Title
The Influences of Spatial and Motion Properties on Auditory Grouping

Author
Robart, Ryan

Publication Date
2010

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

The Influences of Spatial and Motion Properties on Auditory Grouping

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Psychology

by

Ryan Leslie Robart

March 2011

Dissertation Committee:
Dr. Lawrence D. Rosenblum, Chairperson
Dr. Aaron Seitz
Dr. Khaleel Razak

Signature Approval

The Dissertation of Ryan Leslie Robart is approved:

Committee Chairperson

University of California, Riverside

I would like to thank my advisor and mentor, Lawrence D. Rosenblum for his invaluable instruction, guidance, patience and support that he has so gracefully bestowed upon myself over the years. I truly feel blessed to have had so much interaction with him during my time at the University of California, Riverside. Also, I would like to thank Aaron Seitz, Khaleel Razak, and Curt Burgess for providing so much support during the course of the project. I also would like to express my gratitude to the Psychology Department faculty and administration for welcoming me into the program, aiding in my academic endeavors as well as my general wellbeing. I especially would like to thank Faye Harmer for her dedication to helping me stay focused and grounded during my progression. Finally, I would like to express my undying gratitude and appreciation for the support of my family throughout these many years of education. This project is a testament to all of you, and I hope you feel as proud as I do in its completion. I thank you all from the bottom of my heart.

ABSTRACT OF THE DISSERTATION

The Influences of Spatial and Motion Properties on Auditory Grouping

by

Ryan Leslie Robart

Doctor of Philosophy, Graduate Program in Psychology
University of California, Riverside, March 2011
Dr. Lawrence D. Rosenblum, Chairperson

Despite a plethora of research and theory on spatial hearing and the perceptual organization of sound, the question of the relative importance of spatial and frequency relations in low level auditory grouping remains (Bregman, 1990; Rogers & Bregman, 1993; 1998; Strybel and Neale 1998; Kubovy & VanValkenberg, 2001). Most researchers share the assumption that frequency relations of sounds dominate spatial relations in the perceptual organization of sound. In a natural context, sound sources (e.g. automobiles) create sound that has a broad range of frequency content, but a limited –and often predictable–- range of motion. However, in a natural context, the coherent movement of sound components may be as important, if not more, in forming auditory groups than frequency similarity among those components. The research conducted in this project tested a new theory that assumes spatial relations can sometimes dominate frequency relations in grouping (dynamic spatial assessment after initial localization [DSAIL]). Based on the free-field methods used by Rogers & Bregman (1993; 1998), listeners' in these experiments reported the extent of their perceptual grouping of simple tone sequences under various spatial manipulations. In many cases, the results showed that grouping varied systematically with coherent motion-type [opposite trajectory vs shared trajectory] and angular separation [far (>90°) vs near (<90°) among the tones. The implications of these results are discussed.

Table of Contents

List of Figures

Chapter 1:

Introduction

Theory and research on low-level (primitive) auditory perception has built upon the results of localization experiments (see Strybel 1994 for a review) to those that focus on how and why we organize simple sounds into meaningful patterns (e.g. stream segregation and auditory grouping) (Rogers & Bregman, 1993; 1998). Like any empirical framework, research on auditory perceptual grouping (grouping) relies on rigorously tested concepts and assumptions of one or more theories. The theoretical approach of the current project is to challenge the prevailing view that when segregating sounds, the auditory system will overwhelmingly favor frequency attributes (pitch) over spatial attributes (spatial location) (Bregman, 1990). This assumption (frequency dominance assumption), is generally accepted in the field, but is arguably based on evidence from studies that used inadequate spatial manipulations such as headphone presentations. The purpose of the current research is to test this assumption under conditions of motion and various angular separations between sound sources.

This dissertation will begin with a summary of the theoretical framework behind the frequency dominance assumption (Bregman, 1990). This review will generally focus on auditory stream segregation and grouping as well as the central principle thought to govern the process as outlined by Bregman (1990). The evidence that spawned the frequency dominance assumption will then be reviewed. In turn, some evidence that suggests that spatial location can dominate frequency in these processes will also be reviewed.

The research conducted in this project used the conceptual tools and components of many established methods to revisit the frequency dominance assumption. Primarily, the methods were built from Rogers & Bregman's (1993,1998) induction paradigm to approach the question of the importance of spatial location in stream segregation and grouping. Based on arguments that call for more naturalistic stimulus presentations (Handel 1998; Gaver, 1993; Neuhoff 2001), it was predicted that spatial cues such as spatial separation and smooth motion could dominate frequency similarity in group formation. The next chapter reviews an overarching theoretical framework known as Auditory Scene Analysis, which is at the forefront of explanatory theories of auditory perception.

Chapter 2:

Auditory scene analysis

When a listener hears a sequence of ringing bells from a bell tower, they perceive those sounds as a part of a single event. The listener can easily perceive the ringing bells even if they are in a crowded area filled with people talking, ringing cell phones, ticking bicycles, and traffic. Explaining the ability to 'hear out' certain collections of sounds from a cacophony of sonic events has been one of the primary goals of Auditory Scene Analysis (hereafter ASA) (Bregman,1990).

From the perspective of ASA, the ability of a listener to perceive certain collections of sounds as groups is due to a process of perceptual organization known as stream segregation and/or *grouping* (Bregman, 1990). In the bell-tower example, the listener hears the ringing bells as an *auditory group* (group) so that they may stand out against the conglomerate of other sonic events (figure 2.1). The bells are *grouped* together into one auditory object. This can only happen if the bells are *segregated* from the other sound components that surround the listener.

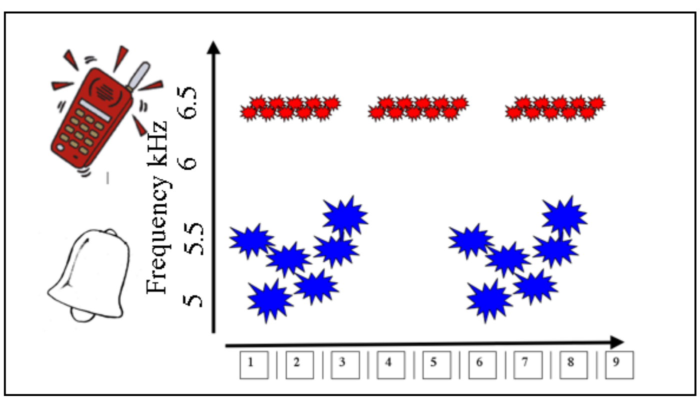

Figure 2.1. The ringing of the bells(blue flashes) are heard as a group (*grouping*), and are segregated from the the rings of cell phone (red flashes).

Even though each bell in the tower may ring at a different frequency and each bell ring may be distinct in time, the listener hears the collection of sounds as if they are united—i.e. a group. Presumably, this is what allows humans to listen to music and have a conversation simultaneously without confusion (Bregman, 1990).

Bregman and colleagues (1990) have proposed that there are two types of grouping. Simultaneous grouping occurs with sounds that have overlapping envelope structure. On the other hand, sequential grouping occurs when sounds are heard as distinct in time with one sound following another in sequence. Both the bells in the bell tower and the cell phone ring in the earlier example would be most associated with sequential grouping.

According to ASA, grouping occurs as a result of the auditory system's Gestalt processes (Bregman,1990). The Gestalt principles were initially developed to describe visual perceptual grouping of stimulus elements. ASA uses the same principles to explain auditory grouping. Two examples of Gestalt grouping principles noted are: *the proximity principle* and *the principle of good continuation*.

Proximity principle: Consider the bell tower example. The grouping of the bells is undisturbed by conversation, traffic, or even the ringing of a cell phone. ASA holds that this can be explained by using the frequency *proximity principle*. The proximity principle states that sounds that are similar in frequency, are more likely to be grouped. In the bell tower example, the frequencies emitted from the bell tower are proximate to one-another in frequency, and are generally dissimilar in frequency to extraneous sounds.

This aids a listener in segregating the bell tower sounds from other sounds that may surround them.

Good continuation principle: Imagine a siren used for police cruisers or ambulances. In its simplest form, a siren consists of a sine wave tone that smoothly glides up and down in frequency (figure 2.2).

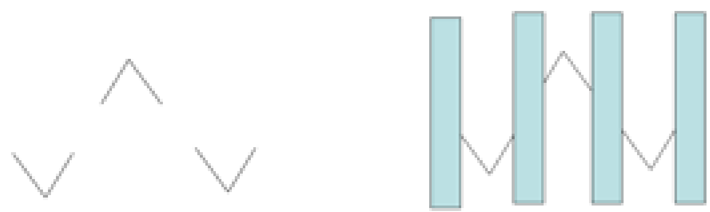

Figure 2.2. An illustration of grouping by frequency considered to be analogous to the visual grouping principle of continuation. The lines represent distinct frequency glides, while the bars represent noise bursts inserted into the silence which separates the frequency glides.

In this case, a listener would hear 3 distinct tones. However, if a researcher would insert noise bursts in those silent spots as demonstrated in the 'burst' portion of fig 1, the listener would fuse the tones. While the listener would hear the noise bursts and the tones, the listener would hear the siren type glide continuing behind the noise. This effect, Bregman (1990) has asserted, exemplifies that the principle of good continuation.

ASA: differences in primary information for visual and auditory Gestalts

According to proponents of ASA, there is a critical difference between the visual and auditory Gestalt principles (Bregman, 1990; Kubovy and VanValkenberg, 2001). The visual principles are based on the *spatial relations* among elements, while the auditory counterparts are based on *frequency relations* among elements (Bregman, 1990). In other words, proponents of ASA assert that physical space is vital for visual grouping and is not primary for audition. Instead, Bregman (1990) proposes that the analogous mapping should be made in frequency space. Bregman (1990) extends this mapping to include apparent motion. This is evident in Bregman's (1990 p 176) statement, "Horizontal position in space [visual apparent motion] is made to correspond to the frequency of tones, with time playing the role of the second dimension in both cases." This analogy is noted throughout ASA's definitive text (Bregman, 1990), and this idea is generally accepted in the grouping literature (see Darwin 2005 for a review). The next section will focus on how Bregman (1990) has parametrically described frequency and time's affect on grouping and segregation.

What facilitates grouping and segregation?

The research conducted on auditory grouping has attempted to parametrically define conditions that facilitate and/or diminish such grouping(s) (Bregman, 1990; and see Best, Gallun, Carlile & Shinn-Cunningham, 2006). Recall that in the bell tower example, the frequencies of the extraneous sounds were too dissimilar to those of the bells to be perceived as part of the same group. The critical point when sounds are segregated into separate groups is known as the *temporal coherence boundary* (Bregman, 1990 p. 60).

The temporal coherence boundary is determined by automatic, i.e. primitive, cognitive processes. In other words, the listener cannot help but segregate sounds into two groups –or *streams*– if the sounds are sufficiently dissimilar in frequency, and are presented at either extremely rapid (e.g. 20 per second) or extremely slow rates (1 per second). Also critical, are the durations of the inter-stimulus intervals (ISIs), or the duration of silence between each tone in a sequence. Research has generally shown that irregular ISIs that are either too short (e.g. 5-10 ms) or too long (500 ms) can play a large role in stream segregation (see Bregman, 1990, Chapter 2).On the other hand, if a tone sequence contains tones that are sufficiently similar and/or have a suitable duration(s) (e.g. 50-200ms) along with regular ISIs, a listener will probably hear the sequence as one coherent stream.

Generally, stream segregation and grouping are closely related. In the bell tower example, the listener must *segregate* the ringing from the background to form the group of ringing bells. However, research dating back to 1950 (Miller & Heise, 1950 as cited in Bregman, 1990) has shown that listeners tend to segregate high and low tones in repetitive sequences when heard for longer than 4 seconds. This occurs even if the tones are initially heard as a single stream (galloping). In other words, if while listening to a repetitive high-low-high tone sequence, a listener initially forms these tones into a single group, after a few seconds, there is a tendency for the listener to hear the tones as two separate groups (Bregman 1990; Rogers & Bregman, 1993; 1998). This *segregation effect* has been hypothesized to be due to an accumulation of evidence for more than one stream of sounds that is lacking in stimulus presentations under four seconds. Therefore, the duration of the entire sequence is another example of a critical parameter of the temporal coherence boundary in ASA.

It should be noted that spatial location of a sound source is excluded from any discussion of the critical boundary for grouping. In fact, spatial location is considered to be less important than frequency similarity among tones as well as temporal presentation parameters in stream segregation and the formation of groups (Bregman, 1990; Kubovy & VanValkenberg 2001). The next chapter will provide a brief summary of some of the research that has driven the frequency dominance assumption and the principle of frequency proximity.

Chapter 3:

The frequency dominance assumption

In ASA's definitive text, Bregman (1990) often points to illusions as a basis for theoretical assertions of frequency dominance. Two of these spatial illusions driven by frequency are discussed in the following sections.

The octave illusion

The octave illusion (Deutsch & Roll, 1976) is often cited as an example of the dominance of frequency proximity over spatial proximity (location) in grouping. In the original study, listeners were presented (over headphones) with binaural sequences consisting of 250 ms steady-state sine wave tones at equal amplitude with no gaps between them. The sequences contained a staggered presentation of tones one octave apart (400 Hz and 800 Hz) between ears (figure 3.1).

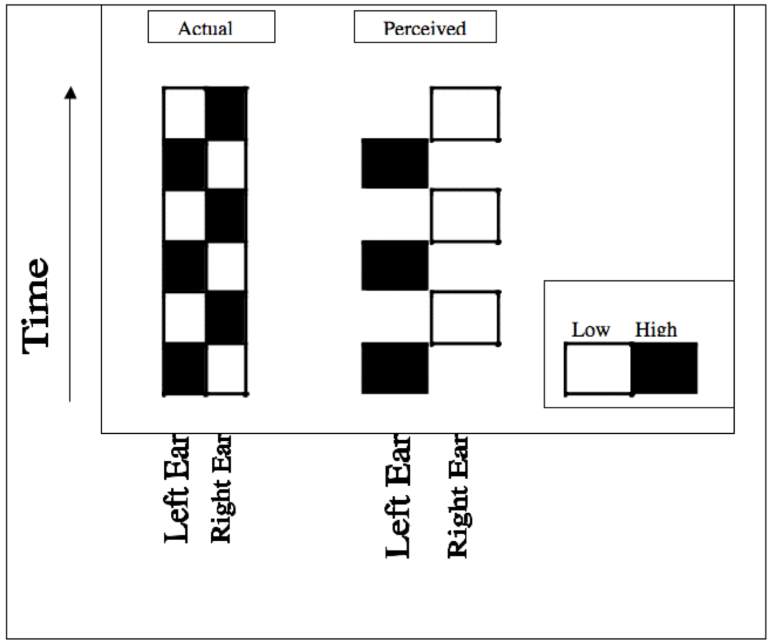

Figure 3.1. An illustration of frequency dominance over spatial location demonstrated by the octave illusion (Deutsch & Roll 1976).

As illustrated in figure 3.1, when the left ear was receiving a 400 Hz tone the right ear was simultaneously receiving an 800 Hz tone and vice versa. The sequence continued to loop in this pattern. Both sequences (left and right channel) began and terminated simultaneously. Results showed that most listeners perceived a repeating sequence of 800 Hz tones in one ear, and a sequence of 400 Hz tones in the other, even though they were receiving both tones in both ears. Listen to an example presentation at:

http://www.brl.ntt.co.jp/IllusionForum/basics/auditory/onkai-e.html

It was concluded that listeners exhibited this spatial "octave illusion" as a result of the dominance of perceptual grouping by frequency proximity over spatial location (Deutsch

& Roll, 1974; also see Bregman, 1990). This interpretation of the illusion persists in the more recent literature (Kubovy and VanValkenberg, 2001).

The spatial continuity illusion

Bregman (1990) often refers to another spatial illusion determined by frequency known as the *spatial continuity illusion*. This illusion was first described by Warren, Osucek & Farmer (1969) based on the results of an experiment on auditory masking. The original stimuli were two 'gliding' sine wave tones that were played to listeners over headphones. One tone was presented to one ear (left or right), and its pitch glided downward in frequency (figure 3.2).

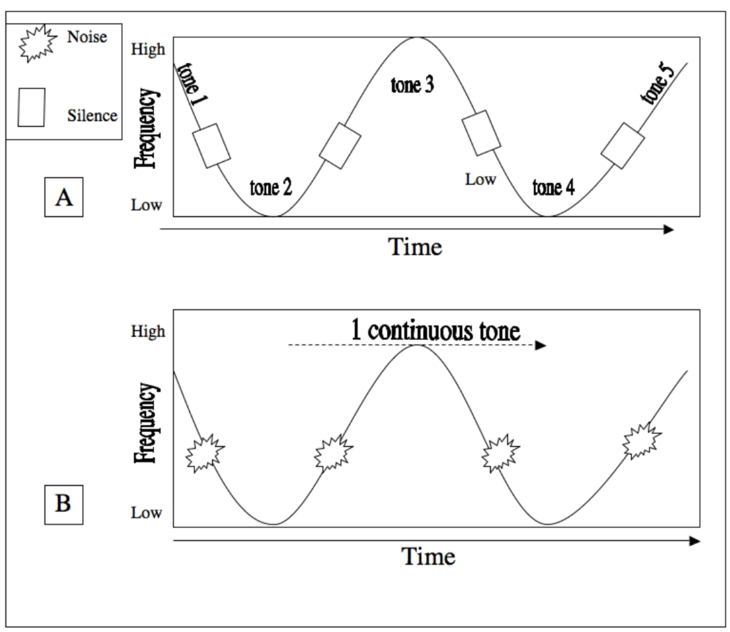

Figure 3.2. The *spatial* continuity illusion demonstrated by Warren, Osucek & Farmer (1969).

When this first tone approached a critical frequency on its downward glide, a 250 ms white noise burst was emitted to both ears. At that time, the tone abruptly switched ears, and began an upward glide pattern. The stimuli were presented to listeners with and without noise filling the silence in the 250 ms cuts (listen to an example at http://www.psych.mcgill.ca/labs/auditory/29.wav.

Listeners reported hearing two separate tones that alternated between the two ears when no noise was presented. But when the noise was added to the 250 ms cuts, listeners reported hearing one continuous gliding tone that moved in a *smooth, continuous trajectory* from ear to ear. In other words, listeners seemed to fill-in the missing segments if the gaps were replaced by noise, and this made the tone's location sound as if it were continuously changing.

Bregman (1990) has concluded that this phenomenon is an example of grouping by good continuation in frequency. Listeners reported hearing one sound moving continuously from ear to ear, even though they were actually hearing distinct tones. The preference for hearing one sound move between the ears over hearing distinct tones in separate ears was taken to show that listeners use frequency similarity (good continuation in frequency—not space) to group sounds over space (Warren, Osucek & Farmer, 1969; Dannenbring, 1976, Bregman, 1990). However, it seems that listeners could also be interpreted to show a preference for smooth spatial motion, as they showed a preference for perceptual persistence of the spatial motion of the sound source. Still, because slight changes in the frequency of the gliding tones, as well as the timing of the events seemed to cause segregation, frequency dominance continues to be assumed.

The evidence reviewed above is often cited to support ASA's theoretical assumption that frequency proximity of sounds in a sequence will dominate spatial proximity in the formation of stream segregation and the formation of groups. Generally these effects (octave & continuity illusions) are thought of as illusions of space that are driven by frequency relationships among tones in the sequences. It should be noted, however, that these effects were originally demonstrated with headphone presentations. It has been argued that one of the major problems with the conceptual development in ASA has been the reliance of unnatural methods (see Walker & Kramer, 2007). The next section covers the criticisms surrounding the assumption of space as a less important dimension in ASA's view of stream segregation and grouping.

Criticisms of the visual space/auditory frequency analogy

Many readily accept the analogy that states, just as physical space is primary for visual grouping: frequency is primary for auditory grouping (Bregman, 1990; Kubovy & VanValkenberg 2001; Palmer, 2008). Still, this analogy has been a point of controversy (Handel, 1988; Kubovy & VanValkenberg, 2001; Kubovy, 2003; Neuhoff, 2003). Handel (1988) has proposed that such an analogy is not useful for a theory regarding perception. According to Handel (1988), perceptual experience is based on dynamic variables that are spatial and temporal in nature. Furthermore, Handel (1988) argues that visual and auditory information changes spatially, and that the analogy is a departure from this fact.

Neuhoff (2003) has also argued against the analogy by citing examples that are based on real-world listening. In an example offered by Kubovy & VanValkenberg (2001), frequency's dominance over spatial cues can be demonstrated by two people playing the exact same note (A 440 Hz) on similar instruments (e.g. guitar) from two separate locations. Presumably, the listener should not hear two sound sources, but instead the listener will hear one auditory object (the A note). In this example, it is assumed that despite evidence that there are two sound sources, the auditory system will automatically group the two sounds into one based on frequency similarity. Neuhoff (2003) suggests that such spatial ambiguities could easily be resolved by moving the head or the body. Furthermore, Neuhoff (2003) argued that it is rare that listeners (and sound sources) are stationary, and it is actually more likely than not that spatial cues are underestimated as determinants of groups. For example, Neuhoff (2003) suggests that sources (e.g. bumblebees) can be emitting sound with the same fundamental frequency and still be separated according to their motion trajectory (Neuhoff, 2003).

These criticisms of the visual space-frequency space analogy have centered on real-world examples, i.e. ecological validity (Handel, 1988; Neuhoff, 2003). Based on these arguments, it seems that one avenue to test the frequency dominance assumption may be to use a spatial sound context with more ecological validity such as one involving sound source motion across *real* physical space. Recall that the majority of the foundational evidence for the frequency dominance assumption was conducted using headphone presentations and/or uninformative spatial manipulations (see Bregman, 1990; Strybel, 1994, Kubovy & VanValkenberg, 2001). Useful stimuli presentations for such research could include free-field loudspeaker arrays as opposed to headphones. The next chapter will give an overview of the relevant psychoacoustic work done in a free-field context.

Chapter 4:

Spatial tasks and grouping in the free-field

Most psychoacoustic research that examines low-level perception in the free-field is designed to assess listener performance of sound localization as spatial variables are manipulated (see Stybel 1994, Middlebrooks & Green 1991; and Zahorik et. al., 2005 for reviews). As in the auditory grouping research, free-field sound localization research often uses simple stimuli such as tones or noise bursts. In the auditory motion research, stimulus presentations are such that the sound source is smoothly moved around a listener as a sound source may move in nature using an array composed of a number of loudspeakers that surround the listener in a semicircular manner (Briggs & Perrot, 1972; Saberi & Perrot, 1990; Chandler & Granatham, 1992; Strybel and Neale, 1998). Sounds are emitted through the loudspeakers sequentially around the semicircle. The important variables in these studies usually include trajectory (direction, smoothness) and velocity (Strybel, 1994). These variables can be seen as parameters of natural spatial motion and could be an effective platform for research that directly tests the importance of frequency and space in grouping. For this reason, additional detail of these methods will be discussed.

Trajectory is established by having either a continuous or broken path of motion. In continuous trajectories, the sound source moves in the same direction (right or left) for the duration of the presentation (Briggs & Perrot, 1972). In broken trajectories, the trajectory starts in one direction and then reverses at some point during the presentation. It has been established that listeners respond faster and are more accurate at identifying trajectories with stimuli that have a continuous trajectory (Briggs & Perrot, 1972). These results can be interpreted to reflect a tendency to attribute a single trajectory to a single sound source. Moreover, they could suggest that listeners are especially sensitive to coherent trajectory, and that perhaps grouping may depend on such properties. Examples of this might be observed in situations wherein listeners group several sounds (dissimilar in frequency) as a result of a shared trajectory, such as a moving car with squeaky breaks and a rattling tailpipe. If true, sound source trajectory may be an important feature of spatial hearing that could outweigh frequency similarity in listeners' grouping and segregation of sound sources.

In a natural context, if motion is indeed an important factor for grouping or segregating sound sources, the velocity of moving sound sources could also be a spatial cue that could potentially outweigh frequency similarity. In the sound localization literature, velocity can be manipulated by emitting sound across a different number of loudspeakers in the same amount of time (Perrot & Tucker, 1988). For example, a sound would seem to be moving faster if it moved across the entire array (from left to right) in 3 sec than if it moved half that distance in 3 sec (Perrot & Tucker, 1988; and see Strybel, 1994). Since these studies typically rely on direction discrimination as the dependant variable, it is unclear how velocity might affect grouping differently than trajectory. However, these methods seem to provide the tools for providing informative spatial presentations (i.e. coherent motion) as opposed to the ear-switching techniques that has driven much of theory supporting the prevailing view (e.g. Deutsch & Roll, 1975; Rogers & Bregman, 1998).

It would seem that before the frequency dominance assumption is accepted, spatial manipulations used in the relevant psychoacoustic work should be tested. It is imperative that basic spatial functions are implemented when testing the importance of frequency and space in primitive grouping.

The manipulations used in the psychoacoustic research on apparent motion can provide a foundation for the design of stimuli that use natural spatial manipulations. However, these methods have rarely been used to directly test the differential effects of space and frequency on grouping using tone sequences (Rogers & Bregman, 1993; 1998). The next subsection will cover the few experiments that have tested grouping in the free-field using presentations that include motion.

The spatial octave illusion

The octave illusion has also been tested in stereophonic sound free-field with the sound being emitted by two loudspeakers (Deutsch & Roll, 1976; Deutsch, 1983). In these studies, one loudspeaker was placed at 45° (in front, off center right) on the horizontal plane, while the other was placed at -45° (in front, off center left). In this free-field context, the illusion was maintained as long as the listener's orientation was held constant relative to the speakers. In this orientation, the listeners heard the low tone coming from one loudspeaker and the high tone coming from the other. However, the illusion failed abruptly when the listener was moved (circularly) so that they were positioned directly facing one of the speakers. Interestingly, in this position, the listeners reported hearing one complex tone rather than two pulsating sequences.

It can be concluded from these results that the octave illusion is fragile in free-field context. These results may indicate the limitations of the frequency dominance assumption. To this author's knowledge however, these findings have not been discussed as evidence for the limitations of the frequency dominance or as support for the importance of spatial features as cues for grouping.

<u>Induction of stream segregation and grouping in the free-field</u>

Bregman and his colleagues (1993; 1998) have used apparent motion methods to test the role of spatial location in the induction of sequential grouping of triplet sequences (figure 4.1).

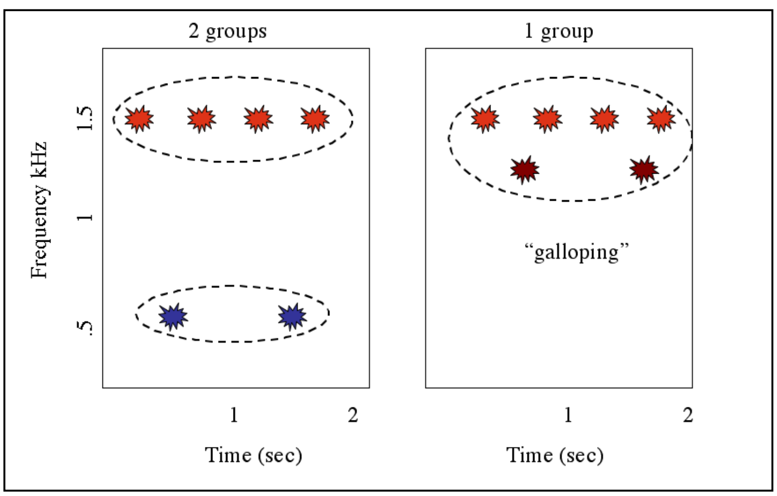

Figure 4.1. This figure illustrates both segregation (left) and grouping (right) of the triple tone sequences used by Rogers & Bregman (1993; 1998).

Because this general method will be used for the current experiments, some detail of the technique will be discussed.

As illustrated in figure 4.1, a single cycle of the triple-tone sequence consisted of three short tones (100 ms) followed by a 100 ms gap of silence (totaling 400 ms) (Rogers & Bregman, 1993; 1998). These sequences were looped repeatedly (4.8 sec) in order to *induce* segregation (*segregation effect* see chapter 2). Recall that if a sequence (such as the triple-tone sequence) is repeated for more than four seconds, the listener has a tendency to segregate the high tones from the low tones, even though they were heard as a single group initially (Bregman, 1990). Based on this fact, Rogers & Bregman (1993; 1998) referred to the first 4.8 seconds of the triple tone sequences as the *induction sequence*. In any trial, the *induction* sequence of would be followed by a *test* sequence. The test sequences would contain the same triple-tone sequence as the induction sequence, but would only be looped three times (1.2 sec), and would often be shifted in spatial position relative to the induction sequence (figure 4.2).

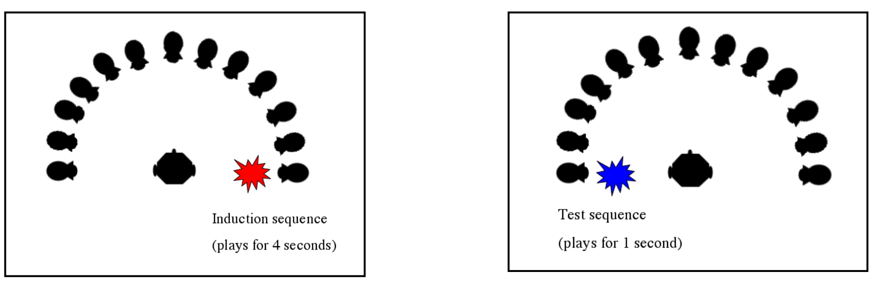

Figure 4.2. Illustration of the spatial aspects of the presentations used by Rogers & Bregman (1993; 1998).

A listener's primary task in these studies was to report if they heard the tones in the test sequence (*only*) as a single stream (grouped) or as two streams (segregated). The spatial location of the test sequence was always directly to the left of the participant

(-90°). However, the spatial location of the induction sequence was conditionally manipulated. There were six variations for the spatial location of the induction sequence (table 4.1).

Center- in front of the listener (0°)
Right- to the right of the listener (90°)
No change- left of the listener (-90°)
Random- each triple tone sequence could be at any position
Gradual- slowly moved from right to left (see fig)
Control- white noise played in front of the listener (0°)

Table 4.1.This table lists the spatial manipulations used in one of the Rogers & Bregman's (1998) induction experiments.

Before the experiment, the participants were provided with example stimuli in order to ensure the listeners understood the difference between galloping (one group) and non-galloping (two groups) stimuli. Stimuli were presented in nine randomized blocks so that each presentation type was presented nine times.

At the beginning of the experiment, the high and low tones in each condition differed in frequency by a certain amount (always nine semitones). In subsequent trials, the frequency relationship between the target and standard tones was automatically changed (+ or – 1 semitone) based on how the participant responded in the previous trial (adaptive procedure). For example, if the participant responded 'grouped' to the first trial of a particular condition, the frequencies in the sequences in the next trial in that condition would be made to be less similar (semitone separation at trial 1+ 1 semitone = semitone separation of trial 2 of that condition).

Generally, it was predicted that when the induction and test sequences did not vary in spatial location (*no change*; see table 4.1) would have the strongest induction effect because of the tendency for listeners to segregate singular tone sequences for longer than 4 sec (Bregman, 1990, Bregman & Rogers, 1993). Also, it was predicted that the induction would be similar when the induction sequence moved smoothly to the test location (apparent motion), as in the *gradual* condition (see table 4.1). All other conditions were expected to induce less segregation because the frequency build-up process would be reset by *sudden changes* in spatial location (see table 4.1).

Overall, these predictions were focused on the idea that frequency information builds up in cases where there are no sudden, large spatial changes in the location of the sequences (across the midline) (Rogers & Bregman, 1998). When, on the other hand, there are sudden changes in frequency stimulation across the midline (across ears), it was predicted that the streaming process would reset, causing listeners to show less of an induction effect. Thus, the predictions were focused on the concept of frequency build-up, as opposed to natural motion and localization as primary factors when listeners group or segregate sound sources.

The results of these studies showed that, as predicted, the *no change* and *gradual* conditions produced the most similar patterns of grouping (Rogers & Bregman, 1993; 1998). The prediction that the conditions with *sudden changes* in frequency content across the midline would produce less segregation was also confirmed. These results led Roger's and Bregman (1993; 1998) to generally conclude that the induction of stream segregation is the greatest when there is no change in spatial location during the sequence(s). This is taken as evidence that this sequence was heard as a unitary sound source that did not reset the streaming process. Gradual changes (smooth movement) cause a slight loss of stream segregation in the test sequences and the sudden change conditions caused a full resetting of segregation (Rogers & Bregman, 1993; 1998). Taken alone, this set of conclusions would seem to be a concession to the importance of spatial location in stream segregation and grouping. However, this was not the case.

Rogers & Bregman (1993; 1998) concluded that the effects seen in the sudden change conditions were not driven solely by differences in spatial location but by a lack of frequency information build-up at the location of the test portion of the sequence. In their concluding remarks, they say, " when changes are gradual [apparent motion], the accumulated evidence for a stream [frequency] may be transferred to the location of the next stimulus with little loss in strength, as evidenced by the strong induction observed in the gradual change conditions of the present experiments." Furthermore, Rogers & Bregman (1998) promote the frequency dominance assumption in their conclusion by saying, "When a newly arriving stimulus [frequency] represents a sudden change in the properties of the incoming sound, this suddenness may act to define an acoustic boundary (much in the way that visual boundaries were said to be established by the Gestalt psychologists)."

However, given that only one apparent motion condition was tested it is unclear if stream segregation is best induced by a stationary sound source. If more stream segregation can be induced when sound sources appear, say, to be moving away from each other, it could suggest that spatial location and motion properties can play a more primary role in grouping as measured by stream segregation.

The results from these induction studies potentially have important implications for the visual space-frequency space controversy. Specifically, the result that motion can facilitate stream segregation based grouping in a similar fashion to stationary sound sources clearly suggests that spatial location [motion] may play a greater role than has been seen in past studies. This suggests that a common trajectory of motion through space could be a powerful way to induce grouping as measured by stream segregation. If this premise is true, then it would be reasonable to assume that common trajectory could sometimes outweigh frequency similarity in stream segregation and grouping.

Chapter 5:

Grouping in the free-field context: Is there an alternative to frequency dominance?

It could be that in a real-world context (free-field), a listener first ascertains their position in reference to any components of sound(s) before grouping sounds based on frequency similarity. In other words, localization of any sound component might precede grouping by frequency similarity. Of course, any sound must have some frequency content, but in a free-field context, listeners could group sounds by location, in order to localize sources. . For example, when a pedestrian is exiting a building to a busy sidewalk near a busy street, they find themselves in a sonic environment that is entirely different from the previous environment (inside the building). The cacophony of moving sound sources must be segregated by location—not frequency content, in order to ascertain which sources might present potential collisions, or might afford the listener safe passage. In this case, grouping by location occurs before a process like grouping or segregation based on frequency relations.

If listeners segregated sounds by frequency *before* or *more often* than by spatial location, then ascertaining an accurate impression of an environment based on sound would be a risky or even a futile prospect. Sound sources such as multiple cars may not be dissimilar enough in frequency for the listener to segregate the sound sources based on frequency content. Also, if frequency dominates spatial location in group formation, then sudden changes in frequency, like the squeaking of brakes or a horn honking, should cause a resetting in grouping. But when a listener hears brakes squeaking or a horn honking in a busy street, it is unlikely they attribute that sound to a new source simply because the sounds can be grouped according to common location or motion properties like trajectory and/or velocity. Of course, it could be argued that the context of a busy street alone may be enough for a listener to abandon primitive processes of grouping. Yet, if this type of everyday listening cannot be described with primitive processes, then what are the parameters of the conditions that separate the primitive processes from the higher levels of auditory grouping? Furthermore, if effects like the octave illusion (Deutsch & Roll, 1976) can only be demonstrated under very precise conditions, then how can these effects be generalized to explain auditory perception in the context of everyday listening?

It could be that with some form of listener-centered dynamic spatial assessment (as it pertains to the spatial location of sound sources), listeners can maintain the percept of multiple objects (vehicles in this example) while they are moving. This perceptual persistence continues even if there are large changes in the frequency of sound components (brakes and horns) as the source is moving. If true, then source segregation and grouping are dominated by a listener's dynamic spatial assessment of sound source location as opposed to simple frequency similarity.

This new theory of sound source segregation and grouping will be referred to as D.S.A.I.L. [dynamic spatial assessment after initial localization]. DSAIL can be considered an alternative to the frequency dominance assumption when explaining sound source segregation and grouping in a naturalistic context. DSAIL considers both frequency and spatial location as potentially dominant cues in an *auditory flow field* (e.g. Gibson, 1966) that surrounds a listener. DSAIL assumes that in an environment where there are multiple sound sources, large spatial separations between the components of the sounds may lead to segregation due to physical differences such as interaural time differences and/or interaural level differences, as well as patterns of change along these dimensions. On the other hand, in cases when sources are stationary (*no change* condition, see chapter 4: table 4.1), or provide little spatial information (octave illusion, see chapter 3), DSAIL assumes that a listener might abandon (unconsciously perhaps) assessments of space and henceforth become heavily influenced by the frequency content of the components of the sounds. To test these issues, large spatial separations can be emulated in a speaker array similar to that used in the psychoacoustics work (see Strybel & Neale 1994) as well as Rogers & Bregman (1993; 1998).

Arguments against the frequency dominance assumption have used examples from everyday listening (e.g. segregating sound components which are similar), which may support this type [DSAIL] of listener & environment-focused conceptualization (Handel, 1988; Neuhoff, 2003). The research conducted in the current project was designed to test the relative strength of spatial and frequency cues by manipulating factors associated with spatial separation including motion (Briggs & Perrot, 1972; Perrot & Tucker, 1988; Saberi & Perrot, 1990; Chandler & Grantham, 1992; Strybel & Neale, 1994).

The methodology largely relied on established techniques for examining the effects of sudden spatial changes on stream segregation of a tone sequence (Rogers & Bregman 1993; 1998) in the free-field context. The general prediction was that if frequency dominates spatial location in stream segregation and group formation, then spatial cues such as continuous motion trajectory and/or velocity should not be shown to dominate stream segregation or grouping. If, on the other hand, these spatial properties sometimes do dominate frequency similarity in grouping and stream segregation, segregation would be systematically influenced by natural motion patterns.

Chapter 6:

Overview of research conducted

Statement of problem:

Auditory Scene Analysis (Bregman, 1990) is a popular, far-reaching theory of auditory perception. The influence of its conceptual framework on theoretical and research development has become fundamental to perceptual psychology (see Palmer, 2008). However, the assumption that frequency attributes of sounds are dominant over spatial attributes may be overstated and under researched. The importance of space seems undervalued by ASA, and for some of ASA's proponents (e.g. Kubovy & Van Valkenberg, 2001) it is clearly viewed as secondary to frequency as a cue for grouping. The evidence used as support for this assumption has been based primarily on laboratory demonstrations using headphone presentations (e.g. Deutsch & Roll, 1976). Alternatively, techniques used to test grouping can incorporate free-field presentations wherein sound sources can be spatially separated without *sudden changes* as defined by Rogers & Bregman (1993; 1998). In a free-field context, additional manipulations of simulated sound trajectory and velocity of the tones within the sequences can be used to test the frequency dominance assumption and the alternative assumption of spatial dominance held by DSAIL. A description of the general methodology and techniques for stimulus design and presentations used in all of the current experiments follows in the next section.

<u>General methodology</u>

<u>Materials and stimuli:</u>

Room: All experiments were conducted in a small laboratory room. The room's dimensions were approximately 8 ft x 13 ft x 13 ft. The walls were constructed of painted gypsum and there were two small observation windows (unused) on one wall. The floor was constructed of blocked tile, and the ceiling was made of sections of foam acoustic tile.

Loudspeakers: Fifteen loudspeakers (Peerless 2.5" full-range) were chosen as the sound sources because of their flat frequency responses, high quality construction, and their small size. Each of the 15 loudspeakers was enclosed in separate, sound absorbing enclosures. The rear sections of the enclosures were airtight, and spherical. Each of the enclosures was covered with 4" polyeofolin heat shrink tubing so that only the outer cone of each speaker was exposed. This was done to avoid acoustic artifacts produced as a result of vibration, enclosure resonance, and sound reflection from other loudspeakers in the array. All speakers were matched within a + or – 3dB range of frequencies that were used in the experiments (300 – 1000 Hz).**additional details available upon request.**

Array design: The loudspeakers were mounted on two sections of 1" aluminum tube that formed a half circle around a central point. Each of the 15 loudspeakers was measured to be 1 m from a central point where the listeners sat during the experiments (Rogers & Bregman, 1993; 1998). One loudspeaker was mounted at every 12° interval along the frontal horizontal azimuth (180°). Each section of the half circle was attached to a single steel rod, which was threaded into a heavy composite base that rested on the floor. All wiring was attached to the back of the aluminum tube so that the only exposed wire ran out of one side of each of the sections of tube. The entire array was fixed at a height of 4 feet. Participants were seated in a rolling chair that had a pneumatic piston for height adjustments. All participants were positioned so that the height of the ear canal was equal to the height of the center of the loudspeakers.

Audio equipment: Each loudspeaker was connected to a single channel in one of two amplifiers (Niles Apx 30 12 channel) with high quality stereo jacks. The two amplifiers were located on the floor in separate positions. Each one was positioned at one side of the room, one at -120°, and one at 120° in reference to the listeners' position in the array. However, the amplifiers were positioned approximately 18" behind the line of the rostral azimuth, and each rested against a wall. All audio was fed to the amplifiers via shielded RCA cable from a Motu I/O 24 multi-channel digital to analog audio interface. The interface used a Motu pci-424 express sound card to communicate with a Macbook pro 3,1 Intel Core 2 Duo 2.2 GHz. A Magma Expressbox Pro extension box was used to house the sound card, and the data transfer was done from the express slot on the Macbook, to the Magma. This set-up allowed for very little computer hardware, all of which was stowed against the rear wall of the room, away from the array.

Software: All stimuli were generated with a program designed using Max/Msp 5 software. The program allowed for real-time monitoring of all aspects of the stimulus presentations, including the frequency content, sound pressure levels, and locations of all sounds in the presentations. Behavioral data was collected within the Max/Msp environment by way of a Shuttlepro v2 connected to the computer through a USB connection. Based on the participant's response to stimuli in any given condition, the computer program would adjust the semitone separation in the following trial of that condition to maintain the adaptive procedure used in Rogers & Bregman (1993; 1998). However, the computer program would randomize the frequencies of the stimuli before each trial, while maintaining the semitone separation required according to the adaptive procedure. This was done to avoid any confounds associated with habituation that may have carried over from trial to trial. The program recorded all tone locations and frequencies, as well as participant responses.

Stimuli: The triple-tone sequences presented in all experiments, consisted of discrete sine wave tones. As in the previous research (Rogers & Bregman, 1993; 1998), one cycle of these repeating sequences contained a high-low-high tone structure (see figure 4.1 in chapter 4). The duration of each tone was fixed at 100 ms with 10 ms onset and termination fade envelopes and there was a 100 ms interstimulus interval between each repetition of one triple-tone sequence. All tones were lowpass filtered so that only frequencies below 8 kHz were generated (Rogers & Bregman, 1993; 1998). The dB levels of all stimuli were set at ~65 dB as measured with a handheld dB meter at 1 m (approximate position of participants). The sequences in all experiments had a fixed duration of 4.5 sec to produce the segregation effects facilitated in Rogers & Bregman (1993; 1998). The frequencies of the tones used in the trials were set to always fall between 312-699 Hz (Rogers & Bregman, 1993; 1998). All stimuli were generated at a 48 kHz sampling rate.

Procedures & design:

Pre-experiment instructions: In all experiments, participants first underwent intensive instructions to familiarize them with the grouped/segregated listening task (see appendix A for full instructions). The listeners were first introduced to the phenomenon of auditory grouping by way of example stimuli. The initial example stimuli were made to be very clear examples of what listeners typically hear as a group (3 semitone separations) and as segregated streams (14 semitone separations). Once listeners could demonstrate (vocally) that they understood the concepts of 'grouped' and 'not grouped' (segregated), the researcher played them examples that were more difficult (4-8 semitone separation). When the participants conveyed (vocally) that in some examples, it was difficult to decide if they heard the sequence as grouped more often than segregated, they were informed that they were experiencing perceptual ambiguity (e.g. as in the necker cube illusion). They were informed that the concept of ambiguity would be important later in the experiment, but to remember that the most ambiguous sequences are heard as 'grouped' half of the time, and heard as 'segregated' the other half. This was done to ready them for their task of reporting the percentage of time that they heard the sequence as ‘grouped’ later in the experiment. All example stimuli were played for 3.7 sec to avoid the segregation effects known to occur after 4 sec of repetitive stimuli (Bregman, 1990; Rogers & Bregman, 1993; 1998).

Adjustment phase instructions and procedures: After the initial instructions, grouping thresholds were estimated for each participant using a method of adjustment. Participants were instructed to face the far left loudspeaker in the array while listening to repeating triple-tone sequences. Over 18 of these *adjustment* trials, participants adjusted the semitone separation between the high and low tones by rotating a jog wheel (clockwise +, counter clockwise -) to find the most ambiguous separation. Participants indicated what they felt was the most ambiguous semitone separation by nodding their head or speaking. The participants were informed that they should try to find the semitone separation that sounded grouped 50% of the time. This was an important instruction, as it was the listeners introduction to the task of estimating a percentage of time, while they were listening, that they heard the sequence as grouped. The task in the final part of the experiment was similar (see the following subsection). The semitone separation of each subsequent trial in the *adjustment* phase followed a staggered 2+, 2- pattern so that after each trial, the responses should form a predictable pattern that centered on what was estimated to be absolute threshold for grouping. These *adjustment* sequences were not given a set duration (e.g. 6 sec) because after each adjustment made by the listener, a new process of segregation and/or grouping could have been initiated (Bregman 1990).

Once a participant had completed all 18 adjustment-trials, the researcher quickly examined the semitone separation data for all adjustment-trials. The researcher then found the semitone separation that was most often chosen by the participant to be the most *ambiguous* in the adjustment-trials (e.g. five semitone separation). The researcher would then add 2 to the estimated threshold and set the initial semitone separation for all of the conditions to be the estimated absolute threshold semitone separation + 2 (e.g. 5 + 2 = 7). This was done to ensure that the initial presentations would exceed that participant's threshold for grouping (Rogers & Bregman, 1993; 1998). This technique was used by Rogers & Bregman (1993; 1998) to initiate a clear pattern of differences between the conditions that may have been difficult to observe if the initial separation was too close to the listeners' thresholds. Even though Rogers & Bregman (1993; 1998) used a similar method, they assumed that a nine-semitone separation was sufficient for all participants, and thresholds were not measured. As in the Rogers & Bregman (1993; 1998) research, once this initial separation was set for all conditions, the critical experiment(s) could be initiated.

Experiment instructions and procedures: In each experiment, after the initial semitone separation was established, the researcher informed the participant that their task in the following trials would be to report the *extent* to which they heard the entire sequence as grouped. The researcher would then inform the participant that they were to listen to the entire sequence until it had terminated.

Task differences from those of Rogers & Bregman (1993; 1998): Recall that in the previous work, participants only reported on the last portion of the stimuli. Based on experimenter observations and comments from participants in pilot studies, it was determined that the induction-test task used by Rogers & Bregman (1993; 1998) was confusing to listeners. This was likely because the stimuli presentations in the current research had no conditions that contained sudden spatial changes as in the Rogers & Bregman (1993; 1998) research. Details of the stimuli used in the experiments are discussed in the following chapters, but it should be noted that most of the current stimuli contained either stationary or smoothly moving stimuli. As a modification to the Rogers & Bregman (1993; 1998) task, participants reported on their perception (grouped/segregated) of the entire sequence. ****This topic is discussed further in the General Discussion section.**

Task instructions: After the adjustment phase, the participants were instructed that once the sequences had stopped playing, they were to indicate what percent of the time they had heard the high and low tones as a group by pushing one of six buttons on the Shuttlepro. Participants were informed that each button represented a percentage: 0%-20%-40%-60%-80%-100% (buttons were not labeled). Once the participant could verbally repeat these instructions, the experiment was initiated.

Post experiment procedure: After a participant completed all the trials in the experiment, the researcher thanked and debriefed the participant. All of the preceding procedures were maintained for all experiments and all participants.

Experiment design: The design of all experiments was generally similar to that used by Rogers & Bregman (1993; 1998). Utilizing a repeated measures design, there were ten trials for each of five conditions. As in the Rogers & Bregman (1993; 1998) studies, trials were presented in randomized blocks, so that every condition was heard in each block. The order of presentation was maintained for each participant in each experiment. An adaptive procedure was used so that the semitone separation between the high and low tones in every subsequent trial of a given condition was increased or decreased according to the previous response (grouped/segregated) in the previous trial of that condition in the preceding block. (Rogers & Bregman, 1993; 1998).

Dependent measures: As in the previous work, the average of the final two semitone separations (trials 9-10) of each condition was calculated and used as a dependant measure. Rogers & Bregman (1993; 1998) defined these per condition averages as the *segregation boundaries* for those conditions. This terminology will be maintained in the subsequent text of this report.

Since thresholds were estimated in our experiments, it was determined that the segregation boundary calculations could potentially be misleading in cases where the estimated thresholds were very different between participants. For example, two participants could have segregation boundaries of five semitones in a given condition at the end of an experiment. In this example, if one participant were measured to have a grouping threshold of seven semitones compared to the second participant's threshold of four semitones, the segregation boundary data would be misleading. For this reason, the difference in semitone separation between the final semitone separation (after trial 10), and the initial semitone separation (trial 1i.e estimated threshold) for each condition was calculated and recorded at the end of all experiments (for every participant). The difference score formula was used to capture the magnitude of change (in semitone separation), from the first trial to the final trial. This degree of change would be captured less accurately by using an averaging process like that used to calculate the segregation boundary variable.

Exclusion of data: The adjustment-trial data also allowed for the exclusion of unreliable data in the analyses. A participant's data was excluded from analysis if the participant did not produce consistent data during the adjustment phase of the experiment. This was determined by calculating the standard deviation (std) of the semitone separation data from the 18 adjustment-trials. If a participant had a std that was more than two standard deviations from the mean of a sample of the std's derived from all participants, then that participant's data was excluded from further analyses. ***further details available upon request.**

Chapter 7:

Experiment 1: Are groupings heard equivalently at locations along the frontal, horizontal azimuth?

This experiment was designed to test if there were differences in stream segregation across the azimuth if the entire sequence was emitted from a single location within a trial. In previous research, presentations of this type were found to produce the most stream segregation (Rogers & Bregman, 1993; 1998). However, it should be considered that the *no change* trials used in the previous research were composed of sequences that were only emitted from a single location, 1 m to the left of the listener (-90°). Based on the results from the previous work, it cannot be determined if all loudspeaker locations affect grouping in the same fashion. The purpose of this first experiment was to directly test this question.

According to the conclusions based on the results of the previous work, if evidence [frequency] is accumulated over time and without sudden changes, then there stream segregation should not vary with the spatial location of the tones (Rogers & Bregman, 1993; 1998). If these conclusions are correct, then all spatial locations can be considered to be equivalent. However, if stream segregation were to vary with spatial location, then the conclusions drawn in the previous work do not necessarily account for the true role of spatial location in the formation of groups. If all locations are found to be equivalent, then the results from this experiment should provide support for a reliable control condition for later experiments.

Experiment 1 Methods

Participants: Fifteen participants volunteered to participate in this study. Thirteen were undergraduates and received course credit for their participation. The other two participants volunteered for the study were not students but were in the same age range as the undergraduates (18-25). All participants reported good hearing and vision.

Stimuli: All stimulus presentations were emitted from a single, stationary location that depended on the presentation condition. Across five conditions, the sequences were presented at –90°, –60°, 0°, 60°, or 90° . All other aspects of the stimulus parameters were unchanged from the description in Chapter 6.

Procedures & Design: The procedures and design were as described in Chapter 6.

Experiment 1 Results: Segregation boundaries analyses

A one-way repeated measures ANOVA showed that there was not a main effect for location condition on the semitone segregation boundaries $F(4,56) = .6$ $p=.67$ (see figure 7.1).

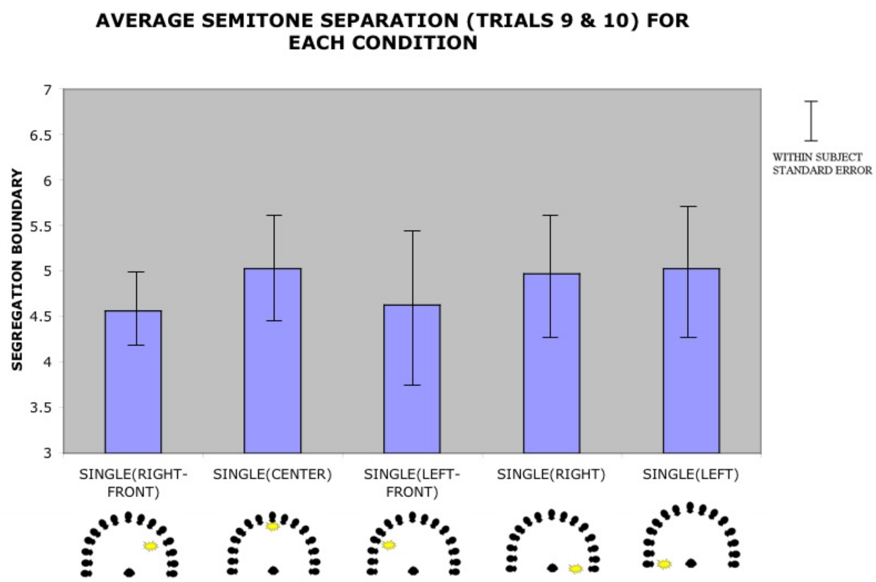

Figure 7.1. The average segregation boundary for all participants in Experiment 1are shown, separated by condition.

Experiment 1 Results: Difference-score analyses

Similarly, a one-way repeated measures ANOVA on the difference-scores (trial 10 semitone separation – trial 1 semitone separation) also showed that there was not a main effect of location $F(4,56) = .38$ $p > .05$ (figure 7.2). These results confirmed the prediction that if a sequence is emitted from a single, stationary location along the frontal hemi field azimuth, a listener does not segregate or group sounds differently based on location.

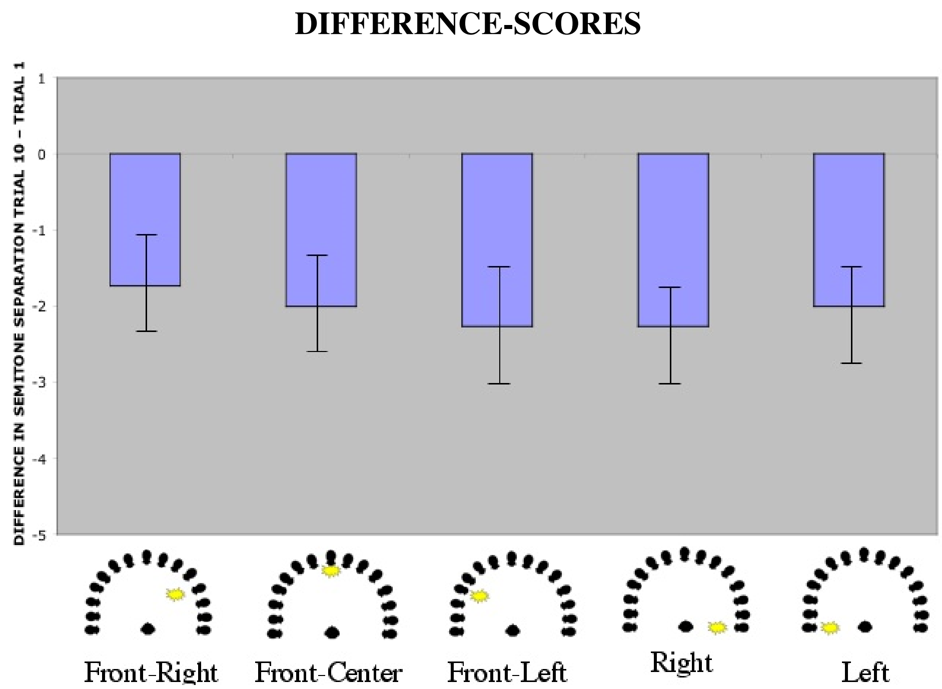

Figure 7.2. This figure shows the average difference-scores for all participants, separated by condition, in Experiment 1.

Discussion

Based on these results, it can be assumed that if sound sources emanate from a single unchanging location, the process of group formation does not vary according to locations along the frontal horizontal plane. This result indicates that any location along the frontal horizontal azimuth could be used to produce reliable data. The next experiment was designed to test the assertion that constant frequency build-up will lead to equivalent patterns of grouping, regardless of the spatial layout of sound sources (Rogers & Bregman, 1993; 1998).

Chapter 8:

Experiment 2: Do listeners use angular separation between sound sources as a basis for grouping or segregation?

In everyday listening, listeners are constantly surrounded with sound sources that are similar in frequency. As in the previous example of a busy street corner, it is easy to imagine that similar cars make similar engine sounds. It is likely that the fundamental frequencies of the engines of similar cars are indistinguishable to most people. Yet, most listeners have little problems distinguishing similar cars if they are on orthogonal streets. This is suggestive that if sound sources are simply separated by a sufficient distance, an angular separation should be enough for listeners to rely on spatial location as a basis for segregating sound sources and forming groups.

According to Rogers & Bregman (1993; 1998), if the spatial location of tone sequences does not suddenly change, stream segregation should be similar to that induced by a presentation emitted from a single stationary location. Based on this assumption, if the high and low tones were emitted from different spatial locations along the azimuth, there should be no differences in stream segregation. This second experiment was designed to test this question by manipulating the spatial proximity of the high and low tones in five conditions. In each condition the high tones were emitted from a separate location than the low tones. However, the position of each tone did not change within a trial. For example, one of these *2-location* conditions will contain trials that separate the high and low tones by 180° along the azimuth during the entire sequence (figure 8.1).

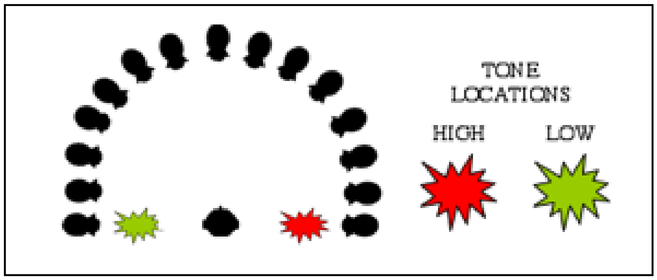

Figure 8.1. This is an example of a *2-location* condition used in Experiment 2.

The condition illustrated in the figure above will be referred to hereafter as *2-180*. The first number (*2*) indicates that there are two separate locations for the high and low tones. The second number (*180*), indicates what the angular separation [degrees] was between the high and low tones. Any further designation of this type of condition in this manuscript will maintain this format of notation.

According to the previous work, because these conditions do not include abrupt spatial changes (frequency build-up in one hemi field), they should induce performance equivalent to the *single location* trials tested in Experiment 1 (Rogers & Bregman, 1993; 1998). However, if stream segregation can be shown to vary according the angular separation between the tones, then it may suggest that stream segregation and grouping may depend on spatial location more than has been proposed by proponents of frequency dominance.

Experiment 2 Methods

Participants: Fifteen undergraduates volunteered to participate in this experiment for course credit. All participants reported good vision and hearing.

Materials: The materials used in this experiment were described in detail in Chapter 6.

Stimuli: In this experiment, there were four *2-location* conditions (see figure above for an example), and one control condition wherein all tones were emitted from a single, stationary location throughout the presentation (as in Experiment 1). The four *2-location* conditions contained angular separations between the high and low tones that increased between conditions. The smallest angular separation was 36° (*2-36),* followed by 60° separation in the *2-60* condition. The largest angular separation used was 180° (*2-180*), with the next largest being 90° (*2-90*). All other aspects of stimulus and sequence design were maintained as described in Chapter 6.

Procedures & design: The design and procedures used in the previous experiment and described in Chapter 6 were maintained.

<u>Experiment 2 Results: Segregation boundary analyses</u>

A one-way repeated measures ANOVA on the segregation boundaries confirmed that there was a main effect for location condition $F(4,56) = 4.76, p < .001$. Mauchly's test of sphericity indicated that the variance between conditions was not homogenous. A Greenhouse-Geisser correction was used to appropriately adjust the degrees of freedom for the omnibus *F*-test. A planned two sample, two-tailed *t*-test indicated that the segregation boundaries in the *2-180* condition were substantially lower than in the single location condition $p < .001$ (figure 8.2).

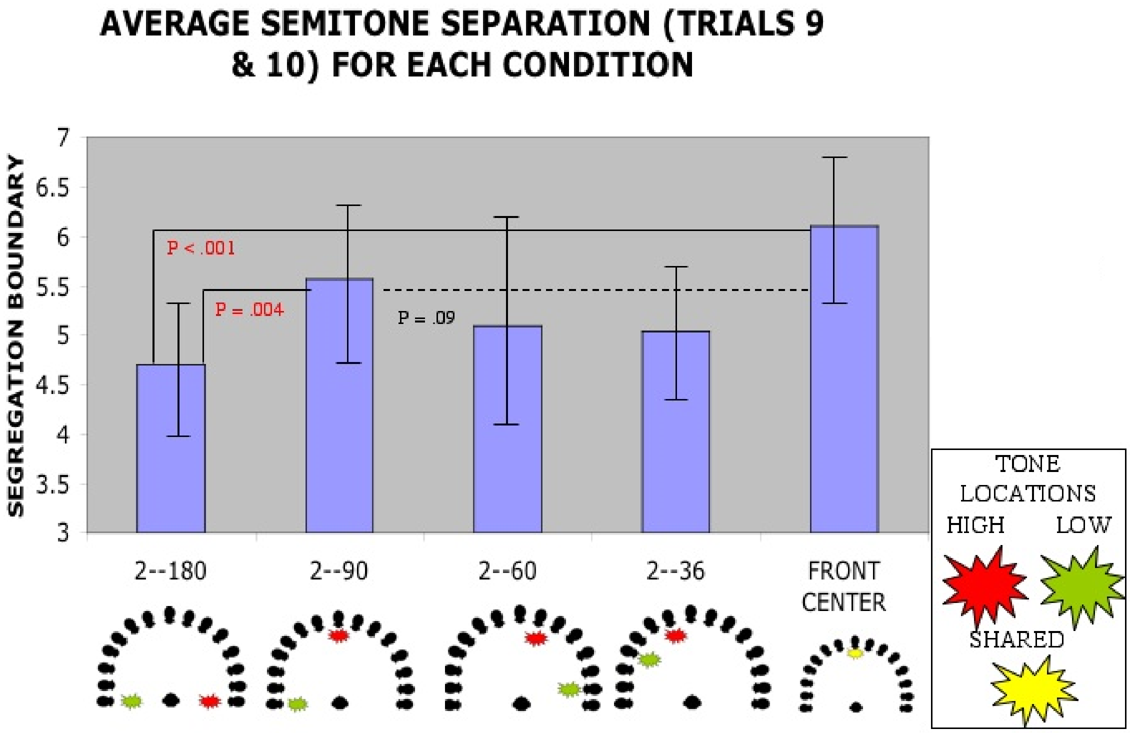

Figure 8.2. The average segregation boundaries for all participants in Experiment 2 are shown separated by condition. Significant differences (p-values) produced by planned two sample *t*-tests (two-tailed) are indicated by the red print and solid lines. Marginal differences (p-values) between conditions are indicated by the bold text and dotted lines.

<u>Experiment 2 Results: Difference-score analyses</u>

The one-way repeated measures ANOVA yielded a main effect for location condition $F(4,56) = 2.63$, $p < .05$. Mauchly's test of sphericity indicated that the variance between conditions was not homogenous. A Greenhouse-Geisser correction was used to appropriately adjust the degrees of freedom for the omnibus F-test. Planned comparisons indicated that the difference scores in the *2-180* condition were significantly larger than in the single location condition $p < .001$. Furthermore, the difference scores in the single location condition were significantly less than those measured in the *2-60* condition $p < .05$. Finally, a planned comparison showed that the difference scores in the *2-90* condition were marginally less than those measured in the single location condition $p = .08$ (figure 8.3).

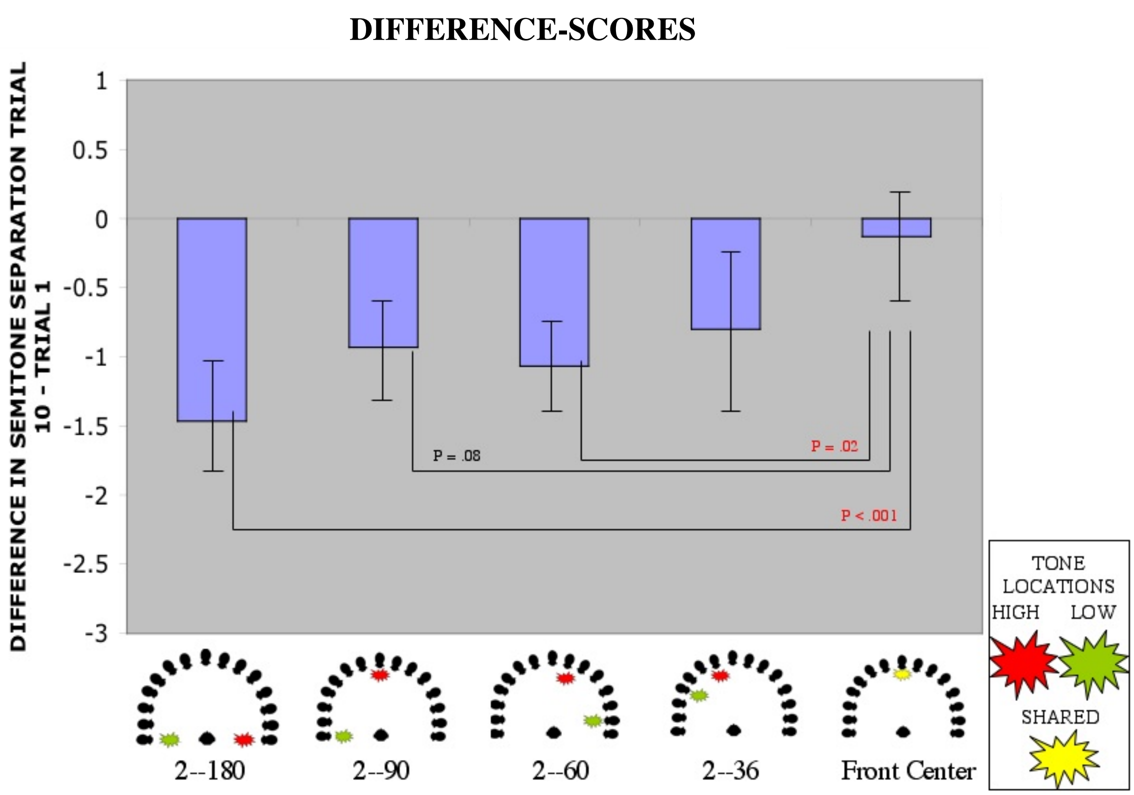

Figure 8.3. The average difference-scores for all participants in Experiment 2 are shown separated by condition. Significant differences (p-values) produced by planned two sample *t*-tests (two-tailed) are indicated by the red print and solid lines. Marginal differences (p-values) between conditions are indicated by the bold text and dotted lines.

Discussion

The segregation boundary analyses showed that there were indeed differences among the location conditions. The planned comparisons show that the *2-180* facilitated significantly more segregation than the control. Furthermore, the *2-90* condition produced marginally more segregation than the control. These results support the [DSAIL] prediction that localization of components of sound can dominate frequency similarity in grouping.

Similarly, the omnibus test of the difference-scores indicated a large difference among the spatial conditions. The planned comparisons of the difference-scores confirmed that the differences in grouping depended on the distance between the high and low tones. All conditions but *2-36* (the closest proximity) produced difference-scores that were at least marginally larger than the control condition. However, the prediction that *2-90* would produce larger difference scores than *2-60* was not confirmed. The difference-scores produced in the *2-90* were only marginally larger than the control, while the *2-60* difference-scores were significantly larger than the control. This result is likely due to between subject variations, and may disappear if more data were collected. This topic will be elaborated on further in the General Discussion.

Generally, these results indicate that listeners sometimes will rely on location as a primary cue for grouping. It can also be concluded from these analyses that as the angular separation between sound sources increases (across conditions), it is more likely that listeners will segregate the tones, despite frequency similarity. However, it could not be concluded from the results from Experiment 2 that listener localization is dynamic and may change depending on the changing location of sound sources, as it occurs when sounds move. Experiment 3 was designed to test this assumption [DSAIL], and further test the frequency dominance assumption under conditions of motion.

Chapter 9:

Experiment 3: If the tones are moving, will listeners still use angular separation between sound-sources as a basis for grouping and segregation?

The results of the previous experiment show that listeners indeed group sound components that emanate from a shared location, and segregate sound sources more as the angular separation increases. In an example, such as a busy intersection, it is not difficult to imagine that listeners can easily distinguish (aurally) between cars that are stopped at one stoplight, and similar cars that are stopped (orthogonally) at a second stoplight. However, it begs the question of the persistence of this ability in instances when the sound sources are moving; as it would if the cars were not stopped. Can listeners still segregate the tones based on the disparity of spatial location between the sound sources? Or, if smooth movement is introduced among similar sounds that are separated, will listeners abandon the reliance on spatial cues for frequency similarity in group formation?

Experiment 3 was designed to test the effects of angular separation between the tones as they move. Based on the frequency dominance assumption, spatial separation between the tones should have minimal effects on the patterns of grouping between conditions. As Rogers and Bregman (1993: 1998) indicated, if the locations of a sound source(s) are unchanging, listeners should produce patterns of segregation as if the sequence were played from a single location. More specifically, if there are no *sudden changes* in the spatial location of sound components in these presentations; reports of grouping should be very similar among all conditions.

DSAIL, on the other hand, asserts that listeners rely on initial localization of sound components as a basis for assessing sound source layout in an environment. If this assumption is true, then the conditions in which the tones have the largest angular separation as they move should produce the least grouping. Also, according to DSAIL, if sound components share a location throughout a presentation, the listener was predicted to produce patterns of grouping that are similar to the control condition. More specifically, the conditions wherein the high and low tones are not separated were predicted to facilitate more grouping than the *separation* conditions. Furthermore, if the listeners' spatial assessments rely first on spatial separation (or a lack thereof), then the patterns of grouping produced by the *gradual-same* conditions were predicted to be similar to the control.

Experiment 3 Methods

Participants: Fifteen undergraduates volunteered to participate in this experiment for course credit. All participants reported good vision and hearing.

Materials: The materials used in this experiment were described in detail in Chapter 6.

Stimuli: In this experiment, there were four *motion* conditions (see figures in the results section of this chapter for examples), and one control condition wherein all tones were emitted from a single, stationary location throughout the presentation. The four *motion* conditions contained smooth gradual motion (Rogers & Bregman, 1993; 1998). In two of these conditions (*gradual-left*, *gradual-right*), the high and low tones were always emitted from the same location during two cycles of the sequence (high-low-high). This would allow for the sequence to smoothly move across eight of the loud speakers from either 0°- -90° (*gradual-left*) or from 0°- 90° (*gradual-right*). The remaining two *motion* conditions (*2-90-left*, *2-90-right*) retained the same patterns of gradual motion. However, the high and low tones in these conditions were constantly separated by 90°. In the *2-90-left* condition, the high tones began at 0° and the low tones began at 90°. They gradually moved eight loudspeakers to end with the high tone at -90°, and the low tone ended at 0°. The inverse was true in the *2-90-right* condition. All other aspects of stimulus and sequence design were maintained as described in Chapter 6.

Procedures & design: All procedures used in the previous experiments and described in Chapter 6 were maintained.

<u>Experiment 3 Results: Segregation boundary analyses</u>

A one-way repeated measures ANOVA indicated that there was a marginal effect for condition $F(3,38) = 2.64$ $p = .07$. Mauchly's test of sphericity indicated that the variance between conditions was not homogenous. A Greenhouse-Geisser correction was used to appropriately adjust the degrees of freedom for the omnibus *F*-test. This marginal effect was driven by significant differences in the average of the last two semitone separations measured in the motion-conditions wherein the tones had a constant angular separation (90°) and the averages of the motion –conditions wherein the tones shared a location t(14 = 2.34 $p = .03$. Further planned comparisons showed that there were large differences between the semitone averages of the final two trials of the following pairs: *2-90 left* vs. control t(14) = 3.06 $p = .009$; *2-90 right* vs. *gradual right* t(14) = 2.97 $p =$

.01; *2-90 left* vs. *gradual right* t(14) = 2.84 *p* = .01; and finally, the *2-90 right* vs. control t(14) = 2.47 *p* = .03 (figure 9.1). These results confirm the prediction that a 90° angular separation between sound components induce more segregation than frequency relations as shown in the research by Rogers & Bregman (1993; 1998).

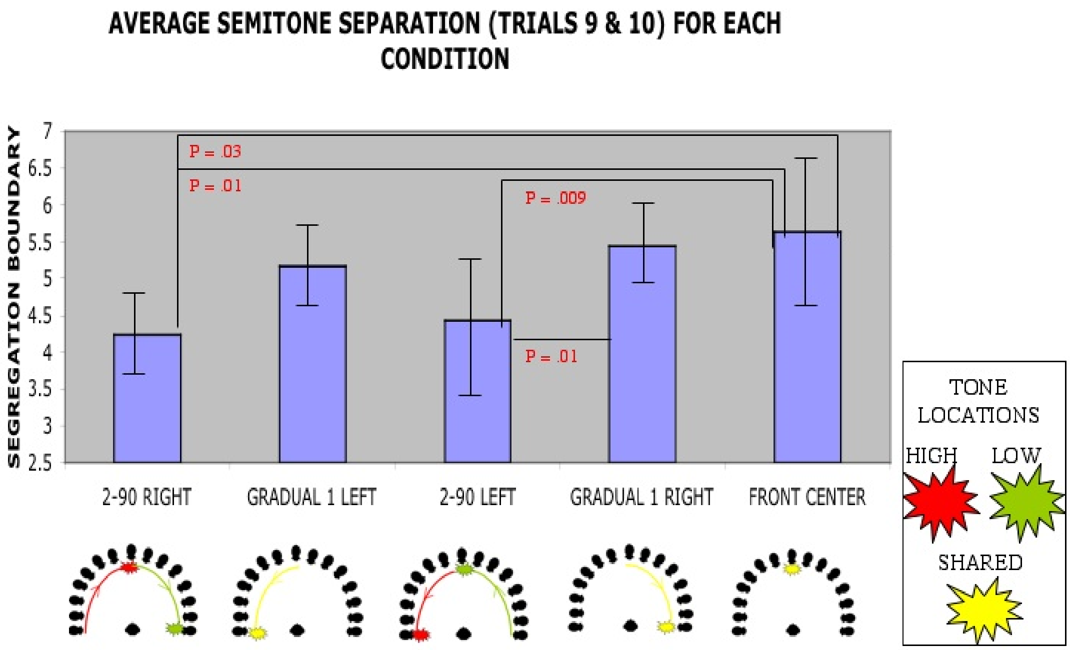

Figure 9.1. The average segregation boundaries for all participants in Experiment 3 are shown separated by condition. Significant differences (p-values) produced by planned, two-sample *t*-tests (two-tailed) are indicated by the red print and solid lines in the figure. Marginal differences (p-values) between conditions are indicated by the bold text and dotted lines.

<u>Experiment 3 Results: Difference-score analyses</u>

A one-way repeated measures ANOVA indicated that there were significant differences between the difference-scores among the conditions, $F(3,40) = 2.84$ $p = .05$. Mauchly’s test of sphericity indicated that the variance between conditions was not homogenous. A Greenhouse-Geisser correction was used to appropriately adjust the degrees of freedom for the omnibus *F*-test. A planned comparison of the averages of the difference scores between the motion conditions showed that the difference-scores of the conditions that contained angular separations (90°) had larger difference-scores than the conditions wherein the tones shared a location, $t(14) = 2.49$ $p = .03$ (figure 9.2). Further planned comparisons showed that these effects were driven by large differences between the following conditions: *2-90 right* vs. *gradual right,* $t(14) = 3.15$ $p = .007$; *2-90 left* vs. *gradual right,* $t(14) = 3.15$ $p = .007$*; 2- 90 left* vs. control, $t(14) = 2.96$ $p = .01$; and finally the *2-90 right* vs. control, $t(14) = 2.22$ $p = .04$. These results confirm the prediction that a 90° angular separation between sound components induce more segregation than frequency similarity.

DIFFERENCE-SCORES

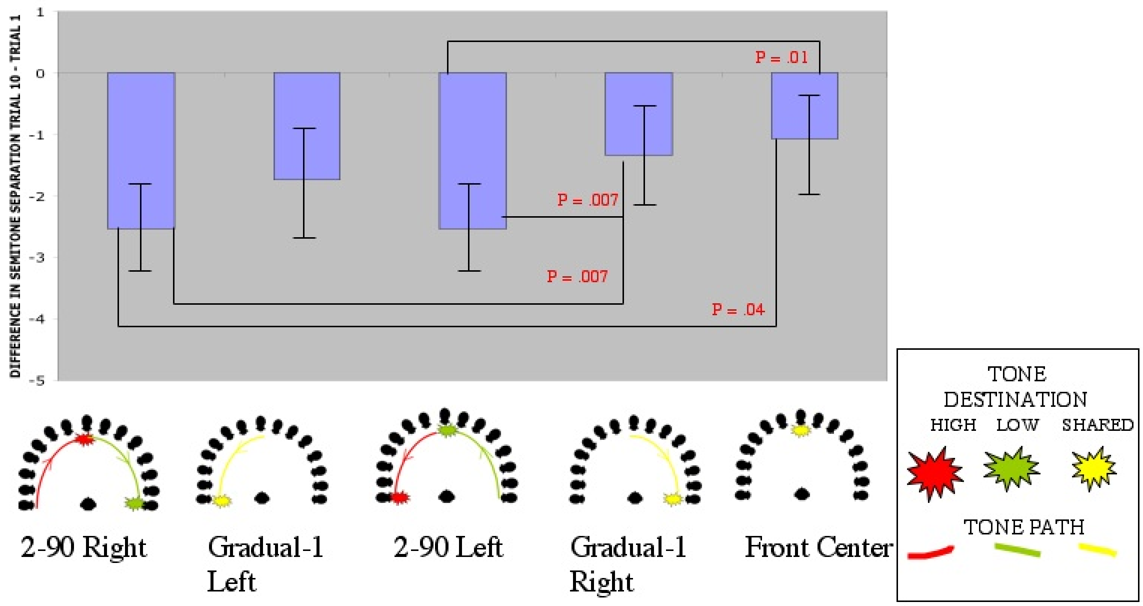

Figure 9.2. The average difference-scores for all participants in Experiment 3 are shown separated by condition. Significant differences (p-values) produced by planned two sample *t*-tests (two-tailed) are indicated by the red print and solid lines. Marginal differences (p-values) between conditions are indicated by the bold text and dotted lines.

Discussion

The results show that segregation varied systematically such that the motion-conditions wherein the high and low tones were constantly separated by 90° produced more segregation than the shared location conditions as measured by both dependant variables. Even though the omnibus test of the average segregation boundaries indicated only a marginal difference, it is clear that the marginal effect was due to the similarity among the shared location conditions. The planned comparisons of the average segregation boundaries among the conditions showed that the *2-90* motion-conditions produced significantly more segregation than all shared location conditions. This was also made clear by the main effect indicated by the omnibus test of the difference-scores among the conditions, as well as the planned comparisons of the difference-scores among the conditions.

These findings confirm the predictions made by DSAIL, and are in conflict with the assumptions of frequency dominance over spatial location (Kubovy & VanValkenberg, 2001). However, it should be noted that these results also replicate the Rogers & Bregman (1993; 1998) finding that gradual motion of sound sources that share a location will produce similar patterns of grouping compared to stationary tone sequences. These results show that if sources are separated by at least 90°, shared motion properties such as trajectory and velocity will not facilitate grouping. The next experiment was designed to test if shared or disparate velocities between sound sources have measurable differences in patterns of grouping.

Chapter 10:

Experiment 4: Do listeners use disparate motion (velocity) among sounds as a basis for segregation and/or grouping?

If a listener can segregate two similar cars based on the rate of change as they pass by them, it would suggest that the velocity of sound sources might be a dominant property of motion for spatial grouping and/or segregation. The results of the previous experiments suggest that the DSAIL assumption of spatial primacy depends mostly on the spatial proximity of the sound components. However, if the listeners' spatial assessment is truly dynamic, it begs the question of the mechanisms associated with the dynamics. Based on a simple, natural example of real-world listening such as the busy intersection, it seems likely that perhaps velocity differences may be informative as to the dynamics of the spatial assessment process.

The goal of Experiment 4 was to examine if velocity can outweigh spatial proximity in listener grouping and segregation. In cases when the high and low tones in these presentations were moving at different speeds, listeners may segregate them based on differential localization. On the other hand, if the tones share a velocity, then the listener should perceive the motion as a spatial feature of a single source i.e. a group. In that case, the segregation boundaries of conditions where the velocities are slightly different or equal should be higher than conditions where the velocities are greatly different. If spatial separation among the tone types (high & low) can cause listeners to segregate sound sources before grouping the tones, then a decrease in grouping was predicted if the high and low tones moved across space at different velocities.

If motion does not play a primary role in stream segregation, then it was predicted that differential velocity among the high and low tones within a sequence should not be seen to effect the segregation boundary among the motion conditions (Rogers & Bregman, 1993; 1998). That is, even if the tones were smoothly moving at different velocities, there would be no difference in the segregation boundaries or difference-scores among the motion conditions as long as the rate of change was smooth.

On the other hand, if the listeners were to distinguish between the velocities, then they may hear two sound sources moving in the same direction (Perrot & Tucker, 1988, Saberi & Perrot, 1990). If true, it was predicted that there would be an increase in stream segregation and a degradation of grouping as measured by the segregation boundaries and the difference-scores. Additionally, if the tones shared a velocity (and trajectory), then the listener may perceive the motion as spatial features of a single source i.e. an group. If so, it was predicted that the segregation boundaries in conditions where the velocities were more similar may be higher and more similar to the *single location* control condition than when the high and low velocities were more different. Finally, it was predicted that all motion conditions should produce a lower segregation boundary than the stationary control if listeners rely on spatial location and motion properties to segregate sound components and form groups.

Experiment 4 Methods

Participants: Fifteen undergraduates volunteered to participate in this experiment for course credit. All participants reported good vision and hearing.

Materials: The materials used in this experiment did not vary from the previous experiments and were described in detail in Chapter 6.

Stimuli: In Experiment 4, there were four *motion* conditions (see figures in the results section of this chapter for examples), and one control condition wherein all tones were emitted from a single, stationary location throughout the presentation. As in the previous experiment, the four *motion* conditions contained smooth gradual motion (Rogers & Bregman, 1993; 1998). However, in all four of these motion conditions, there was a measure of angular separation between the high and low tones. Also, the velocity of the high and low tones was varied between conditions. In the *equal-far* condition, the high and low tones retained a constant angular separation of 90°, as they moved from right to left. Therefore, the velocity between the high and low tones in this condition was equal. In the *little-different-close* condition, the high tones moved slightly faster than the low tones, so that they traveled 12° further than the low tones by the end of the presentation. The label *close* of this condition indicates that the angular separation of the high and low tones never reaches 90° throughout the presentation. Similarly, the high and low tones were never separated by 90° or more in the *more-different-far* condition. However, the velocity difference between the high and low tones in this condition was increased so that the high tones traveled 36° further than the low tones. In the *most-different-far* condition the high tones traveled 48° further than the low tones by the end of the presentations.

Furthermore, the angular separation between the tones exceeded 90° by the end of the presentation (*far)*. All other aspects of stimulus and sequence design were maintained as described in Chapter 6.

Procedures & design: All procedures used in the previous experiments and described in Chapter 6 were maintained. There were no modifications to the experimental design used in the previous experiments, which was described in Chapter 6.

Experiment 4 Results: Segregation boundary analyses

A one-way repeated measures ANOVA produced evidence of a marginal effect (with the green-house geisser correction) of velocity condition $F(2,23) = 2.67$ $p < .10$. Mauchly's test of sphericity indicated that the variance between conditions was not homogenous. A Greenhouse-Geisser correction was used to appropriately adjust the degrees of freedom for the omnibus *F*-test. Planned comparisons showed that only the conditions (*equal-far*, *most-different-far*) where the high and low tones had a far angular separation (> 90°) at the end of the sequence had significantly lower segregation boundaries than the *single location* control condition ($t_{(14)}$ $p < .05$) (figure 10.1). However, the segregation boundaries in the conditions where the high and low tones had a small angular separation (< 90°) were found to be marginally lower than in the *single location* condition. ($t_{(14)}$ $p = .07$).

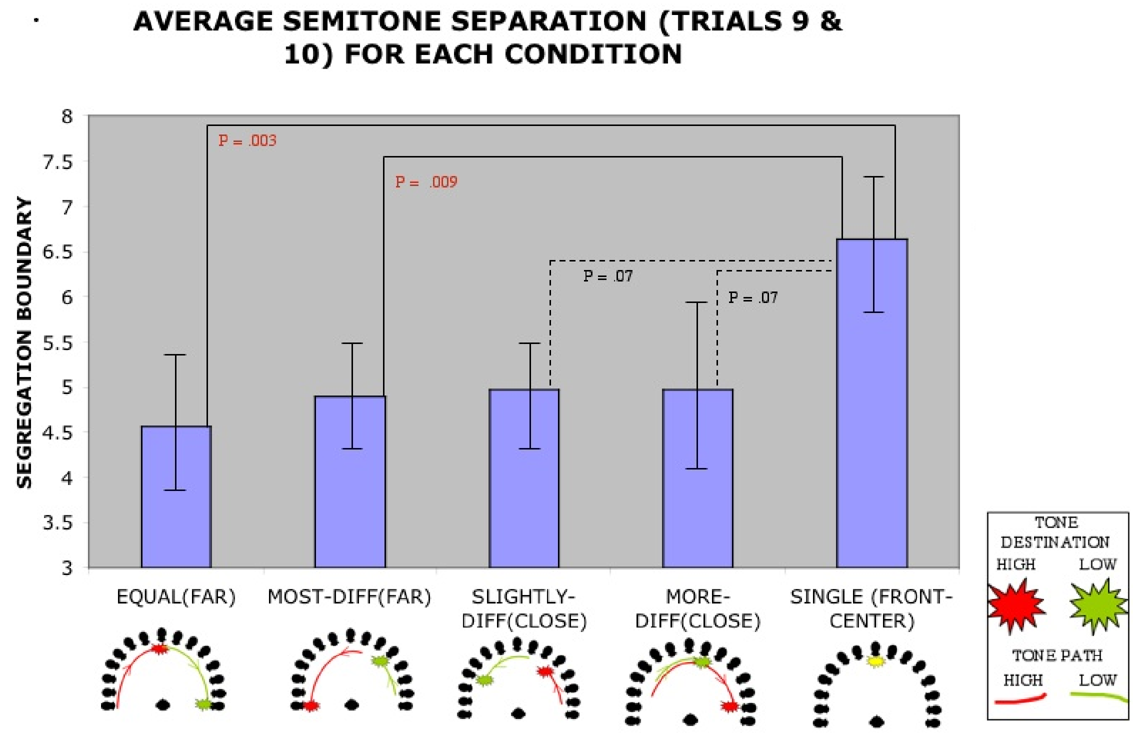

Figure 10.1. The average segregation boundaries for all participants in Experiment 4 are shown separated by condition. Significant differences (p-values) produced by planned, two sample *t*-tests (two-tailed) are indicated by the red print and solid lines. Marginal differences (p-values) between conditions are indicated by the bold text and dotted lines.

Experiment 4 Results: Difference-score analyses

There was a main effect for velocity condition according to a one-way repeated measures ANOVA on the difference-scores (trial 10 semitone separation – trial 1 semitone separation) $F(2,23) = 2.89$ $p < .05$. Mauchly's test of sphericity indicated that the variance between conditions was not homogenous. A Greenhouse-Geisser correction was used to appropriately adjust the degrees of freedom for the omnibus *F*-test. Planned comparisons showed that the difference scores in the *equal-far* condition were significantly larger than both the *more-different-close* and *single location* conditions

($t_{(14)}$ $p < .05$) (figure 10.2). Also, the average difference score between the conditions where the velocity differences between the high and low tones were the smallest were significantly greater than the average difference scores of the conditions where the velocity differences were the greatest ($t_{(14)}$ $p < .05$). A marginal difference was found between the difference scores in the *more-different-close* and the *single location* control condition ($t_{(14)}$ $p = .07$).

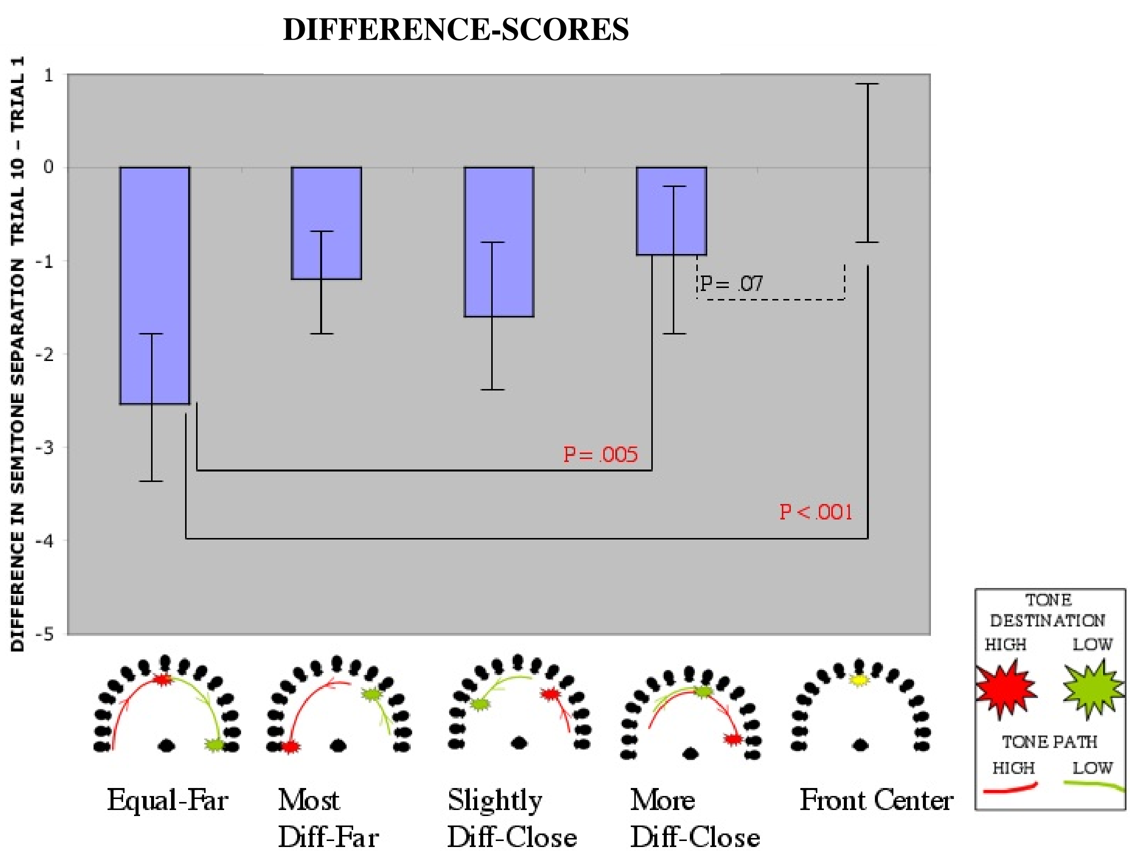

Figure 10.2. The average difference-scores for all participants in Experiment 4 are shown separated by condition. Significant differences (p-values) produced by planned, two sample *t*-tests (two-tailed) are indicated by the red print and solid lines. Marginal differences (p-values) between conditions are indicated by the bold text and dotted lines.

Discussion

Generally, this pattern of results is more in line with the assumptions of DSAIL rather than assumptions of frequency dominance (Kubovy & VanValkenberg, 2001). More specifically, these results suggest that when sound components are moving, listeners segregate the components based on the spatial proximity of the sounds (high and low tones). However the pattern of results did not confirm the prediction that listeners will use velocity differences as a basis of segregation. In fact, it seems that the opposite may be true, given the pattern of results. Still, the conditions that facilitated the most segregation were the conditions that contained the largest angular separations of the high and low tones (> 90°).

As was the case in the previous experiment, these results suggest that when sound components are moving, the spatial proximity of the tones during the sequences plays a larger role in the facilitation of stream segregation than velocity differences. Furthermore, listeners favor spatial proximity as a guiding factor in stream segregation, as opposed to velocity differences. If listeners use dynamic spatial assessment after initial localization, and velocity similarity/differences of sound components could not be shown to differentially effect grouping, then it suggests that motion may not be as important as spatial location in grouping.

These results show that spatial relations can dominate frequency relations in stream segregation (≥ 90°) and grouping (< 90°). This finding suggests that the under-valuing of space by ASA (Bregman, 1990) and its proponents (Kubovy & VanValkenberg, 2001). Still, question remains if a dynamic property of motion can be shown to systematically facilitate grouping and segregation. If true, it would further solidify the DSAIL account of grouping in space, because it would demonstrate that listeners group sounds based on an update-type process of their initial assessment of the spatial location of sound components. Experiment 5 was designed to test if the trajectory of the high and low tones (shared vs opposite) could be shown to systematically effect grouping and segregation.

Chapter 11:

Experiment 5: Do listeners use *direction* of motion [trajectory] of sound sources as a basis for segregation or grouping?

The most critical assumption of DSAIL is that listeners continually update their initial spatial assessment of the location of the components of sounds. In order to test this assumption, it was determined that the trajectories of the high and low tones could be varied in such a way that should produce a predictable pattern of results. The primary goal of Experiment 5 was to test if the importance of angular separation found in the previous experiments is different depending on the relative direction of the high and low tones. This result would suggest that if listeners used an assessment process such as DSAIL, then the spatial assessment may end or change once the listener grouped or segregated sounds based on the trajectory(s) of the localized components(s). On the other, hand, if listeners continued to show a pronounced difference in their patterns of segregation of based on angular separation alone, it would suggest that even smooth motion cannot over come spatial separation in grouping.

Proponents of frequency dominance would predict that stream segregation would be similar across conditions without an abrupt change in source location (Rogers & Bregman, 1993; 1998; Kubovy & VanValkenberg, 2001). On the other hand, if listeners' dynamic spatial assessment is influenced by trajectory, *shared* trajectories were predicted to lead to more grouping (higher segregation boundaries) than *opposite* conditions. Moreover, if *shared* trajectories could be a basis for grouping, the data would be similar among the *same trajectory* and control conditions. However, given the results of

Experiment 4, it was predicted in Experiment 5 that grouping in the *shared-close* condition would be greater than in the *shared-far* condition.

If the assumptions of DSAIL are correct, listeners were predicted to hear the smooth motion of two sound objects if the high and low tones had *opposite* trajectories. Moreover, it was also predicted that that the *opposite* trajectory conditions in Experiment 5 would produce the least grouping. Furthermore, when considering the results of the previous experiments, it was predicted that the *opposite-far* condition would produce the most segregation because of the large angular separation between the tones. However, it was predicted that if motion is a primary spatial cue, then the *opposite-close* condition should still produce more segregation than both *shared* conditions and the control condition.

In sum, the segregation boundaries were predicted to be highest for the *single location* control condition and the s*ame trajectory* conditions. The segregation boundaries were predicted to be the lowest in the *opposite* conditions. Because of the importance of the spatial separation of the components seen in the previous experiments, it was predicted that within the trajectory conditions, grouping would be higher in the *close* conditions, than in the *far* conditions.

Experiment 5 Methods

Participants: Fifteen participants volunteered to participate in this study. Fourteen of these participants were undergraduates who volunteered to participate in this experiment for course credit. One participant volunteered to participate for no compensation. All participants reported good vision and hearing.

Materials: The materials used in this experiment did not vary from the previous experiments and were described in detail in Chapter 6.

Stimuli: In four motion-conditions, the high tones either shared a trajectory with the low tones (*same* conditions), or had an *opposite* trajectory than the low tones (*opposite* conditions). In one of each of these trajectory conditions the angular separation of the high and low tones were small (< 90°) throughout the presentations (see figure below). In the remaining trajectory conditions, the angular separation between the high and low tones was large (≥ 90°) throughout the presentations (figure 11.1). In order to ensure that changes in velocity would not be a factor, it was necessary that the angular separation of the high and low tones would increase in both *opposite* conditions, but remain the same in both *same* conditions.

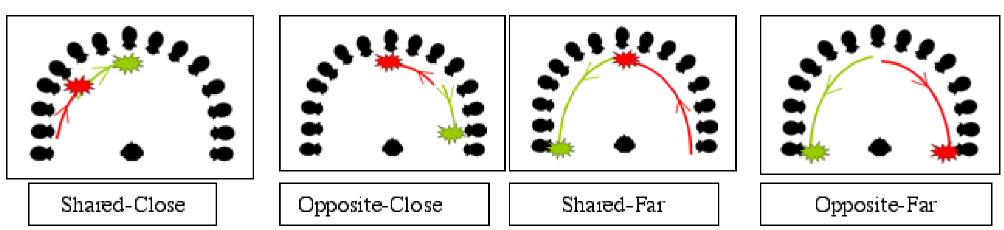

Figure 11.1. These are illustrations of the trajectory manipulations used in Experiment 5. Red flashes are high tones.

As illustrated in figure 11.1, the angular separations of the high and low tones in the *shared-close* and *opposite-close* conditions remained < 90° throughout the presentations (as they moved). So, while the angular separation between the high and low tones increased as they moved in the *opposite-close* condition, it never reached or exceeded 90°. This was done to ensure that angular separations used in both the *shared-close* and the *opposite-close* conditions would always be < 90°.

In order to isolate trajectory in the *opposite-far* condition, it was necessary to set the initial angular separation at < 90°. The angular separation of the high and low tones gradually increased to a separation to 180° throughout the presentation (see fig above). As in the *shared*-close condition, the high and low tones in the *shared-far* condition maintain their angular separation throughout the presentation. However, the angular separation in the *shared-far* condition was always 90°. All other aspects of stimulus and sequence design were maintained as described in Chapter 6.

* All stimuli in experiment 3 moved at the approximately the same rate regardless of the trial type

Procedures & design: All procedures used in the previous experiments and described in Chapter 6 were maintained. There were no modifications to the experimental design used in the previous experiments, which was described in Chapter 6.

<u>Experiment 5 Results: Segregation boundary analyses</u>

There was a main effect for condition on the segregation boundaries according to a one-way repeated measures ANOVA $F(3,42) = 4.24$ $p < .05$. Mauchly's test of sphericity indicated that the variance between conditions was not homogenous. A Greenhouse-Geisser correction was used to appropriately adjust the degrees of freedom for the omnibus *F*-test. Planned comparisons showed that the segregation boundaries for both *opposite* conditions were significantly different from the *single location* condition ($t_{(14)}$ $p < .05$) (figure 11.2) Additionally, the segregation boundaries in the *same-far* condition were significantly lower than those in the *single location* condition ($t_{(14)}$ $p < .05$). The segregation boundaries in the *same-close* condition were found to be marginally different from the *single location* condition ($t_{(14)}$ $p = .07$).

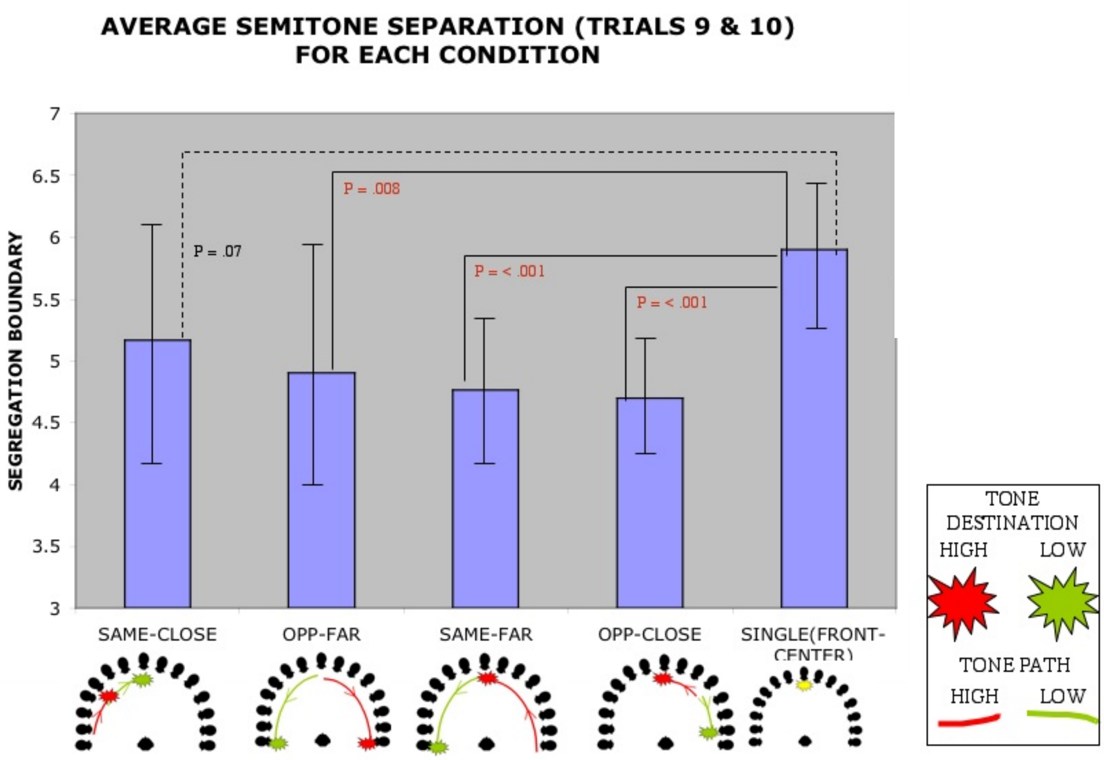

Figure 11.2. The average segregation boundaries for all participants in Experiment 5 are shown separated by condition. Significant differences (p-values) produced by planned, two sample *t*-tests (two-tailed) are indicated by the red print and solid lines. Marginal differences (p-values) between conditions are indicated by the bold text and dotted lines.

Experiment 5 Results: Difference-scores analyses

A one-way repeated measures ANOVA on the difference-scores (final trial semitone separation – first trial semitone separation) did not find a main effect for condition $F(2,33) = 1.79$ $p > .05$. Mauchly's test of sphericity indicated that the variance between conditions was not homogenous. A Greenhouse-Geisser correction was used to appropriately adjust the degrees of freedom for the omnibus *F*-test. However, a planned comparison showed that only the difference-scores in the *opposite-far* condition were significantly larger than the difference-scores in the *single location* condition ($t_{(14)}$ $p < .05$) (figure 11.3). The difference-scores in the *same-far* condition were marginally larger than the *single location* difference-scores ($t_{(14)}$ $p = .06$). The *same-close* difference-scores were marginally less than the opposite-far difference-scores ($t_{(14)}$ $p = .10$).

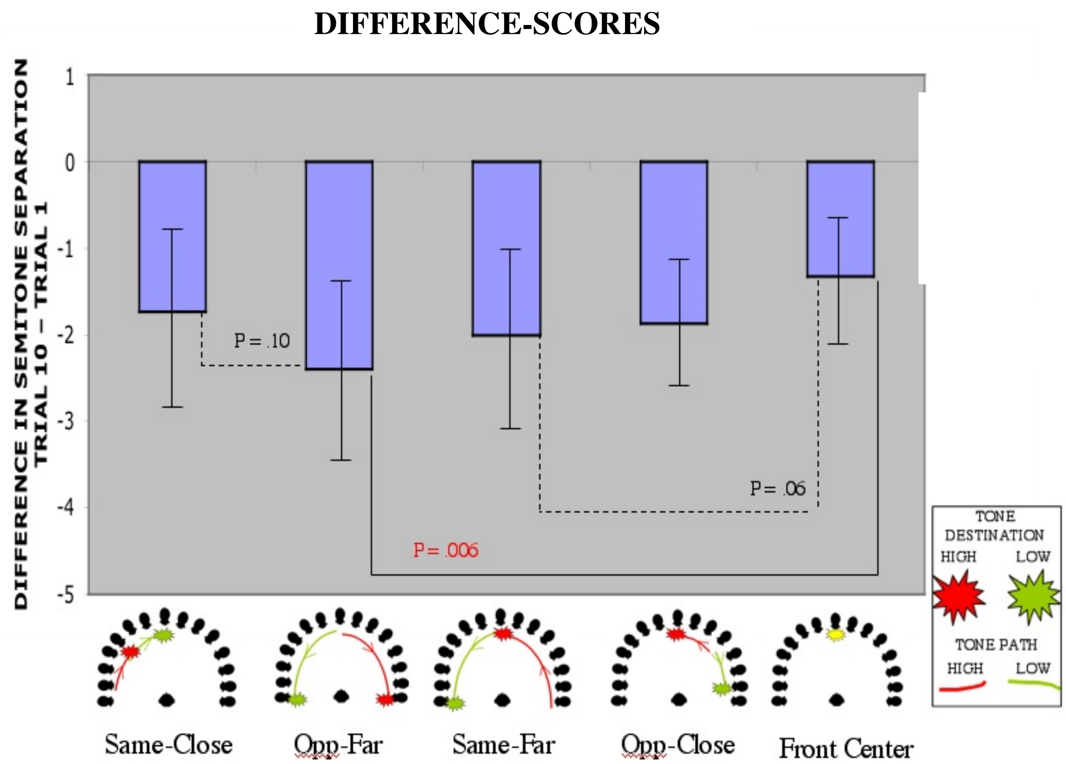

Figure 11.3. The average difference-scores for all participants in Experiment 5 are shown separated by condition. Significant differences (p-values) produced by planned,two-sample *t*-tests (two-tailed) are indicated by the red print and solid lines. Marginal differences (p-values) between conditions are indicated by the bold text and dotted lines.

Discussion

The segregation boundary results suggest that listeners tend to segregate sound components that are moving, especially if they are separated by 90° or more, and/or if the components are moving in opposite trajectories. However, even the *shared-close* condition produced marginally less grouping than the control condition as measured by the segregation boundaries. Based on these results, the prediction that listeners will rely on trajectory to group [shared trajectories] or segregate [opposite trajectories] cannot be confirmed. Even though the *shared-close* condition was only marginally lower than the control, the segregation boundaries produced by the *shared-close* condition was not found to be different from the other conditions. Once again, these results indicate that spatial location plays a larger role than what is assumed by proponents of frequency dominance (Kubovy & VanValkenberg, 2001).

The omnibus test of the difference-score data suggested that there were not differences among the conditions, since there was not a main effect for condition. Even still, the planned comparisons conducted showed that the difference-scores in the *opposite-far* condition were significantly larger than measured in the control (as predicted). However, the *opposite-close* difference scores were not found to be different from the control, indicating that trajectory does not play as large of a role in the facilitation of segregation as spatial separation.

In line with the results from the previous studies, both of these analyses indicate that disparate spatial separation among sources is the primary factor for segregation, even if the tones are similar in frequency. Specifically, this experiment showed that this effect is magnified if the components of sound are moving. Furthermore, based on these results, it seems that trajectory can facilitate segregation if the sound components have opposite trajectories that lead to large angular separations. Even though the prediction that *shared* trajectories would facilitate grouping, were not confirmed, these data were more in line with the assumptions of DSAIL (spatial primacy) than assumptions of frequency dominance (Bregman, 1990; Kubovy & VanValkenberg, 2001).

Chapter 12:

General Discussion

These experiments have shown that listeners most often segregated sound sources that were separated by angular extents of 90° or larger, even if the frequencies were similar (5 or less semitone difference), and the locations are unchanging. Grouping was facilitated if sound components were separated by 36° or less, especially when the sounds shared a location. Interestingly, Experiments 3-5 suggested that if the tones were moving, the reliance on spatial proximity among the sound components was magnified. For example, in Experiment 5 it was found that even if the high and low tones were constantly separated by 36° (*shared-close)*, segregation occurred as often as in conditions where the angular separations were large. This is in contrast to the finding in Experiment 3 that if the high and low tones shared a location while they were moving (*gradual* conditions), the tones were grouped as if they were stationary. Based on this result, it can be concluded that properties of motion such as velocity and trajectory could not overcome angular separation as the dominant cue for segregation and grouping. The general conclusion is that, in this free-field context, the spatial proximity of the components of sound dominated properties of motion and frequency similarity in stream segregation and grouping of repeating triple-tone sequences. This conclusion is contrary to the assumption that frequency dominates spatial cues in grouping (Bregman, 1990; Rogers & Bregman, 1993; 1998; Kubovy & VanValkenberg, 2001; 2003). The results were more in line with the alternative hypotheses of DSAIL. The results of the studies are summarized and discussed comparing both perspectives in the following subsections.

Summary and Conclusions: Experiment 2

In Experiment 2, the [DSAIL] prediction that segregation would increase with the angular separation between the high and low tones (between conditions) was confirmed to some extent. The omnibus tests for both dependant variables (segregation boundaries, and difference-scores) indicated that there significant differences among the different spatial conditions. The planned comparisons showed that these effects were driven by the significant differences between the conditions where there were the largest angular separations and the control condition. However, there were not any significant differences in either of the dependant variables between any of the conditions that contained separations.

Based on Rogers & Bregman (1993; 1998), and assumptions of frequency dominance (Kubovy & VanValkenberg, 2001), the ASA prediction that reports of grouping should be similar between conditions in Experiment 2 was not confirmed. Recall that Rogers & Bregman (1993; 1998) assert that stimulus presentations, such as those used in Experiment 2, would require *sudden-changes* to reset the process of streaming. Even though the '*sudden-change*' hypothesis is somewhat nebulous regarding the dimension associated with the *sudden-changes*, it was assumed that the change must be in spatial location (figure 12.1).

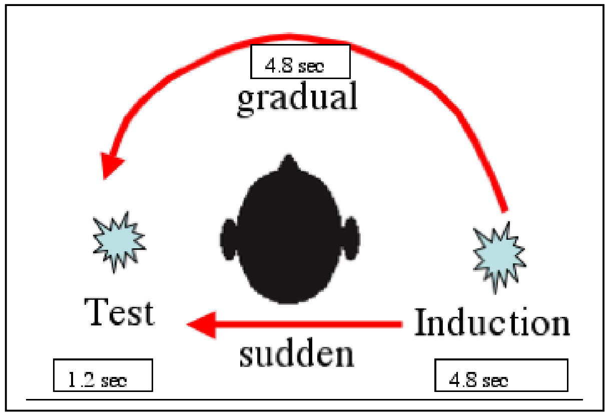

Figure 12.1. This figure demonstrates the difference between gradual motion and sudden motion as explained by Rogers & Bregman (1993; 1998).

Summary and Conclusions: Experiment 3

The results of Experiment 3 showed that the detrimental affect of spatial separation in reports of grouping seen in Experiment 2 was robust under conditions of motion. If there was a 90° angular separation between the tones, them listeners were more likely to segregate them, even if they were moving gradually (see figures 3.1 and 3.2). As in Experiment 2, the omnibus tests of both dependant variables (segregation boundaries, difference-scores) revealed that there was a difference based on presentation condition.

The planned comparisons indicated that the main effect was driven by the large differences (in both dependant measures) between the *separation* conditions and the control condition. Furthermore, similar differences were found between (in both dependant measures) between the *gradual-right* and the *2-90-right* conditions. Also, the gradual-motion conditions that did not contain any angular separation between the high and low tones were grouped as often as the control condition. These result suggests that spatial proximity is dominant to frequency similarity in grouping in a free-field context.

The results of Experiment 3 confirmed the prediction that listeners would maintain a percept of two sound sources throughout the presentations (*separation* conditions) and therefore, were more likely to segregate the tones. As in Experiment 2, the frequency dominance prediction that patterns of grouping would be similar between the conditions was not confirmed. However, grouping was similar among the control and the two *gradual* conditions in which the high and low tones shared a location as they moved. This is in line with the previous work by Rogers & Bregman (1993; 1998), but it does not necessarily refute the assumptions of DSAIL.

General Conclusions: Experiments 2 & 3 (spatial separation & gradual motion)

Based on the results of Experiments 2-3, it can be concluded that in this free-field context, listeners relied on spatial assessments to form groups. Whenever the spatial information differed between the tones, listeners segregated the tones, and whenever the tones shared a location, they were more likely to group them. These results are more in line with the alternative assumptions of DSAIL than the proponents of frequency dominance (Kubovy & VanValkenberg, 2001). Specifically, these results suggest that in a free-field context, listeners will group sounds by location with a method of continual spatial assessment of the relative location(s) of sound components.

Arguments for *frequency build-up* cannot completely account for the patterns of results reviewed so far without acknowledging the primacy of spatial proximity between sound sources (Rogers & Bregman, 1998). Even if the sounds are simple tones, as they were in these studies, frequency similarity has not overcome large angular separations between the tones.

Summary and Conclusions: Experiment 4

Experiment 4 was designed to examine if the perception of a common/disparate property of motion such as velocity, could influence listeners to the extent that spatial proximity becomes less important, or perhaps, would make frequency similarity more important. Contrary to predictions that may undervalue space (Bregman, 1990; Kubovy & VanValkenberg, 2001), it was predicted that if listeners are more sensitive to motion information than spatial proximity, similar velocities should produce more grouping than dissimilar velocities, regardless of the spatial proximity of the tones. This prediction was not confirmed by the results of Experiment 4, even though the omnibus test of the difference-scores produced a main effect of presentation condition (segregation boundaries test was marginal effect). However, this result did not confirm a frequency dominance account, because there were differences based on spatial parameters.

The planned comparisons on the segregation boundaries in Experiment 4 showed that both conditions wherein there were *far* angular separations at the end of the presentations were segregated significantly more often than the control condition. However, both conditions wherein the angular separations among the tones were close at the end of the presentation (*slightly-different-close* and *more-different-close*) were segregated marginally more often than the control condition. There were no differences in the segregation boundaries between the motion-conditions. Hence, velocity similarities among sound components were not found to facilitate grouping. If listeners use a process like that proposed by DSAIL, it cannot be concluded that the velocity of sound components can outweigh spatial separation in the formation of groups.

The planned comparisons of the difference-scores in Experiment 4 indicated that the difference of the estimated threshold and the final semitone separation (after trial 10) of the high and low tones was significantly larger in the *equal-far* condition than in both the *more-different-close* and the control conditions. This was in contrast to the prediction [DSAIL] that if listeners hear that two sounds share a velocity and direction, they would abandon their segregation based on localization. However, these results also show that, frequency similarity is only used to form groups when the sounds share a location. This finding exposes the frailty of the frequency dominance assumption (Kubovy & VanValkenberg, 2001).

The results of Experiment 4 suggest that overall, spatial separation still seems to be the primary cue by which listeners group sound components. While the specific predictions that listeners may abandon their localization(s) in favor of shared velocity were not confirmed, the assumption that listeners continually update their spatial assessments as they listen is still in tact. It would be difficult to explain these results from a frequency dominance perspective without acknowledging that the spatial location of the tones seems to drive these measures of grouping.

Summary and Conclusions: Experiment 5

The omnibus test of the segregation boundaries in Experiment 5 showed that there were differences in patterns of grouping among the presentation conditions. The planned comparisons indicated that only the *same-close* condition did not facilitate significantly more segregation than the control. However, the *same-close* condition did produce marginally more segregation than the control. As in Experiment 4, these results suggest that if components of sounds are separated, even common trajectory will not facilitate grouping, but will likely cause listeners segregate the sounds.

The difference-score results of Experiment 5 were quite different than the segregation boundary results. According to the omnibus test there were not differences in the patterns of grouping among the conditions. However, the paired comparisons showed that only the *opposite-far* condition had significantly larger difference scores than the control. Given that the high and low tones were near each other for nearly half of the time (see figure 11.1), this result is suggestive that listeners may, in some cases, rely heavily on opposite trajectories to segregate sound sources. Furthermore, analysis showed that the *same-close* difference scores were marginally lower than the *opposite-far* difference-scores. Finally, the *same-far* difference scores were marginally greater than the control, which is somewhat consistent with the results of Experiments 3 and 4. However, based on these results the [DSAIL] prediction that shared and opposite trajectories have different effects on grouping could not be confirmed.

As in Experiment 4, it could be concluded from the results of Experiment 5 that listeners will most likely use spatial disparity as a basis for segregation. If components of sound share a location, listeners are more likely to group them. Based on these results, even if the tones are very similar in frequency (5 semitones or less), even small angular separations are enough to facilitate segregation. This conclusion is in direct contrast to the assertions of frequency dominance over spatial location in grouping (Bregman, 1990; Kubovy & VanValkenberg, 2001). However, it could be that motion combined with angular separation of sound sources might lead to a bias for segregation. Still, it is clear that frequency similarity cannot outweigh these spatial properties in free-field listening.

Potential Shortcomings and Future Research

It could be argued that the results of these studies are not comparable to the research conducted by Rogers & Bregman (1993; 1998) because of the methodological differences. The listener's task in the current research was to report on their perception of the entire sequence, not just a small portion of it, as in the Rogers & Bregman (1993; 1998). As noted in Chapter 6, it was determined that it was insufficient to only test grouping in the final 1.2 sec of the presentation. Based on listener reports, the induction-test task was very confusing, primarily because there was nothing to indicate when to start 'listening'. Recall that most conditions in the Rogers & Bregman (1993; 1998) research contained abrupt changes in stimulus locations, which might haves implicitly indicated that the listener should get ready to respond. This implicit cueing could be argued to be a confound in that line of work, wherein the observed *segregation effect* might be a demand characteristic associated with the instructions. Furthermore, it made more sense theoretically for participants to report on their perception for the whole sequence because spatial separation and natural motion were specifically being tested in the current studies. It was imperative that the listeners task be as natural as possible to achieve comfortable listening, so in this sense, the induction-test task was inappropriate for measurements of listening in a natural context.

Further arguments against the current methods might focus on the estimations of listener thresholds for grouping at the onset of the experiments. Recall that in the Rogers & Bregman (1993; 1998) studies, there was a 9 semitone separation between the high and low tones in the first of 10 trials in each of the five presentation conditions for every participant. In our initial pilot studies this method was maintained. The resultant data was extremely variable to the point that clean analyses could not be conducted. As was the case for the induction-test task, some listeners in our pilot studies were confused by the 9 semitone separation. However, the abandonment of this technique was primarily due to a lack of explanation by Rogers & Bregman (1993; 1998) as to the reason for this arbitrary initial semitone separation. Based on the general principles of psychophysics (see Green, 1960), it was decided that establishing thresholds at the onset of the experiment would provide much more useable data and afford more flexibility in analyses.

Arguments against this modification could claim that listeners may have been primed to group the single location conditions, because the location of the sequences during threshold measurements was always at a single location, directly in front of the listener. It should be noted that the listener was positioned at 3 ft in front of the loudspeaker (as opposed to 1 m). Also during this phase, and the loudspeaker location was not located in a position that would be used in any condition. That is, the listener simply faced the far left loudspeaker in the array, a position they would not experience during the actual experiment. Also, in the previous work by Rogers & Bregman (1993; 1998), listeners were familiarized with the *actual* stimuli that they would hear in the experiment while doing practice trials.

Listeners in our studies were not exposed to spatial manipulations until the actual experiment, as they were in the previous work. Because of this fact, it might be true that there was some kind of practice effect that carried over into the critical portion of the experiment. However, it is doubtful that a practice effect could carry over to a motion-condition, such as the *gradual-right* condition (Experiment 3). In fact, both of the single location, *gradual* conditions produced patterns of grouping that were in line with the previous work. That is, those conditions produce similar patterns of grouping to the control condition. If a carry over effect was influencing participants, it should have been visible in differences between the *gradual* and the control conditions. This was not the case.

In future research, it would be useful to incorporate at least some spatial presentations during threshold measurements or training. Any sort of spatial manipulations would have to be carefully chosen to avoid demand characteristics or the facilitation of bias. One method may be to test listeners on a neutral task before and after training, to ensure that their judgments are not influenced by the presentations used in training. Such a task may include visual tasks that are associated with gestalt grouping principles (e.g. proximity, good continuation etc). If biases are created, it could be observable across modalities, although this would have to be validated as well. At any rate, it is potentially a very important issue that should not be ignored in future work. Still, based on the previous work, and the results of our studies, carry-over effects are improbable as confounding.

Another argument against the general conclusion [DSAIL] that spatial relations in some cases, dominate frequency relations could be that the stimulus presentations used in our studies are not comparable to those used in the previous work by Rogers & Bregman (1993; 1998). More specifically, the *sudden-change* hypothesis offered by Rogers & Bregman (1993; 1998) asserts that *sudden-changes* were the driving factor in their results. If *sudden* location changes are a primary cue for segregation, it could be argued that the data and results from Experiment 4 are due to the fact that there were sudden changes for every tone change (high-low-high) in the separation conditions.

It must be considered that the *sudden-change* argument itself demonstrates that spatial location is in fact primary to frequency similarity. More importantly, Rogers & Bregman (1993; 1998) did not fully develop the concept of *sudden-change*. Therefore, it is difficult to understand what justifies any change as 'sudden', based on the conclusions and discussions offered by Rogers & Bregman (1993; 1998). Furthermore, it was certainly clear that *frequency build-up* was considered most important for listeners to group the sequences. If Rogers and Bregman (1993; 1998) were correct about the primacy of *frequency build-up*, then the patterns of grouping should have been similar between all conditions in Experiment 2 because frequency was allowed to 'build-up' for the entire sequence in all cases. The only difference between the conditions must have been due to the angular separation between the high and low tones (see figures 8.2 &

8.3). These results must be concluded to be the result of listener sensitivity to the spatial location of sound components.

While strong conclusions can be made regarding the primacy of spatial location in group formation in the free-field, few strong conclusions can be made concerning properties of motion. While it seemed evident from the results of Experiments 4-5 that frequency similarity does not outweigh motion in group formation, an argument that motion may facilitate a bias for segregation cannot be rejected based on these analyses. However, if one adopts the ‘bias’ argument, one must admit that these spatial manipulations are indeed strong enough to create such a bias. If frequency were dominant to spatial features, would it not be expected to create a bias in favor of frequency similarity in cases when the listener is uncertain?

Based on researcher observations, participants in Experiments 4-5 often seemed confused when responding initially. Then, according to the reports, the participants would often seem to decide on a response (grouped or segregated) according to their recognition of the pattern. This observation was actually confirmed by a few post-experiment interviews with participants. According to the participants, they resolved their initial confusion by simply remembering how the pattern sounded, and responding according to their previous response. Data from these participants was not used in the analyses, but could be revealing as to why the data was so variable in these studies. Still, there is no question that listeners did not resort to frequency similarity as a basis to form groups when motion was introduced.

Another potential shortcoming could be that frequency cues are only considered at the location of the stimulus. There were no corrections made for head movements or other factors associated with receptor location cues. That is, there were no manipulations that controlled interaural time differences or interaural level differences. In defense of the methodology, these measures were not taken because it was a priority to attain a natural listening experience. However, it must be considered that frequency cues can be changed at the receptor level by simply moving one's head. It could be argued that by not considering such cues, frequency is actually undervalued by conclusions of the primacy of spatial location based on these results. As a rebuttal, it can be argued that head movements and the like are actually attempts by the listener to understand their spatial layout in order to group sounds as opposed to relying on the similarity of the sound components. Still, in future studies it would be interesting to monitor the head movements in order to quantify such changes in the stimulus at the receptor level. This could be done by tracking a visible mark placed on the participant's head within video footage recorded in each experiment. This would allow for further analysis of head movements in order to ensure that differences among conditions are not driven by such changes caused by the listener shifting their head position.

Improvements could be made to the general methodology or the stimulus presentations to get a better understanding of how motion effects grouping in presentations where spatial separations exist between sound sources. Further research focused on the effects of velocity on grouping may include conditions where the angular separations are smaller in conditions where the velocities are equal or similar. This could help to reveal the seemingly complex process by which listeners seem to be assessing the spatial layout of surrounding sonic components. It may be that when components of sound are close, velocity matters more in their interpretations. However, when considering real-world listening, it is certainly unusual for sound sources that are separated in space to be apart of a single sound source. A train, for example, can emit sounds that have an angular separation of 180°, but it is doubtful that these sounds are useful to the listener for any other reason than to avoid stepping into the sound stream. Perhaps, more useful presentations might focus on situations in nature when velocity could be most important to a listener.

Similarly, further research that examines the effects of trajectory on grouping should be more precise and would benefit from natural examples. It is probably most important, regarding DSAIL, to attempt to expose the points in time and space when listeners shift abandon a unitary perception for a two sound source percept and vice versa. Based on the results of Experiment 5, it seems likely that listeners will not group components of sound if they are separated in space, even if they are near each other and share a trajectory. However, it may be possible that there are instances when this effect could be facilitated. For example, if the high and low tones were moving in the same direction and constantly separated by 36°, the tones might be made to alternate locations as they move so the average frequency of the tones would actually fall between them in space. Incremental variations in the angular separation of the tones would be a trivial manipulation. A natural example of this kind of sound source may be a flock of geese or a swarm of bugs. Again, future work must consider natural examples when designing stimuli.

Theoretical Implications

The general result that grouping in the free-field is dependant upon the spatial proximity of the sound sources has implications for a theoretical framework such as ASA (Bregman 1990). The most important implication is that the prevailing view that spatial features, such as location, in fact can sometimes *dominate* frequency similarity in low-level grouping. However, because the current formulation of ASA does little to explain how listeners might weigh stimulus features such as frequency content or spatial proximity, this finding implies that such a formulation should be made. While the primacy of spatial location has been overlooked in the foundational research of ASA (see Bregman 1990), these findings can be viewed as evidence to strengthen ASA as an explanatory theory of human sound perception.

Spatial illusions such as the octave and good continuation used as evidence for the primacy of frequency (Bregman, 1990; Kubovy & VanValkenberg, 2001), are actually exceptions to the primacy of spatial location. This does not discount the importance of such examples, but it does imply that much more research is needed to better understand why there are exceptions, and when they occur. Also, these findings suggest that the research cited by the proponents of frequency dominance should be revisited to ensure that reliable frequency dominance effects have been thoroughly tested in a free-field context.

While the spatial primacy effect found in these studies may seem isolated, recent research on spatial release from masking suggests that space plays a much more important role in neural processing than previously thought (Stecker et al. 2005; 2005). In short, this work has provided evidence for a spatial coding among auditory neurons, even though it has long been assumed (by most) that auditory neurons only have frequency codes. Along with new evidence for a spatial code in auditory cortex, the spatial primacy effects shown in our studies are converging evidence for re-evaluating the importance of space in auditory perception.

Moreover, it would be beneficial to further develop the concept of *sudden-change* in free-field context. Recent research by Snyder and colleagues (2005; 2009) has shown that there is a continued interest in the neural correlates of adaptation as it pertains to auditory grouping and stream segregation. The results and method of the current research could provide a whole new set of questions and line(s) of research to examine the neural correlates of adaptation and auditory grouping in the free-field context. Based on the results of the current research, it is imperative that spatial parameters be considered to fully understand how the brain uses sound.

The spatial primacy effect produced in our studies cannot be meaningfully extended to properties of motion. This could be because if there are multiple sounds emanating from disparate locations, listener localization leads to a robust segregation of the sources that properties of motion cannot overcome. However, the work here is far from exhaustive regarding a potentially subtle, but dynamic relationship among spatial and frequency cues that a listener may rely on when discerning their auditory scene.

Even though the results of Experiments 4-5 provided limited evidence for the importance of motion, the results suggest that more rigorous psychophysical methods could be useful to test the question further. It can be difficult to find methods that utilize the rigor of psychophysics in the study of perceptual organization. This is primarily due to the potential confounds associated with higher-level cognitive functions such as memory and strategic reasoning. Still, the stimulus presentations used in this research might be simple enough to pursue questions of auditory organization using classic psychophysical/psychoacoustical techniques.

Practical Considerations

The work presented here was conducted with minimal resources, aside from the abundance of participants. Technically, this work demonstrates that even with minimal programming and equipment, it is now possible to easily conduct tests of human spatial hearing with complex spatial presentations. Furthermore, this work also shows that with the current technology available, complex real-time data can be gathered fairly easily. While the results are not intended to be directly applicable to multi-sensory/multi-modal research, the project itself is certainly suggestive that future work could easily involve multimodal presentations along with several potential sources of data such as reaction time, eye movements, ERPs etc.

The theoretical implications of the research findings are primarily limited to auditory perception, but with recent developments of laboratory equipment and computer processing, the implications might easily transfer to stimulus presentations in a virtual reality context. Specifically, the methods used here might be useful in studies of immersion that use visual, auditory and tactile stimulation. Replicating our effects in an immersive environment may be a useful validation tool for system calibration. Also, laboratories with such facilities may find it trivial to extend our findings to continue the search for answers concerning the dynamic nature of spatial perception. It seems likely that given the evidence of visual dominance over sound in spatial perception, such as the ventriloquist effect, that a preference for properties of motion might be easily demonstrated.

DSAIL: A new theory of dynamic spatial sound perception

This research has shown that in a free-field context, listeners will rely on dynamic localization of sonic components when forming groups. The source localization literature provides a platform by which an account of dynamic, spatial grouping can be formulated with the methods used in the studies conducted in this research project. Based on the sound localization research, it has long been known that listeners are sensitive to small changes in the spatial location of sound sources along the horizontal azimuth (minimum audible angles) (Perrot & Tucker, 1988). Furthermore, psychoacoustical research has also provided numerous data involving human sensitivity to motion and distance changes in stimuli. However, until now, there have been few attempts (Rogers & Bregman, 1993; 1998), to extend the breadth of that work to higher level processes such as low-level grouping.

Based on the results found in our studies, it seems that the psychoacoustical work can be extended into the domain of perceptual organization. DSAIL asserts that when listeners are immersed in a free-field sonic environment, their perceptual organization(s) could be based on a continuous localization process (spatial assessment). The results of these studies suggest that this assessment process can dominate frequency similarity in group formation in nearly all cases *except* when the sounds are emanated from a single location. An account such as DSAIL, provides a basis for hypotheses that include space as at least potentially primary in spatial auditory perceptual organization.

References

Best, V., Gallun, F.J., Carlile, S., and Shinn-Cunningham, B. (2006). "Binaural interference and auditory grouping," J. Acoust. Soc. Am. 1070-1076.

Bregman, A. (1990). *Auditory scene analysis: The perceptual organization of sound.* CMA: The MITP.

Briggs, R.M. & Perrot, D.R. (1972). Auditory Apparent Movement Under Dichotoic Listening Conditions. *Journal of Experimental Psychology, 92(*1), 83-91.

Butler, D. (1979). A further study of melodic channeling. *Perception and Psychophysics, 25,* 264-268.

Chandler, D.W., & Grantham, D.W. (1992). Minimum audible movement angle in the horizontal plane. *Journal of the Acoustical Society of America, 79,* 1939-1949.

Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In B. C. J. Moore (Eds.), *Hearing*. *Handbook of perception and cognition* (2nd ed.). 387-424. San Diego, CA: Academic Press.

Deutsch, D. (1983). Auditory illusions, handedness, and the spatial environment. *Journal of the Audio Engineering Society*, *31,* 607-622.

Deutsch D., & Roll, P. (1976). Separate "what" and "where" decision mechanisms in processing a dichotic tonal sequence. *Journal of Experimental Psychology: Human Perception and Performance*, *2,* 23-29.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, Ma, USA: Houghton Mifflin Company.

Grantham, D.W., Hornsby, B.W.Y., Erpenbeck, E.A. (2003). Auditory spatial resolution in vertical and diagonal planes. *Journal of the Acoustical Society of America, 114(*2), 1009-1022.

Green, D.M. (1960). Psychoacoustics and detection theory. *Journal of the Acoustical Society of America, 32,* 1189-1203.

Handel, S. (1988). Space is to time as vision is to audition: seductive but misleading. *Journal of Experimental Psychology: Human Perception & Performance, 14* (2), 315-317.

Kubovy, M., & Van Valkenberg, D. (2001). Auditory and visual objects. *Cognition, 80,* 97-126.

Middlebrooks, J.C., & Green, D.M. (1991). Sound localization by human listeners. *Annual Review of Psychology, 42,* 135-159.

Neuhoff, J. G. (2003). Pitch variation is unnecessary (and sometimes insufficient) for the formation of auditory objects. *Cognition, 87,* 219-224.

Palmer, E. (2009). Hearing in the environment. In Wolf eds. *Sensation and Perception.*

Perrot, D.R., & Tucker, J. (1988). Minimum audible movement angle as a function of signal frequency and the velocity of the source. *Journal of the Acoustical Society of America, 83,* 1522-1527.

Rogers, W. L., & Bregman, A.S. (1993). An experimental evaluation of three theories of auditory stream segregation. *Perception & Psychophysics*, 53, 179-189.

Rogers, W.L., & Bregman, A.S. (1998). Cumulation of the tendency to segregate auditory streams: Resetting by changes in location and loudness. *Perception & Psychophysics*. 1216-1227.

Snyder, J.S., Carter, O.L., Hannon, E., Alain, C. (2009). Adaptation reveals multiple levels of representation in auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance, 35*(4), 1232-1244.

Stecker, G.C., Harrington, I.A., Middlebrooks, J.C. (2005). Location coding by neural populations in the auditory cortex. *PloS. Biol., 3*(3), 520-528.

Strybel, T.Z., and Neale, W. (1994). The effect of burst duration, interstimulus onset interval, and loudspeaker arrangement on auditory apparent motion in the free field. *Journal of the Acoustical Society of America, 96* (6). 3463-3475.

Saberi, K., and Perrott, D.R. (1990). Minimum audible movement as a function of sound source trajectory. *Journal of the Acoustical Society of America, 88,* 2639-2644.

Van Valkenberg, D., & Kubovy, M. (2003). In defense of indispensable attributes. *Cognition. 87*, 225-233.

Walker, B.N., and Kramer, G. (2007). Ecological psychoacoustics and auditory displays: hearing, grouping, and meaning making. In Neuhoff, J.G. eds. *Ecological Psychoacoustics.*

Warren, R.M., Osucek, C. J., Farmer, R. M., Warren, R. P. (1969). Auditory sequence: confusion of patterns other than speech or music. *Science, 164*. 586-587.

Zahorik, P., Brungart, D.S., and Bronkhorst, A.W. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acustica, 91,* 409-420.

Appendix A

<u>Instructions for all experiments</u>

All instructions below shown in bold face type and noted with * are instructions to the research assistants and were not said to the participants. All other text was to be read to the participants by the researchers as instructions.

In this experiment we are interested in examining the effect that moving sound sources has on the perception of a three tone sequence. We will refer to the sequences as gallops or non-gallops. Your task is going to involve reporting to what extent the tones sound like a gallop or a non-gallop during the trials.

***Show them the picture**

This short sequence will sound like three pitches which alternate with a triple-beat rhythm. Let me tap it to show you...

This example stimulus should help clarify this to you.

***Play them stimuli 'low 3'**

Do you hear the gallop?

Here's an example where the gallop falls apart.

***Play them stimuli '19'**

Do you hear a gallop?

What do you hear instead?

Lets listen again to compare and make sure you understand.

***Play them stimuli 'low 3'**

Does it sound like a gallop?

P*lay them stimuli '19'

Do you hear a gallop?

(If they don't get it repeat the process until they do)

Okay. The reason the sequences sometimes are galloping and sometimes are not galloping is because of the pitch differences between the tones in the sequence.

Show them the picture and point out that as the frequencies get closer it is galloping and when they are further apart they aren't galloping.

***Play them stimuli 'low 3'**

The frequencies are close enough in this one. **(point to the picture)**

***Play them stimuli '19'**

The frequencies are too far in this one. **(point to the picture)**
This next one is a bit more like the ones you will hear in the experiment.

***Play them stimuli 9**

Does this sound like a gallop?

***(The answer should be no, if it is yes, play them stimuli low 3 again and alternate between them until they get it right)**
I think you understand.
Do you have any questions before we move on?

Lets get started then.

PARTICIPANT ADJUST INSTRUCTIONS

In the first part of the experiment we need to find that crucial point where the galloping begins and ends for you. Its different for everyone so its necessary for you to be sure about how you respond. The way this first part is going to work is that you are going to actually adjust the frequencies of the tones to control your perception of the gallop. The instructions will vary from trial to trial in this part so its very important that you are sure you understand what to do. Now, in this first group of trials, when the sequence starts, you will either turn the (silver) dial on the mouse clockwise to make the gallop fall apart, or turn the (silver) dial counter clockwise to bring it back together.

***(have them turn it clockwise and watch the little line move up, then tell them to move it counter clockwise until the line comes back down to one).**

Good. We are just about ready, but there are just a couple more very important instructions I have to give you regarding your responses.

First, I need to tell you that to report that you hear a non-gallop, you need to press the little black button on the left of the mouse. This tells me that you hear a non-gallop. Later in the experiment you will use this button to report '0%' galloping, so this button is thought of as the 0% button. To report that you do hear the gallop, push little black button on the right side of the mouse. This will be known as the 100% button. For these first few trials, these two buttons are the only ones you need to worry about.

Secondly, and this is very important, you will notice that during the sequences that at first the sequence does not sound like a gallop at all, but after only a few seconds, it suddenly sounds like a gallop! You can expect this to happen, and so during these first few trials, I want you to wait a few seconds before you respond just to be sure that it is either a non-gallop or a gallop.

Understand?

Any questions?

Okay. On this first trial, the sequence is going to begin as a gallop and I want you to turn the dial clockwise until the sequence becomes a non-gallop. Remember, as you turn the dial, it will initially sound like a non-gallop cause you are used to hearing it the other way. I need you to wait just a few seconds to make sure it is no longer galloping. When you are absolutely sure it is not galloping, press the 0% button.

Now that you've heard how the sounds within the sequences can actually make the sequence seem different, I'm going to play you some examples to demonstrate when the sequences sound as '1' coherent gallop pattern, or when they seem to split apart into a non-gallop pattern. These examples will more closely resemble the actual stimuli, in that they will be exactly the same length (approximately 4.5 sec). Pay special attention to the last 3 cycles of the triple beat rhythm. Any questions?

After the experiment:

Thank you for participating in our experiment. Before I tell you about how this research may be used, is there anything you can tell me about your experience while performing these tasks?

Did you feel like you heard the gallop more easily when the sounds were not moving?

Write down any responses they have to these questions

Thanks for that information. This research can be used to help design more useful human/machine interfaces, like the controls of your car or a jet cockpit. They can also be used to aid in the development of more useful architecture and urban planning; designing building and rooms that are more acoustically informative. Also, results from these studies can aid in the development of machines that can hear like humans by modeling performance in such tasks as you've just completed. These studies can also aid in the development of more realistic virtual environments for training simulations or entertainment, like video games. If there aren't anymore questions, I'd like to thank you again for your participation. You will receive credit electronically, so please contact me if you do not receive your credit within 3 days. Thanks again.