for feature assignment to objects in auditory figure. One rule that applied across a variety of different paradigms (Experiments 1b, 3, and 4) was that a feature is more likely to contribute to an object in figure if that object needs the feature in order to be an identifiable object. This rule may explain one way in which the auditory perceptual system maintains object stability in a constantly changing auditory world.

The rule that a feature contributes to an object in figure when that object needs the feature applied across a variety of two-object contexts. The feature contributed to the perception of the object when the two were presented at different spatial locations, even when a second object that could group with the feature was also presented at that location (Experiment 1b). In addition, the feature contributed to the object in figure even when the pitch of the feature was mistuned by 32 percent and masked by white noise presented at the same amplitude as the feature (Experiment 2). The feature also contributed to the object in figure regardless of whether the feature could contribute to the object in ground when the spatial location of the feature and the object in figure were the same (Experiment 3). Thus, it appears that feature assignment depends on whether the feature is needed to produce an identifiable auditory object.

However, the rule for feature assignment to auditory figure did not consistently apply in more complicated perceptual scenarios. The auditory system appears to need more evidence to group a feature with an object when spatial cues are ambiguous, and when there is an object in the background that could potentially “take” the feature (the feature was less likely to contribute to the object in figure in such a scenario in Experiment 4). The system also needs more evidence to group two things together when the number of objects in an auditory scene increases (the feature only contributed to the object in figure when the feature and object in figure were presented at the same spatial location in Experiment 4). These limitations on feature assignment presumably are due to increased competition for the feature. This finding supports the claim that more evidence is needed to pull a feature into perceptual figure than to reject it to the ground (Shinn-Cunningham et al., 2007), with this asymmetry being more likely to guide perception in multiple-object scenes.

When the object in figure does *not* need the feature, the feature is not always allocated to an object in ground (Shinn-Cunningham et al., 2007;

Experiment 1a). One explanation for this odd finding of feature nonallocation has been that features in the unattended background may float around in a perceptual limbo (Shinn-Cunningham et al., 2007). The goal of the second part of this project was to test the validity of this idea. Three experiments examined the fate of features in perceptual ground. Experiment 5 provided behavioral evidence and Experiment 6 provided physiological evidence that features are assigned to objects in perceptual ground if the objects are meaningful to the listener. Experiment 7 revealed that features are not assigned to perceptual groups in the background.

The results of Experiments 5-7 suggest that although a perceptual limbo may exist, it is not a completely free-floating feature field. Instead, the perceptual limbo may be loosely structured. For example, features may group with similar features to form objects, but not with other similar features to form perceptual groups. This finding may occur because of decreased competition for features among perceptual groups than among objects.

*Results and Current ASA Theory.* The task of establishing auditory figure from ground has been studied within the framework of Auditory Scene Analysis, most prominently by Bregman and his colleagues (see Bregman, 1990 for a review). According to Bregman's theory, both auditory feature integration and segregation are results of a process that combines physical components that belong together to construct streams of sounds for all sources of sound in a scene. The analysis produces multiple co-occurring auditory objects. Whether the auditory features in any given listening situation are deemed to belong together is determined by two factors: low-level acoustic grouping principles (e.g., frequency similarity), and higher-level cognitive processes (e.g., prior knowledge and attention).

Bregman's (1990) theory has been foundational in explaining how auditory features are combined into objects. However, there are several phenomena that the theory cannot explain, including the results of the current project. The key result of Experiment 1a, and of Shinn-Cunningham et al. (2007) was that a feature disappeared from perception. The principles of ASA theory cannot explain this "ungrouping" of a feature and object. ASA theory would predict that the feature should group with the object that is the best fit, given the physical evidence available.

Bregman's (1990) ASA theory could potentially explain the reappearance of the feature in Experiment 1b. The low-level grouping principles of ASA would predict that the feature will be taken by either the tone sequence or the vowel. The competition between the two objects would probably result in the system executing higher-level processes in the feature allocation process, rather than low-level processes. However, these higher-level processes would have to include a mechanism that considers the structure of the object in figure in relation to the objects in ground. No such mechanism has been described by Bregman and colleagues. A revised theory of auditory scene analysis that considers how the properties of the object in figure and the object in ground affect perceptual coherence would provide a more comprehensive picture of how auditory objects of interest are extracted from perceptual ground.

In summary, this project seeks to further our understanding of perceptual coherence in audition by providing insight into the nature of auditory feature assignment in perceptual figure and in perceptual ground. By examining the characteristics of 1) objects in figure, 2) objects in ground, and 3) features that affect the likelihood of feature and object binding, this project provides a more complete picture of how auditory scene analysis is accomplished.

References


Alain, C., & Woods, D.L. (1997). Attention modulates auditory pattern memory as indexed by event-related brain potentials. *Psychophysiology, 34*, 534-546.

Anstis, S. & Saida, S. (1985). Adaptation to auditory streaming of frequency modulated tones. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 257-271.

Bentin, S., & Mann, V.A. (1990). Masking and stimulus intensity effects on duplex perception: A confirmation of the dissociation between speech and nonspeech modes. *Journal of the Acoustical Society of America, 88,* 64-74.

Boersma, P., & Weenik, D. (2007). Praat: Doing phonetics by computer (Version 4.6.01 [Computer program]. Retrieved May 16, 2007, from http://www.praat.org/.

Botte, M.C., Drake, C., Brochard, R., & McAdams, S. (1997). Perceptual attenuation of nonfocused auditory streams. *Perception & Psychophysics, 59,* 419-425.

Bregman, A.S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 380-387.

Bregman, A.S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.



Bregman, A.S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 380-387.



Cherry, C. (1953). Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America, 25,* 975-979.

Ciocca, V., & Bregman, A.S. (1989). The effects of auditory streaming on duplex perception. *Perception & Psychophysics, 46,* 39-48.

Cusack, R., Deers, J., Aikman, G., & Carlyon, R.P. Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 643-656.

Cutting, J.E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review, 83,* 114-140.

Darwin, C.J. (1983). Auditory processing and speech perception. In H. Bouma and D.G. Bouwhuis (Eds.), *Attention and performance X.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Darwin, C.J. (1984). Perceiving vowels in the presence of another sound: Constraints on formant perception. *Journal of the Acoustical Society of America, 76,* 1636-1647.

Darwin, C.J., & Gardner, R.B. (1986). Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality. *Journal of the Acoustical Society of America, 79,* 838-845.

Darwin, C.J., & Hukin, R.W. (1997). Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity. *Journal of the Acoustical Society of America, 102,* 2316-2324.

Darwin, C.J., & Hukin, R.W. (1998). Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony. *Journal of the Acoustical Society of America, 103,* 1080-1084.

Darwin, C.J., Pattison, H., & Gardner, R.B. (1989). Vowel quality changes produced by surrounding tone sequences. *Perception & Psychophysics, 45,* 333-342.

Darwin, C.J., & Sutherland, N.S. (1984). Grouping frequency components of vowels: When is a harmonic not a harmonic? *The Quarterly Journal of Experimental Psychology, 36A,* 193-208.

Deouell, L.Y., Bentin, S., & Giard, M. (1998). Mismatch negativity in dichotic listening: Evidence for interhemispheric differences and multiple generators. *Psychophysiology, 35,* 355-365.

Deutsch, D. (1972). Octave generalization and tune recognition. *Perception & Psychophysics, 11,* 411-412.

Duncan, J., & Humphreys, G.W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433-458.

Eramudugolla, R., Irvine, D.R.F., McAnally, K.I., Martin, R.L., & Mattingley, J.B. (2005). Directed attention eliminates 'change deafness' in complex auditory scenes. *Current Biology, 15,* 1108-1113.

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,* 742-754.

Gregg, M.K., & Samuel, A.G. (2008). Change deafness and the organizational properties of sounds. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 974-991.

Gregg, M.K., & Samuel, A.G. (2009). The importance of semantics in auditory representations. *Attention, Perception, & Psychophysics, 71,* 607-619.

Hall, M.D., & Pastore, R.E. (1992). Musical duplex perception: Perception of figurally good chords with subliminal distinguishing tones. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 752-762.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events.* Cambridge, MA: MIT Press.

Hukin, R.W., & Darwin, C.J. (1995). Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel. *Journal of the Acoustical Society of America, 98,* 1380-1387.

James, W. (1890). *Principles of psychology.* Cambridge, MA: Harvard Unviersity Press.

Koffka, A. (1935). *Principles of gestalt psychology.* New York: Harcourt Brace.

Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology.* New York: Liveright.

Luck, S.J. (2005). *An introduction to the event-related potential technique.* Cambridge, MA: MIT Press.

Mack, A., Tang, B., Tuma, R., Kahn, S., & Rock, I. (1992). Perceptual organization and attention. *Cognitive Psychology, 24,* 475-501.

Nygaard, L.C. (1993). Phonetic coherence in duplex perception: Effects of acoustic differences and lexical status. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 268-286.

Nygaard, L.C., & Eimas, P.D. (1990). A new version of duplex perception: Evidence for phonetic and nonphonetic fusion. *Journal of the Acoustical Society of America, 88,* 75-86.

O'Regan, J.K., Deubel, H., Clark, J.J., & Rensink, R. (2000). Picture changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition, 7,* 191-211.

Pastore, R.E., Schmuckler, M.A., Rosenblum, L., & Szczesiul, R. (1983). Duplex perception with musical stimuli. *Perception & Psychophysics, 33,* 469-474.

Peterson, G.E., & Barney, H.L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24,* 175-184.

Rand, R.C. (1974). Dichotic release from masking for speech. *Journal of the Acoustical Society of America, 55,* 678-680.

Rensink, R.A., O'Regan, J.K., & Clark, J.J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*, 368-373.

Repp, B.H., & Bentin, S. (1984). Parameters of spectral/temporal fusion in speech perception. *Perception & Psychophysics, 36,* 523-530.

Ritter, W., Sussman, E., & Molholm, S. (2000). Evidence that the mismatch negativity system works on the basis of objects. *Neuroreport, 11*, 61-63.

Rubin, E. (1921). *Visuaell wahrgenommene figuren [figure and ground]*. Copenhagen: Gyldendalske. (Excerpt reprinted in Yantis, S. (Ed.) Visual Perception, Taylor & Francis, Philadelphia, 2001).

Shinn-Cunningham, B.G., Lee, A.K.C., & Oxenham, A.J. (2007). A sound element gets lost in perceptual competition. *Proceedings of the National Academy of Sciences, 104,* 12223-12227.

Shinn-Cunningham, B.G., & Wang, D. (2008). Influences of auditory object formation on phonemic restoration. *Journal of the Acoustical Society of America, 123,* 295-301.

Simons, D.J., & Chabris, C.F. (1999). Gorillas in our midst : Sustained inattentional blindness for dynamic events. *Perception, 28,* 1059-1074.

Simons, D.J., & Levin, D.T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review, 5*, 644-649.

Simons, D.J., & Rensink, R.A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences, 9*, 16-20.

Sussman, E.S., Horváth, J., Winkler, I., & Orr, M. (2007). The role of attention in the formation of auditory streams. *Perception & Psychophysics, 69*, 136-152.

Treisman, A. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology, 12*, 242-248.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97-136.

Treisman, A., & Schmidt, N. (1982). Illusory conjunction in the perception of objects. *Cognitive Psychology, 14,* 107-141.

Vitevitch, M. S. (2003). Change deafness: the inability to detect changes between two voices. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 333-342.

Wertheimer, M. (1923). Untersuchungen zür Lehre von der Gestalt, II. (Translated as Laws of organization in perceptual forms.) In W.D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71-88). London: Routledge & Kegan Paul.

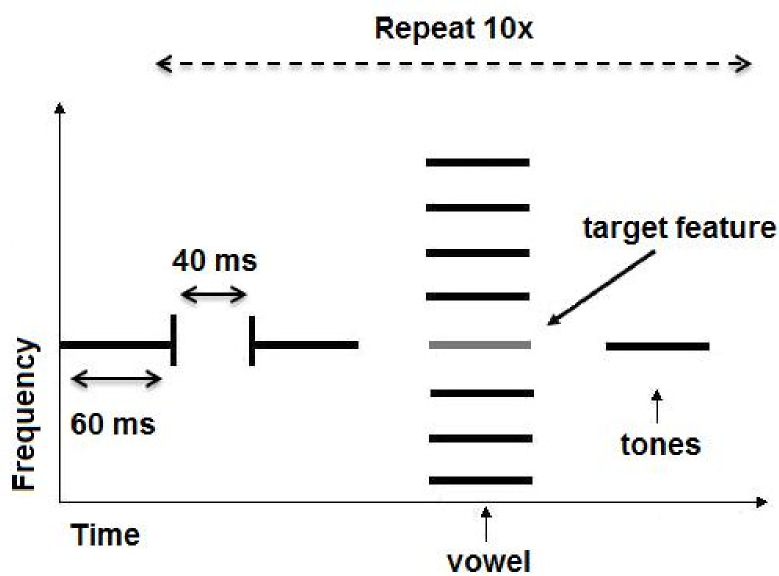
Whalen, D.H., & Liberman, A.M. (1987). Speech perception takes precedence over nonspeech perception. *Science, 237,* 169-171.

Wolfe, J.M., Cave, K.R., & Franzel, S.L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception & Performance, 15,* 419-433.

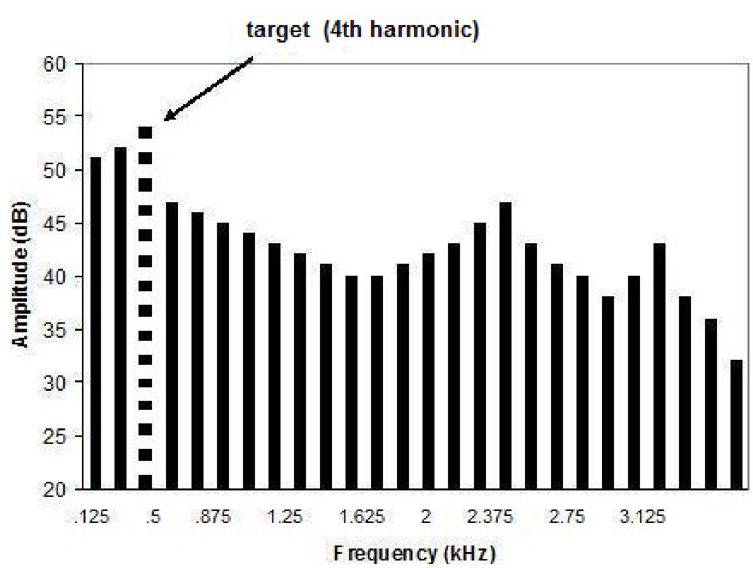
Ylinen, S, Shestakova, A, Huotilainen, M, Alku, P., & Näätänen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: A cross-linguistic study. Brain Research, 1072, 175-185.

Appendix

*Figure 1.* Stimuli and conditions that were used in Experiment 1.
(A) Stimuli in the two-object condition consisted of a 3 second long repeating sequence of tones, a vowel, and a target feature that can group with either object.




(B) Spectrum of the synthetic vowel that was used in Experiment 1a. If the feature is perceived as part of the vowel, the percept is /ɛ/, otherwise the percept is /I/.



(C) Spatial configuration of the stimuli in both the Single- and Two-Object conditions.

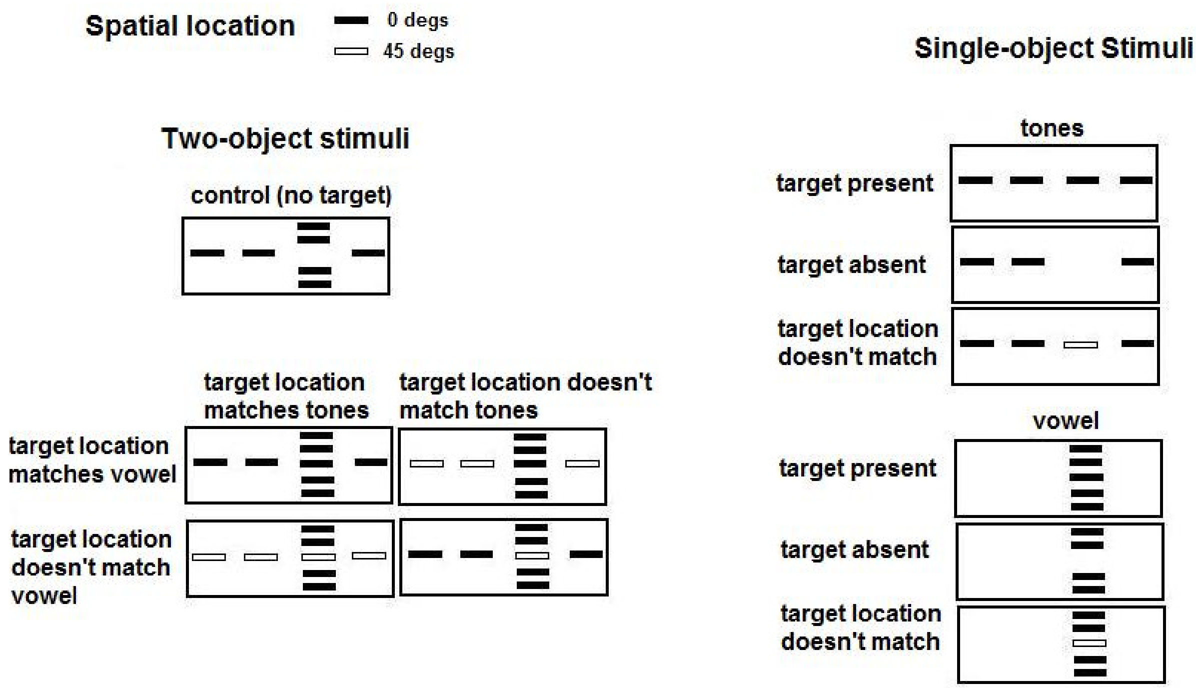
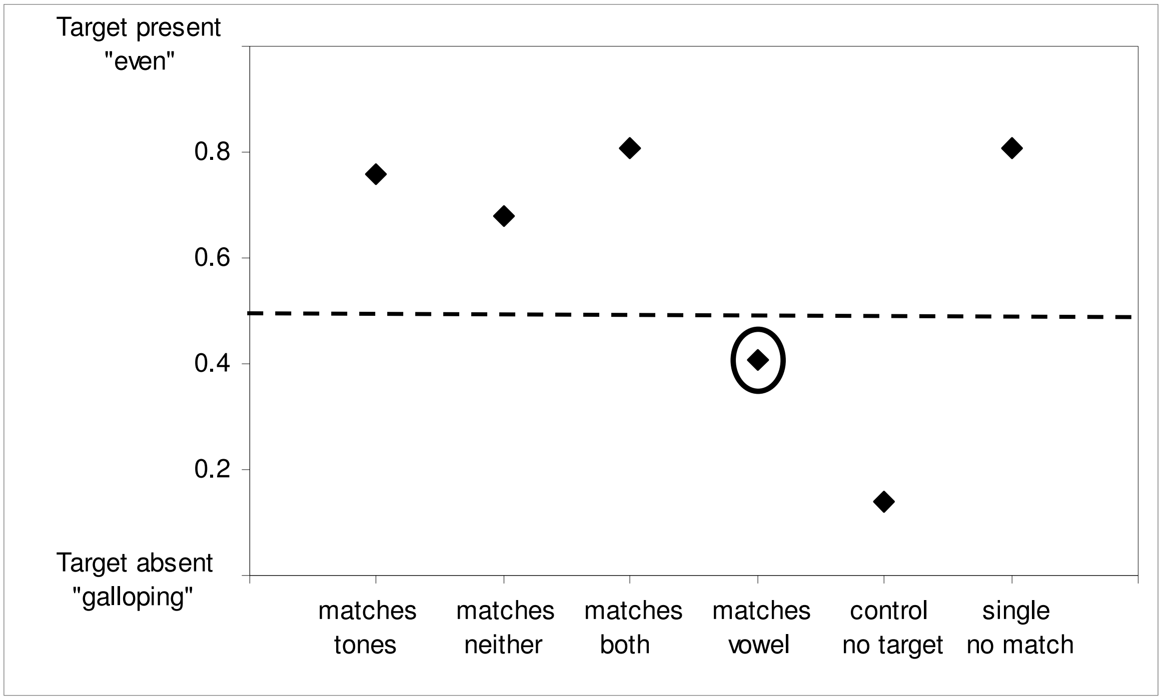


*Figure 2*. Experiment 1a results. The diamond markers represent δ scores.

Attend-tones




Attend-vowel

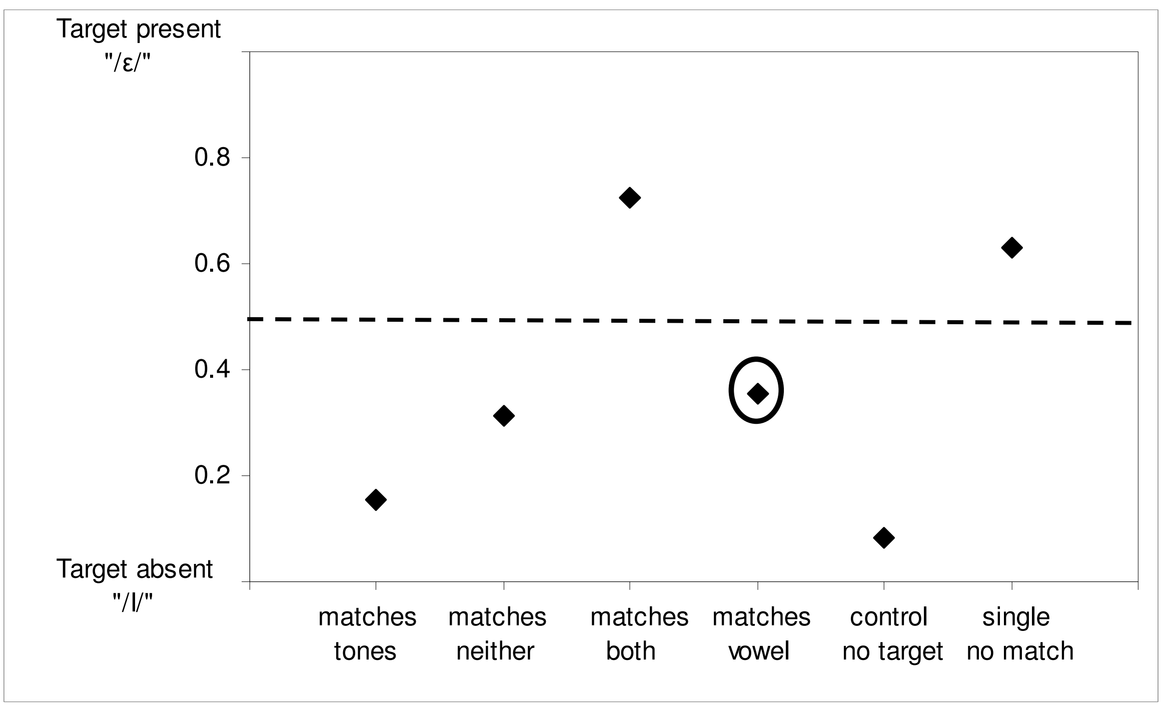


*Figure 3.* Spectrum of the synthetic vowel that was used in Experiment 1b. If the feature is perceived as part of the vowel, the percept is /I/, otherwise the percept is a non-identifiable speech sound.

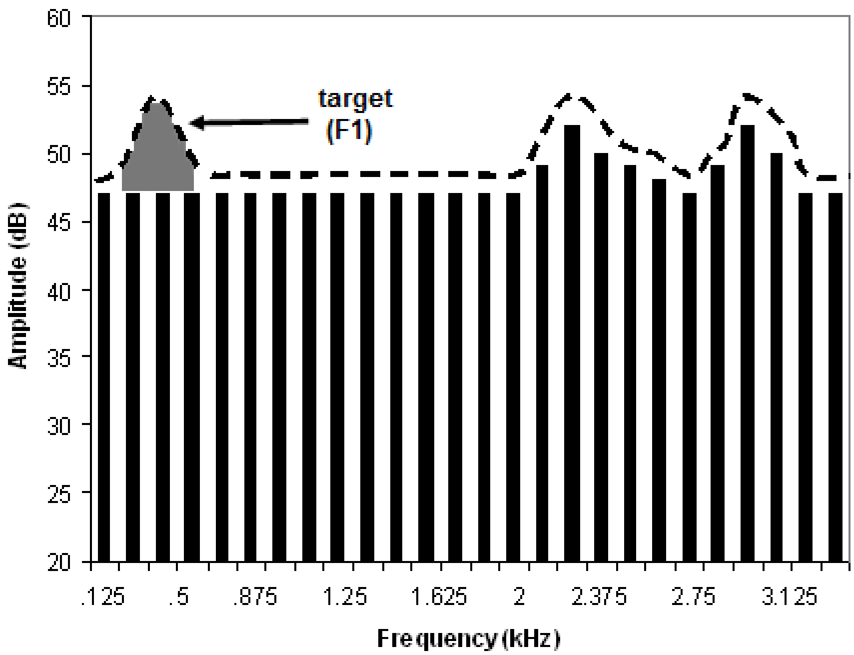
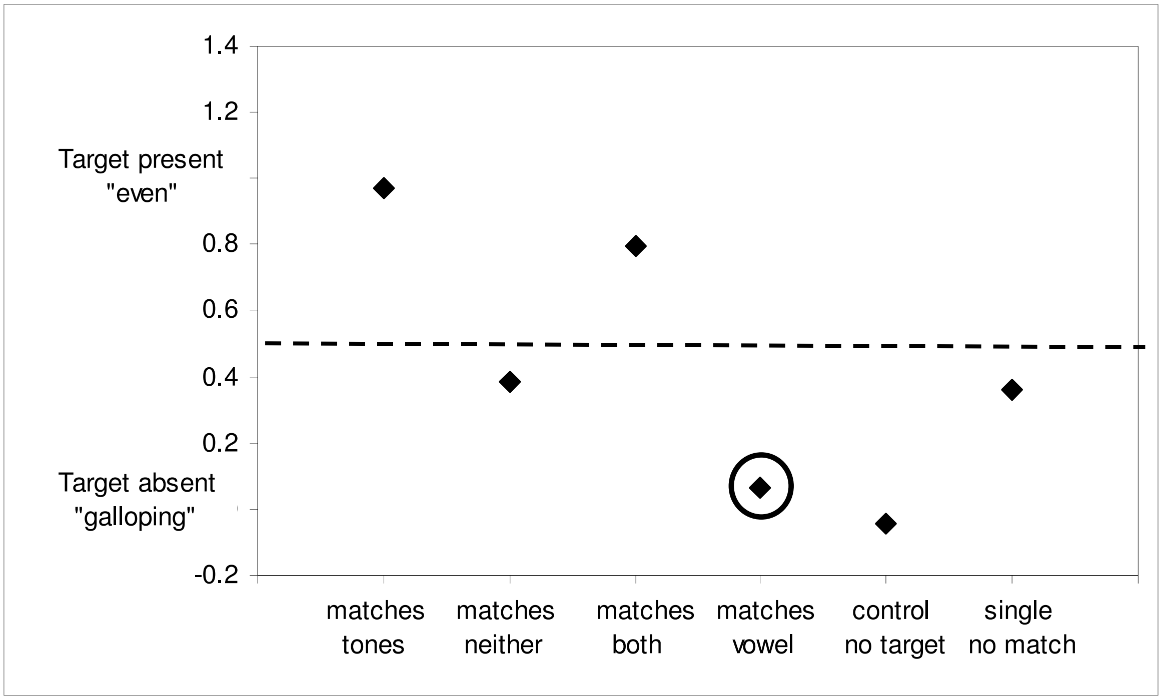


*Figure 4*. Experiment 1b results. The diamond markers represent δ scores.

Attend Tones




Attend Vowel

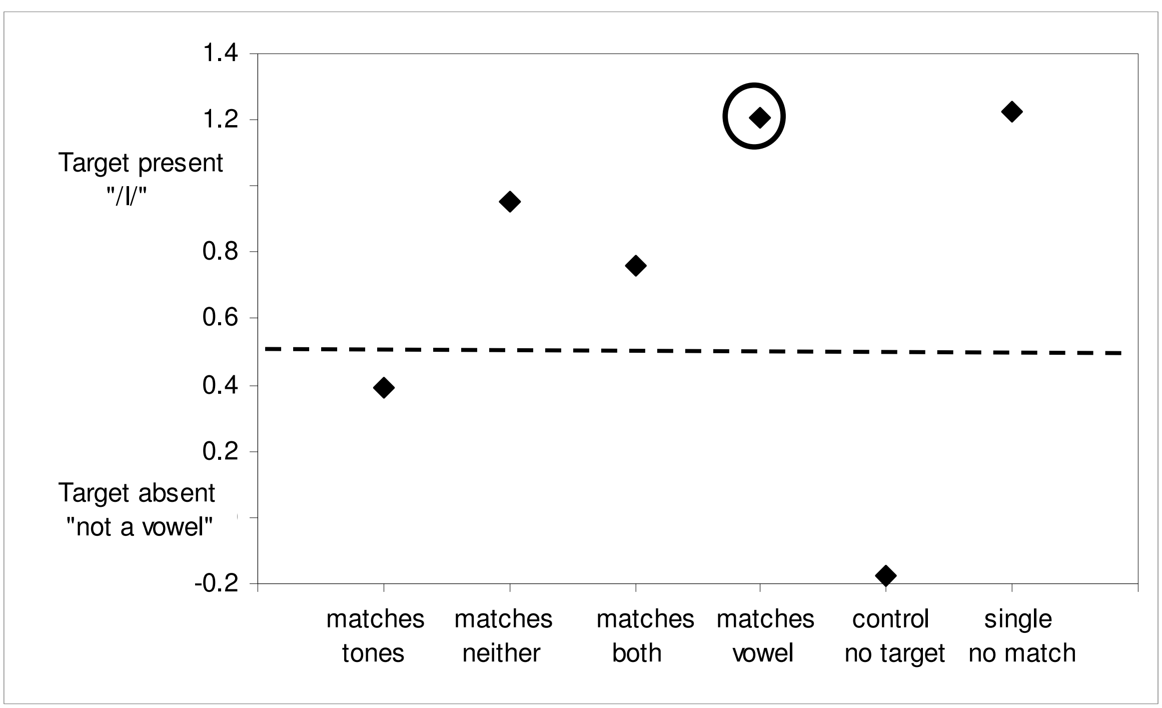


*Figure 5A*. Noise masking conditions used in Experiment 2.

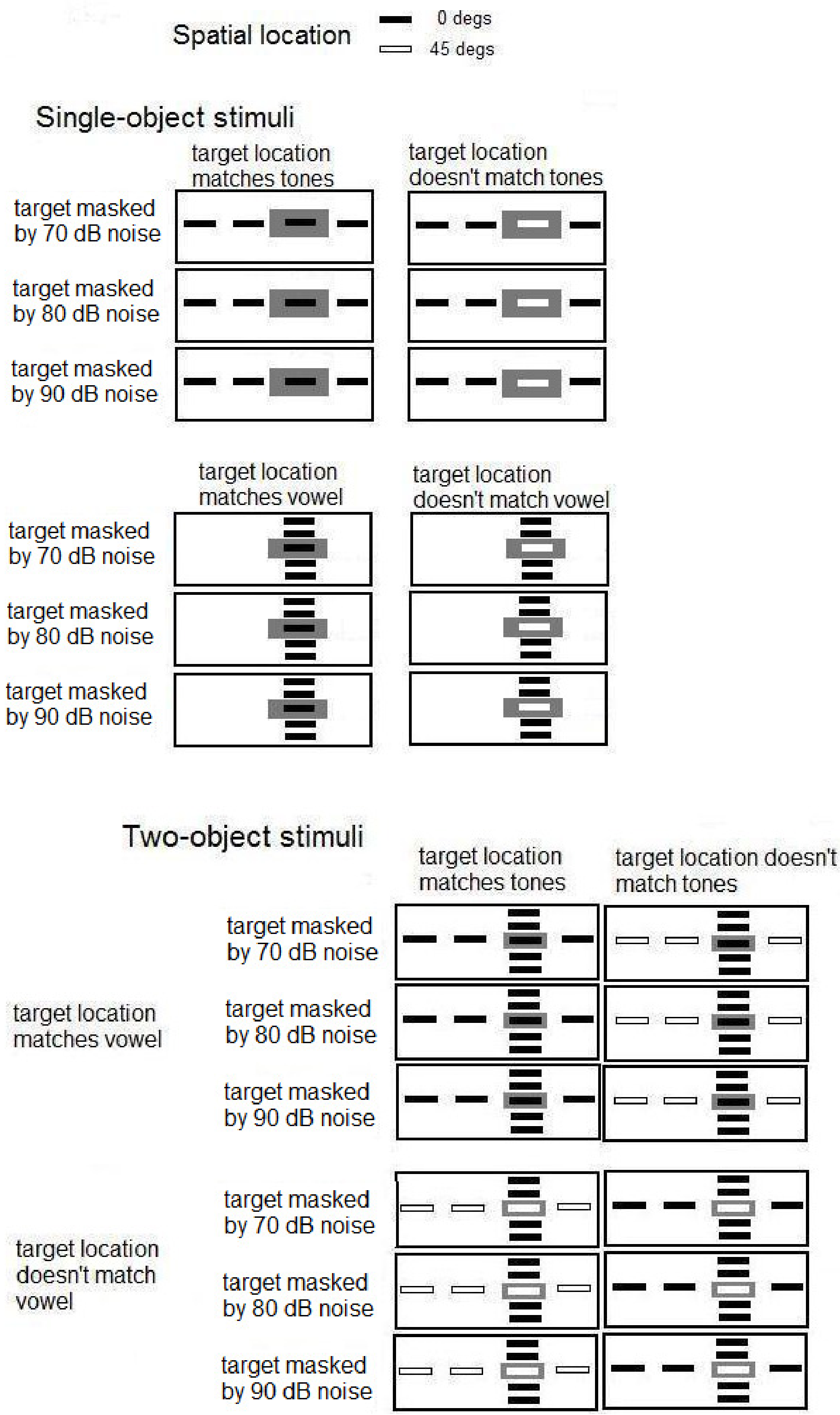


*Figure 5B*. Mistuning conditions used in Experiment 2. The frequency mistuning of the feature is not represented in the figure.

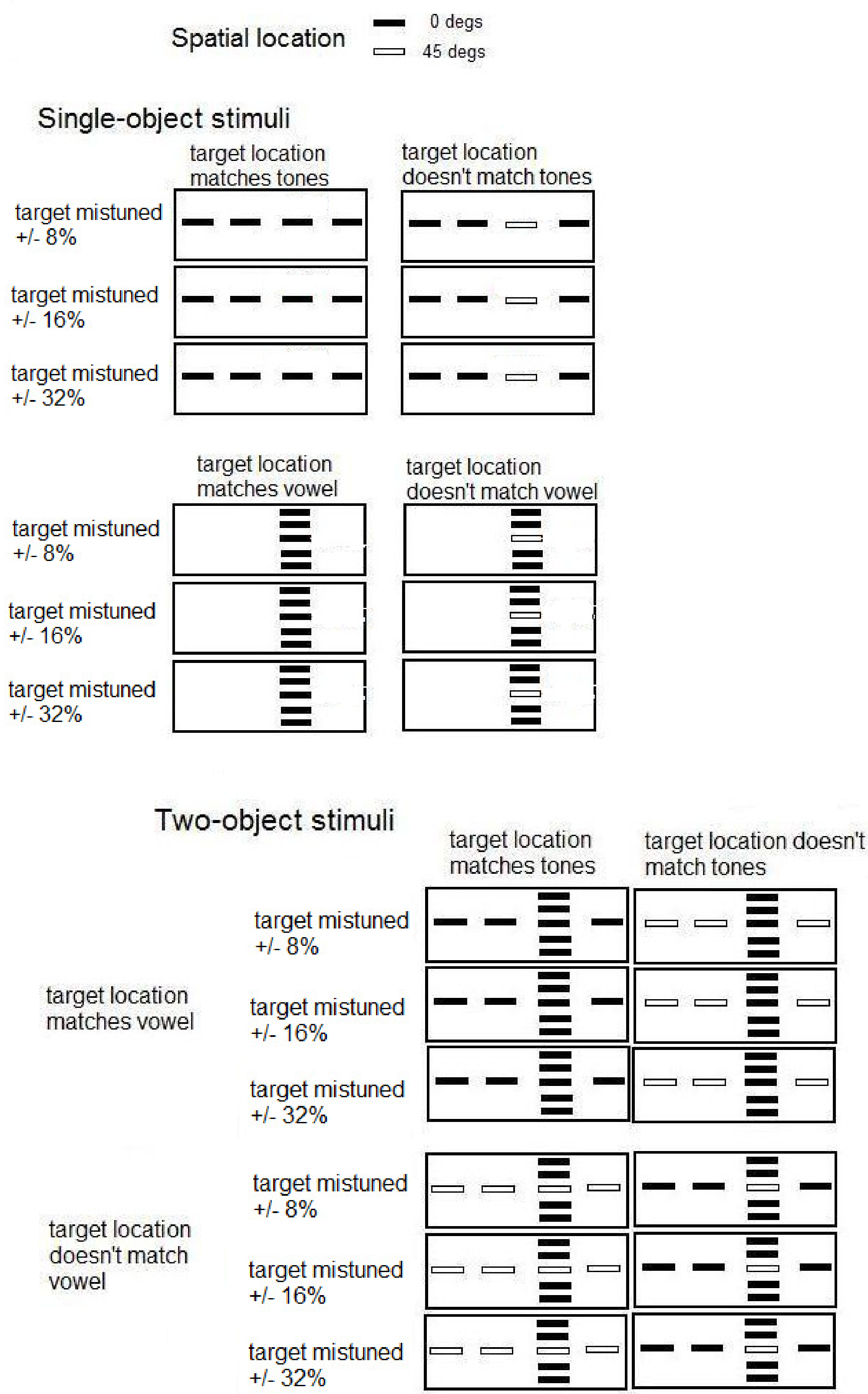


*Figure 6A.* Experiment 2 results of the noise masking manipulation (collapsed across the low, medium, and high levels of noise.

Noise

Attend-Tones | Attend-Vowel

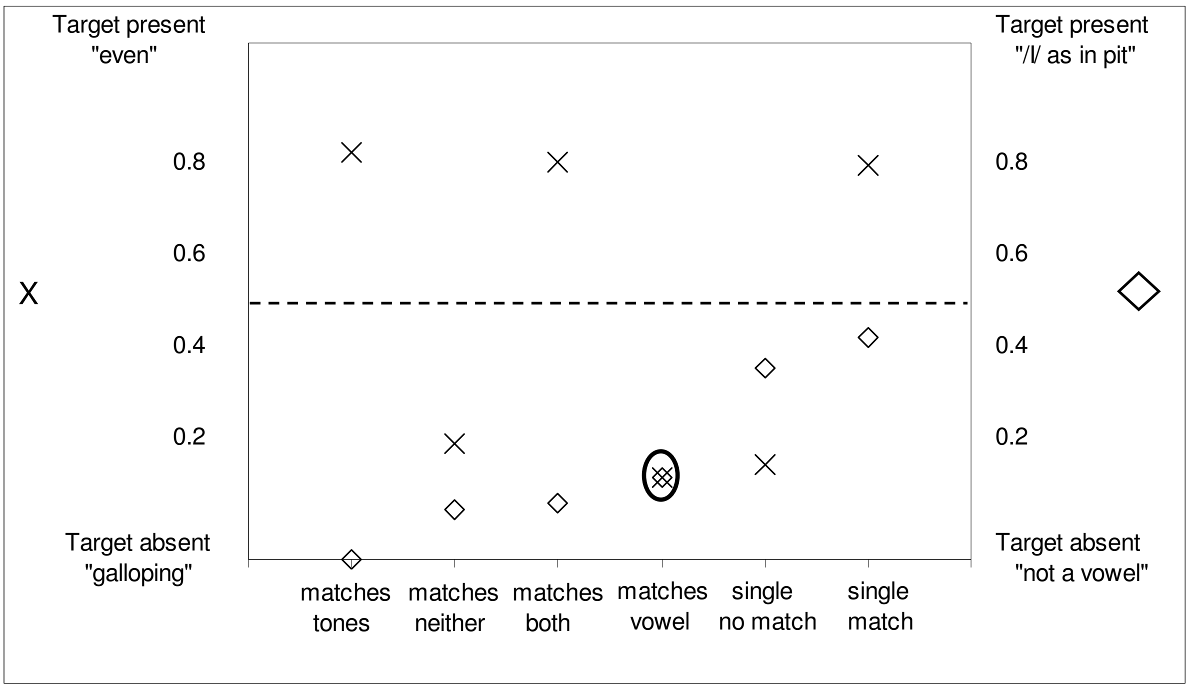


*Figure 6B.* Experiment 2 results of the mistuning manipulation (collapsed across the low, medium, and high levels of mistuning. The diamond markers represent δ scores.

Mistuning

Attend-Tones Attend-Vowel

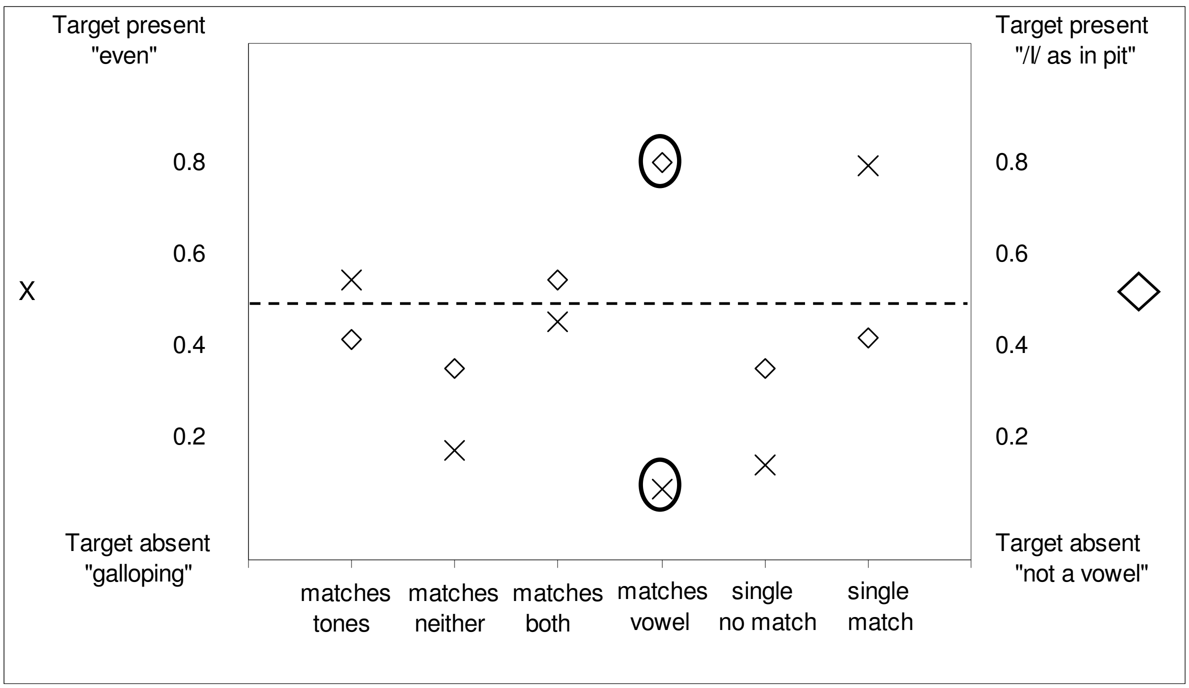


*Figure 7A.* Single-object conditions that were used in Experiment 3.

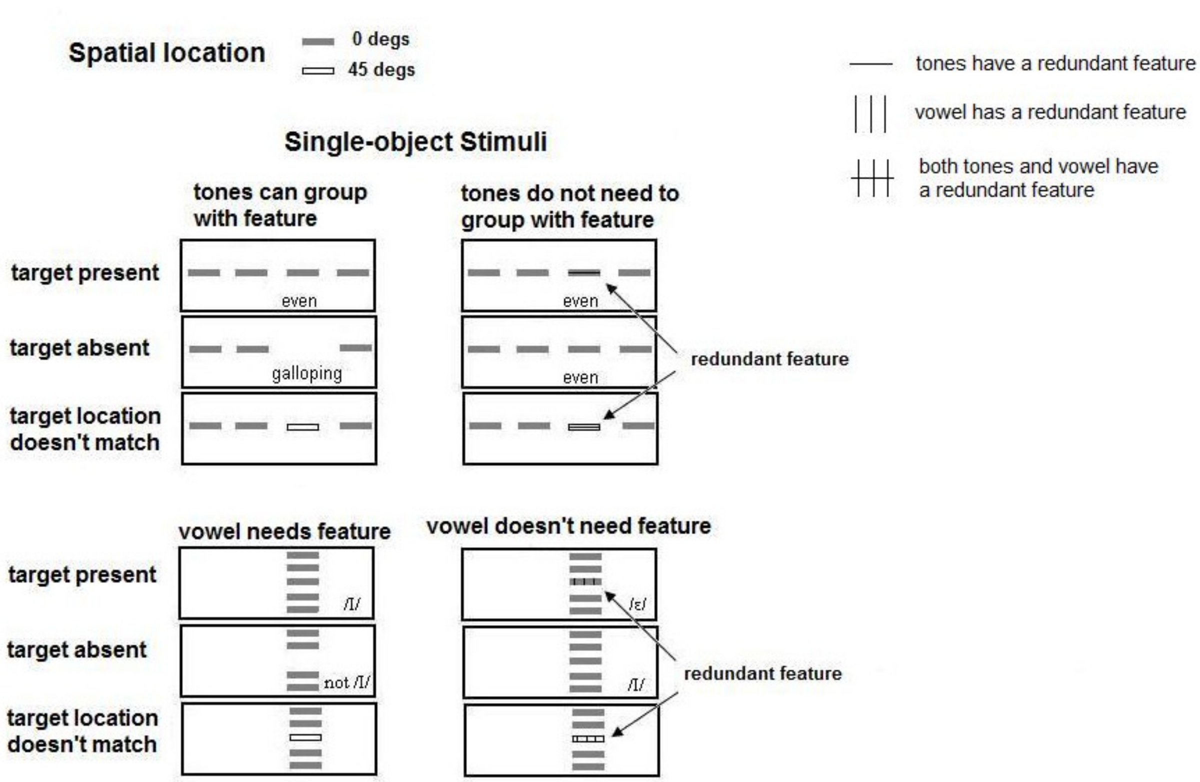


*Figure 7B.* Two-object conditions that were used in Experiment 3.

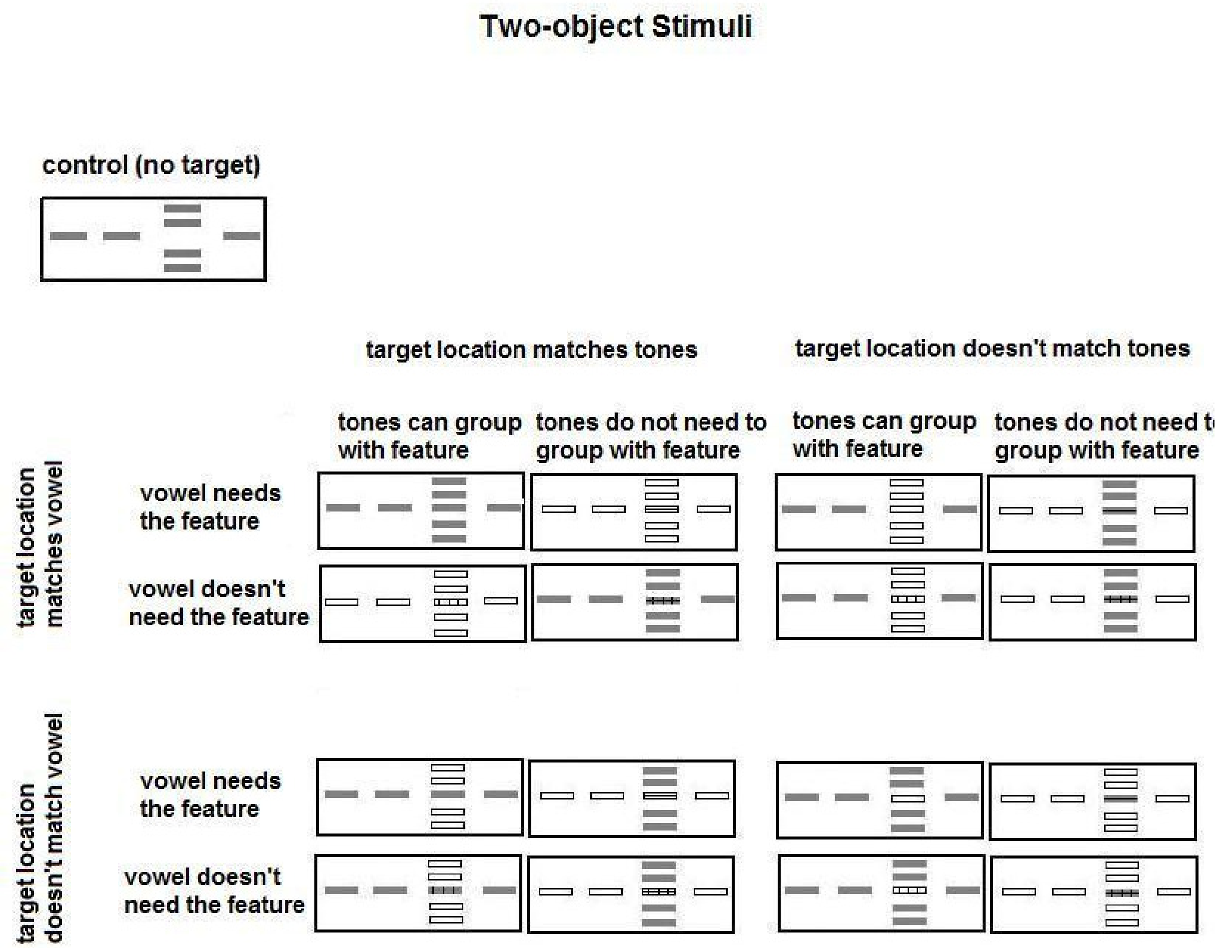
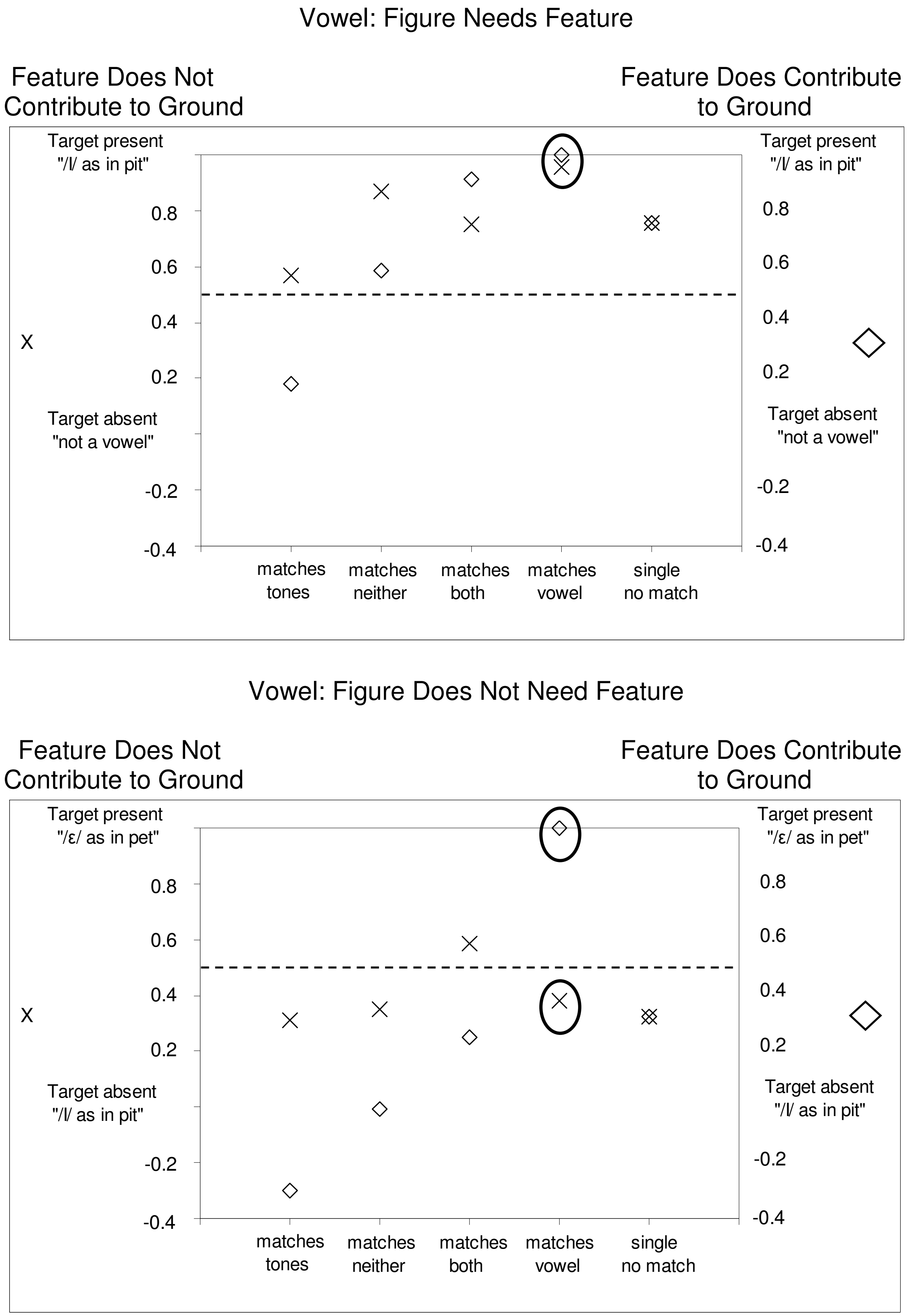
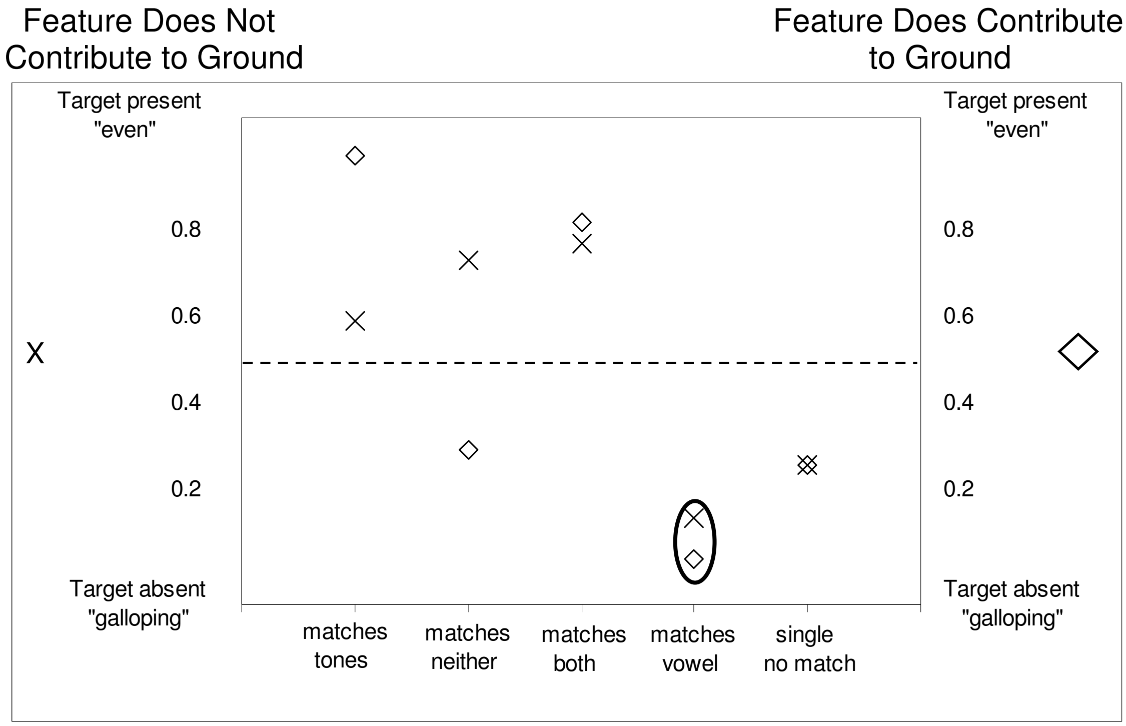


*Figure 8.* Experiment 3 results. The diamond markers represent δ scores.



Tones: Figure Needs Feature




Tones: Figure Does Not Need Feature

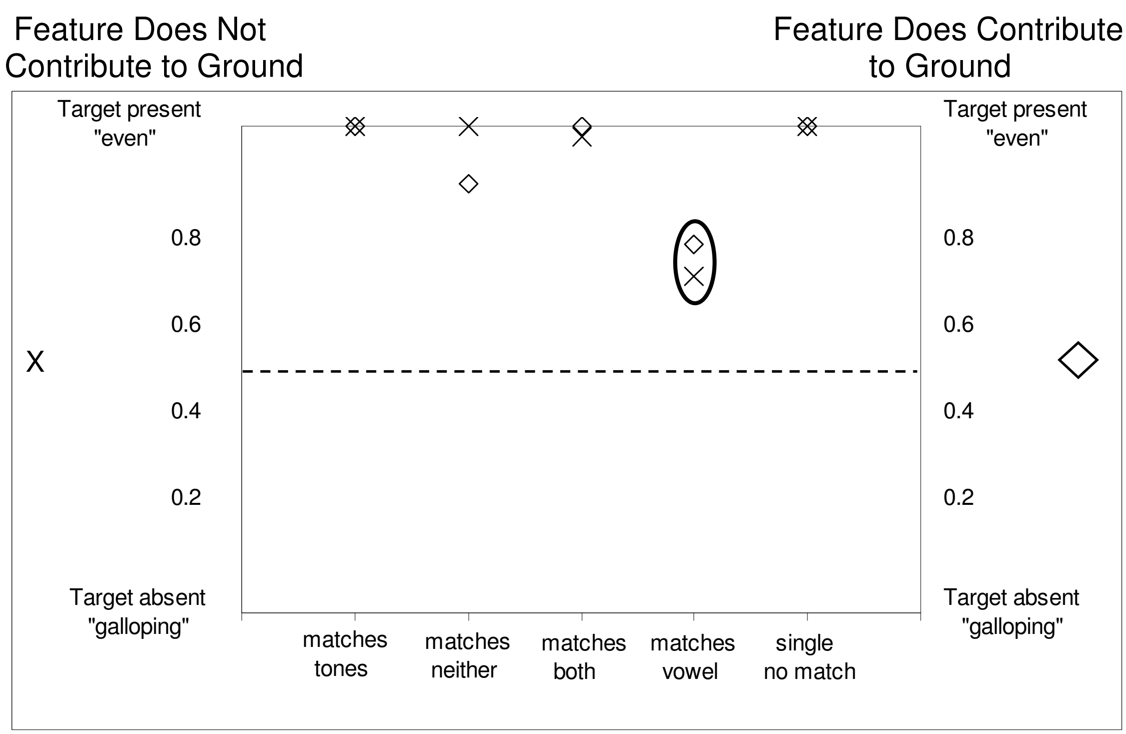


*Figure 9.* Conditions that were used in Experiment 4.

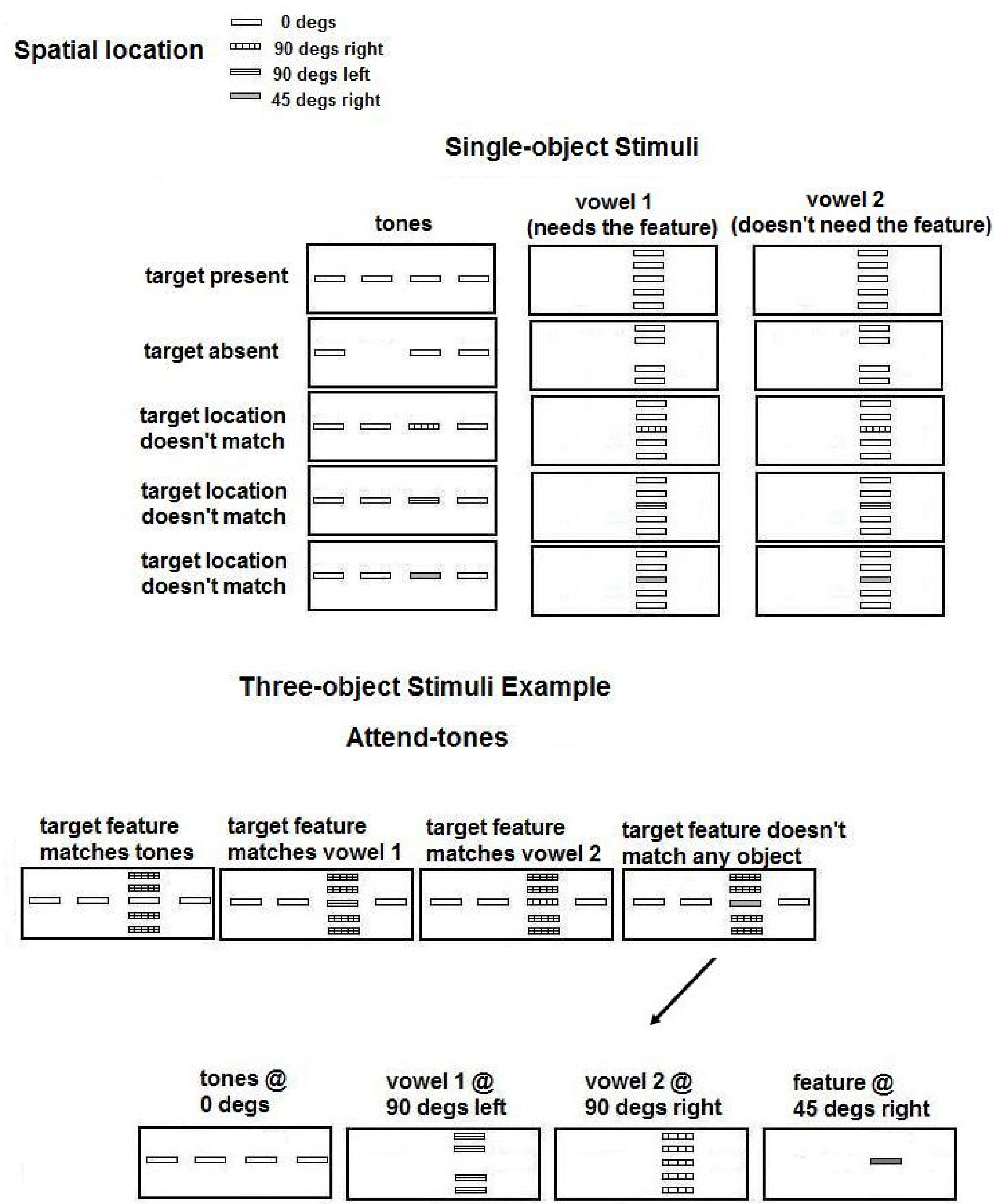
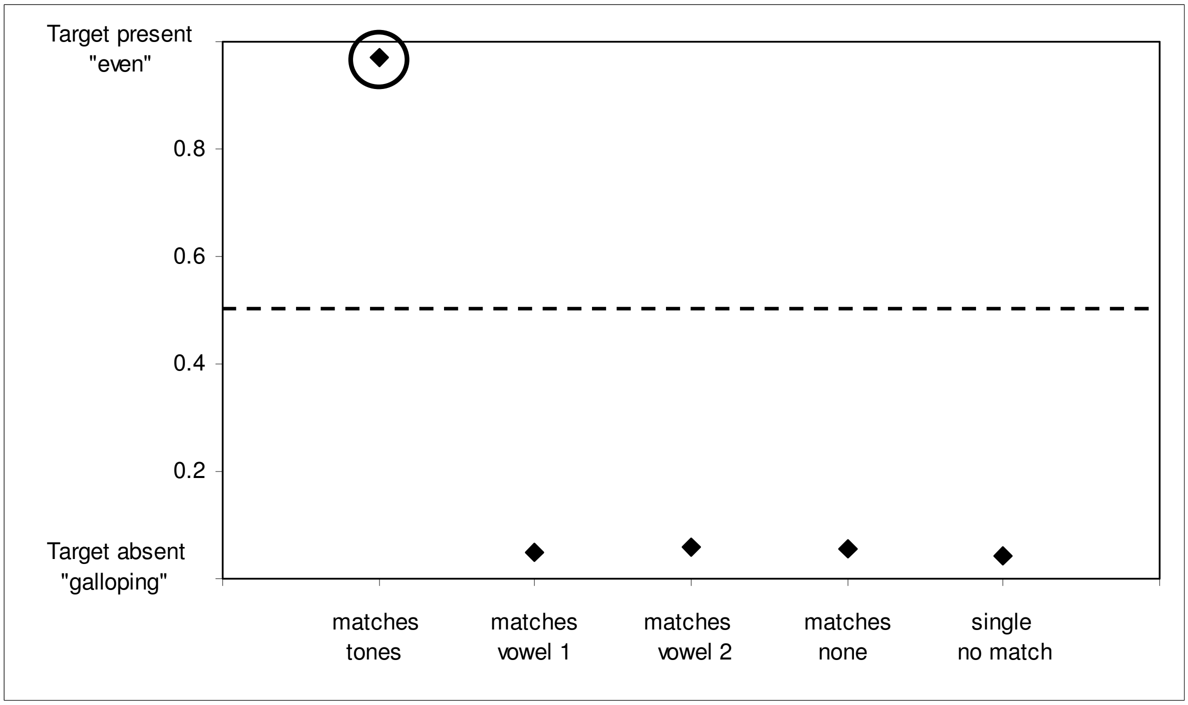
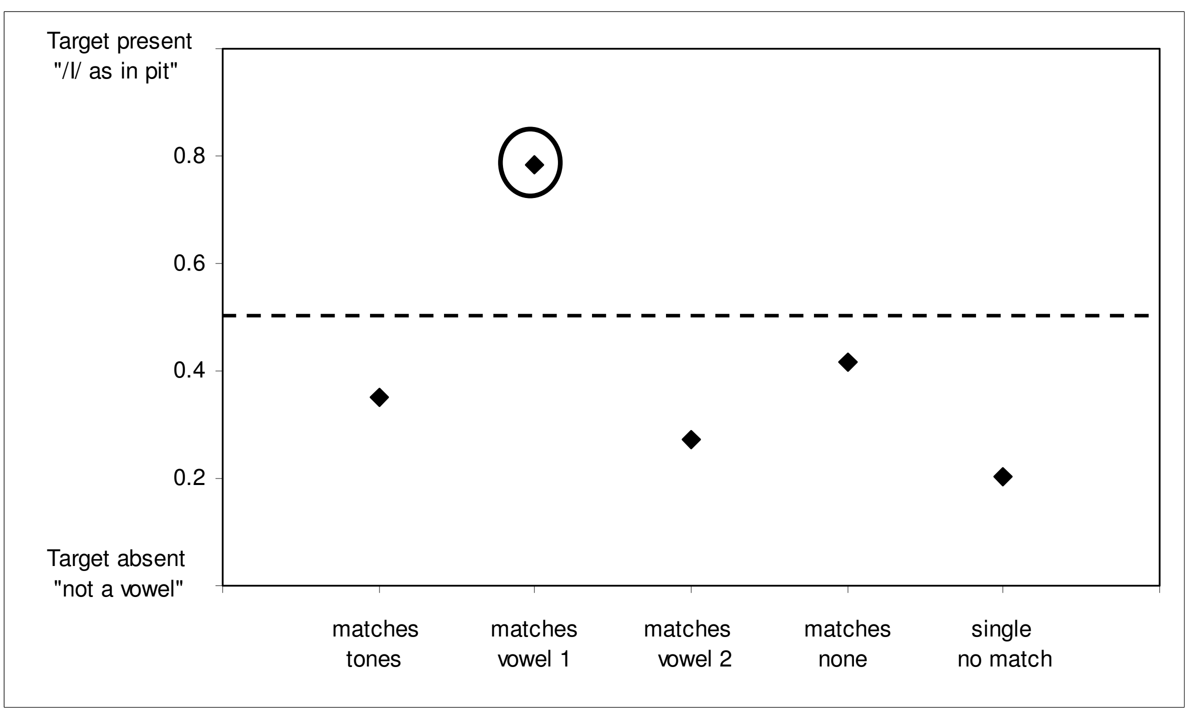


*Figure 10.* Experiment 4 results. The diamond markers represent δ scores.

Attend-Tones




Attend-Vowel 1



Attend-Vowel 2

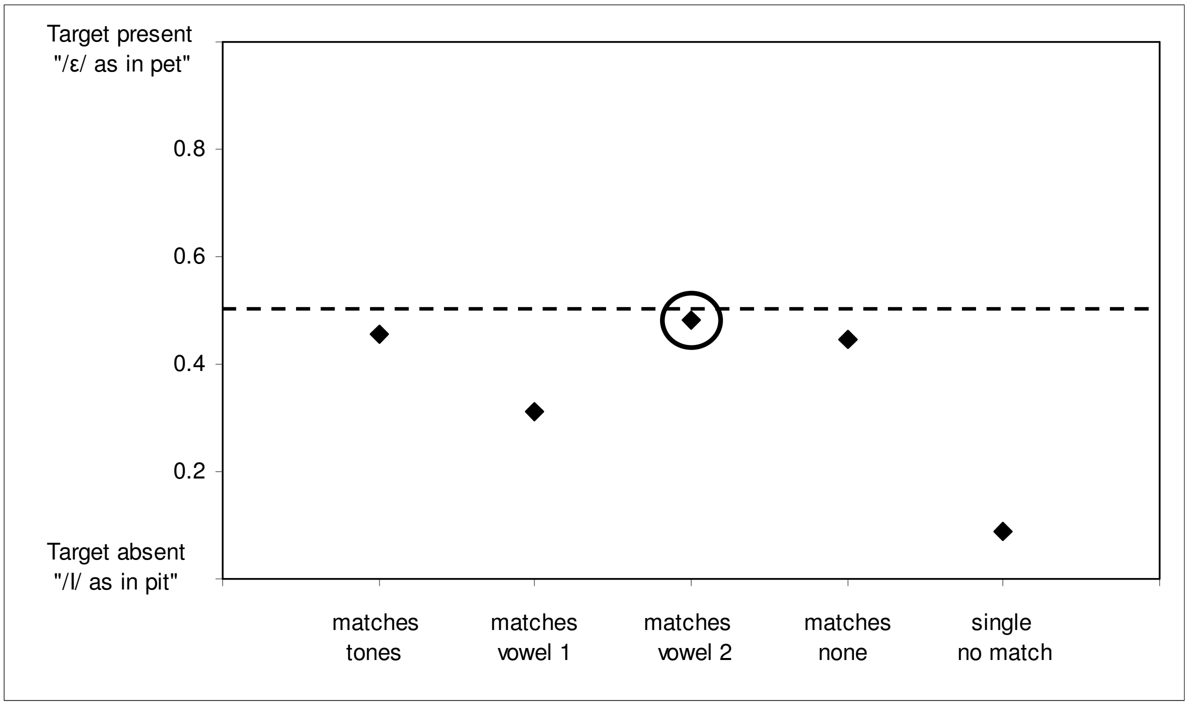


*Figure 11.* Primary task and background stimuli that were used in Experiment 5.

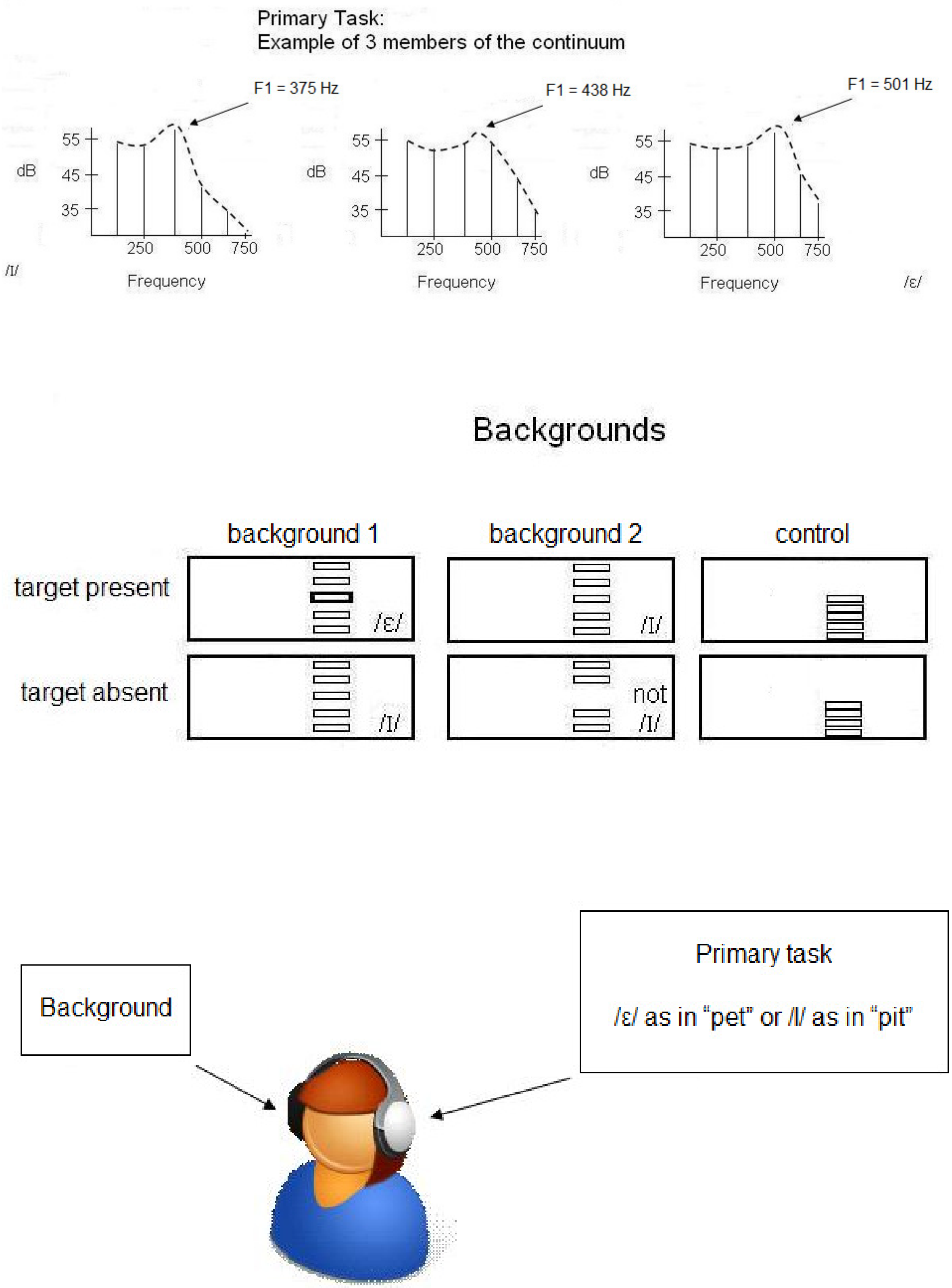
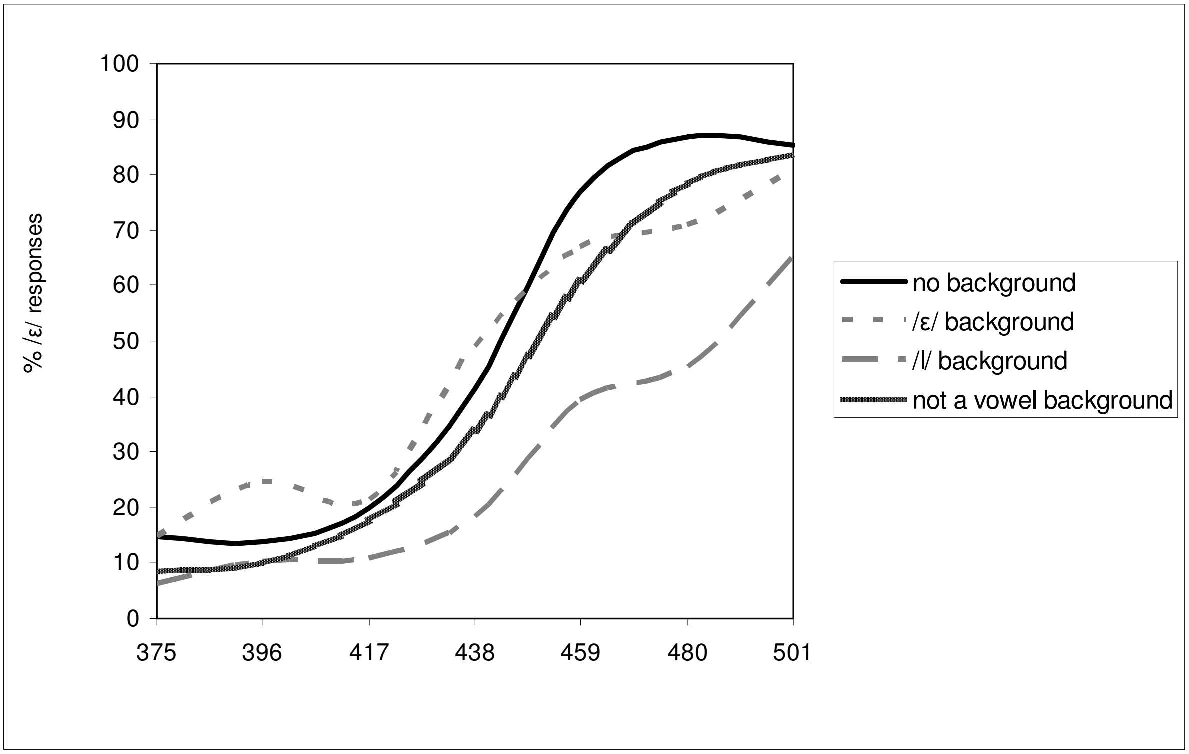


*Figure 12.* Experiment 5 results. Labeling functions are for the group of listeners who could distinguish between the backgrounds before training.



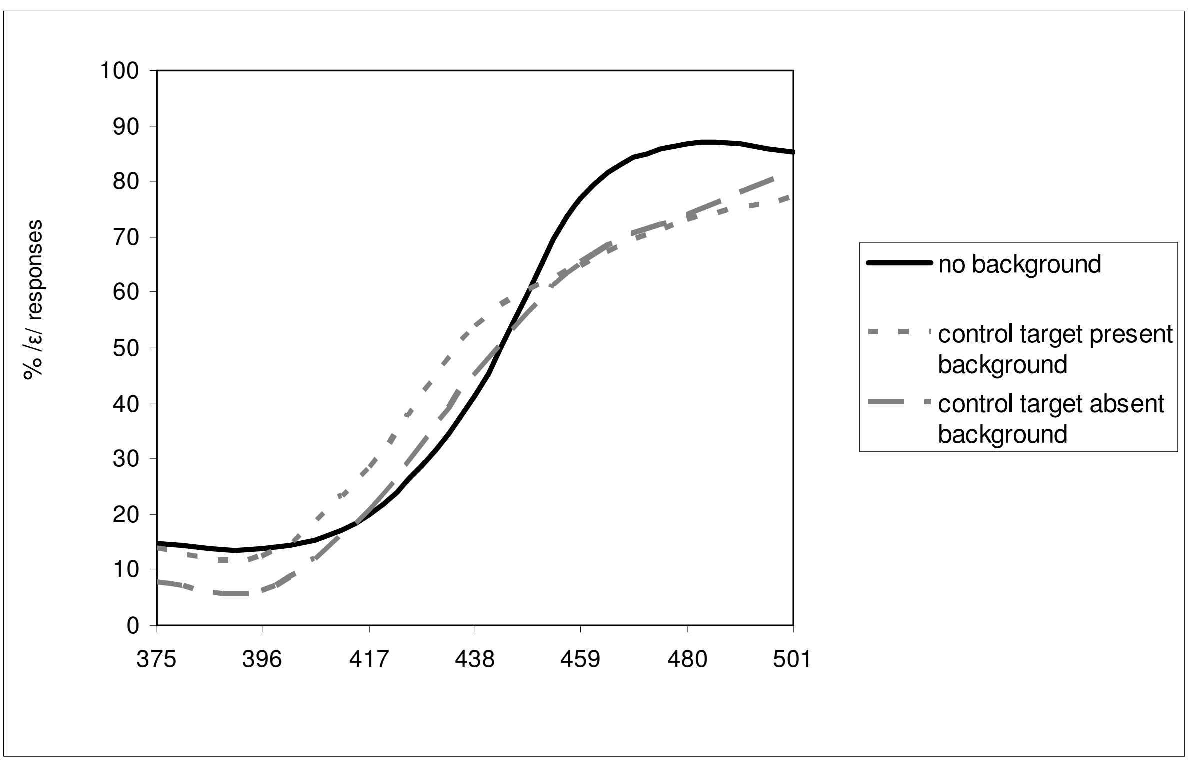



*Figure 13.* Background stimuli that were used in Experiment 6.

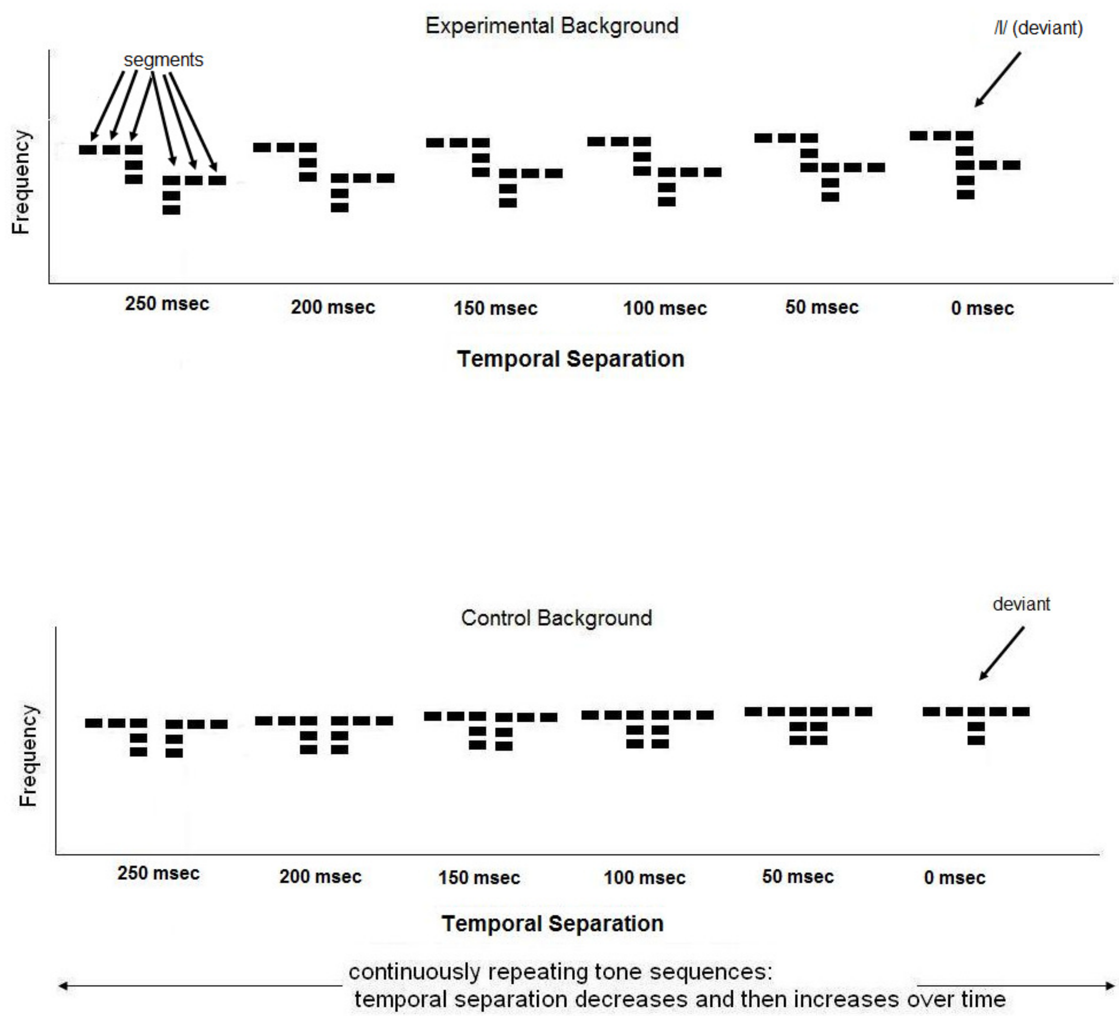
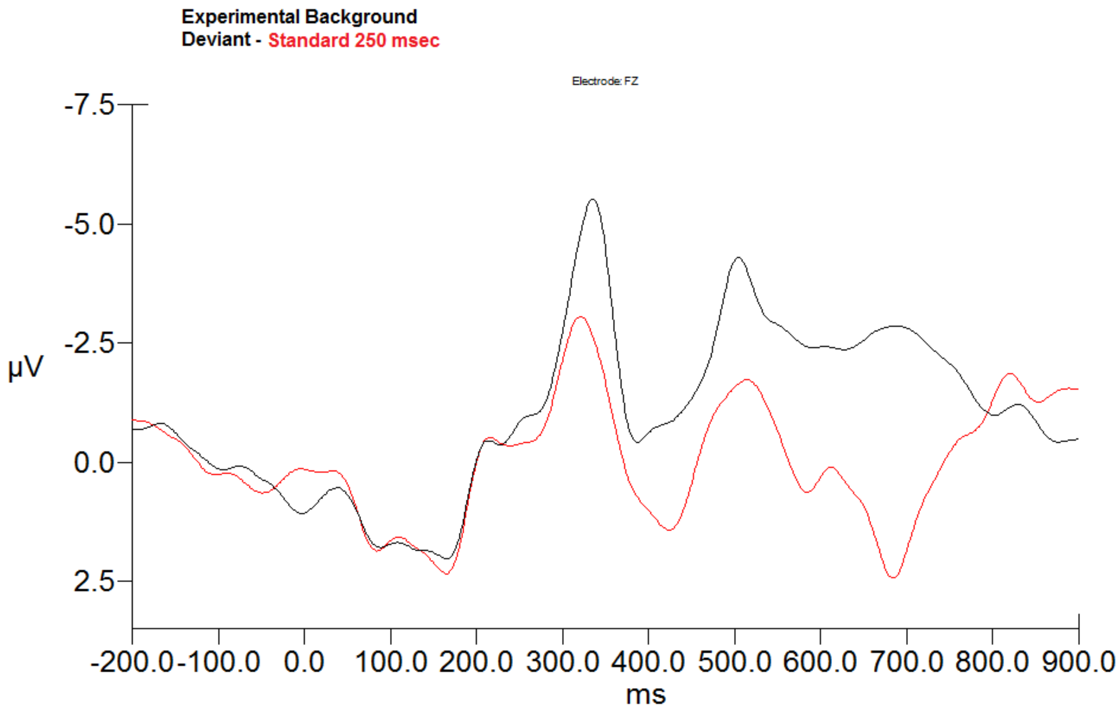
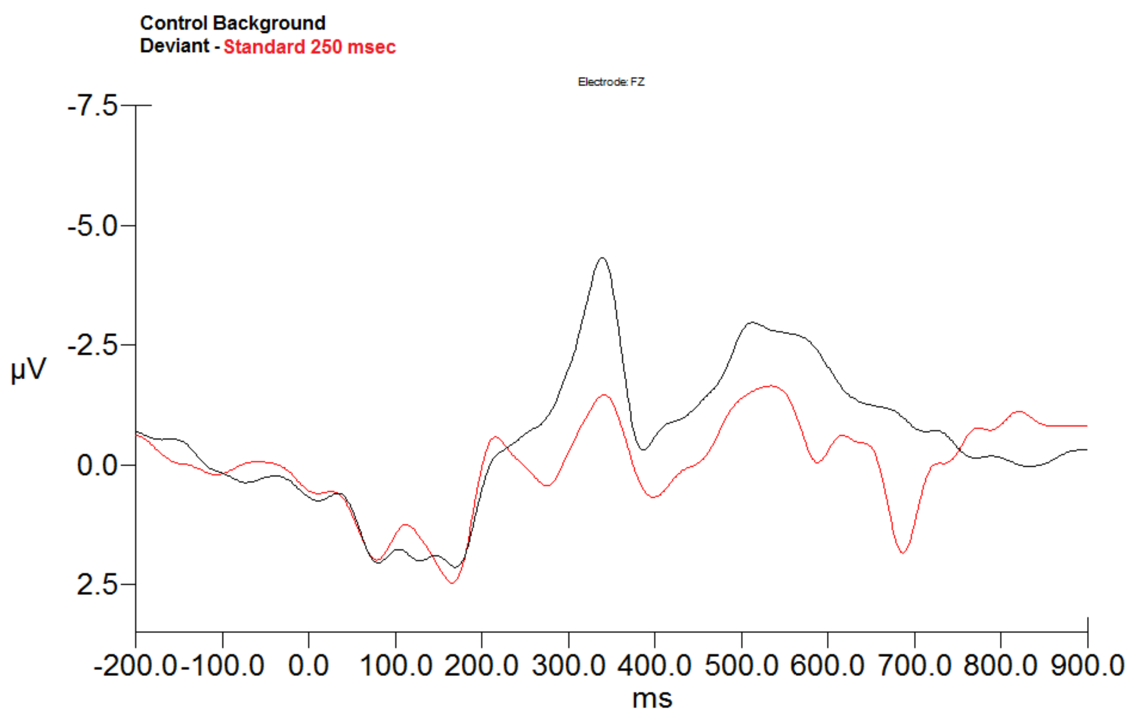
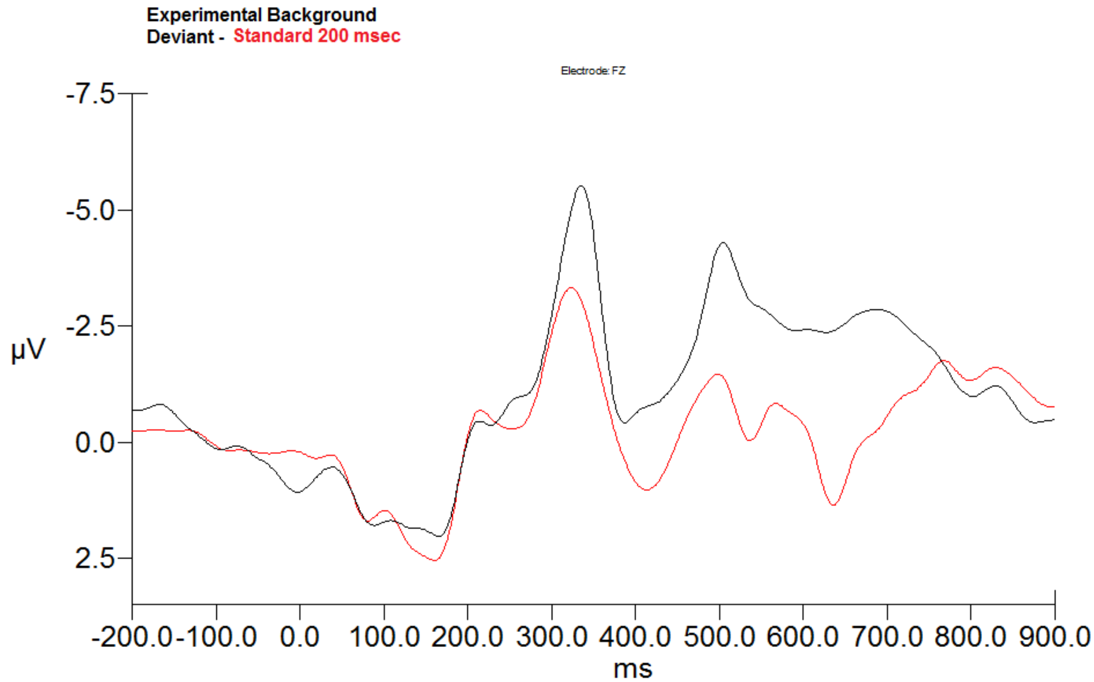


*Figure 14.* Experiment 6 results.

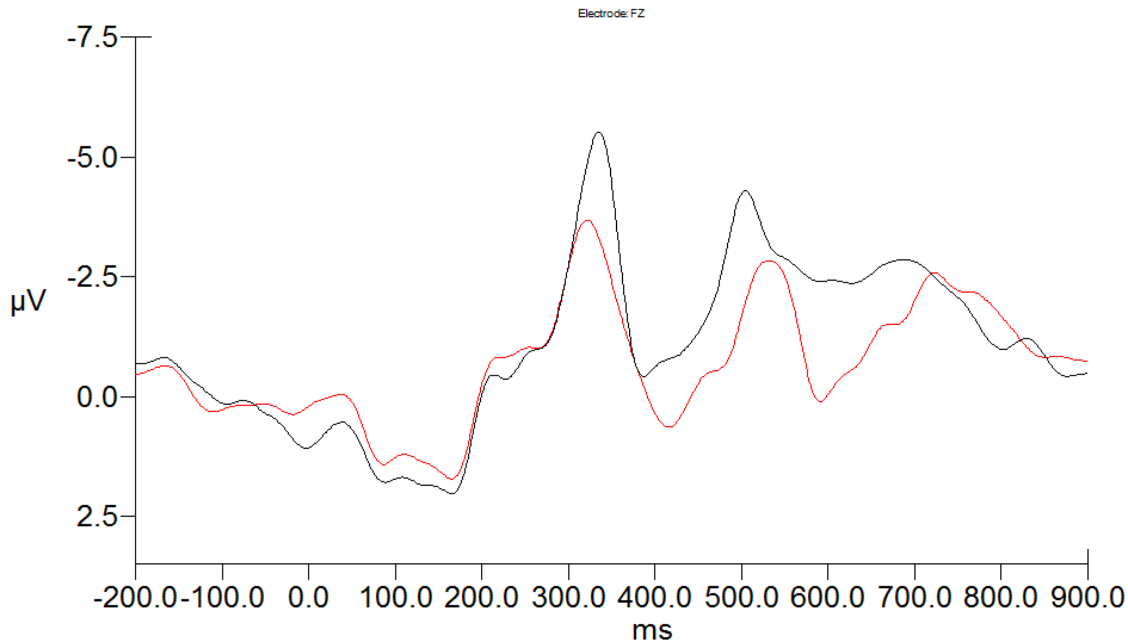


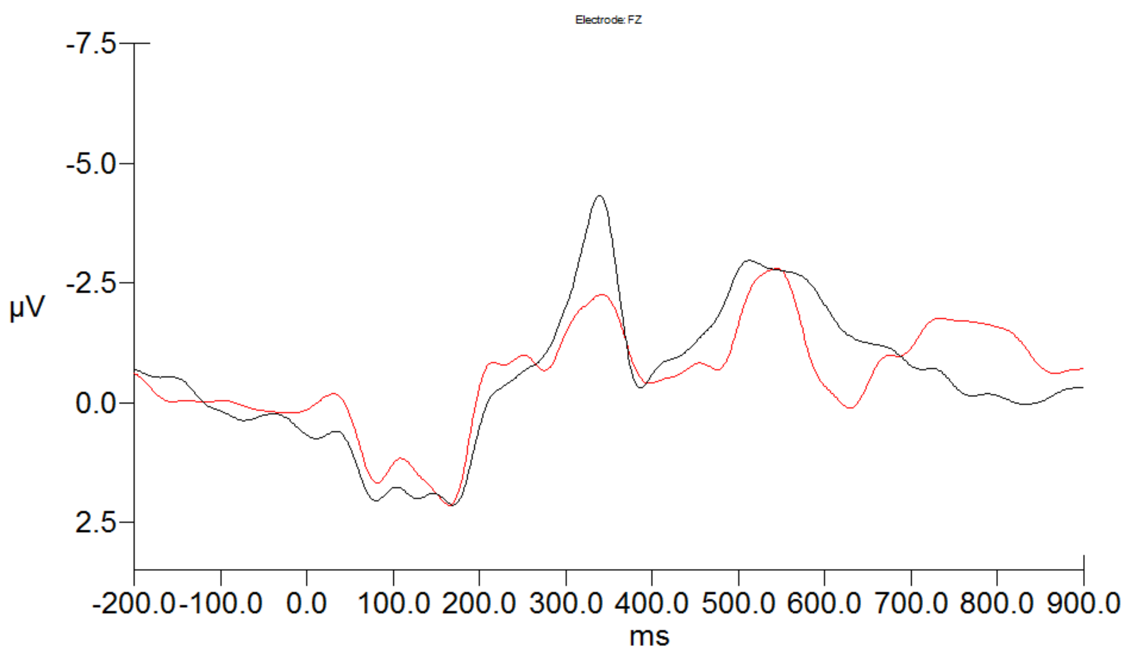

Experimental Background
Deviant - Standard 200 msec
Electrode: FZ
-7.5
-5.0
-2.5
0.0
2.5
µV
-200.0 -100.0 0.0 100.0 200.0 300.0 400.0 500.0 600.0 700.0 800.0 900.0
ms

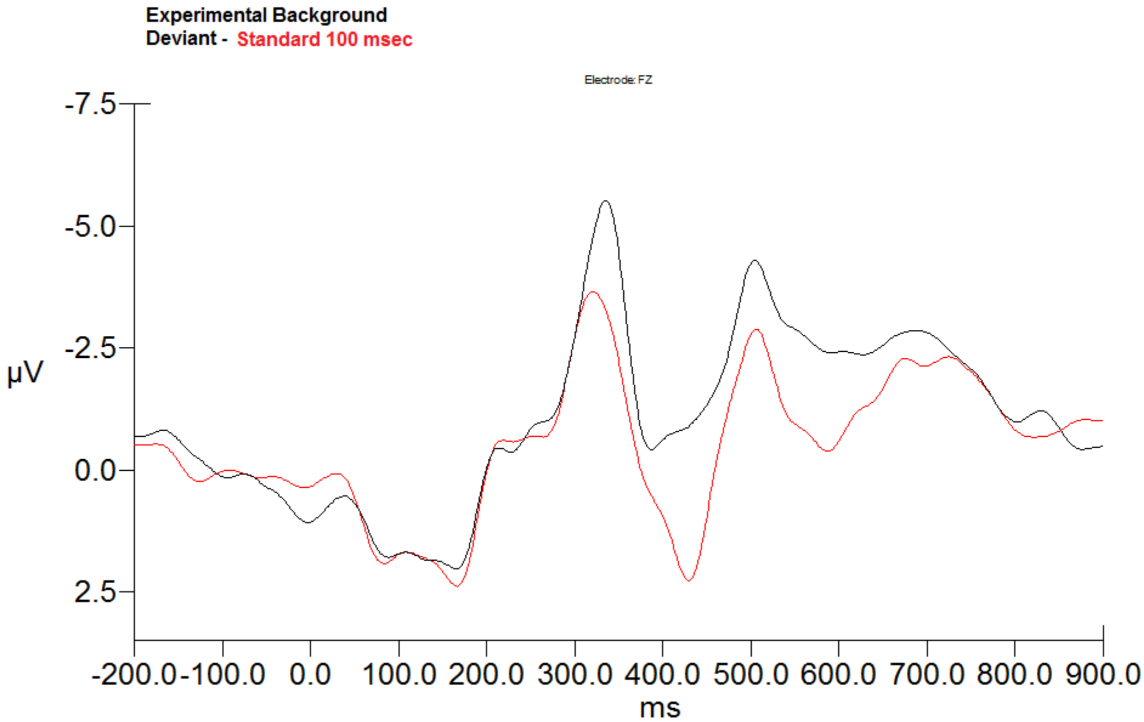

Control Background
Deviant - Standard 200 msec
Electrode: FZ
-7.5
-5.0
-2.5
0.0
2.5
µV
-200.0 -100.0 0.0 100.0 200.0 300.0 400.0 500.0 600.0 700.0 800.0 900.0
ms

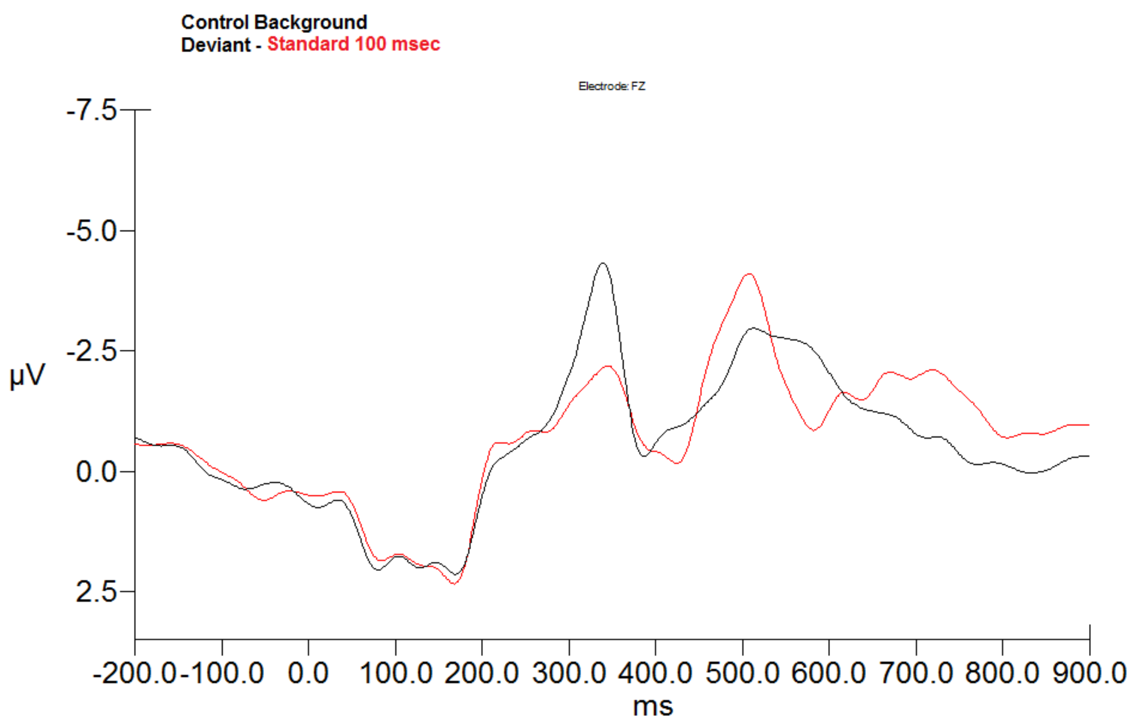
**Experimental Background**
**Deviant - Standard 150 msec**

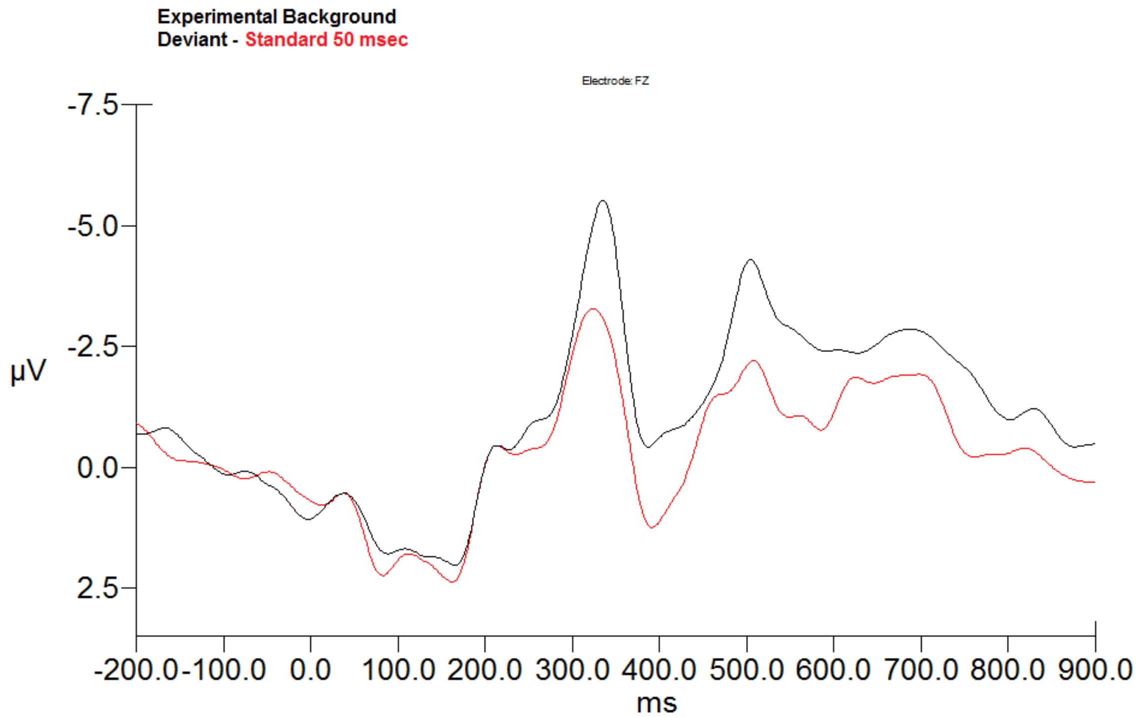



**Control Background**
**Deviant - Standard 150 msec**




Experimental Background
Deviant - Standard 100 msec
Electrode: FZ
-7.5
-5.0
-2.5
µV
0.0
2.5
-200.0
-100.0
0.0
100.0
200.0
300.0
400.0
500.0
600.0
700.0
800.0
900.0
ms


Control Background
Deviant - Standard 100 msec
Electrode: FZ
-7.5
-5.0
-2.5
µV
0.0
2.5
-200.0
-100.0
0.0
100.0
200.0
300.0
400.0
500.0
600.0
700.0
800.0
900.0
ms


Experimental Background
Deviant - Standard 50 msec
Electrode: FZ
-7.5
-5.0
-2.5
µV
0.0
2.5
-200.0 -100.0 0.0 100.0 200.0 300.0 400.0 500.0 600.0 700.0 800.0 900.0
ms

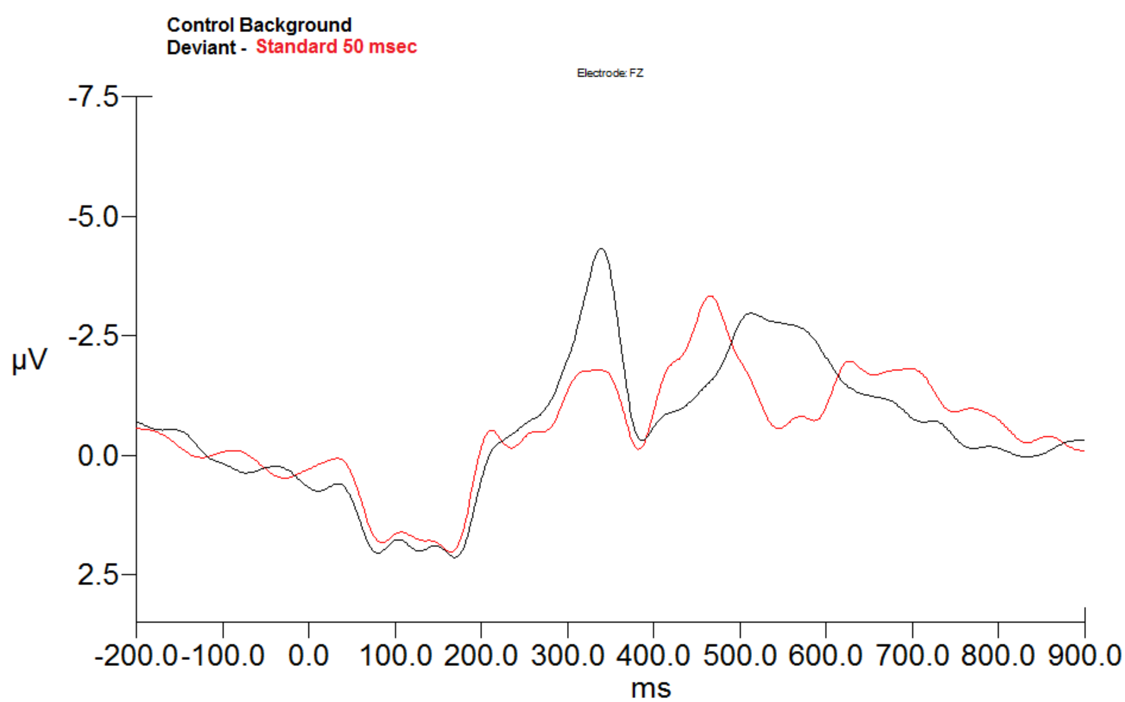

Control Background
Deviant - Standard 50 msec
Electrode: FZ
-7.5
-5.0
-2.5
µV
0.0
2.5
-200.0 -100.0 0.0 100.0 200.0 300.0 400.0 500.0 600.0 700.0 800.0 900.0
ms

*Figure 15*. Primary task and background stimuli that were used in Experiment 7.

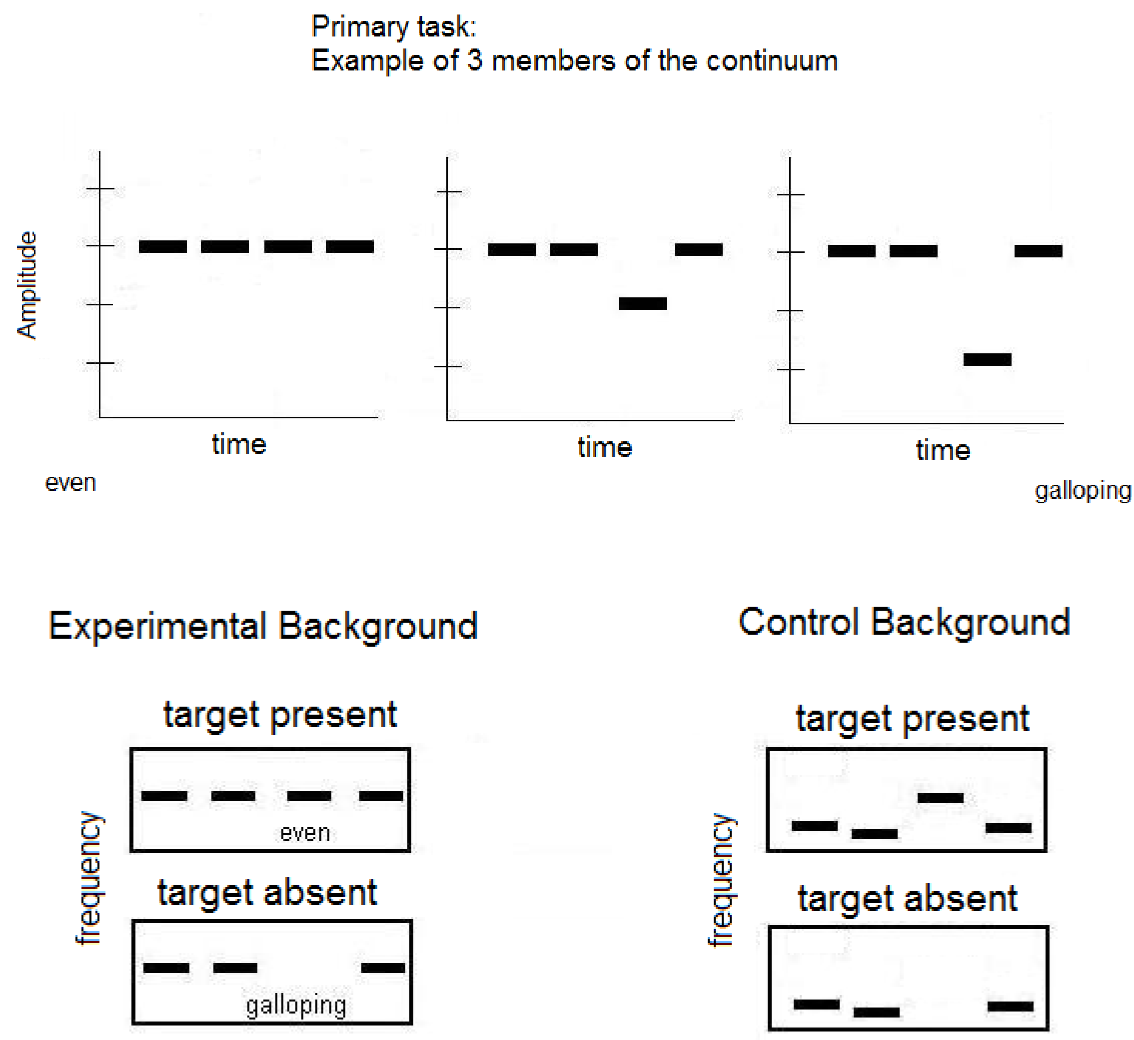


*Figure 16.* Experiment 7 results. Labeling functions are for the group of listeners who could distinguish between the backgrounds before training.

Labeling: No Background