1993

Factors Affecting the Time-Course of Auditory Stream Segregation.

Lawrence Lee Mendoza III
Louisiana State University and Agricultural & Mechanical College

Order Number 9419912

Factors affecting the time-course of auditory stream segregation

Mendoza, Lawrence Lee, III, Ph.D.

The Louisiana State University and Agricultural and Mechanical Col., 1993

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

FACTORS AFFECTING THE TIME-COURSE OF AUDITORY STREAM SEGREGATION

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Communication Sciences and Disorders

by
Lawrence L. Mendoza, III
B.S., University of Southwestern Louisiana, 1986
M.S., Florida State University, 1988
December 1993

ACKNOWLEDGMENTS

A list of people who served as sounding boards and/or offered moral support would add significantly to the length of this document. Here is a partial list of people important to the completion of this dissertation.

This dissertation was supported by a Grant-in-Aid of Research from Sigma Xi, The Scientific Research Society.

Dr. M. Jane Collins. Her direction during my time at LSU, and particularly during the course of this dissertation, was invaluable. Alex couldn't have a better aunt.

Dr. John K. Cullen, Jr. His knowledge of the auditory system and hearing science is only exceeded by his patience with his students.

Drs. Joseph Hall, III and John Grose. They provided helpful comments, hospitality, and patience. A special thanks goes to John Grose for saving this dissertation from digital oblivion.

Vijay Ratham, M.S. Vijay contributed to the software used to develop the signals used in this study.

Dr. Patrick Rappold. Discussions with Pat are always informative (and entertaining). Earlier discussions with Pat got this idea off the

ground, and his discussions and encouragement along the way were quite helpful.

Dr. Tom Rigo. Tom got the ball rolling many years ago.

The subjects in both the dissertation and the pilot, who will go unnamed, gave unselfishly of their time.

My parents, Larry, Sandy, Herb, and Adele. Without their support and encouragement, this dissertation could not have been attempted, much less completed.

Finally, my family, Alisa & Alex. Alisa spent many hours listening to endless ramblings while working on her own dissertation, and on Alex. I respect her greatly. Alex joined us along the way, and continues to put everything into perspective.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Auditory scene analysis refers to the process through which sounds are heard as either belonging to separate sources or as perceptually grouped together and arising from a single source. Stream segregation describes the perceptual separation of sounds, while sequential integration is a term used to describe the linking together of sounds following one another in time. This study examined the minimum duration necessary for the perception of sequentially-presented sounds to change from sequential integration to stream segregation. The experimental stimuli were sequences of alternating sounds. Auditory scene analysis was examined under conditions in which the sequential sounds differed only in frequency (pure-tone sounds) or center frequency (amplitude-modulated noise-bands). Other experimental conditions were included in which the alternating sounds differed also in intensity (pure-tone sounds) or temporal envelope (amplitude-modulated noise-bands). This study revealed that stream segregation was increased by increasing the duration of the sequences. Frequency separation also had a significant effect on auditory scene analysis. Larger frequency separations led to more rapid stream segregation. Additionally, less segregation was observed for the amplitude-modulated sounds with similar temporal envelopes. Finally, the results indicated that intensity difference and envelope similarity were not processed independently of the frequency separations used in this study.

INTRODUCTION

For the past forty-odd years, researchers have investigated the means by which the auditory system perceptually organizes complex acoustic signals. Their experimental inquiries have arisen from diverse questions concerning auditory perception as it relates to auditory figure-ground perception (Bregman, 1990), or the perception of speech signals (Chalikia and Bregman, 1989; Repp, 1987).

Auditory perceptual organization, in the above contexts, refers to the ways in which the auditory system structures complex acoustic signals. The result of auditory perceptual organization is the capability to isolate or "hear out" the individual constituents that make up a complex sound (e.g., speech in noise).

There are several prerequisites for the auditory perception of a speech signal. An obvious requirement is that the speech must be audible to the listener. A second necessary step is the separation of the speech signal from other co-occurring sounds that may act to hinder the perception of speech. To separate the components of a sound, however, the auditory system must first perform an analysis of the entire sound. In this analysis, the perceptual system must accurately parse a complex pressure wave that was generated by more than one source into two or more perceptual entities. The analysis could be considered an attempt to increase the signal-to-noise ratio. Parsing the complex wave into constituent sounds is analogous to deciding which of the components belong to a signal (or is generated by a particular source) and which are noise (or <u>not</u>

produced by the source) and should be disregarded. The component deemed signal and those deemed noise is a matter of choice. For instance, when listening to an orchestra, an individual may wish to pay particular attention to the violins (in this case the signal) and not the trumpets (noise). Thus, through perceptual organization, listeners "tune in" to a constituent of a complex sound they consider as being important.

The perceptual organization of an acoustic environment is termed "auditory scene analysis (ASA)" (Bregman, 1990). ASA can be thought of as operating through three processes, the *fusion of simultaneous sounds*, *stream segregation*, and *sequential integration*. The fusion of simultaneous sounds refers to the perceptual grouping of sounds that co-occur. The effects of fusion can be heard, for example, when one perceives a musical chord instead of the individual components making up the chord. Stream segregation operates on both simultaneous and sequential sounds. When stream segregation occurs, the sounds in an acoustic environment appear to be generated by more than one source. Once the acoustic environment is parsed into constituents, each constituent (or, alternatively, the constituent under scrutiny) is called an "auditory stream," or simply a "stream." After stream segregation in the orchestral example described above, the violin sounds would form an auditory stream separate from that associated with the trumpet sounds. Sequential integration describes the association of sounds that follow one another in time. The act of hearing a melody from a series of sequential notes is an example of

sequential integration. These processes organize an acoustic environment, presumably to relate the constituents of the environment back to their source or sources of generation.

Factors known to affect stream segregation

Historically, attempts to define the acoustic characteristics that aid the parsing of the acoustic environment have been based on the perception of continuous tone sequences. The reasons for using tone sequences are threefold: (1) tonal stimuli are easily generated and calibrated; (2) they are free from the influence of linguistic processes; and (3) extraneous variables that could affect or confound experimental results are more easily controlled than in experiments using complex stimuli. Miller and Heise (1950), for example, used a repeating sequence of tones that alternated in frequency. They observed that listeners perceived a single sequence in some instances and two separate sequences in other instances. When two sequences were heard, one consisted of only the high-frequency tones; the remaining sequence consisted of the low-frequency tones. The percept that listeners reported was dependent on the frequency ratio of the alternating tones. For sequences in which the lower-frequency tone ranged from 125 Hz to 7000 Hz, a minimum frequency ratio [$(F_H - F_L) / F_L$, where F_H is the higher frequency and F_L is the lower frequency] of roughly 0.15 was necessary for segregation when the entire sequence was presented at a rate of 10 tone-bursts per second. Twenty-five years later, van Noorden (1975) found that the segregation of a continuous sequence of alternating tones was affected by both the frequency ratio of the

tones in a sequence, and the rate of presentation of the tones in a sequence. As the speed of a sequence increased, the frequency difference required for the sequence to segregate into two separate tonal groups decreased.

Intensity variation of tones in a sequence will also affect stream segregation. For example, a sequence consisting of a repeating pure tone may segregate if the intensity of the tone alternates by as little as 3 dB (van Noorden, 1977).

Jones (1976) proposes a model of stream segregation and sequential integration based on frequency and intensity differences between alternating sounds. The dimensions of frequency and intensity differences are mapped in perceptual space, with Δf on one axis and ΔI on another. The *serial integration region* (SIR) is a subset of this space, corresponding to small differences in frequency and intensity between two sounds. If the frequency or intensity differences of the sounds in a sequence are small enough to fall within the SIR, stream segregation does not occur. Stream segregation occurs only when differences in frequency or intensity are large enough to exceed the SIR boundaries in either dimension. According to Jones, the SIR boundaries change to decrease the space as the rate of the sequence increases. This aspect of the model is necessary to explain the rate-frequency tradeoff observed by van Noorden (1975).

While no research has been conducted on sequential integration based on temporal envelope similarity, Bregman, Levitan, and Liao (1990) examined the fusion of simultaneous sounds as a function of this parameter. In their experiment, an amplitude-modulated (AM)

tone was alternated with a replicate AM tone; both tones had the same center frequency. The replicate tone was presented simultaneously with an AM tone having a different center frequency. Listeners were to rate the "decomposition", or lack of fusion, of the two simultaneous sounds. The modulation frequency of the first AM tone and its replicate was held constant, while the modulation frequency of the off-frequency tone was varied from below to above the modulation frequency of the other two sounds. Ratings of decomposition decreased as the modulation frequency of the off-frequency tone approached that of the other tones, reaching a minimum when all three tones were modulated at the same rate. The authors suggested that the lack of common modulation rate among the simultaneous sounds decreased the fusion of the two sounds. This left the sound identical to the singly-presented sound open for sequential integration. These results were consistent with earlier work (Bregman, Abramson, Doehring, and Darwin, 1985) which examined the effects of across-signal AM rate on fusion over a smaller range of amplitude-modulation frequencies.

The effects of AM on the perceptual grouping of simultaneous sounds can also be seen to operate in the speech domain. Using a synthetic analog of speech known as time-varying sinusoidal (TVS) speech, Carrell and Opie (1992) examined intelligibility under three conditions: 1) with no modulation of the TVS components, 2) with each of the TVS components modulated at 100 Hz (coherent modulation), and 3) with each of the three TVS components modulated at different rates (incoherent modulation). The results showed that the coherently-modulated TVS speech was more

intelligible than the unmodulated TVS speech. Additionally, the intelligibility of the coherently modulated TVS speech was superior to that of the TVS speech with incoherent modulation. Carrell and Opie speculated that if the incoherently-modulated sounds were not highly fused, then the likelihood that the sounds arose from the same source would be low. Thus, the incoherently-modulated sounds would be less likely to be heard as speech than would the coherently-modulated (and thus fused) TVS components. This resulted in superior intelligibility for the coherently modulated TVS speech than for the other conditions. While the studies demonstrating fusion as a function of common modulation patterns were performed using simultaneous sounds, it might be possible for the auditory system to use common modulation patterns in the perceptual organization of sequential sounds as well.

The time-course of stream segregation

If the auditory system is to extract specific portions of a complex acoustic environment, the system must need some finite amount of time to perform an analysis of the environment. A set of constituents designated as "different" from the remainder of those in the environment form a separate stream and become dissociated from the remainder of the sounds. If the components of interest are highly dissimilar from other components, it is reasonable to assume that the two sets of sounds will segregate from each other rapidly. For example, the segregation of a string bass and piccolo played simultaneously should occur after only a brief period because these instruments generate strikingly different sounds. In contrast,

experience suggests that segregation of the voices of multiple talkers may require considerable effort and possibly more time. Presumably, the difficulty in isolating one speech signal from the total sound is due to similarities among speech sounds, regardless of the speaker.

The question that arises is "Does a relationship exist between the similarity of the sounds in an acoustic environment and the duration necessary for the auditory system to segregate the sounds?". It is hypothesized that the minimum interval of time that is necessary to separate two sounds perceptually is directly related to the degree of similarity of the two sounds.

It is assumed that stream segregation operates to relate components of a complex acoustic environment back to their original sources, segregating the sounds arising from one source from the remaining sounds. However, the acoustic environment must be observed for some period before the decision is made that more than one sound source exists.

Bregman (1978a) investigated the time-course of segregation indirectly and concluded that the tendency toward segregation increased as the number of components in a sequence increased. He used sequences made of a fixed number of low-frequency tones (330 Hz) alternated with high-frequency tones (784 Hz and 831 Hz). Following a silent interval of four seconds, the sequence (330 Hz, 784 Hz, 330 Hz, 831 Hz) was repeated. Subjects were allowed to listen to as many sequence repetitions as they wished and were instructed to adjust the rate of the sequence until segregation of the components occurred. For sequences consisting of only four

tones, segregation occurred at a presentation rate of 6.5 tones per second. Thus, segregation resulted when the duration of the four-tone sequence was 620 msec. If the number of tones in a sequence was increased, the perception of segregation was possible at a lower sequence rate than if there were fewer tones in a sequence. For sequences consisting of 16 tones, segregation occurred at a sequence rate of 4.5 tones per second, i.e., when the duration of the 16-tone sequences was 3556 msec. The segregation of components at lower sequence rates has been associated with easier segregation of sounds (see, for example, van Noorden, 1975; Rappold, Mendoza, & Collins, 1993). Thus, it could be inferred that the longer sequences (16 vs. 4 tones) were more likely to segregate than shorter sequences, as measured by a lower rate (4.5 versus 6.5 per sec) necessary for the perception of segregation. The interpretation offered by Bregman was that stream segregation occurred through an accumulation, over time, of evidence indicating that more than one sound source was present. Segregation only occurred after the auditory system had sufficient time to examine the acoustic environment.

Previous research has indicated that the temporal length of a sequence eliciting segregation changes as a function of the frequency difference of the tones in the sequence. French-St. George and Bregman (1989) had subjects listen to alternating pure-tone sequences lasting 30 seconds. They showed that the perception of segregation occurred at seven seconds with a five-semitone frequency difference of the tones and at approximately 25 seconds with a frequency difference of one-semitone. That is, stream

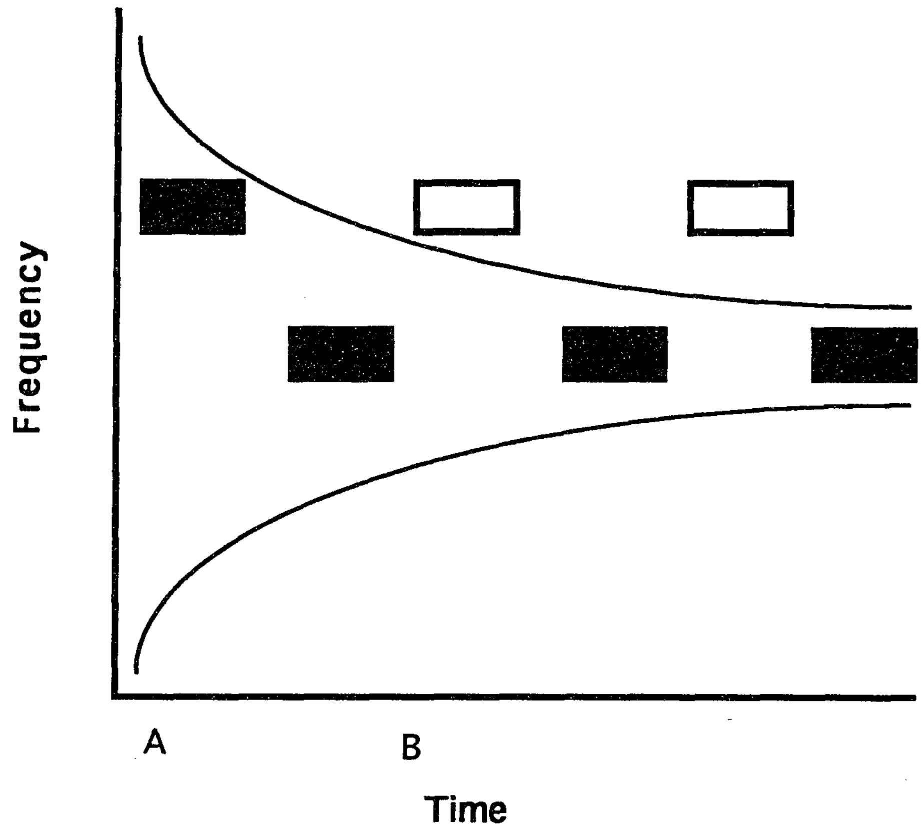

Figure 1. The formation of an auditory stream. Frequency is represented on the Y-axis. The X-axis shows acoustic event duration as increasing from left to right. Filled boxes represent sounds in the stream: open boxes are rejected sounds. See text for explanation of points A and B.

segregation only occurred after some duration of the sequences. The requisite duration was dependent on the frequency difference of the tones in a particular sequences.

Anstis and Saida (1985) also supported the idea that the frequency difference necessary for stream segregation changes over time. During presentation of the sequences, subjects were to adjust the frequency separation of the tones in order to maintain the perception of segregation. Subjects then heard an alternating series of tones over a period of 30 seconds. Anstis and Saida reported that as the sequences increased in duration, the frequency difference between the tones necessary for the perception of segregation decreased.

The two studies reviewed above provide evidence that stream segregation is only accomplished after examining the attributes of sounds in the acoustic environment over some period of time. In addition, as sounds are increasingly separated in frequency, segregation occurs more rapidly.

Jones's model (discussed earlier), however, assumes that segregation does not change during the course of a sequence. In addition, this model does not consider the effects of differences in stimulus attributes other than those associated with frequency and intensity. If the aspects of the model to be proposed here are incorporated into Jones's theory, then the serial integration region (SIR) should be maximal at the beginning of a sequence, and decrease over short durations. In other words, the SIR is dynamic, not static.

Additionally, the number of axes of the SIR should be expanded, corresponding to the similarity of the sounds along other dimensions.

Figure 1 shows schematically the changes, over time, in an auditory stream based on frequency differences between sounds. All sounds in the acoustic environment are initially assumed to be generated by a single source. Thus, all sounds are assigned to the stream (point A). However, sounds from the same source should be similar in terms of their attributes. As time goes by, the frequency range of the auditory stream decreases, and sounds differing greatly in frequency are separated from the stream (point B). Therefore, as an acoustic event continues, sounds with smaller frequency differences are rejected until the auditory stream encompasses a stable range of frequency difference. Only one auditory stream exists at a given time; sounds not in the auditory stream form a background.

The process proposed here is more complex than the above representation. Stream segregation is based not only on differences in frequency, but also on differences in other dimensions, including intensity and temporal envelope. The overall similarity of two sounds is assessed through a comparison of the attributes of each sound. Those attributes that differ between sounds contribute to the perceptual distance between sounds. The perceptual system has the opportunity to refine the analysis of the environment as long as a particular acoustic environment (or an auditory event) exists. Over time, the extent to which two sounds can differ and still be

associated with the same source diminishes. Thus, the range over which components of an acoustic environment can differ and not become segregated decreases over time.

Stream segregation should occur rather slowly for an environment in which the sounds have a given difference in an attribute compared to (1) an environment containing sounds with a larger difference in the attribute, and (2) an environment in which the sounds differ in more than one stimulus attribute. <u>It is hypothesized that differences in more than one attribute lead to faster stream segregation compared to the speed of segregation for sounds with differences in only one attribute</u>.

A sound can be described as having a set of attributes <A, B, C, ...>, with the attributes corresponding to pitch, loudness, timbre, and so forth. The subjective percept a particular sound evokes depends on the values of these attributes. For example, let sound 1 have a set of attributes with the values $<A_1, B_1, C_1>$, and sound 2 the attribute values $<A_2, B_2, C_2>$. (Only three attributes are represented here for simplicity.) In comparing the attributes of the two sounds, let the values of attributes A and C of the first sound equal the values of the second sound's A and C. That is, $A_1=A_2$ and $C_1=C_2$. The value of the first sound's B_1, however, is different from the value of B_2 of the second sound. If B represents the attribute of loudness, then the first sound appears to be identical to the second sound, except that the first sound is physically more intense than the second sound. Since attributes A and C of sound 1 and sound 2 are the same, these attributes do not encourage the segregation of the

two sounds. Only the difference in attribute B provides evidence that the two sounds might have been generated by different sources, and thus contributes to segregation.

The representation can be simplified if the attribute values of a sound (<A_X, B_X, C_X>) are used to form a single *combined attribute value*, or CAV, for the sound. The CAV of a sound having the attribute values <A_X, B_X, C_X> can be represented by S_X. When we say that an acoustic environment containing two sounds (having CAVs of S_X and S_Y) is analyzed, a comparison of the CAVs of the two sounds is implied.

It is assumed that the relative influence, or weight, of each attribute in the combination process is unequal. In addition, the attributes may not be orthogonal. That is, attributes of a sound may interact, affecting the CAV in a multiplicative fashion. Furthermore, the weights assigned to attributes may differ from individual to individual.

For the purposes of description, the action of segregation may be viewed in terms of a probability distribution that only one sound source is contributing to the acoustic environment. For example, consider the case for the perceptual organization of a sequence consisting of a repeating sound. Figures 2A, B, and C illustrate the perceptual organization, with CAVs on the X-axis. The Y-axis represents the probability that a particular sound with a corresponding CAV of S_X arose from the same source as the preceding S_X, given all previous S_X. The horizontal line represents a criterion probability that must be reached by CAVs for their corresponding sounds to be included in the auditory stream.

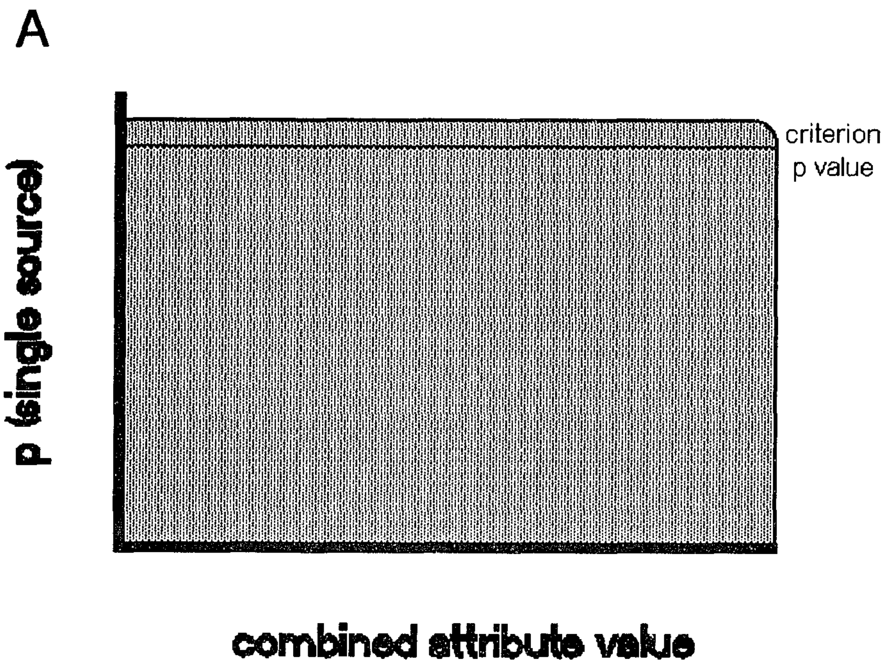

Figure 2A. Representation of the CAVs spanned by an auditory stream. The shaded area shows the CAVs that reach a criterion probability of arising from the same source. Before a stimulus exists, no CAVs are excluded from the stream.

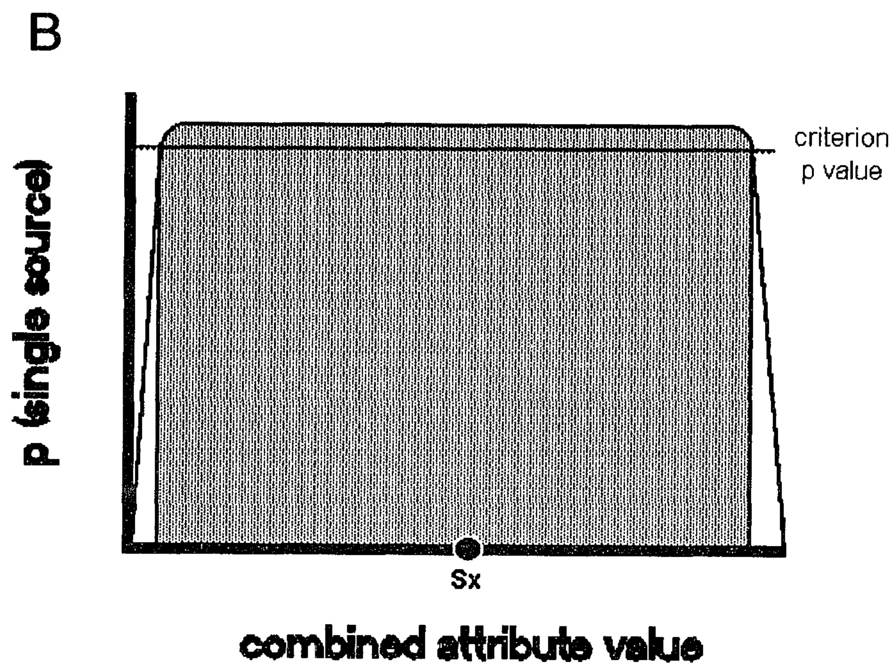

Figure 2B. The range of CAVs when a stimulus is first presented. The range of allowable CAVs is reduced compared to Figure 2A. See text for explanation of S_X.

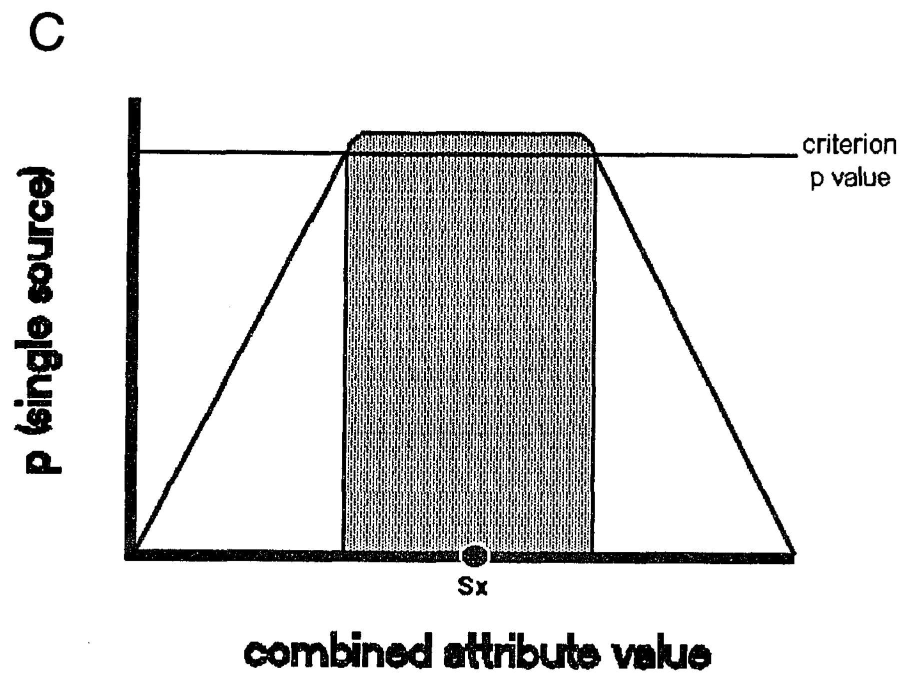

Figure 2C. The range of CAVs that reach criterion probability of being generated by the same source is reduced as the stimulus repeats.

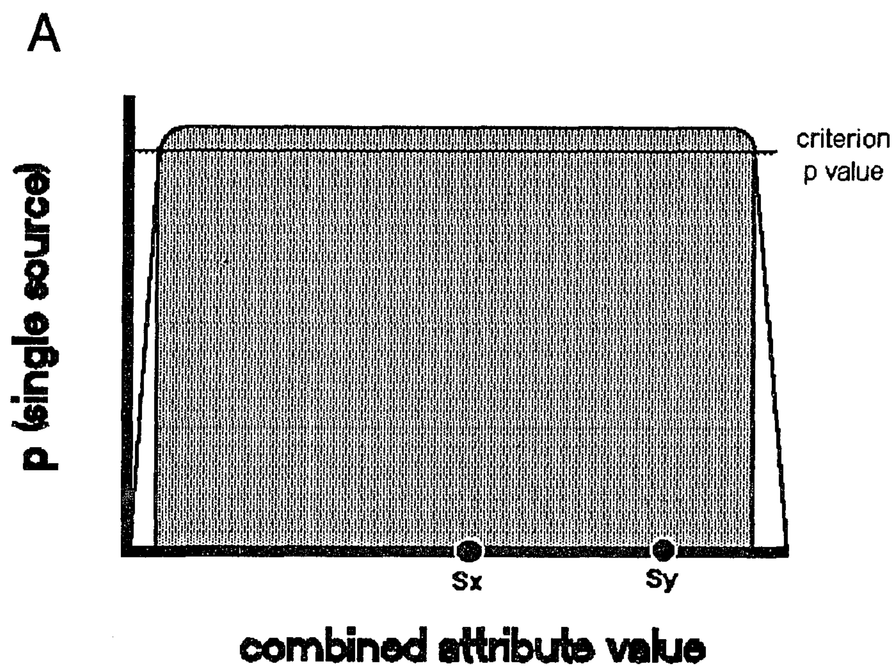

Figure 3A. CAVs encompassed by an auditory stream for an alternating sequence of tones. At the beginning of the sequence, both tones are in the stream. See text for explanations of S_X and S_Y.

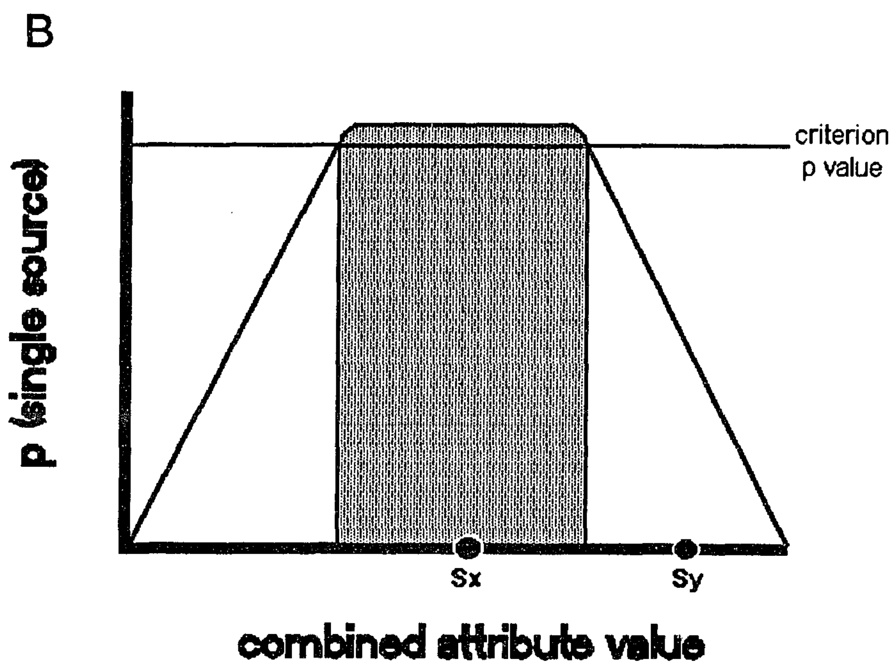

Figure 3B. After some duration of a sequence, S_y does not reach criterion probability of being generated by the same source as S_x. Segregation of S_x and S_y has occurred.

Before the listener hears the repeating sound, the only information available for analysis by the auditory system is noise. Figure 2A shows that the noise is assumed to be generated from a single source, i.e., all CAVs fall into the auditory stream, and no segregation occurs.

In Figure 2B the first of the repeating sounds with a combined attribute value of S_x is present. A wide band surrounds the sound, and sounds with CAVs falling within the band are allowed in the auditory stream with S_x. As S_x repeats for some duration, the range of allowable CAV differences decreases (Figure 2C). This has the effect of suppressing noise for a repeating sound.

The description may be extended to show the case of a train of two alternating sounds, a reference sound (S_x) and a second sound (S_y), forming a sequence of S_x, S_y, S_x, S_y, ... In this case, the Y-axis represents the probability that an S_y, at a particular moment, is similar enough to the preceding S_x to have arisen from the same source, given all previous S_x and S_y. Simply put, the Y-axis shows the likelihood of the two sounds being integrated into the same stream at a given point in time. The distance between CAVs on the X-axis reflects the attribute differences between the two sounds. The shaded regions of Figures 3A and B indicate the range of CAVs reaching a criterion probability of arising from a single source. Initially, as in the single-sound case, the range of allowable CAVs is large, and the sounds associated with S_x and S_y are assessed as probably coming from the same source (Figure 3A). The listener would hear a sequence of alternating sounds at this point in time. Only sounds vastly different from the sound corresponding to S_x will

be segregated from the sound. However, the CAV range of the auditory stream decreases over time, and the likelihood that S_y arose from the same source as S_x decreases accordingly (Figure 3B). At this point, the listener would hear two perceptually independent sequences, one containing the repeating sounds corresponding to S_x, the other corresponding to S_y. Only sounds having attributes closely similar to S_x would be included in the stream with S_x.

In summary, the time-course of stream segregation, as described here, allows the discussion of auditory scene analysis as two hypothetical processes. The first process is the setting of perceptual distance between sounds in the acoustic environment. This is accomplished by comparing the CAV of each sound in the environment. The second step reduces the range of allowable CAV values, eliminating sounds unlikely to have arisen from the reference source. This effectively partitions the CAV axis: sounds falling on opposite sides of the partition are segregated from each other.

Norman's (1967) data can be explained on the basis of this framework. He used a sequence of ten alternating tones, "with frequency separation well within Miller and Heise's trill threshold" (p 296), and a probe presented once in the sequence. The frequency of a probe tone was varied below, between, and above the frequencies of the tones in the sequence. Subjects were required to report the order of the probe and surrounding tones (e.g., low tone, probe, high tone). The correct identification of the position of the probe in a sequence occurred only when the frequency of the probe was between the frequencies of the tones in the sequence. [Note:

Other investigators have also observed that the temporal order of segregated sounds is difficult to accurately report. See, for example, Bregman and Campbell (1976), and Bregman (1978b).].
In terms of the previous description of auditory scene analysis, an auditory stream was formed, spanning the range of CAVs of the tones in the sequence. The CAV of the probe, however, exceeded this range when the frequency separation of the sequence tones and probe was large. As a consequence, the probe was segregated from the sequence, and the temporal relationship between the probe and tone sequence was lost. In contrast, when the probe frequency was between those of the tones in the sequence, the CAV of the probe was within the range of the auditory stream. Segregation of the probe from the sequences did not occur, and the temporal relationship between the sequence tones and the probe was maintained.

The attribute values of individual sounds are static with respect to time. Changes in perceptual organization over time observed in earlier studies are not due to changes in the dissimilarities of the sounds in the environment, but are the result of decreasing the allowable CAV differences. That is, it is the range of CAV values encompassed by an auditory stream which decreases over the duration of an auditory event.

As an example of the second hypothetical process, let sequence A consist of a 1000 Hz pure-tone alternating with a 1750 Hz pure-tone. Let sequence B consist of the same tones; however, the 1750 Hz pure tone has an intensity 12 dB higher than the 1000 Hz tone. According to the model proposed here, the high- and low-frequency

tones in sequence B should segregate more rapidly than those in sequence A, even though the sounds in each sequence have the same frequency separation. The CAVs of the sounds in sequence B should be perceptually distant compared to the CAVs of sequence A. In other words, the elements of sequence B should be farther apart on the CAV axis than those of sequence A, since the elements in sequence B differ in more than one attribute. An orderly reduction in the range of CAVs would separate the high-frequency/high-intensity tones from the low-frequency/low-intensity tones in sequence B sooner than the separation of the high-frequency and low-frequency tones in sequence A.

To substantiate the idea that the minimum interval of time necessary to segregate two sounds is directly related to the degree of similarity of the two sounds, it is necessary to re-examine the time-course of perceptual organization. French-St. George and Bregman (1989), and Anstis and Saida (1985) show that perceptual organization of sequences changes over time. Bregman (1978a) indicates that segregation is accomplished more easily as the duration of a sequence increases. These studies, however, show segregation changing over a period of 2-30 seconds. Everyday experience suggests that stream segregation can occur much more rapidly. One only has to listen to speech mixed with a background noise to realize that changes in the perceptual organization of the acoustic environment (the segregation of the speech from the noise) can occur over a much shorter interval than previous research suggests. Furthermore, there currently is no research addressing possible changes in the perception of acoustic events lasting less

than two seconds. A possible basis for rapid stream segregation is that, in many cases similar to the speech-in-noise example, the constituent sounds of a given acoustic environment differ in more than one attribute.

Aside from frequency and intensity, a characteristic of sounds on which segregation might be based is temporal envelope. In addition to Bregman, Levitan, and Liao (1990), Bregman *et.al.* 1985, and Carrell and Opie (1992), other research has shown that intensity fluctuations of sounds have an effect on auditory perception. One example is comodulation masking release, wherein a lower threshold is obtained for a tone when a masker and spectrally-removed flankers fluctuate in synchrony than when only the masker and tone are present (Moore, Hall, Grose, and Schooneveldt, 1990; Hall, 1986; Hall, Haggard, and Fernandes, 1984). Modulation detection interference (Yost and Sheft, 1989; Yost, Sheft, and Opie, 1989) results when the amplitude-modulation of one sound reduces the detectability of modulation (at a similar rate) in a spectrally-distant target. These examples suggest that, in certain cases, the percept evoked by a complex sound depends on the relationships of its amplitude-fluctuating constituents to each other.

The auditory system appears to make use of temporal envelope information in the perceptual organization of simultaneous sounds when the information is available. It might be possible for the auditory system to use temporal envelope information in the perceptual organization of sequential sounds as well. In addition, contrasting short-term fluctuations in intensity may be utilized differently by the perceptual system than are intensity

dissimilarities over the entire duration of sounds (i.e., sequences in which alternating sounds differ in intensity). Therefore, the effectiveness of temporal envelope correlation on the time-course of stream segregation will be explored in this investigation.

The theoretical framework for stream segregation, as proposed here, consists of three components. First, segregation is based on the dissimilarity of sounds in multiple perceptual dimensions. Second, stream segregation does not occur instantaneously, rather, segregation operates on an acoustic event of some finite duration, and is, in a sense, limited by the duration of that event. The third component of the hypothesis to be tested in this study is related to the first two considerations. That is, the period necessary for stream segregation is directly related to the similarity of the sounds in an acoustic event. This project investigated all three components of the proposed hypothesis.

In order to obtain as unbiased data as possible in this study, there is a need to use a relatively objective measure of stream segregation, hopefully leaving as little as possible open to interpretation by the subjects. Rather than explain stream segregation and sequential integration to subjects, it would be more desirable to use a response mode with a physical, not perceptual, reference. This is because, under some circumstances, reports of segregation are dependent on the bias of the listener. Van Noorden (1975) showed that the frequency-difference threshold of stream segregation could differ depending on the instructions to listeners. If listeners are instructed to indicate the minimum frequency separation necessary for *segregation*, the frequency separation is

smaller than the maximum separation that allowed the subjects to hear the tones as sequentially integrated. Informal listening conducted previously in this laboratory suggests that the perceptions of segregation and integration are not always dichotomous: listeners can consciously modify their perceptions. It is assumed that naive subjects, using a response other than a segregated/not segregated distinction, would avoid the bias reported by van Noorden and observed by others.

Research conducted by Dannenbring and Bregman (1976) suggests a methodology for studying segregation that avoids direct judgments of stream segregation. Using sequences of alternating pure-tones, they found that as the frequency separation of the tones was increased, subjects no longer perceived the sequences as strictly alternating. Listeners heard the tones from one frequency range overlap in time with the remaining tones. Further analysis revealed that judgments of overlap were highly correlated with judgments of segregation using the same stimuli in separate tasks. The perception of overlap might be a useful criterion for listeners who have no previous experience in perceptual organization experiments.

If judgments of overlap and separation are used as responses, it is possible to use sequences with the elements physically separated by a small period of time. Perception of the elements as overlapping could then be interpreted as a consequence of segregation of the alternating sounds. A potential difficulty with this stimulus configuration is that gap discrimination varies as a function of frequency separation (Fitzgibbons, Pollatsek, & Thomas,

(1974), Collyer (1974), Perrot and Williams (1971)). A finding common to these studies is that gap detection decreases as the frequency difference between sounds surrounding the gap increases. Any results found using stimuli arranged in the above manner might be contaminated by differences in the differential ability of the auditory system to detect gaps as a function of frequency separation. For example, the temporal separation between the tones with a small frequency difference might be more striking than the gap between tones with larger frequency separations. This might bias listeners against reporting elements with large frequency separations as being separated in time. As a result, subjects might report more overlap in sequences with larger frequency separations due to the absence of a perceptible gap. One solution would be to use a temporal separation between successive sequence elements below the gap threshold for the smallest frequency difference. However, since threshold for gap discrimination is usually equated with 70% - 80% correct, the processing of the gap between successive sequence elements might still be better (although less than optimal) for sequences with smaller frequency separations than for larger frequency separations. To avoid this problem, a condition can be included where no gap, and alternatively, a physical overlap, is present in the sequences. A comparison of the responses to the two types of sequences (where the sounds are physically separated or are physically overlapped) can be used to examine the use of gaps in

the sequences to judge overlap and separation. If the pattern of responses to the physically separated and physically overlapped sequences are the same for all frequency separations, then this would argue against contamination of the results by gap detection.

GENERAL METHODS

The experiment designed to test the hypotheses consisted of four conditions (see Table 1). Data collection was conducted for the four conditions simultaneously. While the specific stimuli were unique to each condition, the sequences used in each were constructed in the same way.

The first condition in the experiment examines the time-course of segregation for stimuli differing along a single perceptual dimension, namely, pitch. The second condition probes segregation as a function of differences in the two perceptual dimensions of pitch and loudness. The effect of the physical characteristic of temporal envelope similarity is considered in Conditions 3 and 4. Amplitude-modulated sounds with identical temporal envelopes and differing center frequencies are used in Condition 3, while Condition 4 employs amplitude-modulated sounds with dissimilar temporal envelopes and center frequencies. The effects of variation of a single attribute are examined in Conditions 1 and 3. Conditions 2 and 4 were designed to examine the time-course of stream segregation with differences in two stimulus attributes. A summary of the conditions is provided in Table 1.

Subjects

Eleven adult subjects with normal peripheral hearing sensitivity participated in the study (see Appendix A). They were recruited from the faculty and students of the Department of

Communication Sciences and Disorders, LSU. No subject had previously served as a subject in perceptual experiments.

Table 1. Experimental conditions

Condition	Type of sound	Differences
1	Pure-tones	Frequency
2	Pure-tones	Frequency, intensity
3	AM noise-bands	Center frequency
4	AM noise-bands	Center frequency, temporal envelope

Stimuli

Each experiment employed sequences of sounds varied on specific parameters. A sequence can be considered as a repetition of "pairs" of stimuli. A stimulus "pair" refers to two adjacent sounds in a sequence. For example, the first and second sounds formed a pair, as did the third and fourth, fifth and sixth, etc. Each sequence consisted of repetitions of a stimulus pair. The members of a stimulus pair differed either in frequency, frequency and intensity, center frequency, or center frequency and temporal envelope, depending on the experimental condition.

Two types of stimulus pairs were used to create the sequences (see Figure 4). Separated pairs consisted of higher- (1250 Hz, 1750 Hz, or 2250 Hz) and lower-frequency (1000 Hz) sounds. The duration of each member of a pair was 123 msec, with a 2-msec interval between the members of a pair. Thus, the interval from the onset of the higher-frequency sound to the offset of the lower-frequency

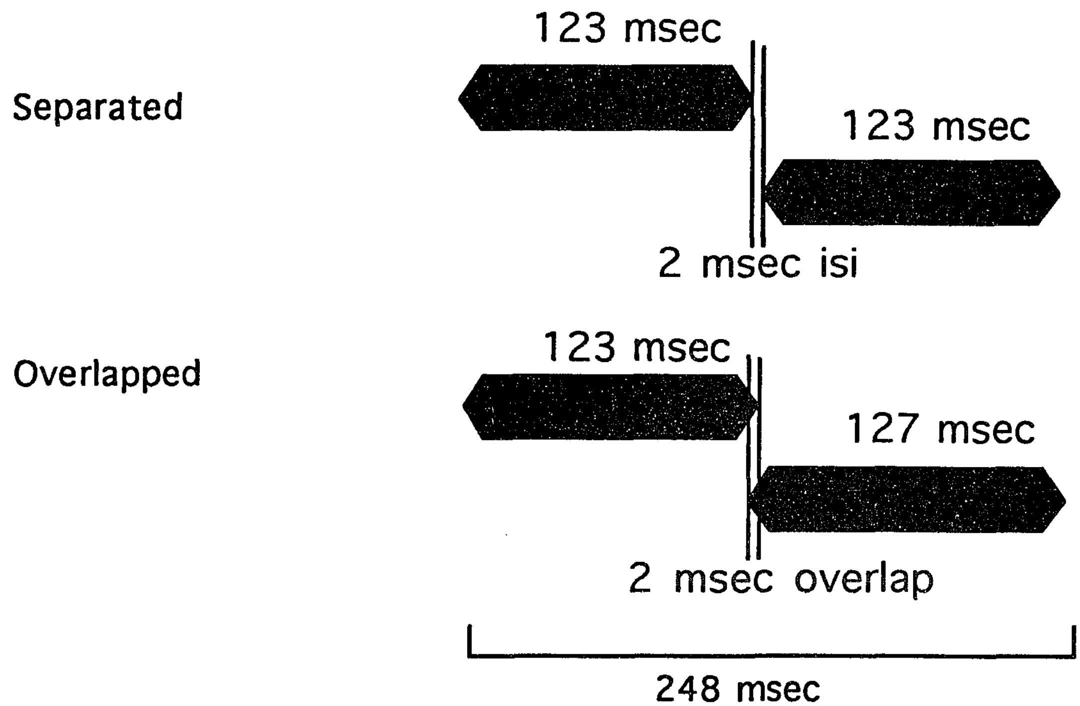

Figure 4. Signal arrangement showing separated and overlapped pure tone pairs.

sound was 248 msec. Overlapped pairs also were made up of a 123-msec higher-frequency sound, however, the lower-frequency sound was 127 msec in duration. The onset of the lower-frequency sound occurred 2 msec before the offset of the higher-frequency sound; thus, the sounds in overlapped stimulus pairs overlapped by 2 msec. The interval from the onset of the higher-frequency stimulus to the offset of the lower-frequency stimulus was 248 msec, as in the separated pairs. For both types of sequences (containing either separated or overlapped pairs), a 2-msec interval separated repetitions of the stimulus pairs. Stimulus pairs were repeated to generate sequences of 500, 1000, 1500, 2000, 2500, and 3000 msec in duration. The combination of 6 durations x 4 conditions x 3 frequency separations x 2 classes of sequences (separated and overlapped) resulted in 144 unique sequences.

Pure tone stimuli: Under conditions using pure tone sequences with constant envelopes, the stimulus pairs consisted of higher- and lower- frequency pure tones; higher and lower defined as above. The pure tones were generated (10 kHz digitization rate) using locally developed software and stored in digital form. Linear rise- and fall-times were 10 msec. Conditions 1 and 2 employed the sequences made of pure tones (pure-tone, 0 dB ΔI and pure-tone, 12 dB ΔI sequences, respectively). The stimuli used in Condition 2 were the same as those used in Condition 1. The higher-frequency tones in the sequences of Condition 2, however, were made 12 dB more intense than the lower-frequency tones by decreasing attenuation during presentation using a computer-controlled attenuator (MI^2 108).

AM noise bands: Under conditions using narrow-band noises with time-varying envelopes, the stimulus pairs were constructed using sounds with the same center frequencies as those in the pure tone sequences. Random noise was low-pass filtered at 16 Hz (Wavetek Rockland Brickwall) and converted to digital form at a 10 kHz conversion rate. A 127-msec segment of the noise with minimal onset and offset transients was extracted. This segment was used to create the lower-frequency sound for overlapped stimulus pairs. The final 123 msec of this segment was used to create the lower-frequency sound for separated stimulus pairs. It was also used to create the higher-frequency sounds in stimulus pairs where the envelope correlation of the high- and low-frequency sounds was 1.0 (used in correlated AM noise-band sequences).

A second segment was selected from the original noise sample such that the correlation coefficient with the previously obtained 123-msec noise segment was 0.0. A second set of higher-frequency sounds were created using this second segment. When alternated with the lower-frequency sound created for the correlated AM noise-band condition, sequences were generated where the envelope correlation of the higher- and lower-frequency sounds was 0.0 (the uncorrelated AM noise-band sequences).

The noise segments were used to amplitude-modulate carriers at the previously mentioned pure-tone frequencies with a modulation index of 1. The resulting stimuli were narrow-bands of noise 32 Hz wide. These signals were then linearly ramped (10 msec rise/fall times).

Instrumentation

The sounds making up each sequence were stored digitally. These were converted to analog form (MI^2 108), with the high- and low- frequency sounds in a sequence converted by separate digital-to-analog converters (DACs). The output of each DAC was connected to a computer-controlled attenuator (MI^2 208). The output of each attenuator was then routed into a custom mixer, and the resulting signal was low-pass filtered (Wavetek Rockland 752A Brickwall, 115 dB/octave attenuation slope) at 4900 Hz. The sequences were again attenuated (Hewlett-Packard 350D), and output to headphones (Sennheiser HD 430) for diotic presentation.

Level

The output level of each sound used to construct the sequences was adjusted using computer-controlled attenuation (MI^2 208). This resulted in each sequence measuring 58 dB SPL (Bruel & Kjær Type 2609, fast setting) at the output of one headphone during the experimental sessions. For the sequences used in Experiment 2 (where the sounds differed in frequency and intensity), the attenuation of the higher-frequency sound was decreased by 12 dB. This was accomplished using the computer-controlled attenuator connected to the DAC converting the higher-frequency sound.

Task familiarization

Subjects were allowed to hear an alternating pair of separated pure-tones (123 msec in length, 2-msec ISI, $(F_H-F_L) / F_L = 9.625$, 0 dB intensity difference) presented continuously. Due to the

frequency separation of the tones and sequence rate, segregation of these tones is likely (van Noorden, 1975). Subjects were permitted to continue the experiment if they described the tones in the sequence as overlapped. Only one potential subject (Subject 8) was eliminated on this basis. Ten subjects were allowed to participate in the experiment. As mentioned earlier, prior to data collection all subjects completed a practice block of trials. The results from these trials were not used as experimental data.

Procedure

Subjects were tested under headphones in a single-walled listening chamber. Timing, signal output, and attenuation were under computer control. The subjects received instructions and responded via a response terminal also controlled by the computer.

Before testing began, each subject read instructions on the response terminal. The instructions stated that they would hear sequences of sounds, and that their task would be to decide whether the sounds in the sequence "do not overlap, and sound separate" or "sound overlapped." Subjects were also instructed to listen to a sequence in its entirety before making a decision; responses were not accepted by the computer until a given sequence was terminated. The subjects indicated their decision by pressing one of two keys on the response terminal. An experimental trial proceeded as follows: ready prompt, 200-msec interval, sequence presentation, 200-msec interval, response prompt. The response interval was unlimited, and at least five seconds of silence separated each trial to avoid any cumulative effects of perceptual organization (Bregman, 1978a). No

feedback was given. Each subject completed six blocks of trials; data from the first block were discarded. Every combination of sequence type (separated and overlapped), condition, duration, and frequency separation was randomly presented during a block. A subject's calculated d' (the z-transformation of the hit rate - the z-transformation of the false alarm rate) and criterion measure (-0.5 * [the z-transformation of the hit rate + the z-transformation of the false alarm rate]) were averaged across the last five blocks of trials. The d' measure was an indication of the discriminability of the overlapped and separated pairs. A d' value near 0.0 would indicate that subjects could not discern a difference between the physically overlapped and physically separated pairs. The criterion measure indicated the partition of the subject's responses. Due to the number of trials, the criterion measure ranged from 2.5 to -2.5 (see Appendix B for calculation of d' and criterion measure). Positive criterion values indicated sequential integration of the higher-frequency sequence elements and the lower-frequency sequence elements, while negative criterion values indicated stream segregation.

RESULTS

Although ten subjects completed the experiment, the performances of two subjects (6 and 10) were sufficiently different from the remaining subjects to warrant separate discussion. As an overall index of performance, each subject's criterion measure (-0.5[z(hit rate) + z(false alarm rate)]) was averaged across conditions, frequency separations, and sequence durations. Two post-hoc subject groups were formed on the basis of this index. Subjects 1, 2, 3, 7, and 11 formed a high-segregation group (mean criterion measure less than 0.0), while Subjects 4, 5, and 9 formed a low-segregation group (mean criterion measure greater than or only slightly less than 0.0). The mean criterion measures are shown in Figures 5 and 6. A one-tailed t-test indicated a significant difference between these two groups ($t = 8.45$, $df = 549.2$, $p < 0.01$ [see Appendix C for details]). Criterion measures and d' [z(hit rate) - z(false alarm rate)] were analyzed for the two groups separately using a multifactor (number of subjects x 4 x 6 x 3), mixed-effects analysis of variance.

High-segregation group

For the criterion measure, the factors condition, frequency separation, and sequence duration were significant beyond the 0.05 level. In addition, a significant interaction of condition and frequency separation was observed. No factor was significant on the d' measure (see Table 2).

Newman-Keuls post-hoc tests of significant differences were carried out for the significant criterion main effects of the high-segregation group (Appendix D). In terms of sequence duration, the 500 msec sequences were significantly different from sequences 1500 msec - 3000 msec in duration. There was no difference between 500 msec and 1000 msec sequences. 1000 msec sequences differed from the 3000 msec sequences to a significant extent. Also, there were no significant differences between sequences 1500 msec - 3000 msec in duration. The mean criterion measure for each sequence duration is shown in Figure 7. In general, as sequence duration increased, stream segregation increased.

For the high-segregation group, the mean criterion for the 1000 Hz - 1250 Hz sequences was significantly different ($p < 0.05$) from both the 1000 Hz - 1750 Hz and the 1000 Hz - 2250 Hz sequences. Figure 13 shows that the mean criterion for the 1000 Hz - 1250 Hz sequences are above the criterion for the other sequences, indicating less segregation for the smaller frequency separations than for the larger frequency separations. No significant difference was found between the mean criterion for 1000 Hz - 1750 Hz sequences and 1000 Hz - 2250 Hz sequences.

The criterion measure for the correlated AM noise-band sequences was significantly different from the responses to each of the other conditions. Inspection of the mean criterion for each condition showed the average criterion for this condition to be above the mean criterion for each of the other conditions (Figure 18). This indicated less segregation of the elements in the correlated AM

noise-band sequences compared to the other conditions. No other differences between the conditions were significant.

Low-segregation group

No significant main effects or interactions were observed for this group on the criterion measure. The d' measure also did not reach significance for any factor (Table 3).

Table 2. Anova tables for high-segregation group

High-Segregation Group: Criterion Measure

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUMS OF SQUARES	MEAN SQUARES	F-RATIO	PROBABILITY LEVEL
Condition	3	14.28173	4.76058	5.86730	.01063*
ERROR	12	9.73649	.81137		
Frequency separation	2	309.60333	154.80170	10.32624	.00647**
ERROR	8	119.92879	14.99110		
C x F	6	4.51851	.75308	2.76316	.03433*
ERROR	24	6.54107	.27254		
Duration	5	117.06889	23.41378	6.24807	.00149**
ERROR	20	74.94721	3.74736		
C x D	15	2.14687	.14312	.69412	.78057
ERROR	60	12.37182	.20620		
F x D	10	6.28823	.62882	1.00098	.45963
ERROR	40	25.12831	.62821		
C x F x D	30	6.39932	.21331	1.20495	.23773
ERROR	120	21.24347	.17703		
TOTAL	359	776.38129			

* SIGNIFICANT BEYOND THE 5-PERCENT LEVEL
** SIGNIFICANT BEYOND THE 1-PERCENT LEVEL

High-Segregation Group: d'

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUMS OF SQUARES	MEAN SQUARES	F-RATIO	PROBABILITY LEVEL
Condition	3	6.59490	2.19830	3.30765	.05684
ERROR	12	7.97533	.66461		
Frequency separation	2	.72202	.36101	.37290	.70364
ERROR	8	7.74494	.96812		
C x F	6	1.37651	.22942	.48722	.81222
ERROR	24	11.30091	.47087		
Duration	5	1.35134	.27027	.49420	.77824
ERROR	20	10.93750	.54687		
C x D	15	17.23155	1.14877	1.63347	.09150
ERROR	60	42.19615	.70327		
F x D	10	4.30128	.43013	.43098	.92251
ERROR	40	39.92086	.99802		
C x F x D	30	20.2106	.67369	1.19532	.24666
ERROR	120	67.63266	.56361		
TOTAL	359	240.94471			

Table 3. Anova tables for low-segregation group.

Low-Segregation Group: Criterion Measure

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUMS OF SQUARES	MEAN SQUARES	F-RATIO	PROBABILITY LEVEL
Condition	3	13.33550	4.44517	1.07271	.42950
ERROR	6	24.86320	4.14387		
Frequency separation	2	47.75490	23.87745	1.71543	.28977
ERROR	4	55.67693	13.91923		
C x F	6	4.11476	.68579	2.55483	.07852
ERROR	12	3.22116	.26843		
Duration	5	24.86673	4.97335	2.33356	.11899
ERROR	10	21.31227	2.13123		
C x D	15	4.31886	.28792	1.26668	.28098
ERROR	30	6.81916	.22731		
F x D	10	3.09744	.30974	1.33956	.27597
ERROR	20	4.62458	.23123		
C x F x D	30	8.21022	.27367	1.27497	.20934
ERROR	60	12.87909	.21465		
TOTAL	215	252.37820			

Low-Segregation Group: d'

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUMS OF SQUARES	MEAN SQUARES	F-RATIO	PROBABILITY LEVEL
Condition	3	.60348	.20116	.36231	.78416
ERROR	6	3.33130	.55522		
Frequency separation	2	3.45742	1.72871	2.15656	.23150
ERROR	4	3.20643	.80161		
C x F	6	3.36755	.56126	1.02147	.45727
ERROR	12	6.59351	.54946		
Duration	5	5.78065	1.15613	1.19074	.37946
ERROR	10	9.70931	.97093		
C x D	15	14.37140	.95809	.96800	.50875
ERROR	30	29.69299	.98977		
F x D	10	7.08208	.70821	.59588	.79932
ERROR	20	23.77030	1.18852		
C x F x D	30	31.42729	1.04758	.93516	.56929
ERROR	60	67.21270	1.12021		
TOTAL	215	218.37343			

DISCUSSION

The purpose of this experiment was to (1) investigate short-term changes in the segregation and sequential integration of sequences; (2) examine stream segregation in sequences with differences in single and multiple attributes; and (3) study the effects of attribute differences on the time-course of perceptual organization. It was hypothesized that the duration of a sequence necessary for stream segregation relates directly to the similarity of the sounds in the sequence.

In the present study, a criterion measure of 0.0 indicates that the proportion of "separated" and "overlapped" responses is equal. When criterion is equal to 0.0, across trials, the perception of a particular sequence as sequentially integrated or segregated is equally likely. The shift from a positive to a negative criterion measure indicates a change from sequential integration to stream segregation.

The criterion measure used in this study was observed to be variable across subjects. Figures 5 and 6 show the mean criterion and standard error averaged over frequency separations, conditions, and durations for the high- and low-segregation subjects. While the main effects of frequency, condition, and duration were not significant for the three low- segregation subjects, their data are presented along with the high-segregation group for comparison.

The high inter-subject variability observed in this study is not unusual. Neff, Jesteadt, and Brown (1985) attempted to relate gap discrimination and stream segregation by using the same stimuli in

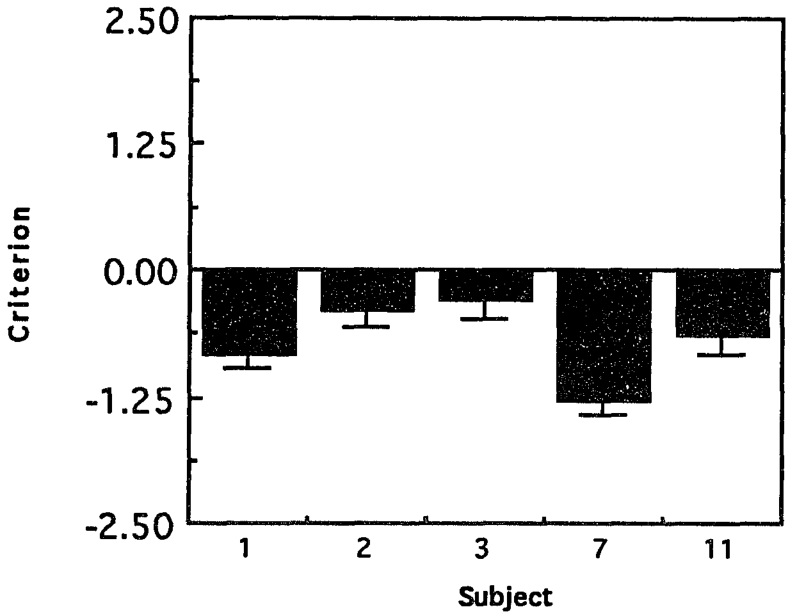

Figure 5. Mean criterion for subjects in the high-segregation group. Bars in all Figures represent 1 standard error.

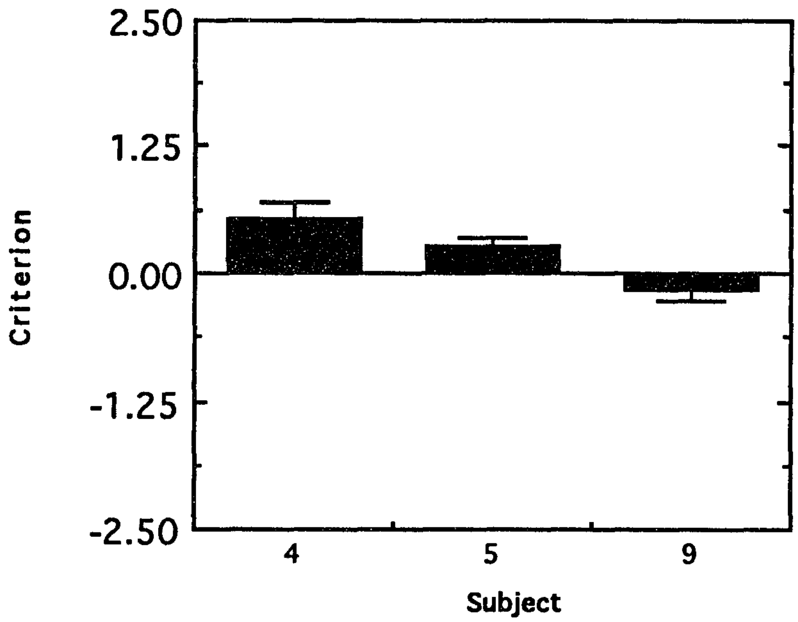

Figure 6. Mean criterion for subjects in the low-segregation group.

tasks designed to measure each phenomenon. In the gap discrimination task, the performance of their four listeners was fairly homogenous. Performance in the stream segregation task, however, was quite different between subjects.

Other investigators of stream segregation have used various approaches in dealing with inter-subject variability. In one study of stream segregation, Bregman and Rudnicky (1975) simply eliminated a total of 18 out of 31 potential subjects. In another experiment, Bregman (1978a) rejected 1 out of 12 listeners. From an initial pool of 28 subjects, Bregman (1978b) dismissed 7 who could not perform a screening procedure. Bregman, Abramson, and Doehring (1985) rejected 1/3 of their potential listeners. Starting with 32 subjects, Massaro (1975) eliminated 2, and divided the remaining 30 subjects into 3 groups based on their overall performances. Because the data from the present study also varied across subjects, dividing the subjects into two groups seemed to be a reasonable approach. In this way, trends might be observed that would otherwise be obscured by the inter-subject differences.

Two subjects of the original ten were excluded from the data analysis. The performances of Subjects 6 and 10 were unlike either the high- or low-segregation groups. Criterion measures for these two subjects are shown in Appendix E. The striking aspect of these data is that the responses seem to be the opposite of what might be expected based on previous research.

After all subjects had completed the study, these two subjects were recalled and separately completed another half-block of trials. Afterwards, the subjects were asked how they arrived at their

responses. Subject 6 stated that she responded "separate" when the high- and low-frequency sounds appeared to be made by separate things. What this subject reported was, in fact, a good description of the perceptual effect of stream segregation. It appeared that the subject experienced the same percepts as the majority of the subjects, but due to a misunderstanding of the instructions, used a response opposite to the other subjects. This explanation was more convincing since the subject's criterion was reversed relative to the majority of subjects for each manipulation designed to increase stream segregation. The responses of the second excluded subject are more difficult to explain. Subject 10 stated that for large frequency separations, and the longer sequences, the sounds appeared to be more discrete, as if they had nothing to do with one another. Recalling the Dannenbring and Bregman (1976) research relating stream segregation and the perception of overlap, the correlation between these two factors was not perfect.

The lack of complete correlation between factors intended to induce stream segregation and perceived overlap was also evident in the pilot phase of the present study. Five listeners participated, of which, four reported increased overlap in response to manipulations predicted to increase segregation. One subject, however, reported the same perceptual experience as Subject 10 (described above). It would seem that, in general, not all listeners hear overlap when stream segregation occurs.

Despite the inter-subject variability, the data from the majority of subjects (1, 2, 3, 7, and 11) will be used to test the hypothesis presented earlier. In the discussion that follows, each

aspect of the hypothesis will be examined in light of the data. In addition, the description of the segregation process forwarded earlier will be considered in reference to the responses of the subjects.

Effect of sequence duration

Averaged across frequency separations and conditions, the perception of the sequences changes from sequentially integrated to segregated, on average, between 500 and 1000 msec for the high-segregation group (Figure 7). Again, the effect of sequence duration is statistically significant ($p < 0.001$) for only the high-segregation group. Although the effect of sequence duration lacks statistical significance for the low-segregation group, a trend in the data (Figure 8) is apparent.

A change from sequential integration to stream segregation over time is not obvious for all subjects. Subject 1 (a high-segregation subject), for example, shows no clear trend for stream segregation to increase with increasing sequence duration (Figure 9; data were averaged over conditions and durations). The average criterion for Subject 4 (a low-segregation subject) remains positive; however, the criterion does decrease from +1.117 at 500 msec to +0.063 at 3000 msec (Figure 10; averaged data). While it is not entirely accurate to say that Subject 4's stream segregation increases with increasing sequence duration, sequential integration of the sequences decreases over time.

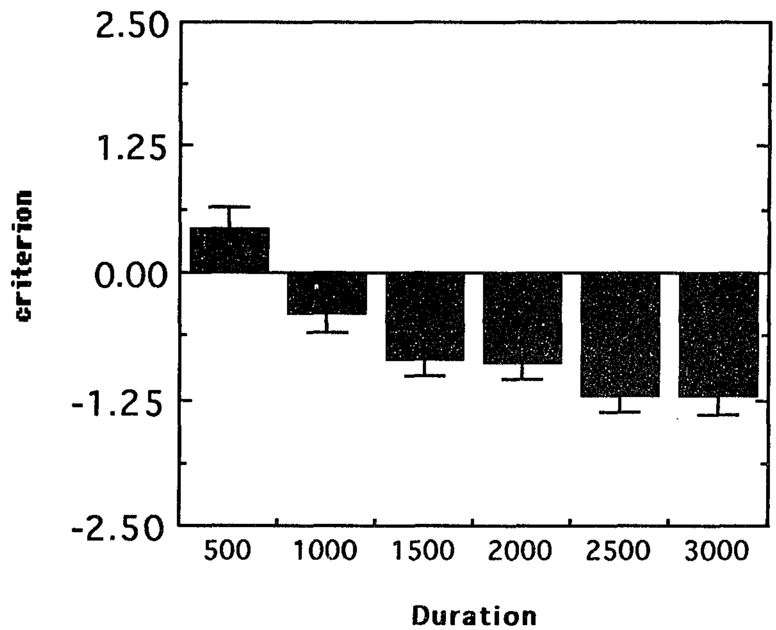

Figure 7. Criterion for the high-segregation group as a function of sequence duration in msec.

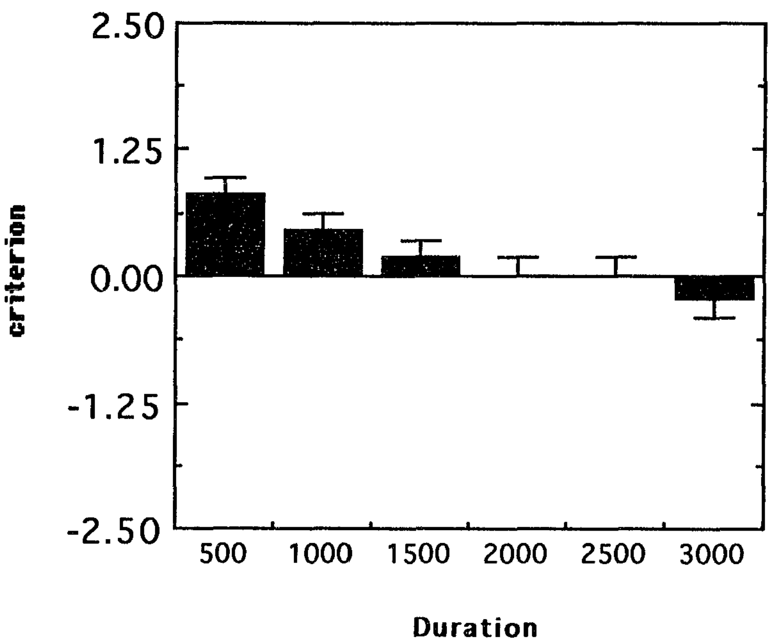

Figure 8. Criterion for the low-segregation group as a function of sequence duration in msec.

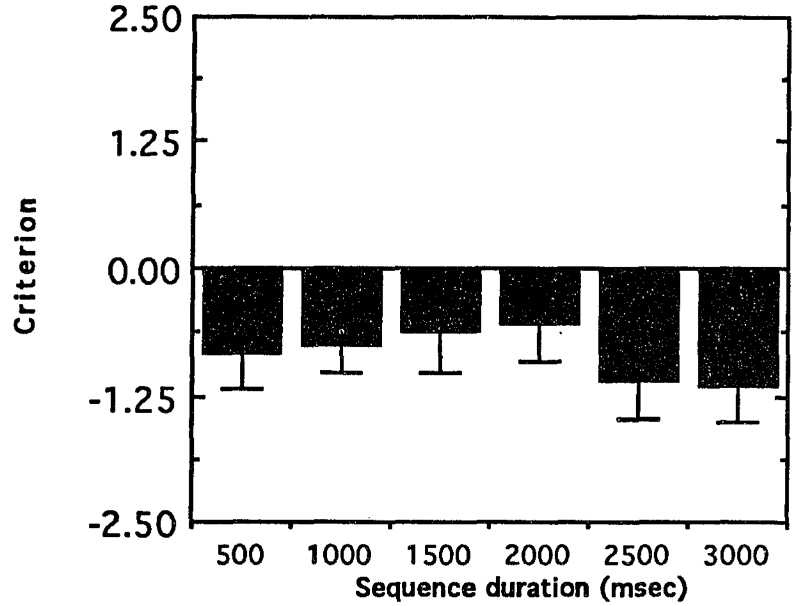

Figure 9. Mean criterion for Subject 1 as a function of sequence duration. Segregation does not increase with increasing duration.

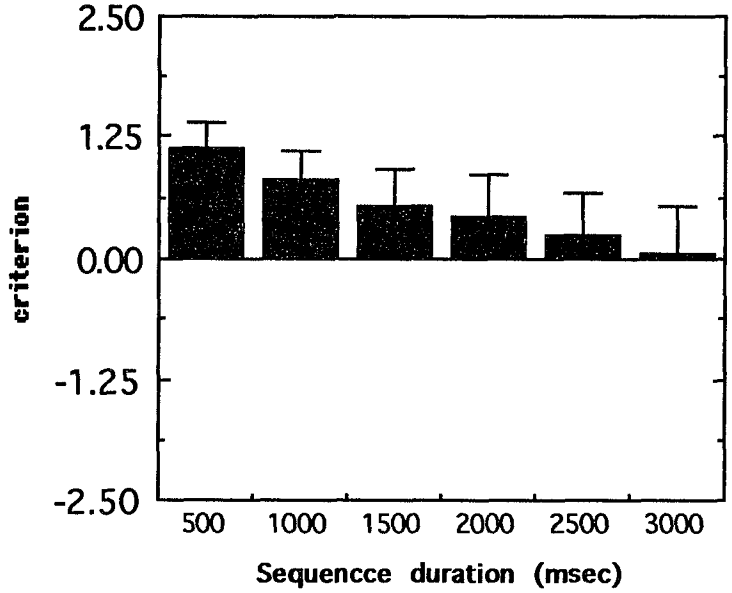

Figure 10. Mean criterion for Subject 4 as a function of sequence duration. Sequential integration decreases with increasing sequence duration for Subject 4.

Anstis and Saida (1985) and French-St. George and Bregman (1989) have demonstrated that increases in sequence duration improves the likelihood of stream segregation. The present data are at least qualitatively like those available in the literature. In the experiment reported here, the sequences made of pure-tones with no difference in intensities are the most comparable to the stimuli used by Anstis and Saida (1985). Their Figure 2 (page 260) shows the probability of coherence (sequential integration) plotted as a function of sequence duration for sequences with a one-half octave frequency separation between elements. When Anstis and Saida's sequences were presented at 8 tones per second (the rate used for the sequences in this dissertation), segregation occurred at approximately 7.5 seconds for Anstis and Saida's (1985) subjects. In the present data, the duration associated with a change in criterion from positive to negative indicated the duration necessary for stream segregation. The data in the present experiment revealed that when pure-tones had a frequency separation of less than one-half octave, they were not heard as segregated at the longest sequence duration examined (3 seconds). This was true for both the high- and low-segregation subjects (Figures 11 and 12). Pure-tones with frequency separations larger than one-half octave segregated in less than 7.5 seconds. In other words, sounds with a smaller frequency separation than that used by Anstis and Saida (1985) did not segregate. Those with frequency separations larger than that used by Anstis and Saida segregated sooner than in the earlier study.

Thus, the basic idea of the first component of the hypothesis stands. For the majority of listeners, perception of the sequences

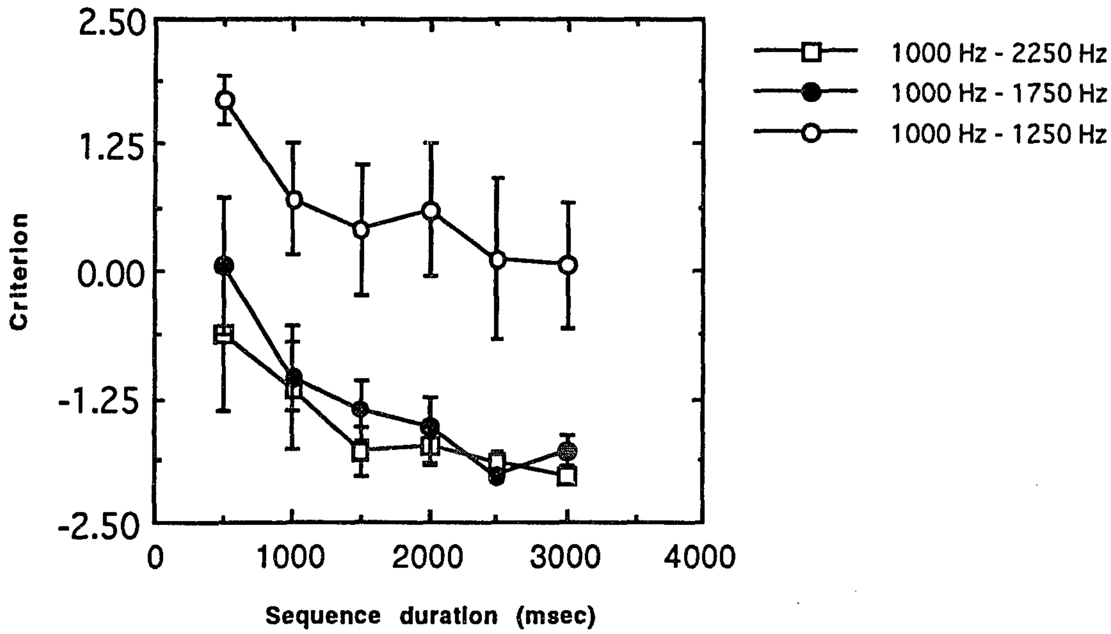

Figure 11. The effect of sequence duration on stream segregation in the high-segregation group. The parameter is frequency separation of the pure-tones in the sequences.

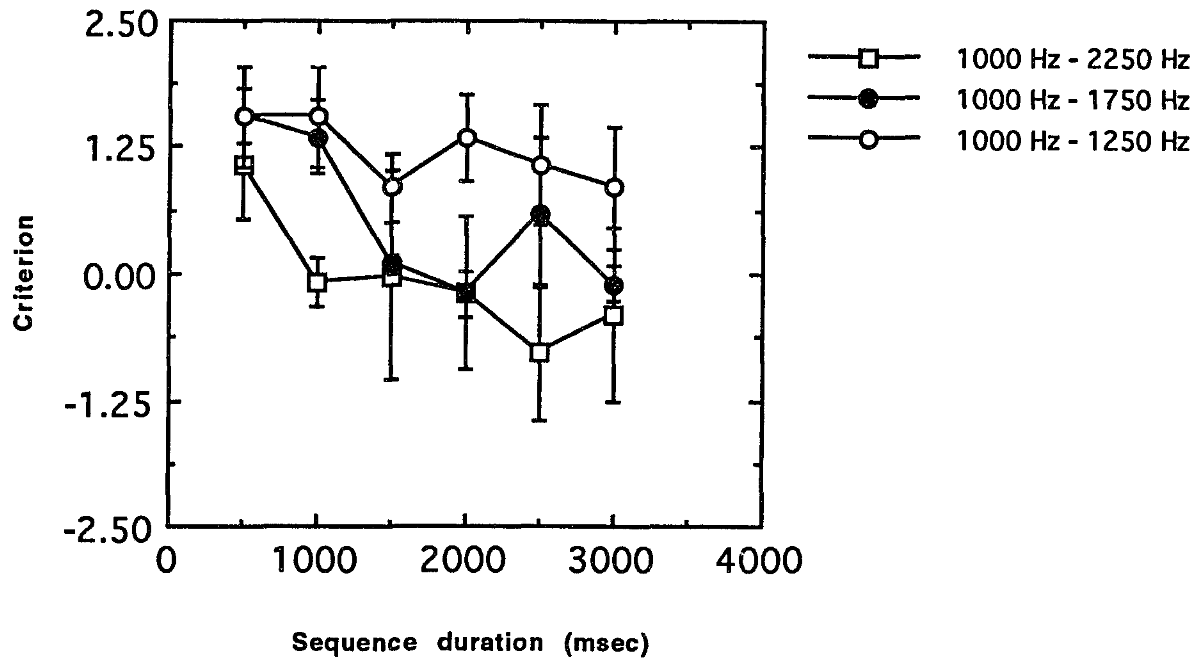

Figure 12. The effect of sequence duration on stream segregation for each frequency separation in the low-segregation group.

tends to favor stream segregation as sequence duration increases. In some cases, this is seen as a reduction in sequential integration over time.

The effect of sequence duration shows that stream segregation can occur at durations less than 3 seconds. In order to more directly address the hypothesis, however, it is necessary to examine the effects of similarity of the sounds in the sequences.

The similarity of the sounds making up the sequences was varied in two ways. The frequency separation between the sounds was varied by using 1000 Hz and 1250 Hz, 1000 Hz and 1750 Hz, or 1000 Hz and 2250 Hz as the frequencies (in pure-tone sequences) or center frequencies (in AM noise-band sequences) for the sounds. The second way of manipulating similarity was a function of condition. Conditions differed on the basis of (a) an intensity alternation of 0 dB or 12 dB in pure-tone sequences and (b) the similarity or difference in temporal envelope for the AM noise-band sequences (see Table 1 in Methods). The manipulations of frequency separation and intensity alternation in the pure tone sequences affected the stream segregation of alternating sequences in previous studies (van Noorden, 1974; 1977). The manipulation of envelope similarity in AM sequences has been shown only to affect the segregation and fusion of simultaneous sounds (Rappold, Mendoza, and Collins, 1993; Carrell and Opie, 1992; Bregman, et al, 1990; Bregman, et al, 1985).

<u>Effect of frequency separation</u>

A significant effect of frequency separation ($p < 0.01$) was noted for the high-segregation subjects (Figure 13). A trend for

segregation to increase with increasing frequency separation was observable for the low-segregation subjects (Figure 14). The trend was not, however, statistically significant.

The concern about the use of the gap in sequences as a contaminating factor (discussed in the Methods section) was apparently unwarranted. There was no difference in responses to the physically separated or physically overlapped stimulus pairs for any frequency separation, as indicated by a d' near 0.0 for each frequency separation, averaged over condition and sequence duration (Figures 15 and 16). That is, sequences containing physically separated and physically overlapped sounds were indistinguishable for all three frequency separations. Thus, any perception of overlap must have been related to stream segregation of the sounds in the sequences.

Averaged across sequence durations and across conditions, the requisite frequency separation for segregation (designated by a criterion measure of 0.0) is between 0.25 and 0.75 octaves in the high-segregation group. The results of van Noorden (1975), however, show that a one-octave frequency separation is necessary for the segregation of pure-tones in sequences presented at 8 tones per second (the rate of the tone sequences in the present study). On the other hand, the data of the low-segregation subjects (averaged across sequence durations and conditions) appear to fall in line with the results of van Noorden (1975). That is, for the low-segregation subjects, the frequency separation necessary for segregation is between 0.75 and 1.25 octaves.

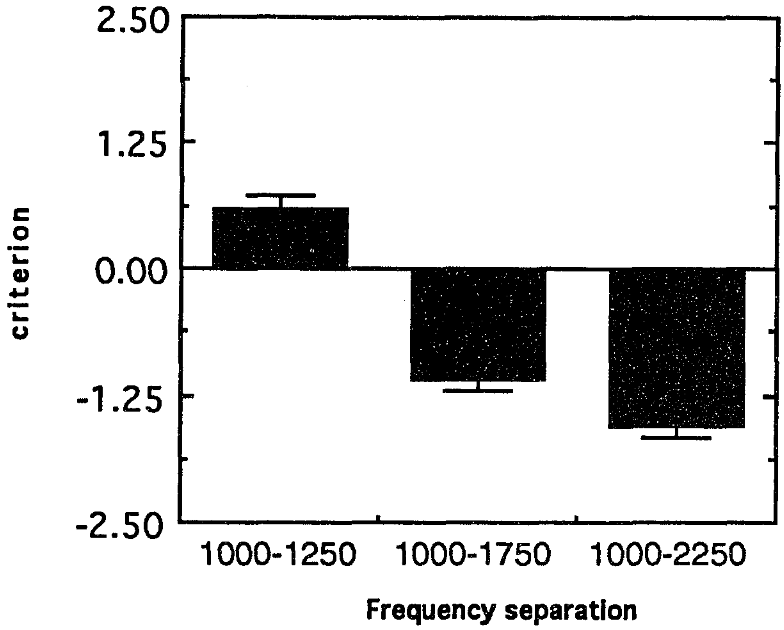

Figure 13. The effect of frequency separation, averaged over conditions and durations, on stream segregation. High-segregation group.

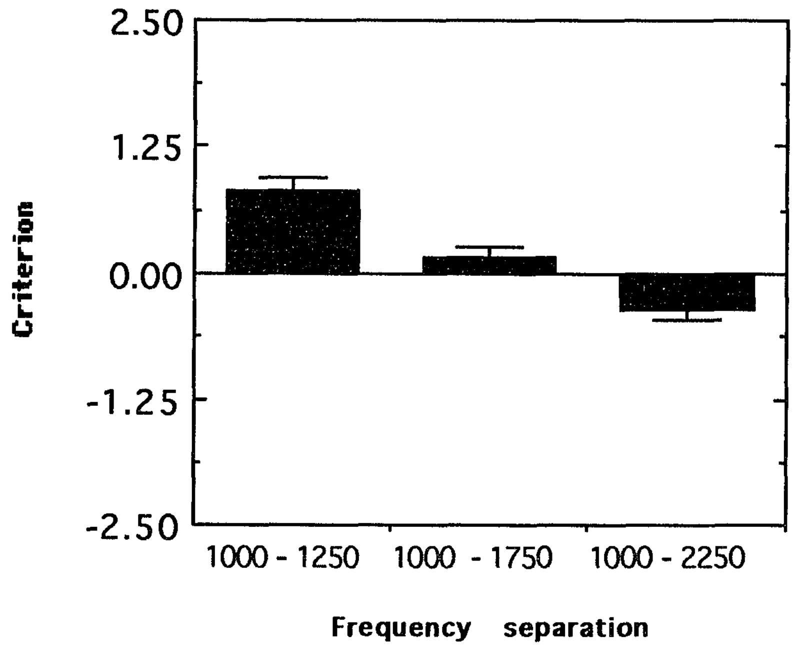

Figure 14. The effect of frequency separation averaged over conditions and durations for the low-segregation group.

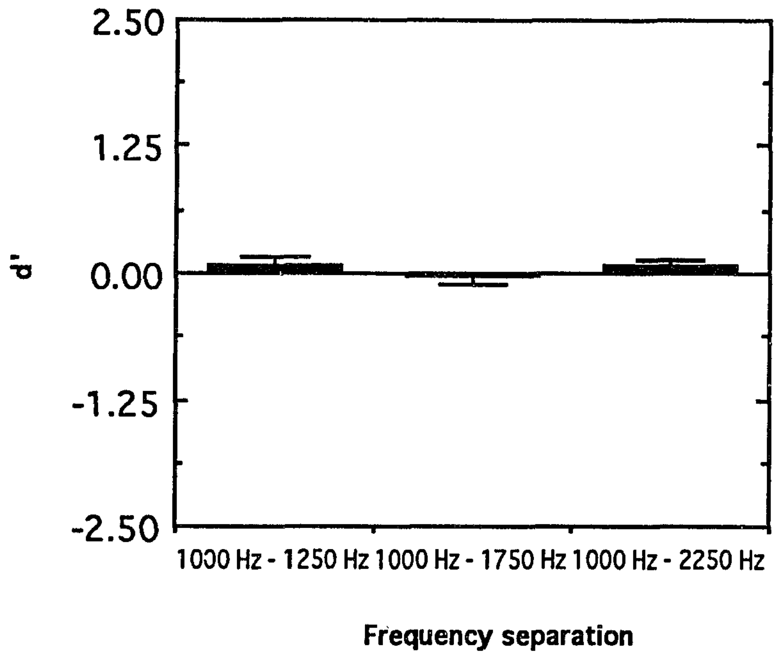

Figure 15. The discriminability index (d') of the physically separated and physically overlapped sequences for the high-segregation group.

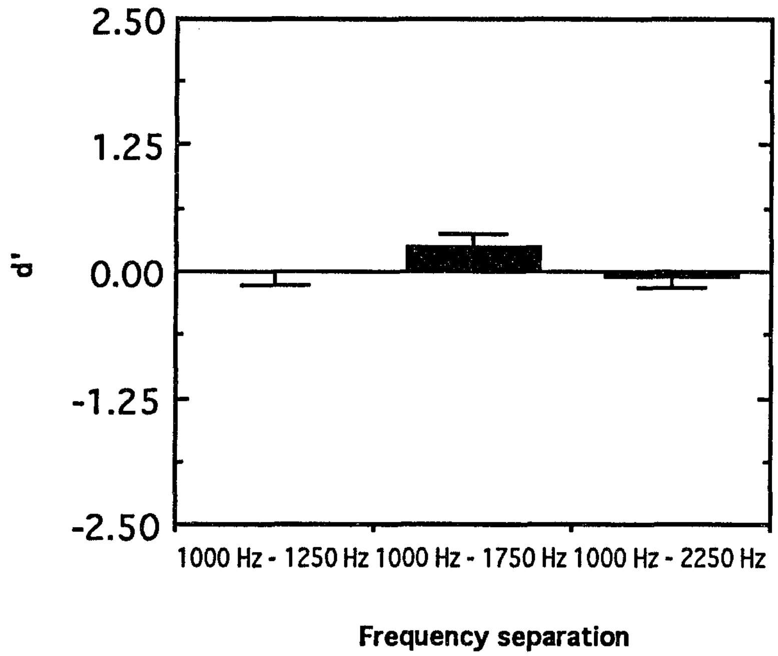

Figure 16. The d' of the physically separated and physically overlapped sequences for the low-segregation group.

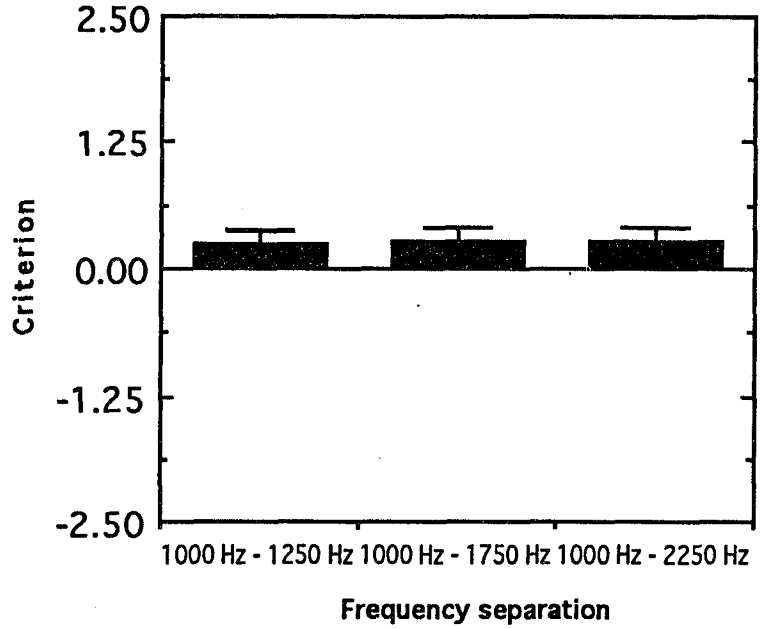

Figure 17. Subject 5 shows no difference in segregation as a function of frequency separation (averaged over conditions and durations).

The subjects for the present study, with few exceptions, showed increased segregation with increased frequency separation. Frequency separation appears *not* to have been a dominant factor for Subject 5 (Figure 17). This is somewhat surprising, given that frequency separation is usually a primary determinant of sequential integration and stream segregation (see, for example, van Noorden, 1975, and Bregman, 1990). It is possible, though unlikely, that the effects of duration and/or condition over-whelmed the effects of frequency separation. This would mean, for these subjects, frequency separation was a less effective cue than duration and condition. This, however, was clearly not the case for Subject 5, who showed little difference in criterion measures to manipulations of frequency, separation, condition, or duration.

Effect of condition

The similarity between the alternating sounds was also varied in terms of overall intensity and temporal envelope. Figure 18 shows that the high-segregation group reported the least segregation for the correlated AM noise-band sequences. This condition was significantly different from the other conditions ($p < 0.05$). The criterion responses for the other conditions were not statistically different from one another. The low-segregation group, though, exhibited the most sequential integration for the pure-tone sequences with the same overall intensity for all elements. Pure-tone, 12 dB ΔI sequences had similar sequential integration as correlated AM sequences, while the uncorrelated AM sequences appeared to be segregated (Figure 19).

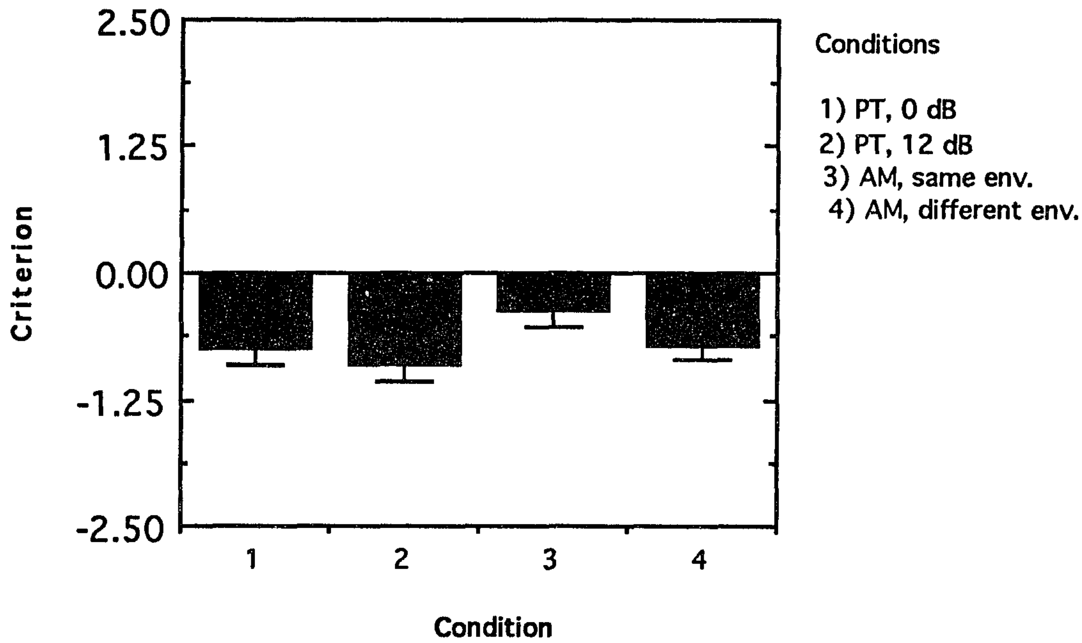

Figure 18. Criterion in each condition for the high-segregation group. Data are averaged over frequency separation and duration.

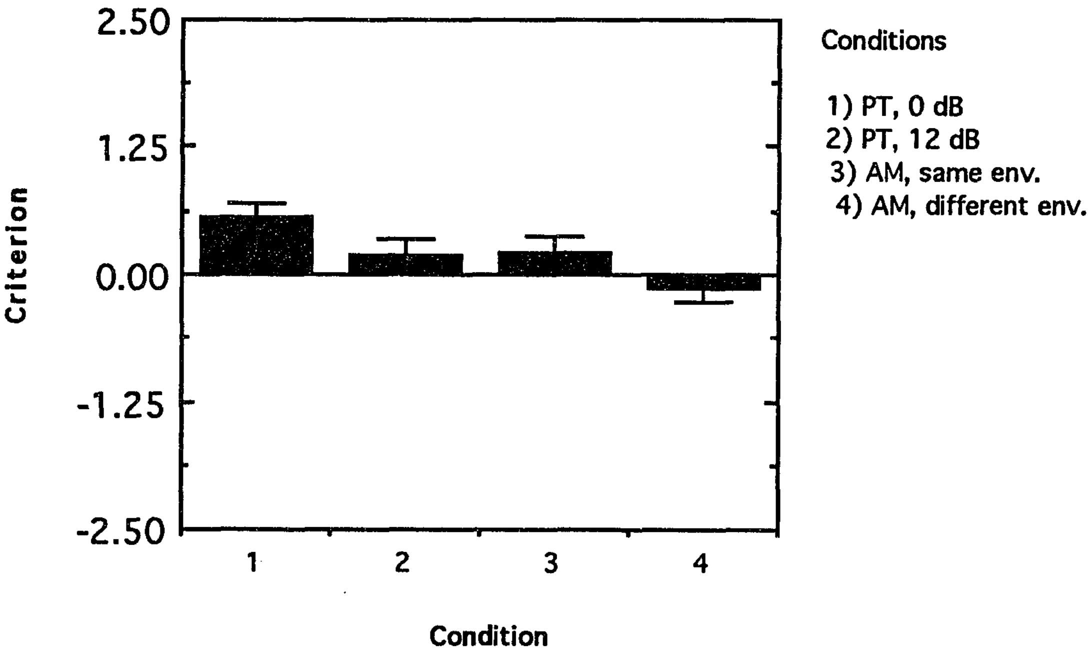

Figure 19. Criterion per condition for the low-segregation group.

Differences between conditions for the low-segregation group were, however, not significantly different from one another. Averaged across subjects, frequency separations, and durations, the 12 dB difference in intensity had no additional effect on segregation compared to pure-tone sequences with no alternations in intensity. This is surprising in two respects. First, it is logical to assume that a difference in alternating sounds as great as 12 dB should signify to the auditory system that the alternating sounds were generated by separate sources. Thus, segregation should occur. Second, van Noorden (1977) observed stream segregation based on a difference in intensity as small as 3 dB. However, the 12 dB intensity difference used in the present study did not affect stream segregation. The differences between van Noorden's experiment and the present study may explain the discrepancy. Van Noorden based his findings on only two subjects. Figure 20 shows the difference between the average criterion for the 12 dB ΔI sequences and the average criterion for the 0 dB ΔI sequences for each of the eight subjects in the present study, averaged over frequency separations and conditions. In spite of the lack of statistical significance, it appears that marginally more segregation occurred for the sequences with an intensity difference between alternating sounds than for sequences with no intensity differences. Subject 9, however, appeared to have employed this difference perceptually to assign the alternating sounds to different sources. This suggests that the subjective "weights" applied to the differences between sounds varied from subject to subject. Thus, if van Noorden had used a larger number of subjects, he might have

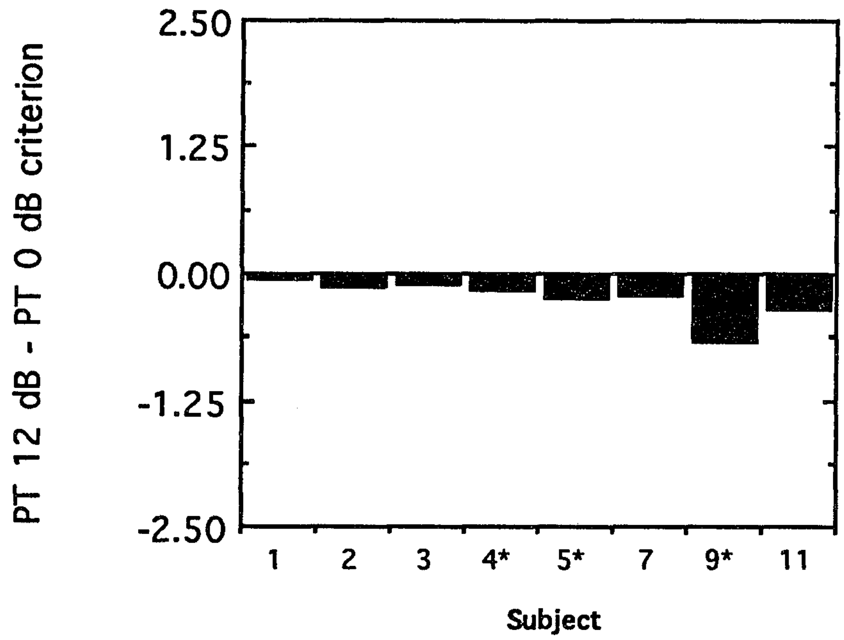

Figure 20. Criterion for pure-tone sequences with 12 dB alternating intensities minus criterion for pure-tone sequences with the same intensity. Asterisks denote subjects from the low-segregation group.

found that not all of his subjects took advantage of intensity alternations. A possibility for future research would be to relate ratings of similarity of the sounds making up sequences with the criterion measure as used in the present experiment. A relationship between stream segregation and direct judgments of similarity might help explain differences between individuals found in this study.

Another difference between the van Noorden experiment and the present investigation is that van Noorden used single-frequency sequences; the tones alternated only in intensity. It could be the case that, in the face of the frequency separations used in this experiment, alternating intensity was relatively unimportant in deciding whether or not more than one source was contributing to the acoustic environment (at least in the averaged data). Another possible area for future research would be to use sequences with smaller frequency separations along with differences in intensity to examine further the relationship between these two factors. A difference in mean criterion (for the high-segregation group) was also evident in the sequences consisting of AM noise-bands. As mentioned earlier, these conditions were significantly different ($p < 0.05$), with the lesser degree of segregation seen for the correlated AM noise-band sequences. Not all subjects, however, showed this pattern of results. Figure 21 shows the difference between mean criterion for the uncorrelated AM noise-band sequences and the mean criterion for the correlated AM noise-band sequences averaged across frequency separation and sequence duration. Only Subject 4 appears not to have segregated the AM noise-band sequences any

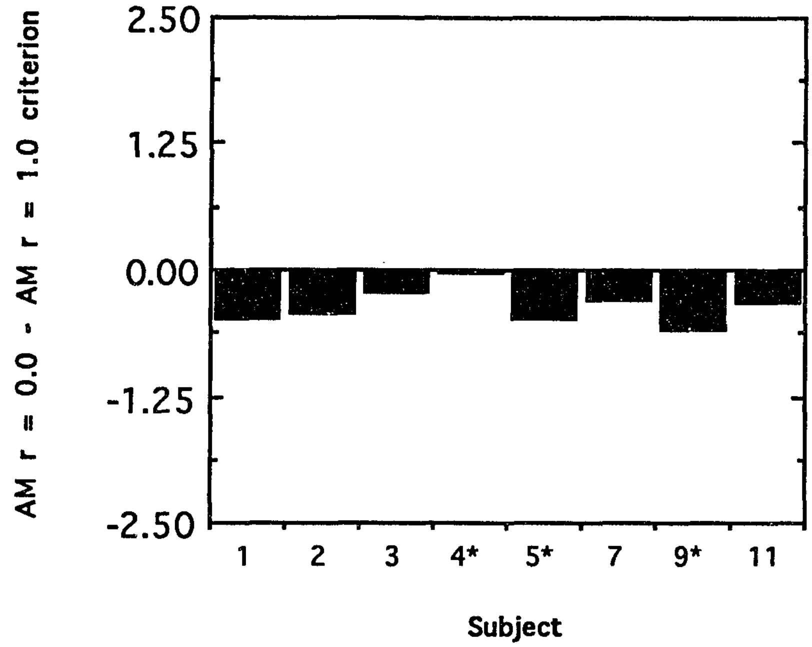

Figure 21. The difference in criterion between the uncorrelated AM noise-band and correlated AM noise-band sequences. Low-segregation subjects are marked with an asterisk.

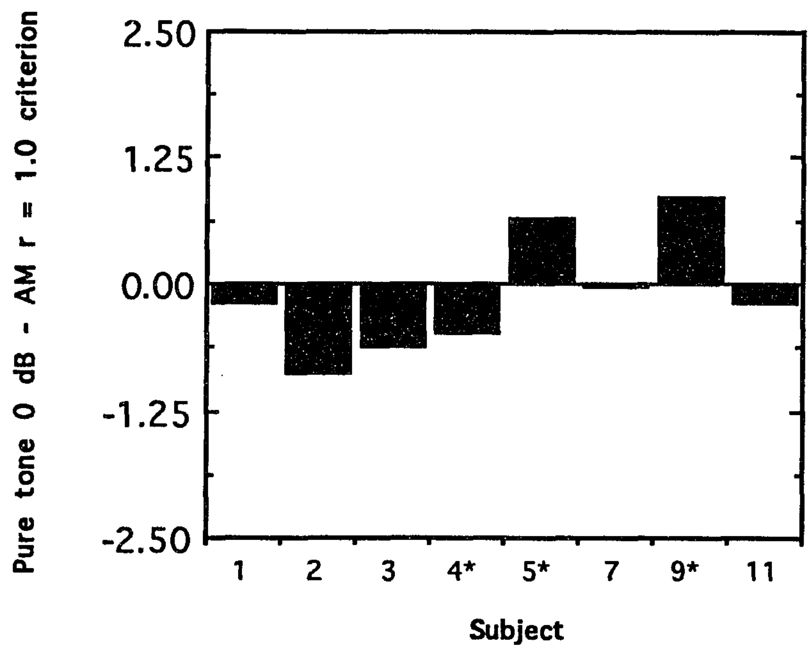

Figure 22. The difference in criterion of the pure-tone sequences with no intensity alternations and the correlated AM noise-band sequences.

differently with respect to envelope correlation. For the average subject, however, differences in the temporal envelopes of alternating sounds increased the tendency of the auditory system to segregate the sounds in the sequences. While other studies have shown that segregation and fusion can be affected by the envelope correlation of two or more co-occurring sounds, this result is the first indication that envelope similarity may play a role in the sequential integration and stream segregation of non-simultaneous sounds. This leaves open the possibility that temporal envelope plays a role in the perceptual organization of other sequential sounds, linking those sounds with similar temporal envelopes, and segregating sounds with different temporal envelopes. Another comparison of interest is that of the segregation of the pure-tone, 0 dB ΔI sequences to that of the correlated AM sequences (Figure 22) It might be supposed that these two conditions would result in the same degree of segregation. In both cases, the sequences are composed of alternating sounds with identical envelopes, and the alternating sounds in the sequences have equal intensities. In other words, the sounds in the sequences differ in only one attribute. However, the correlated AM noise-band sequences are less segregated than the pure-tone, 0 dB ΔI sequences to a statistically significant degree for the high-segregation group ($p < 0.05$). It seems that amplitude fluctuations in the sounds making up the sequences reduces the stream segregation of the alternating sounds.

Comparing the constituent sounds of the sequences in these two conditions gives rise to two possible reasons for the difference in segregation. First, the AM noise-bands fluctuate in intensity over

their durations. That is, the modulated noise bands used in this study have a certain depth of modulation (1.0), corresponding to changes in intensity over the durations of the sounds. It might be supposed that modulation depth plays a role in segregation. The pure-tones used as sequence elements did not fluctuate in intensity over their durations (i.e., were *un*modulated). Since more sequential integration was found for the correlated AM noise sequences than for the pure-tone sequences, one hypothesis would be that sequential integration increases as the modulation depth of the sounds making up the sequences increases. Preliminary work, however, shows this not to be the case. For sequences consisting of sinusoidally amplitude-modulated pure-tones modulated at 20 Hz, sequential integration does not differ with modulation depths of 0.25, 0.5, and 1.0 (Mendoza, 1993). These data are limited in terms of number of subjects and number of trials; further work is required for a definitive test of the hypothesis.

Another difference between the sounds in the pure-tone and AM sequences is the bandwidth of the sounds. The pure-tone sounds have a nominal bandwidth of 1 Hz, while the AM noise-bands had a bandwidth of 32 Hz. It was assumed at the outset that the factor of frequency separation would be related to the center frequencies of the sounds in the sequences. However, it might be the case that auditory scene analysis operates on the *minimum* difference between sounds. For example, the frequency ratio of the sounds in the 1000 Hz - 1750 Hz pure tone sequences is 0.75. Since the AM sounds have a bandwidth of 32 Hz, and have center frequencies corresponding to the frequencies of the sounds in the pure-tone

sequences, the minimum frequency ratio in the 1000 Hz - 1750 Hz AM sequences ([1734 - 1016]/1016) is 0.71. It is conceivable that the smaller minimum frequency ratio in the correlated AM noise-band sequences is responsible for the lesser degree of segregation compared to that with the pure-tone sequences.

If the decreased frequency ratio between sounds enhance sequential integration, and envelope differences between sounds promote segregation, then the equivalence of the mean criterion of the pure-tone, 0 dB ΔI sequences and the uncorrelated AM noise-band sequences (averaged across frequency separation and duration; see Figure 18) might be explained by an interaction of these factors. By virtue of the increased bandwidth (and decreased minimum frequency separation), the segregation of the AM sequences is reduced compared to the sequences with (steady-state) pure-tone elements. However, the temporal envelope differences in the uncorrelated AM noise-band sequences is a cue that the sounds were generated by different sources, thus increasing the segregation of the alternating sounds. These two factors may interact in the uncorrelated AM sequences in such a way as to cancel each other, resulting in the same degree of segregation in the uncorrelated AM sequences as in the pure-tone sequences. Thus, while these two conditions show the same degree of segregation, the underlying processes leading to that degree of segregation may be different in the two conditions.

Interaction of frequency separation and condition

Figure 23 shows mean criterion measure as a function of frequency separation, with condition as the parameter. At the smallest frequency separation, it appears that the least sequential integration occurs for the pure-tone sequences with alternating intensities. The difference in sequential integration between the alternating intensity sequences and the other conditions decreases, however, at larger frequency separations. This lends credence to the idea proposed earlier: a significant difference was not observed for the alternating intensity sequences in the averaged data because of the interaction between effects averaged over frequency separation. Thus, it might be expected that the differences between the 0 dB ΔI and 12 dB ΔI conditions would increase with even smaller frequency separations than those used here. A statistically significant difference in segregation might then be observed, as was demonstrated by van Noorden (1977).

The difference in criterion between the correlated AM noise-band sequences and the remaining conditions also varied according to frequency separation. It appears that more sequential integration occurs for 1000 Hz - 1250 Hz sequences, and less segregation for 1000 Hz - 1750 Hz, for the correlated AM noise-band sequences than for the other conditions at the same frequency separations. At the largest frequency separation, the criterion measures of the sequences are equivalent, regardless of condition. This contradicts the bandwidth explanation offered earlier. The minimum frequency ratios for the 1000 Hz - 1250 Hz, 1000 Hz - 1750 Hz, and 1000 Hz -

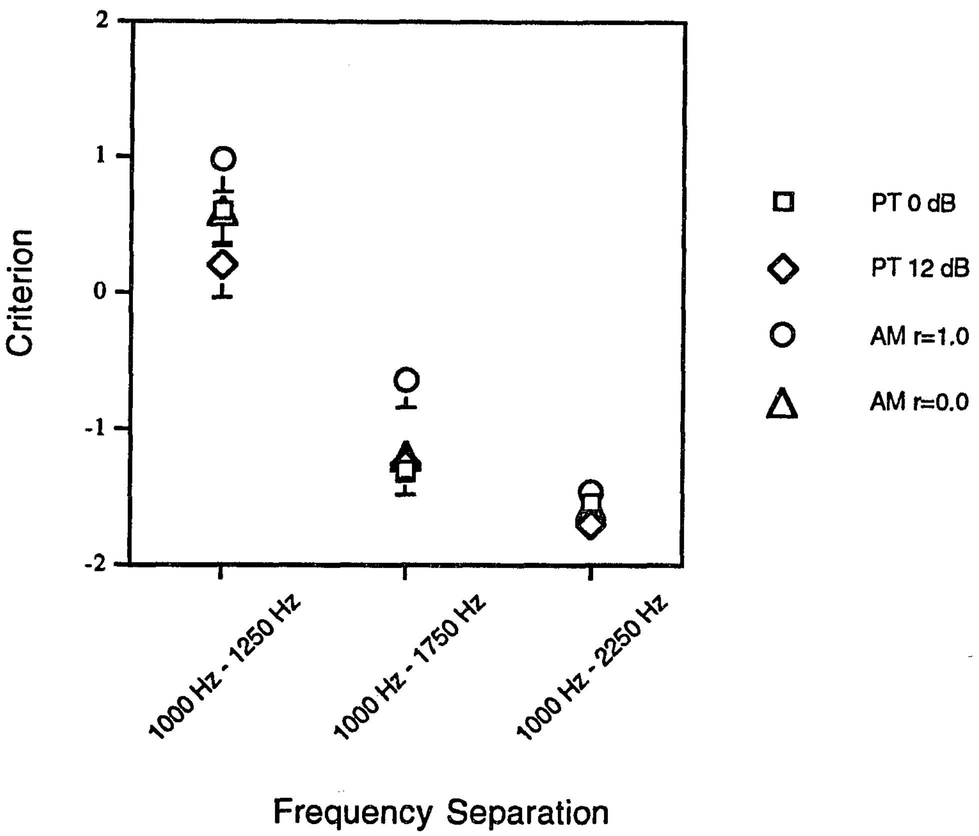

Figure 23. Interaction of frequency separation and condition. Data for high-segregation subjects averaged over duration.

2250 Hz AM noise band sequences are 0.22, 0.71, and 1.20, respectively. The minimum frequency ratio of the AM noise-band sequences becomes increasingly smaller relative to the pure-tone frequency ratios of 0.25, 0.75, and 1.25 as frequency separation increases. One would predict that the difference between the criterion measures of the two conditions would *increase* as frequency separation increases. This was, in fact, not the case. Thus, the explanation based on the minimum frequency separation between complex sounds offered earlier does not hold.

Relating these findings to the hypothesis proposed earlier, the effects of differences in attributes other than frequency separation depends on the specific attributes, as well as the frequency separation of the sounds. Differences in intensity and frequency appear to lead to less sequential integration compared to only a difference in frequency, but only when that frequency difference is small. Differences in center frequency and temporal envelope lead to greater segregation than differences in only center frequency, but, again, this effect is limited by the frequency separation of the sounds. The effects of envelope similarity appear to operate only for sounds with frequency separations of less than one octave. Thus, there is no evidence of orthogonal effects of attribute differences.

Another way of viewing the results found here is simply that when the frequency separation of the sequences provides inconclusive evidence as to "source" for auditory scene analysis, the auditory system falls back on envelope information in assigning

sounds to sources. This would imply that the process of auditory scene analysis employs a hierarchy of cues, with frequency separation carrying the greatest weight, followed by temporal envelope similarity, then alternations in intensity.

Rather than auditory scene analysis disregarding temporal envelope information at large frequency separations, it might be the case that the auditory system was unable to discriminate between temporal envelopes at large frequency separations. If this were true, then temporal envelope similarity would have been useless as a factor in the perceptual organization of the sequences. Sheft (1990) showed that the discrimination of the temporal envelope of sequentially-presented sounds decreased as the frequency separation between the sounds increased from 0 to one-half octave. Research using simultaneous sounds also indicated that the discrimination of envelope correlation might decrease with increased frequency separation between the two sounds (Richards, 1987, Hall and Grose, 1993). Thus, the subjects may not have been able to make use of envelope similarity at the largest frequency separation used in the present study (1.25 octaves). Work is currently underway to further clarify the relationship between frequency separation and temporal envelope discrimination in sequentially-presented sounds.

To summarize the interactions found in the present data, it seems that auditory scene analysis uses a hierarchy of cues in determining whether one or more sources are responsible for a given acoustic environment. Frequency separation appears to have the greatest influence of the factors tested here. At the largest

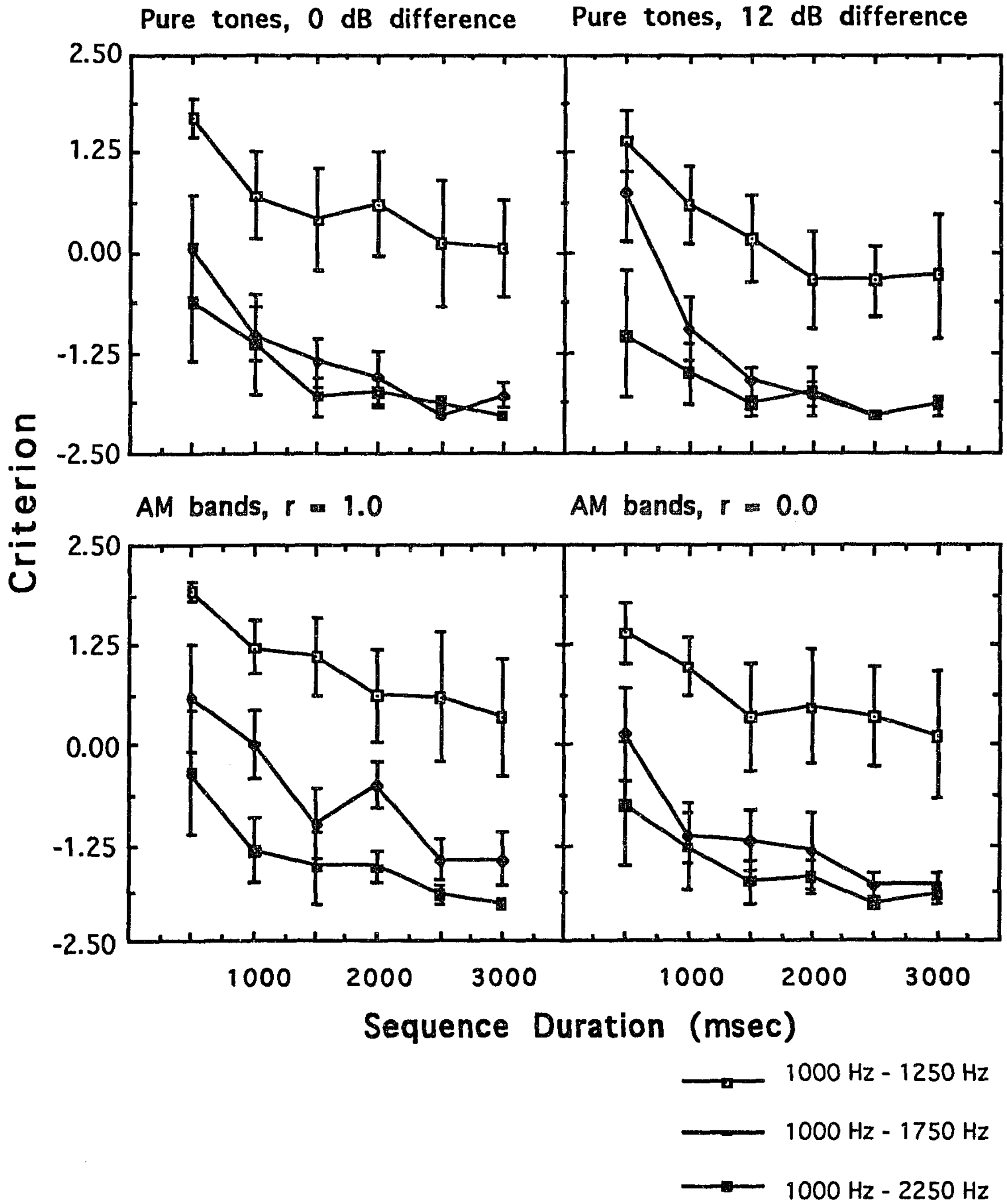

Figure 24. Criterion as a function of sequence duration. Each condition is represented in a separate panel. Parameter is frequency separation. High-segregation group.

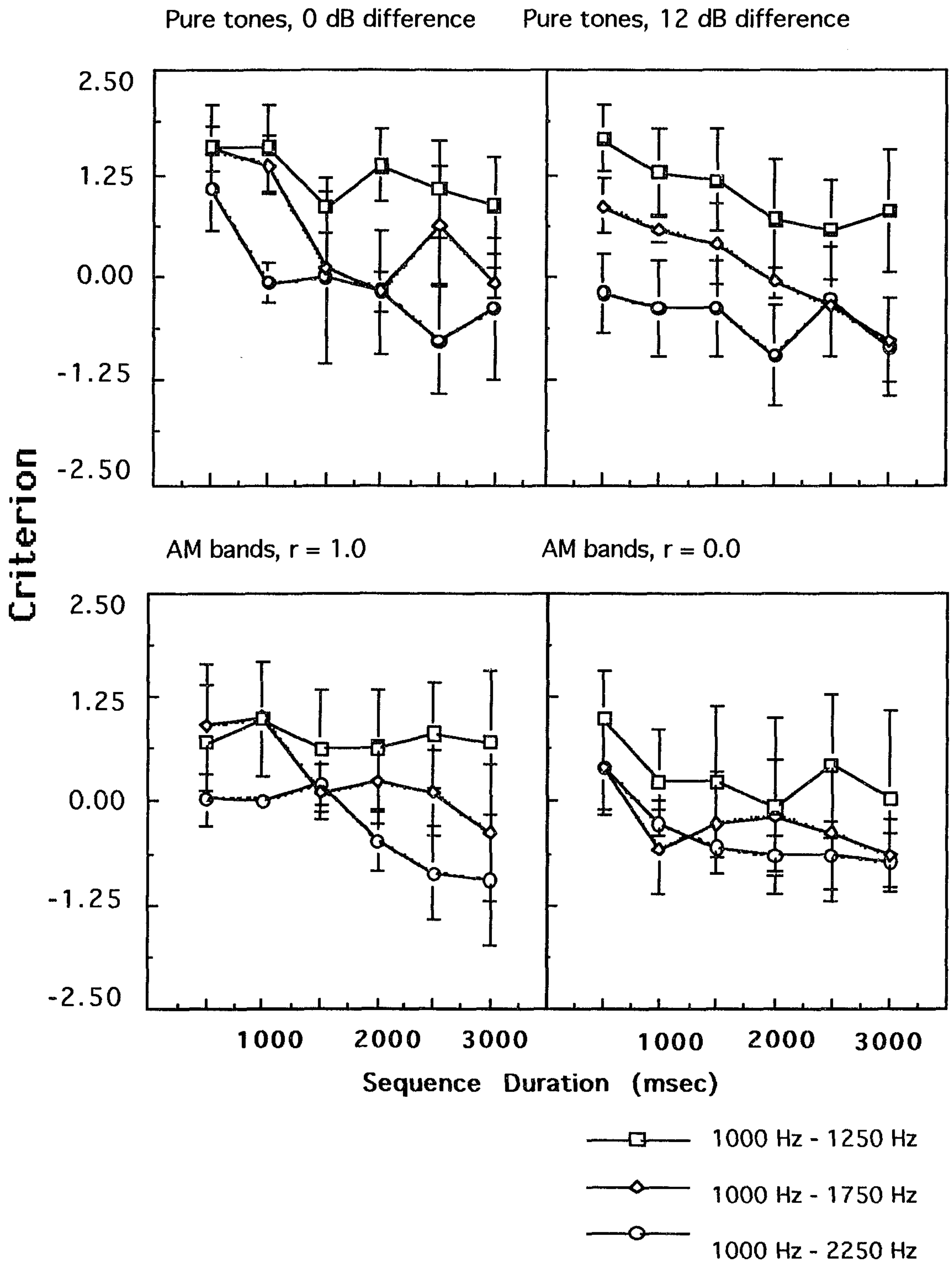

Figure 25. Criterion as a function of sequence duration. Conditions are represented in separate panels. The parameter is frequency separation. Low-segregation group.

frequency separation, the amount of segregation is independent of the differences in other attributes. As frequency separation decreases, however, other factors come into play. Thus, the effectiveness of a particular attribute difference in auditory scene analysis appears to be dependent on the differences in other attributes.

The effects of attribute differences on the time-course of stream segregation

Figures 24 and 25 show criterion measures as a function of sequence duration for each frequency separation and condition. For the high-segregation group, an effect of frequency separation is obvious. The data show that sequences with smaller frequency separations are segregated (criterion measures go from positive to negative) later than sequences with larger frequency separations, if at all. A similar trend occurs for the low-segregation group; however, the data from this group are more variable than those for the high-segregation group. This suggests that stream segregation occurs sooner for sounds with a larger difference in an attribute compared to sounds with a smaller difference in the attribute, if that attribute is frequency.

There appeared to be evidence of sounds with differences in two attributes to segregate sooner than sounds with a difference in only one attribute. This was evident comparing 0 dB ΔI pure-tone sequences to the 12 ΔI pure-tone sequences for 1000 Hz - 1250 Hz alternating tones. Stream segregation also occurred sooner in the uncorrelated AM noise-band sequences than in the correlated AM

noise-band sequences for the 1000 Hz - 1750 Hz case. One comparison in conflict with this trend was that of the pure-tone, 0 dB ΔI sequences and the pure-tone, 12 ΔI sequences in the 1000 Hz - 1750 Hz case. Also, other factors did not seem to play a role in the duration necessary for segregation of the 1000 Hz - 2250 Hz sequences. The lack of a statistically significant interaction of frequency separation x condition x duration suggests that the relationship between these factors, if one exists, is complex.

Implications for the description of the process of stream segregation

Earlier discussion offers a description to help conceptualize the process of auditory scene analysis. The description revolves around two separate processes. The first step is the assessment of the similarity of the sounds in the environment. This is determined by comparing the attributes of each sound. The idea of a combined attribute value (CAV) is used to describe the characteristics of a sound, in order to simplify the representation of this process. By using a CAV, the characteristics of any particular sound can be represented by a single value instead of representing the sound in multidimensional space. Thus, complex differences between sounds can be represented on a single axis. The distance between the CAVs of two sounds is an indication of their overall similarity. Sounds differing greatly in their attributes would be farther apart on the CAV axis than sounds with smaller differences between their attributes. The second process involves partitioning the CAV axis over time. Sounds with CAVs falling on the same side of the

partition are sequentially integrated, while those whose CAVs fall on opposite sides of the partition are segregated. This is intended only as a conceptual model of the process of stream segregation, similar in intent to Jones's (1976) model.

Concerning the time-course of stream segregation, the lack of an interaction between sequence duration and either frequency separation or condition indicates that the decrease in criterion occurring from 500 msec to 3 seconds was about the same for all combinations of frequency separations and conditions. If one assumes that each sequence started with the same criterion, that is, the sounds in each sequence were originally sequentially integrated to the same degree, then substantial changes in perceptual organization must have occurred below 500 msec. The methodology used in the current work did not allow the examination of auditory scene analysis for sequences below 500 msec in duration. Moore, Glasberg, and Peters (1986) have shown that the ability to "hear out" a component from a complex increases as the duration of the sounds increase from 50 msec to 1610 msec. It may be the case that changes in sequential integration also occur at such durations. Further work, however, is needed to address this question.

The effects of intensity differences and temporal envelope were seen to vary, depending on the frequency separations used in the present experiment. Other researchers have demonstrated "context" effects in auditory scene analysis. For example, with sequences consisting of pure tones and noises, McNally and Handel (1977) showed that 6000 Hz and 750 Hz pure tones can avoid segregation when they are included in sequences with a sample of

white noise and a 40 Hz "buzz" segment. Thus, a given difference between two sounds can segregate or not depending on other sounds in the environment, as well as the differences in the attributes of the sounds themselves. Thus, one finding of this research is that no evidence exists that the attributes of sounds are treated orthogonally in auditory scene analysis.

SUMMARY

In summary, then, stream segregation was seen to be variable across the subjects used in this experiment, in spite of attempts to reduce listener bias. Substantial changes in auditory scene analysis were observed for sequences 500 msec - 3 sec in duration. The duration necessary for stream segregation was dependent primarily on the frequency separation of the sounds in the sequences. Similarity of temporal envelope was seen to reduce the segregation of sounds compared to the segregation seen for pure-tone sounds and alternating sounds with different temporal envelopes. The combined effects of frequency separation with either intensity or temporal envelope differences were more powerful than frequency separation alone, but only in sequences in which the frequency separation of the sounds was less than one octave.

REFERENCES

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

Bregman, AS (1990). Auditory Scene Analysis. Cambridge: MIT Press.

Bregman, AS (1978a). Auditory streaming is cumulative. Journal of Experimental Psychology: Human Perception & Performance, 4, 380-387.

Bregman, AS (1978b). Auditory streaming: Competition among alternative organizations. Perception & Psychophysics, 32, 391-398.

Bregman, AS, Abramson, J, Doehring, P, & Darwin, CJ (1985). Spectral integration based on common amplitude modulation. Perception & Psychophysics, 37, 483-493.

Bregman, AS, & Campbell, J (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. Journal of Experimental Psychology, 89, 244-249.

Bregman, AS, & Rudnicky, A (1975). Auditory segregation: stream or streams? Journal of Experimental Psychology: Human Perception and Performance, 1, 263-267.

Carrell, TD, & Opie, JM (1992). The effect of amplitude comodulation on auditory object formation in sentence perception. Perception & Psychophysics, 52, 437-445.

Chalikia, MH & Bregman, AS (1989). The perceptual segregation of simultaneous auditory signals: Pulse train segregation and vowel segregation. Perception & Psychophysics, 46, 487-496.

Collyer, CE (1974). The detection of a temporal gap between two disparate stimuli. Perception & Psychophysics, 16, 96-100.

Dannenbring, GL, & Bregman, AS (1976). Stream segregation and the illusion of overlap. Journal of Experimental Psychology: Human Perception & Performance, 2, 544-555.

Fitzgibbons, PJ, Pollatsek, A, & Thomas, IB (1974). Detection of temporal gaps within and between perceptual tonal groups. Perception and Psychophysics, 16, 522-528.

French-St. George, M, & Bregman, AS (1989). Role of predictability of sequence in auditory stream segregation. Perception & Psychophysics, 46, 384-386.

Hall, JW (1986). The effect of across-frequency differences in masking level on spectro-temporal pattern analysis. Journal of the Acoustical Society of America, 79, 781-787.

Hall, JW, & Grose, JH (1993). Monaural envelope correlation perception in listeners with normal hearing and cochlear impairments. Journal of Speech and Hearing Research, in press.

Hall, JW, Haggard, MP, & Fernandes, MA (1984). Detection in noise by spectro-temporal pattern analysis. Journal of the Acoustical Society of America, 76, 50-56.

Jones, MR (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. Psychological Review, 83, 323-355.

MacMillan, NA, & Creelman, CD (1991). Detection Theory: A Users Guide. Cambridge University Press.

Massaro, DW (1977). Perceptual grouping in audition. Perception, 6, 541-553.

McNally, KA, & Handel, S (1977). Effect of element composition on streaming and the ordering of repeating sequences. Journal of Experimental Psychology: Human Perception & Performance, 3, 451-460

Mendoza, L (1993). Unpublished data.

Miller, GA, & Heise, GA (1950). The trill threshold. Journal of the Acoustical Society of America, 22, 637-638.

Moore, BCJ, Glasberg, BR, & Peters, RW (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. Journal of the Acoustical Society of America, 80, 479-483.

Moore, BJC, Hall, JW, Grose, JH, & Schooneveldt, GP (1990). Some factors affecting the magnitude of comodulation masking release. Journal of the Acoustical Society of America, 88, 1694-1702.

Neff, DL, Jesteadt, W, & Brown, EL (1982). The relation between gap discrimination and auditory stream segregation. Perception & Psychophysics, 31, 493-501.

Norman, DA (1967). Temporal confusions and limited capacity processors. Acta Psychologica, 27, 293-297.

Perrot, DR, & Williams, KH (1970). Auditory temporal resolution as a function of interpulse frequency disparity. Psychonomic Science, 25, 73-74.

Rappold, PW, Mendoza, LL, & Collins, MJ (1993). Measuring the strength of auditory fusion for synchronously and nonsynchronously fluctuation noise-band pairs. Journal of the Acoustical Society of America, 93, 1196-1199.

Repp, BH (1987). Integration and segregation in speech perception. Haskins Laboratories Status Report on Speech Research (SR 89/90), 44-75.

Richards, VM (1987). Monaural envelope correlation perception. Journal of the Acoustical Society of America, 82, 1621-1630.

Sheft, S, & Yost, WA (1992). Spectral transposition of envelope modulation. Journal of the Acoustical Society of America, 91, pt 2, 2333.

van Noorden, LPAS (1975). Temporal Coherence in the Perception of Tone Sequences [diss]. Technische Hogeschool Eindhoven, Eindhoven, The Netherlands.

van Noorden, LPAS (1977). Minimum differences of level and frequency for perceptual fission of tone sequences ABAB. Journal of the Acoustical Society of America, 61, 1041-1045.

Yost, WA, & Sheft, S (1989). Across-critical-band processing of amplitude-modulated tones. Journal of the Acoustical Society of America, 85, 848-857.

Yost, WA, Sheft, S, & Opie, J (1989). Modulation interference in detection and discrimination of amplitude modulation. Journal of the Acoustical Society of America, 86, 2138-2147.

APPENDIX A.

PURE-TONE AIR CONDUCTION THRESHOLDS FOR SUBJECTS

Subject	Ear	250	500	1K	2K	4K	8K
1	R	0	5	5	5	-5	0
	L	5	0	5	0	0	10
2	R	5	5	0	10	-5	15
	L	5	5	0	5	5	5
3	R	5	5	5	5	5	0
	L	10	0	0	0	5	0
4*	R	0	5	10	-5	5	0
	L	0	5	10	10	10	5
5*	R	5	5	0	10	5	0
	L	5	10	5	5	-5	10
7	R	5	10	5	0	10	0
	L	10	5	0	0	0	-5
9*	R	15	5	0	5	0	5
	L	10	10	5	0	0	10
11	R	15	10	15	10	5	5
	L	5	15	10	0	-10	15
6**	R	10	10	0	15	10	15
	L	10	10	5	5	0	5
10**	R	15	10	5	5	0	5
	L	10	10	5	5	0	5

* - Low segregation group
** - "Reversed" subjects excluded from study.

Note: Subject 8 was excluded from the study based on performance on pretest (see Methods).

APPENDIX B.

C ALGORITHMS USED TO CALCULATE d' AND CRITERION

```
/* STD.C: Sample code used to calculate d' and criterion measures. Translated to C from
Pascal code in MacMillan & Creelman, 1991. */

#include <stdio.h>
#include <math.h>
#include <conio.h>

main()
	{
	float nhits, nmisses, total, nfa, ncr, n1, n2;
	float hitrate, farate, dp, cr, beta, zh, zf;
	char answer;
	int adjustment;
	float z(float p);
	do {
		adjustment = 0;
		printf ("\n# of trials:  ");
		scanf ("%f", &total);
		printf ("\n# of correct rejections:  ");
		scanf ("%f", &ncr);
		nfa = total - ncr;
		printf ("# of hits:  ");
		scanf ("%f", &nhits);
		nmisses = total - nhits;
		if (nmisses == 0)   {
			nmisses = 1/(2*total);
			nhits = nhits - (1/(2*total));
			adjustment = 1; }

		if (nfa == 0)   {
			nfa = 1/(2*total);
			ncr = ncr - (1/(2*total));
			adjustment = 1; }

		hitrate = nhits/total;
		farate = nfa/total;
		zh = z(hitrate);
		zf = z(farate);
		dp = zh-zf;
		cr = -0.5 * (zh + zf);
		printf  ("\n");
		if (adjustment == 1)
			printf ("Data have been adjusted\n");
		printf ("H = %4.3f     F =    %4.3f\n", hitrate, farate);
		printf ("d' = %4.3f    c =    %4.3f\n",
				dp, cr);

		printf ("continue?  ");
		answer = getche();
```

```
            }   while (answer != 'n');
        }

float z (float p)
        {
        double y;
        double x;

        y = sqrt (-2*log(p));
        x = (y*-1) + ((((0.00004536422l0148 * y +0.0204231210245) * y +
                    0.342242088547) * y +1) * y+0.322232431088) /

        ((((0.0038560700634*y+0.10353775285)*y+0.531103462366)*y+
                    0.588581570495)*y+0.099348462606);

        return ( (float) x);
}
```

APPENDIX C.

DETAILS OF HIGH-AND LOW-SEGREGATION GROUPS

	High	Low
No. Obs	360	216
Average	-0.691375	0.212231
Variance	2.16262	1.17385
Std. Dev.	1.47059	1.08344

The ratio of the variances of the high- and low-segregation groups was 1.84233. The criterion F-ratio (with 359 and 215 degrees of freedom) at the 0.05 level was 1.26. This value was exceeded, indicating that the variances of the two groups were unequal. Therefore, a t-test for groups with unequal variances was conducted. The error term for this procedure is:

$$S\bar{x}_1 - S\bar{x}_2 = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

with

$$df = \frac{(s_1^2 / n_1 + s_2^2 / n_2)^2}{(s_1^2 / n_1)^2 / (n_1 - 1) + (s_2^2 / n_2)^2 / (n_2 - 1)}$$

degrees of freedom.

APPENDIX D.

NEWMAN-KEULS POST-HOC TEST OF SIGNIFICANT DIFFERENCES

High-segregation group: Criterion Measure

Duration	Mean
(1) 500	0.43213
(2) 1000	-0.41467
(3) 1500	-0.85027
(4) 2000	-0.88398
(5) 2500	-1.21003
(6) 3000	-1.22113

* - Significant at the 5 percent level
** - Significant at the 1 percent level

Comparison	Q
(1-2)	3.388
(1-3)	5.131*
(1-4)	5.266*
(1-5)	6.571*
(1-6)	6.615*
(2-3)	1.743
(2-4)	1.878
(2-5)	3.183
(2-6)	3.227*
(3-4)	0.135
(3-5)	1.436
(3-6)	1.484
(4-5)	1.305
(4-6)	1.349
(5-6)	0.044

Frequency Separation	Mean
(1) 1000 Hz - 1250 Hz	0.59173
(2) 1000 Hz - 1750 Hz	-1.09790
(3) 1000 Hz - 2250 Hz	-1.56796

Comparison	Q
(1-2)	4.7804**
(1-3)	6.1103**
(2-3)	1.33

Condition:	Mean
(3) AM, same envelope	-0.37372
(4) AM, different envelope	-0.71908
(1) Pure Tones, I = 0 dB	-0.74981
(2) Pure Tones, I = 12 dB	-0.92289

Comparison	Q
(3-4)	3.637*
(3-1)	3.961*
(3-2)	5.784**
(4-1)	0.323
(4-2)	2.147
(2-1)	1.822

APPENDIX E.

DATA FOR EXCLUDED SUBJECTS

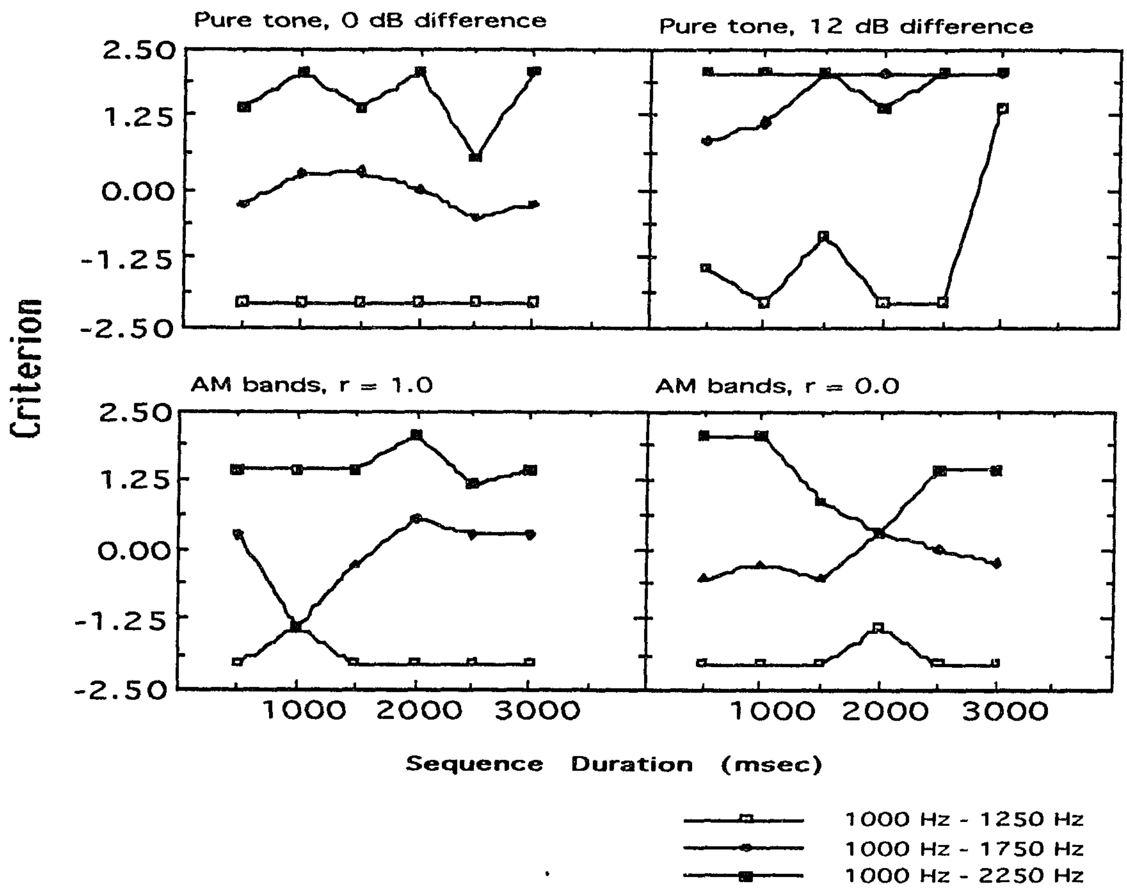

Figure 26. Criterion as a function of sequence duration for Subject 6. Each panel represents a different condition. The parameter is frequency separation.

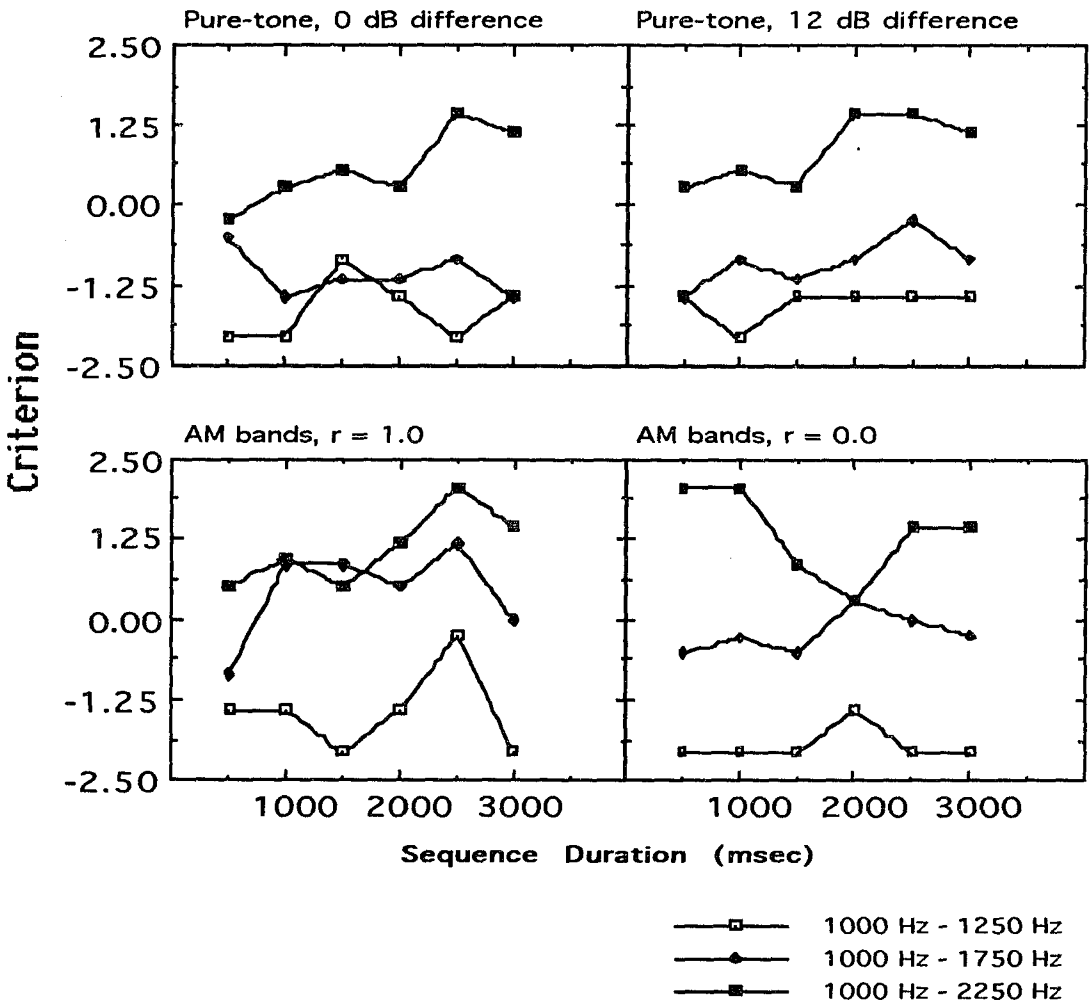

Figure 27. Criterion as a function of duration for Subject 10. Parameter is frequency separation. Each condition is represented in a separate panel.

VITA

Lawrence Lee Mendoza III received his B.A. in Speech Pathology and Audiology in 1986 from the University of Southwestern Louisiana, and his M.S. in Communication Disorders from Florida State University in 1988. His first employment in the field of communication disorders was as an audiologist and Clinical Instructor in Shreveport, LA. In 1989, he moved to Baton Rouge, LA, to begin the doctoral program at LSU under the direction of Dr. M. Jane Collins. During that time, he was employed as a teaching assistant: his duties also included computer programming and general laboratory assistance. His son, Alexander Michael, was born during the completion of his dissertation. He is currently employed as a Post-doctoral Fellow at the University of North Carolina at Chapel Hill, Department of Surgery.

DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Lawrence L. Mendoza, III

Major Field: Communication Disorders

Title of Dissertation: Factors Affecting the Time-Course of Auditory Stream Segregation

Approved:

Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

Date of Examination:

6/10/93