

Summer 5-7-2010

Limits on the number of concurrent auditory streams

Jonathan Henry Schuett
James Madison University

Follow this and additional works at: <https://commons.lib.jmu.edu/master201019>

 Part of the [Psychology Commons](#)

Recommended Citation

Schuett, Jonathan Henry, "Limits on the number of concurrent auditory streams" (2010). *Masters Theses*. 402.
<https://commons.lib.jmu.edu/master201019/402>

This Thesis is brought to you for free and open access by the The Graduate School at JMU Scholarly Commons. It has been accepted for inclusion in Masters Theses by an authorized administrator of JMU Scholarly Commons. For more information, please contact dc_admin@jmu.edu.

Limits on the Number of Concurrent Auditory Streams

Jonathan H. Schuett

A thesis submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Master of Arts

Department of Graduate Psychology

August 2010

Dedication

To my parents who have always supported me in all of my accomplishments.

Acknowledgments

I would like to acknowledge and thank my committee members, Jeffrey T. Andre and Richard F. West, for their commitment and encouragement as they provided feedback to help me through the process of completing my thesis. I would also like to give special thanks to Michael D. Hall, my adviser, for all of his advice and encouragement along the way. His time commitment in reading my drafts and offering feedback has helped me through the process of completing my thesis, while his guidance has helped me through the overall process of completing my degree. I cannot thank him enough for helping shape me into a capable member of the academic community.

Table of Contents

| | |
|---|------|
| Dedication | ii |
| Acknowledgements | iii |
| List of Tables..... | vi |
| List of Figures | vii |
| Abstract | viii |
| I. Introduction..... | 1 |
| General Principles of Stream Segregation..... | 2 |
| Initial Investigations of the Number of Auditory Streams | 6 |
| The Current Investigation..... | 10 |
| II. Experiment 1: Number of Perceived Streams with Complex Tones..... | 13 |
| Method..... | 14 |
| Participant..... | 14 |
| Stimuli | 15 |
| Procedure..... | 17 |
| Complex Tone Number Judgment Task..... | 17 |
| Perceived Sequence Structure Task | 18 |
| Results & Discussion..... | 20 |
| Complex Tone Number Judgment Task..... | 20 |
| Perceived Sequence Structure Task | 25 |
| III. Experiment 2: Stream Segregation by Frequency and Timbre | 27 |
| Method..... | 27 |
| Participants | 27 |
| Stimuli | 27 |
| Procedure..... | 30 |

| | |
|------------------------------|----|
| Results & Discussion..... | 31 |
| IV. General Discussion | 34 |
| V. Footnotes | 46 |
| VI. Appendices..... | 59 |
| VII. References..... | 61 |

List of Tables

| | |
|--|----|
| Table 1: Harmonic Frequencies for the 65 Hz and 124 Hz Fundamental Frequency | |
| Tones | 47 |
| Table 2: Measurable Formants and Corresponding Bandwidths for Each | |
| Instrument that Constituted the Different Spectral Envelopes in Experiment 2 | 48 |

List of Figures

| | |
|--|----|
| Figure 1: Temporal Coherence and Fission | 49 |
| Figure 2: The sine wave and the sawtooth wave represented in amplitude spectra..... | 50 |
| Figure 3: Depiction of the tonal elements within the 5-tone <i>UpDown</i> , <i>Up</i> , and <i>Mixed</i> sequence patterns used in Experiment 1..... | 51 |
| Figure 4: The eight image choices presented to participants in the <i>Sequence Structure Task</i> | 52 |
| Figure 5: Mean number of reported events for the 2- to 5-tones <i>Up</i> and <i>UpDown</i> sequences | 53 |
| Figure 6: Mean number of reported events for the 5-tone <i>Up</i> , <i>UpDown</i> , and <i>Mixed</i> pattern sequences | 54 |
| Figure 7: Depiction of the timbre conditions for Experiment 2..... | 55 |
| Figure 8: Mean number of reported events for the 2- to 5-tone sequences by timbre pattern | 56 |
| Figure 9: Mean number of reported events for the 5-tone sequences by timbre pattern | 57 |
| Figure 10: Mean number of reported events for the 5-tone timbre sequences for participants with the most musical training..... | 58 |

Abstract

Evidence suggests that listeners are limited to perceiving only three streams of auditory information when that information is presented within a rapid sequence of tonal elements. However, this limit has yet to be determined in a controlled setting while using complex tones. Because complex tones are more analogous to naturally occurring tones, finding this limit with complex tones would offer more ecological validity to the suggested perceptual limitation. Thus, Experiment 1 of the current investigation presented listeners with 2- to 5-tone sequences of sawtooth tones at a variety of presentation rates, instructing listeners to report the number of tonal events they perceived within each sequence. Similar to past findings, listeners were limited to only three perceptual streams as presentation rate increased.

Since Experiment 1 found listeners to be limited in the number of auditory streams they could attend when complex tones were included, it begs the question of whether or not a situation could be created in which listeners could ever perceive more than three events at faster presentation rates. Therefore, Experiment 2 attempted to create a situation where additional streams could occur by adding another segregation cue. Timbre has been suggested to be a salient cue that promotes segregation between tones, making it a clear choice for the additional segregation cue in Experiment 2. Participants were presented with different presentation rates of 2- to 5-tone sequences containing either a single timbre or multiple timbres. Following the same procedure as Experiment 1, participants would report the number of perceived events within each sequence. The inclusion of timbre differences between tones in Experiment 2 was not enough to promote segregation of additional streams.

CHAPTER 1

Introduction

The world around us is full of sounds that enter our perceptive field from a variety of sources at any given time. Our auditory system is able to differentiate between sound sources and auditory streams using a variety of techniques, but there may be a cognitive limit on how many of these streams we can attend to at a given time. Mix engineer and recording icon, J.J. Puig, mentioned during a recent interview (see Tingen, 2007) regarding his role during the mixing of the hit song ‘Big Girls Don’t Cry’ that:

“You have to consider the fact that the ear can process only three things at once. When you get to the fourth thing, the attention drops away somewhere. So if you think about this record, what you remember are [Fergie’s] voice, the acoustic guitar, and the bass.” (p. 145)

Although this is an anecdotal account, it reveals an intuitive understanding of a basic perceptual limitation. The relevance of this issue regarding limits of auditory stream segregation becomes clearer with the realization that it has implications far beyond the music industry. Today’s technology presents people in many roles with multiple auditory sources; each source intended to provide the listener with important streams of auditory information. Roles such as air traffic controllers, pilots, doctors in an operating room, or even the average driver are all situations in which listeners are presented with multiple auditory streams of information. Exceeding the number of perceivable streams a listener can attend to at one time could cause a loss of information that could lead to outcomes as drastic as the loss of lives. For example, any given driver could be flooded with auditory

information from their radio, GPS unit, cellular device, passengers, and the outside environment while trying to safely navigate their vehicle. As all this is going on, the driver must determine which piece of auditory information is most important and successfully segregate that information from its source. These anecdotal examples help to highlight the need for an understanding of how many concurrent streams of auditory information listeners can attend to at any given time. The intention of the current investigation was to determine how many concurrent auditory streams listeners are able to attend.

General Principles of Stream Segregation

Although there is plenty of empirical research that explains how listeners group or exclude sound elements into streams (e.g., see Bregman & Campbell, 1971; Miller & Heise, 1950; Rose & Moore, 2000; Singh, 1987; van Noorden, 1975), very few of these studies assess the maximum number of streams that listeners can simultaneously perceive (e.g., see Brochard *et al.*, 1999; Hall & Schuett, 2008; Huron, 1989).

Bregman and Campbell (1971) coined the term *auditory stream segregation* when referring to the perceptual process of segregating auditory elements within a sequence into auditory streams based on origin of the elements. Elements with like frequencies or timbres are more likely to come from the same sound source and are grouped into the same auditory stream. This process is part of *auditory scene analysis*, through which we organize sounds within our environment into meaningful elements (Bregman, 1990).

Some of the earliest studies on stream segregation were concerned with combined effects of frequency and presentation rate (Miller & Heise, 1950; Rose & Moore, 2000; van Noorden, 1975). Elements of an auditory sequence that have similar frequencies are more likely to come from the same sound source and therefore are grouped together. However, this effect is dependent upon the rate at which elements are presented within the sequence. Miller and Heise (1950) found that as the frequency separation between two tones in a sequence becomes greater, the presentation rate of the sequence does not need to be as rapid in order for the two tones to segregate into their own streams. However, as the frequency separation becomes less between the two tones, an increase in the presentation rate of the sequence is needed for the two tones to segregate.

A similar relationship between the components within a sequence based on frequency separation was established by van Noorden (1975; also see Rose & Moore, 2000; Snyder, Carter, Lee, Hannon, & Alain, 2008). Using a sequence of two alternating pure sine wave tones, A and B [such as ABA-ABA-ABA...] presented at different presentation rates, van Noorden (1975) created a task which held the frequency of tone B constant, while the frequency of tone A would be altered by the listener turning a knob. During some trials the listener would alter tone A until A and B fused to one stream, heard as a galloping pattern (Temporal Coherence in Figure 1). On other trials the listener's task would be to alter tone A until the two tones were heard in separate streams (Fission in Figure 1). When the tones are heard as one stream or sequence it is referred to as temporal coherence, while segregation between the two tones into separate streams is called fission. The frequency separation boundary at which the perceptual change occurs

(i.e., changing to either the temporal coherence or the fission boundary), is dependent upon the presentation rate of the tones in the sequence. As presentation rate of the sequence increased, the frequency separation needed between the tones for fission to occur became less. As presentation rate of the sequence decreased, greater frequency separation was needed between the tones before listeners could report perceiving fission of the tones. The results of van Noorden (1975) corroborated with Miller and Heise (1950) to establish the tradeoff between frequency separation and presentation rate.

Rose and Moore (2000) offer a physiological explanation to segregation based on frequency cues. The authors suggest that a contributing factor to frequency separation is the difference of equivalent rectangular bandwidths as the sequence alternates between frequencies. The successive tonal sequence alternating between tones A and B at different frequencies creates an overlap of excitation patterns on the basilar membrane promoting segregation between the two frequencies.

It is important to note that the required frequency separation for fission to occur at a specific presentation rate is contextual. When multiple trials are presented to a listener, trials with large enough frequency separation between tones may be less effective in segregation when following trials that contained a much larger frequency separation. This could be due to physiological causes such as auditory sensory memory or neural adaptation (Snyder *et al.*, 2008).

Another critical factor that impacts stream segregation is the repetition of a sequence, with segregation building over time (Anstis & Saida, 1985; Bregman, 1978). Faster presentation rate is one way to create more repetitions of the sequence. However,

Bregman (1978) also suggests that package size, that is, the number of elements within the sequence, has an effect on segregation as well. Specifically, as he increased the package size from 4 to 8 to 16 tones, the slower presentation rates were required for segregation to occur. This finding indicates that slower rates of presentation are needed for stream segregation as more elements are presented within a given sequence.

Timbre is another fundamental difference between elements within an auditory sequence that contributes to perceptual segregation into different streams (Bregman & Campbell, 1971; Singh, 1987; Warren, Obusek, Farmer, & Warren, 1969). Elements with similar timbres are more likely to come from the same sound source and are therefore grouped together as part of the same stream. Conversely, elements with different timbres should come from different sound sources and thus are more likely to be segregated into separate streams. Warren, *et al.* (1969), presented listeners with non-speech, non-musical elements within a sequence. Each element had a different timbre, such as a hiss, a buzz, the phoneme /i/, and a tone. They discovered that at rapid rates of presentation listeners could not identify the order in which the elements were presented within the sequence. This finding was interpreted as indicating that the elements were forming their own perceptual streams, and, as a result, information about the relationships between elements of the sequence in its entirety, such as their temporal order, was being lost.

The loss of information between streams is better depicted by Bregman and Campbell (1971) as they determined that listeners were able to process information within a stream much better than across streams. Bregman and Campbell (1971) presented listeners with a sequence of six tones, three of which were higher pitch tones,

which were higher in frequency, and the other three were lower pitch tones, due to their lower frequency. In accordance with the Fission Boundary (van Noorden, 1975), the three higher pitch tones segregated into one stream while the three lower pitch tones segregated into the other stream. As the sequence of six tones was presented in different orders, the researchers discovered that listeners could identify changes in the sequence's pattern much better when the change occurred between the higher pitch tones or the lower pitch tones, but struggled in identifying the change when it occurred between a high pitch and a low pitch tone.

Considering the empirical evidence, an auditory sequence of elements with different timbres and large frequency separation between them being presented at a rapid presentation rate should create the optimum likelihood of segregation into different auditory streams. If these properties are implied in real world applications, such as warning tones used to alert pilots, the assumption can be made that auditory information can be presented to the listener in a way that suggests the possibility for multiple concurrent information streams to be attended to at one time.

Initial Investigations of the Number of Auditory Streams

If segregation cues, such as frequency or timbre, are used in real world application to present listeners with multiple streams of information, it is important that we have an understanding of how many concurrent streams listeners can attend. If this limit on the number of simultaneous streams exists, listeners could only attend a number of the auditory streams, causing the information in other streams to be lost.

Very few studies have addressed how many streams listeners can perceive within a single sequence. Huron (1989) instructed his participants to directly report how many streams they perceived at a given time. Concerned with ecological validity, due to the fact that most auditory stream segregation research uses pure tones in a laboratory setting, Huron (1989) presented listeners with Bach's Fugue in E-flat ("St. Anne" fugue). Huron chose the fugue because sections of it contained up to five concurrent melodic 'voices' representing potential streams. Listeners were instructed to report the number of voices they perceived throughout the song. Huron's participants included both musicians and non-musicians. Although musicians performed slightly better than the non-musicians, both groups showed much more error in reporting the number of voices in the 4 and 5 voice sections than the 1, 2, or 3 voice sections. When four or five voices were present, listeners tended to report hearing three or less. Huron concluded that listeners have trouble attending to more than 3 musical 'voices' at a time.

To determine the upper bound on the number of perceivable streams with more controlled stimuli, Hall and Schuett (2008) presented listeners with sequences of 2 to 5 pure tones at different frequencies separated by pitch intervals of sevenths along an A_{\min} scale (440, 784, 1262, 1868, and 2637 Hz). Tone durations also were varied across sequences to insure sufficient speed to establish stream segregation (31.25, 62.5, 125, 250, or 500 ms). On each trial, participants received one sequence of tones at one of the durations. Decreasing tone durations increases the presentation rate of the sequence. As tone durations decreased, the number of tonal events reported for that sequence decreased, not exceeding three events at higher presentation rates. This finding suggests that at presentation rates sufficient for stream segregation, the number of perceivable

events is limited to three auditory streams. This is congruent with Huron's (1989) participants having trouble reporting more than three perceivable voices. By using pure tones instead of the musical stimuli used by Huron (1989), Hall and Schuett (2008) were able to argue that segregation is limited to three streams when only frequency differences are present to promote segregation.

In an attempt to discuss specific tones within a sequence more easily, tones in the sequence with the highest- and lowest-fundamental frequencies will henceforth be referred to as the highest tone and lowest tone, while a middle tone will refer to one of the tones with a fundamental frequency between that of the highest and lowest tones. Streams will be referred to in a similar way. Specifically, reference to a middle-frequency stream will indicate a stream comprised of middle tones in the sequence, while an outer-frequency stream will indicate a stream that includes either the highest or lowest tone from the sequence.

Brochard, Drake, Botte, and McAdams (1999) determined that the outer-frequency tones of a sequence are easier for listeners to attend to than tonal elements in the middle-frequency portion of the sequence. This suggests that listeners may be segregating the highest tone and the lowest tone into their own streams due to frequency separation between the outer tones and the middle tones. If listeners are only able to attend three streams, the remaining middle elements would be perceived together as part of the third stream.

To determine what was happening to the perception of the middle-frequency stream, and to see if any information about the middle components was lost, Hall and

Schuett (2008) created a set of comparison stimuli, each given after a brief delay following the presentation of the five tone sequence. The comparison stimuli were designed to test how listeners perceived the middle three component tones. The comparison stimuli consisted of; (a) the third or middle-fundamental frequency tone being presented by itself, (b) all three middle tones presented at compressed tone durations so that their combined duration matched the duration of just one tone from the stimulus sequence, (c) all three middle tones presented at the same time, (d) a logarithmic glide of all three tones presented at the same time, (e) a logarithmic glide of just the middle tone, and (f) the original three middle tones presented as the control. These stimuli were chosen because they represented different intact variations of the three middle-frequency tones, as well as various transitions encompassing the three middle-frequencies.

Participants correctly rated the control stimulus as most similar to what was presented in the sequence, suggesting that even though listeners are perceptually grouping the middle tones in rapid sequences into a single stream, information about the individual components within the middle-frequency stream is still accessible. This is congruent with what Bregman and Campbell (1971) stated about listeners ability to process information within streams. Information between the streams is lost, such as how many elements comprise the sequence. Yet, within the middle-frequency stream the tones themselves are still identifiable enough to be correctly selected in the comparison task.

The Current Investigation

The previous study by Hall and Schuett (2008) used pure tones to demonstrate that stream segregation is limited to three streams when frequency cues are available for the listener to aid segregation. However, pure tones are very artificial in the sense that they generally do not occur in a natural environment. As a result, it is important to establish if a corresponding perceptual limit on the number of concurrent auditory streams is found when complex tones are involved. Therefore, the current investigation asked if the general findings of Hall and Schuett (2008) can be replicated with complex tones. Complex tones are comprised of multiple harmonics which will overlap in frequency between adjacent tones when played together in a sequence, possibly masking one another.

Masking tends to occur when frequencies fall within the same critical bandwidth. The critical band is a processing unit roughly one third of an octave wide that gets processed by the same area of the basilar membrane (Moore, 2003). Table 1 depicts the harmonic frequencies for tones at pitches of C2 and B2. Based on the third of an octave critical bandwidth, it is apparent from the frequencies listed in Table 1 that some of the harmonic frequencies will blend together. For example, the first harmonic of the C2 tone will fall within the same critical band as the fundamental frequency of the B2 tone. Another example is the 10th harmonic of the C2 tone which will blend with the 5th and 6th harmonic of the B2 tone. These overlapping harmonics could perceptually cause the tones to blend, increasing the likelihood that these tones will stream. However, it was suggested that listeners will rely more heavily on the fundamental frequency of each tone

for perceptual grouping as seen in previous studies (i.e., Bregman *et al.*, 1990; Singh, 1987), which may not affect streaming any differently than when pure tones are used.

Experiment 1 of the current investigation extended the results of Hall and Schuett (2008) by replacing the sine wave tones in the 2- through 5-tone sequences with sawtooth wave tones. The sawtooth wave is a complex tone with a spectral slope similar to the vocal source and bowed instruments like the violin (Askenfelt, 1991). The sawtooth tone was also ideal because it covers a full spectrum with all harmonics present. Figure 2 depicts the difference between the sine wave and the sawtooth tones, displaying the number of harmonics and their corresponding amplitudes for each.

To assess the number of perceived streams, listeners were presented with 2- to 5-sawtooth tone sequences at different presentation rates and asked to report how many tonal events they perceived. A new task additionally presented participants with a series of graphical depictions and instructed them to select the image that best represented their perception of the 5-tone sequences' structure after it was presented to them. The purpose of the second task was to provide an understanding of how the relationship between tones was perceived in the 5-tone sequence where listeners may only have been reporting three streams.

If previous results are replicated with complex tones, it would suggest that the limit on the number of streams in stream segregation is naturally occurring and not just observable under laboratory conditions. If this is the case, it would beg the question of whether or not it is possible for more than three streams to occur simultaneously. To address this issue, a second experiment additionally introduced timbre differences

between tones within the sequence to further promote tone segregation. If the use of complex tones in Experiment 1 fails to limit the number of perceivable streams, the addition of timbre in Experiment 2 is still a reasonable research condition because it creates a situation that highlights the effect of including additional segregation cues in a sequence.

Timbre has been selected as the additional cue because past research (i.e., Bregman *et al.*, 1990; Singh, 1987) has shown that timbre is heavily relied upon during segregation. By directly asking participants to select how tones were grouping, Singh (1987) found that timbre cues are relied on more than frequency cues when the two are put in competition with one another. Inclusion of timbre cues should provide optimal conditions for the segregation of tones into more than three streams. If the conditions of Experiment 2 fail to create more than three concurrent auditory streams, it could be argued that stream segregation is limited to three simultaneous streams, even under optimal conditions for segregation.

CHAPTER 2

Experiment 1: Number of Perceived Streams with Complex Tones

The purpose of Experiment 1 was to determine if listeners would be limited to three perceptual streams when given sequences of complex tones distinguished by frequency. To address this question, this experiment followed the same design as the number judgment task of Hall and Schuett (2008), but with complex tones in place of pure tones. On each trial within this task, participants received a 2- to 5-tone sequence at one of five tone durations. Shorter individual tone durations were equivalent to a faster presentation rate for the sequence. It was expected that at the slower presentation rates, participants would be able to identify the correct number of tonal events present within the sequence. However, it was also predicted that as presentation rate increased, listeners would perceptually group the tones into streams, causing participants to report three or less tonal elements. Congruent with the results from Hall and Schuett (2008), it was hypothesized that the 4- and 5-tone sequences would form no more than three streams during faster presentation rates. The highest and lowest tones were expected to form their own outer-frequency streams, while the middle tones instead would merge into a single, middle-frequency stream. It was also hypothesized that there would be an interaction between presentation rate and the number of tones. Thus, it was believed that as the presentation rate and the number of tones both increased, a decrease in the number of tonal elements reported by listeners would occur.

Method

Participants. Participants were 10 adult listeners from James Madison University, all of whom had self-reported normal hearing. Participants were under the age of 40 to reduce the likelihood of presbycusis, which would reduce the listener's ability to perceive the higher frequencies used in this study. Participants volunteered to complete the experiment in exchange for credit towards meeting one of their requirements for an undergraduate psychology course.

Differences, such as ear dominance for recognizing simple melodies, have been found to occur between musically trained and untrained listeners when presented with musical stimuli. Musically trained listeners tend to use the left hemisphere for analytical processing, in that they perceptually break melodies apart into tonal or instrumental pieces. Non-musically trained listeners tend to be more reliant on the right hemisphere for holistic processing, in that they perceive the elements of a melody together as a whole (Bever & Chiarello, 2009). Huron (1989) suggested that analytical listening due to musical training, could aid segregation tasks when instrument based timbres are involved. In order to determine if musically trained listeners were more accurate in reporting the correct number of tonal elements, participants began the experiment by completing a musical training questionnaire. This questionnaire, which can be found in *Appendix A*, was intended to help identify if aspects of training for musical performance correlated with participants' ability to correctly identify the number of tonal elements within the sequences. Musical training, such as knowledge of musical theory or amount of relevant experience with situations that require the identification of multiple sound sources could

potentially aid listeners when reporting how many streams are present within each sequence. Although musical training may have helped listeners to an extent, it was expected that musically trained listeners would still have difficulty perceiving more than three streams at faster presentation rates, as found by Huron (1989).

Stimuli. All tones in this study were synthesized at a 44.1 kHz sample rate (16-bit resolution). At the time of presentation, all sequenced tones were passed through a low-pass filter with a cut-off frequency of 22 kHz (-24 dB/octave slope) for anti-aliasing purposes. Each tone has a 10 ms linear increase in amplitude from silence at tone onset, as well as a 10 ms linear decrease in amplitude to silence at tone offset, to prevent any perceived abruptness (e.g., as a pop or click) due to changes in intensity. Tones were equated for loudness using Adobe Audition prior to sequencing. Tones in this study were presented at a peak intensity of 80 dB [A] within any given sequence. All stimuli were presented over Sennheiser HD25SP earphones.

The stimuli for this task were presented as sequences constructed from arpeggiated cycles of two to five sawtooth tones, characterized by a -6 dB per octave slope, which were generated using Camel Audio's Alchemy, a VST plug-in based upon spectral modeling synthesis (Serra & Smith, 1990). There was no time separation between adjacent tones. Sequences ended once ten seconds had elapsed and the 2- to 5-tone pattern fully completed its last cycle. Henceforth, the term "sequence" will be used to refer to the explained ten-second cycle of one of the 2- to 5-tone patterns. Sequences contained complex tones with fundamental frequencies (F_0) separated by intervals of sevenths along a C minor scale (65, 124, 233, 440, and 831 Hz, which are C₂, B₂, A[#]₃,

A4, and G[#]5 respectively). To permit a manipulation of presentation rate, the tones within each sequence were given durations of 31.25, 62.5, 125, or 250 ms.

All sequences began with the 65 Hz F_0 tone followed by each successively higher pitched tone (i.e., the tone representing the smallest increase in F_0) until the number of intended tones for that sequence was reached. The tones within each sequence follow one of three patterns, which are depicted for 5-tone sequences in Figure 3. The *UpDown* pattern, in Figure 3, exists for the 2- to 5-tone sequences. This pattern was designed to allot equal pitch separation between tones. In this pattern, each tone was presented in order, until the maximum number of tones in the sequence was reached. The tones were then presented in reverse order until the second to lowest F_0 (124 Hz) tone was reached. The possible disadvantage with this pattern is that it creates increased repetitions of the middle-frequency tones. However, this may actually aid segregation of the middle-frequency tones, because repetition has been shown to favor segregation as it becomes more obvious to the listener over time (Bregman, 1978).

To control for the extra repetition created by the *UpDown* pattern, the *Up* pattern, depicted in Figure 3, was included for the 2-5 tone sequences. This pattern simply ends when the highest F_0 tone is reached. A possible drawback to this pattern is that the pitch separation between the highest F_0 tone and the starting tone becomes too large. The large separation in pitch could lead to stronger segregation between these two tones, than the other tones.

For the 5-tone sequences a *Mixed* pattern was included. This pattern attempts to minimize the pitch separation between the highest F_0 and lowest F_0 tones in the *Up*

pattern, while also controlling for the extra repetitions of the three middle-frequency tones in the *UpDown* pattern. The Mixed pattern consists of the 124 Hz tone, followed by the 65 Hz tone, then the 233 Hz tone, 831 Hz tone, and finally the 440 Hz tone, and is depicted in Figure 3. Based on the results of Hall and Schuett (2008), which used the same sequence patterns for pure tone sequences, it was expected that there would be no significant difference in participants' performance across between sequence patterns.

Procedure.

Complex Tone Number Judgment Task. On each trial of this task, participants were presented with a sequence of complex tones and asked to report how many tonal events they perceived in the sequence. Participants first completed the Musical History Questionnaire. They were then comfortably seated in a single-walled attenuated sound chamber. Participants were instructed to listen to each presentation of the 2- to 5-tone sequences and indicate after its completion how many distinct tonal elements they perceived in the sequence. Responses were made by pressing the number (1 through 9) on the numeric keypad of a PC keyboard that directly corresponded to the number of distinct tones that were perceived. Responses were followed by a 500 ms inter-trial-interval before the next sequence is presented.

Participants received 4 blocks of trials, each containing 64 randomized trials. Each of the conditions was played at each presentation rate 8 times within the experiment. Each block lasted approximately 12 minutes. Upon completing each block, participants took a short rest break before the next block of trials started. Stimuli were

delivered, and responses were collected, using the *Music Experiment Development System* (MEDS 2002-B-1, Kendall, 2002).

Perceived Sequence Structure Task. A subsequent task was expected to provide a better understanding of how the participants perceived structure within the 5-tone sequences. Specifically, this second task was designed to gain a better description of whether or not the middle-frequency tones integrated into a single stream, and if so, the degree to which they might be perceptually altered as a consequence of being conjoined.

Participants received one presentation of the 5-tone *UpDown* sequence containing 62.5 ms tones and were asked to focus on the structure of the sequence. They were then prompted with eight visual representations that displayed different relationships between tones and asked to select the one that best depicts the structure of the sequence. The 62.5 ms tones were selected for this task because the presentation rate is fast enough to assure streaming. The *UpDown* sequence was selected because it provides a fixed distance between adjacent tones.

The eight visual representation options are based on what may happen perceptually to the tones within the presented sequence. These images are depicted in Figure 4. Image *A* correctly represents the relationship between tones, depicting the appropriate pitch and timing for the sequence. Image *B* depicts the 5-tone sequence with the correct pitch represented for each tone, but with the timing of the three middle tones equivalent to that of one of the higher- or lower- frequency streams. The listener would perceive this option if temporal information is lost between the middle tones, causing them to be perceived with much shorter tone durations. Image *C* depicts another situation

where temporal information is lost. In this illustration, pitch differences and tone duration are correct, but the middle tones occur simultaneously. This would be selected by a listener if they perceive the middle-frequency stream to contain more than one component played simultaneously.

Rapid changes in pitch are sometimes perceived as a glide in the direction of the pitch change (Schneider, Parker, & Upenieks, 1982). Image *D* represents this, illustrating the three middle-frequency tones as a glide in pitch. Another possibility is that listeners perceive the rapid pitch change of the middle-frequency tones as a glide, yet perceive that glide to still be a complex stimulus (i.e., in that it is made up of more than one component). To illustrate this, Image *E* depicts three glides in pitch within the middle-frequency stream.

If listeners perceptually average the middle-frequency pitches over the duration of all three tones, it is remotely possible that Image *F* could be selected. Although even more unlikely, it is remotely possible that the middle-frequency pitches would not only be perceptually averaged, but that the timing of the middle-frequency stream would be perceived as equivalent to the highest- and lowest-frequency streams, as depicted by Image *G*.

The final comparison, Image *H*, depicts the sequence without the presence of the middle-fundamental frequency tone. Previous research has shown that perception can be slightly enhanced by preceding or succeeding auditory stimuli (Moore & Welsh, 1969). Therefore, in this option, the highest- and lowest-frequency tones in the sequence enhance the perception of the tones they are played alongside. Although there is a

possibility that the middle-frequency tone could be enhanced as well, it would seem more likely that the two lower-frequency and two higher-frequency tones could be enhanced masking the middle-frequency tone. It was not expected that this option would be preferred by listeners.

Although it was believed that listeners would report hearing fewer tonal events than actually present in the 4- and 5-tone sequences, it was expected that participants would still have access to information within the middle-frequency stream about the individual tones. As demonstrated by Bregman and Campbell (1971), listeners are able to retain information between elements within a stream. Therefore it was believed that more participants would correctly select Image *A* than the other images.

Results & Discussion

Complex Tone Number Judgment Task. The mean numbers of perceived events were calculated for each combination of presentation rate, sequence pattern, and number of tones for each participant. Given the lack of an orthogonal design, it was impossible to include all of the conditions within a single analysis. More specifically, for the 2-tone sequences the *Up* and *UpDown* sequence patterns were the same. Data therefore were submitted to a series of repeated measures ANOVAs.

The first analysis included all manipulated variables, and as a result, will henceforth be referred to as the *general* analysis. This consisted of a 4 x 2 x 3 ANOVA with presentation rate (31.25, 62.5, 125, and 250 ms/tone), sequence pattern (*Up* and *UpDown*), and number of tones within the sequence (3, 4, or 5) as factors. A secondary analysis was conducted to permit a more complete evaluation of potential effects of the

number of tones by including data from 2-tone conditions, and will henceforth be referred to as the *number* analysis. This consisted of a 4 x 4 ANOVA that was restricted to data from the *UpDown* conditions, with presentation rate (as specified for the general analysis above) and number of tones (2-5) as factors. A final analysis was additionally included to better assess potential effects of the direction of pitch changes within the tone sequences on the number of reported events within the 5-tone sequences. It will be referred to as the *pattern* analysis and consisted of a 4 x 3 ANOVA with presentation rate and sequence pattern (*Up*, *UpDown*, and *Mixed*) as factors. Across these analyses, as well as throughout corresponding analyses in Experiment 2, Greenhouse-Geisser adjustments are reported in any cases where the sphericity assumption was violated.

To better understand the nature of any statistically significant effects that were obtained from the indicated ANOVAs, post-hoc pair-wise comparisons of means were included. Throughout both experiments, main effects were clarified using Bonferroni comparisons, while LSD pair-wise comparisons are reported for observed interactions.

An appropriate evaluation of the effects of presentation rate and sequence pattern first requires verification that participants were generally sensitive to the number of presented tonal elements. As expected, ANOVA analyses revealed that participants reported more events for sequences that contained more tonal elements. This finding contributed to a main effect of the number of tones within both the general and number analyses, $F(1.13, 10.21) = 100.52, p < .001$ and $F(1.37, 12.32) = 64.18, p < .001$, respectively. Pair-wise comparison of means confirmed that the mean number of reported events for each number condition was significantly larger for each one-tone increase in

the length of the sequence ($p < .001$). This tendency can be easily seen in Figure 5, which depicts the mean numbers of reported events, along with corresponding standard error bars, as a function of presentation rate, number of tones, and sequence pattern (only data from the 5-tone *Mixed* condition is excluded from this figure). Thus, it is clear that listeners were able, at least under some of the conditions, to report more than three events.

As hypothesized, the mean number of reported tonal events decreased as the presentation rate of the sequence increased. Additionally, the average number of tonal events reported at the fastest presentation rate was no more than three. This can be seen in Figure 5 as the duration of tones along the horizontal axis decreases. This pattern of results contributed to a main effect of presentation rate in both the general and number analyses, $F(3, 27) = 33.60, p < .001$ and $F(1.96, 17.6) = 29.16, p < .001$ respectively. Pair-wise mean comparisons further revealed that the mean number of reported events was significantly lower for the 31.25 ms condition than either the 125 or 250 ms conditions ($p < .05$). Additionally, the mean number of reported events for the 62.5 ms condition was significantly lower than for the 250 ms condition ($p < .001$). It appears that as presentation rate of a sequence increases, listeners lose the ability to segregate tonal elements.

As the presentation rate increased, the mean number of reported events for the 4- and 5-tone sequences decreased to three. However, the mean number of reported events for the 2- and 3-tone sequences did not change as drastically. This tendency can be seen in Figure 5, as it contributes to the interaction between presentation rate and the number

of tonal elements in the general analysis as well as in the number analysis, $F(1.79, 16.14) = 9.65, p < .05$ and $F(2.18, 19.57) = 13.84, p < .001$, respectively. Pair-wise comparison of means confirmed that the number of reported events for the 31.25 ms presentation rate was significantly fewer than the other presentation rates for the 3-tone sequences ($p < .01$). The number of reported events was significantly fewer for the 4-tone, 31.25 and 62.5 ms presentation rates than for the slower two presentation rates ($p < .01$). The mean number of reported tones was significantly lower for the 5-tone sequences, at each increasing level of the presentation rate ($p < .05$). The main effect of presentation rate suggests that participants had an increased difficulty segregating tones as they became shorter in duration and were presented more rapidly within the sequence.

Different sequence patterns were included in the experiment in order to determine if the direction of pitch change between tonal elements affected the number of reported events. It was hypothesized that sequence pattern would not significantly affect the number of reported events. However this hypothesis was not supported by the data. Relative to the other sequence patterns, the mean number of reported events was significantly higher for the *UpDown* pattern at all but the 31.25 ms presentation rate ($p < .01$). This can be seen in Figure 6, which depicts the mean number of reported events and their corresponding standard error bars for the 5-tone *Up*, *UpDown*, and *Mixed* sequence patterns at each presentation rate. A greater mean number of reported events for the *UpDown* pattern can also be seen for the 2- through 4-tone sequences in Figure 5. This tendency contributed to a main effect of sequence pattern within both the general and pattern analyses, $F(1, 9) = 28.52, p < .001$ and $F(1.22, 11.01) = 14.90, p < .05$, respectively.

An explanation for the greater number of reported tonal elements within the *UpDown* sequence is that the extra repetition of the middle-frequency tones led listeners to mistakenly report more tonal events because they counted the repeated middle tones as part of the sequence, even though participants were instructed to report the number of distinct events and not duplicate pitches. Although the hypothesis that no main effect will be found for different sequence patterns was not confirmed by the results, it is important to realize that the *UpDown* pattern actually highlights the strength of the perceptual limit to only three streams. At slower presentation rates for the *Up* and *Mixed* conditions participants reported four or five events, whereas in the *UpDown* condition they reported around seven events. However, at fast presentation rates, the mean number of reported events for all pattern conditions is limited to three streams. Thus, the effect of presentation rate is even more obvious for the *UpDown* condition as the mean number of reported events decreases.

Sequences that contained more tonal elements had a greater mean number of reported events for the *UpDown* pattern, whereas sequences with fewer tones had less variation in participants' responses. This tendency contributed to an interaction between sequence pattern and number of tonal elements for the general analysis, $F(2, 18) = 7.08$, $p < .05$, which can be seen in Figure 5. Pair-wise comparison of means revealed that the mean reported number of events was significantly higher in the *UpDown* condition than the *Up* condition for each consecutive number (3- to 5-tone) condition ($p < .01$). A possible explanation for this interaction is that sequences with fewer tonal elements will not have as many middle-frequency tones that repeat during the *UpDown* sequence pattern.

For the *UpDown* sequence pattern there was a drastic decrease in the mean number of reported events as presentation rate increased. This can be seen in Figure 5 for the 3- to 5-tone sequences and in Figure 6 for the 5-tone sequences. This tendency contributed to an interaction between presentation rate and sequence pattern in the general and pattern analyses, $F(3, 27) = 11.55, p < .001$ and $F(2.97, 26.69) = 7.87, p < .001$, respectively. Pair-wise mean comparisons revealed that the number of reported events was greater for the *UpDown* pattern than the *Up* pattern at the 125 and 250 ms presentation rates, $p < .01$ and $p < .001$, respectively.

Perceived Sequence Structure Task. The number of participants that selected a particular sequence structure for the 62.5 ms 5-tone sequence stimulus was tabulated. These frequencies are reported in Figure 4, in parentheses, next to the corresponding image representing a sequence structure. Only one of the participants selected *Image A*, which depicts all five tonal elements intact, suggesting this participant retained the ability to recognize more than three streams. This was not true for any of the remaining participants, who showed some collapsing of the middle-frequency elements into a stream. Only one of those nine listeners reported hearing a single event in the middle-frequency stream, as this participant chose *Image G*. The other eight participants selected images that still included all three tonal elements within the middle-frequency stream. However, temporal information for the elements within that stream was compressed.

The eight participants that retained the middle events within a single stream responded in a way that suggested a variety of perceptual transformations of temporal information. Although participants were not reporting the same transformation, some sort

of temporal compression is occurring for eight of the ten participants. These results suggest that some of the frequency information remains intact for listeners, even as temporal information is lost within the stream. Three participants selected *Image B* of Figure 4, which shows the middle tones with compressed tone durations. This selection suggests that the listeners perceived the correct frequency information for the middle tones, however streaming caused them to be perceived at shorter tone durations. Two of the participants selected *Image C*, which depicts the middle tones presented simultaneously, but with the proper tonal durations. This choice suggests that frequency information remained intact for the listener as the tones were perceived simultaneously. The remaining three participants selected *Image E*, which shows a pitch glide between the tonal frequencies present. This choice suggests that the listeners perceived the tonal elements as a glide in pitch across the frequencies of the tones and lost temporal information because the glide would have been perceived at the duration of only one of the tones.

CHAPTER 3

Experiment 2: Stream Segregation by Frequency and Timbre

As it was expected that listeners would not perceive more than three streams in Experiment 1, it presents the question of whether more than three concurrent streams can occur. It is logical to test this limitation by adding an additional cue. Timbre has been shown to be a strong segregation cue (e.g., see Bregman, 1990; Singh, 1987; Bregman *et al.*, 1990), making it ideal as an additional cue for this study.

Method

Participants. A set of 14 adult listeners from James Madison University were selected. The criteria for participant selection, such as age and compensation were the same as in Experiment 1. Participants were asked to complete the same Musical History Questionnaire as was described in Experiment 1, to help identify if aspects of training for musical performance (e.g., knowledge of musical theory, amount of relevant experience with instruments that have timbres the same or similar to the instrument timbres in this study) correlated with participants' ability to correctly identify the number of tonal elements within the sequences.

Stimuli. Some aspects of the stimuli for Experiment 2 mirrored the parameters provided for stimuli in Experiment 1. For example, the stimuli in Experiment 2 reflect the same sample rate and low-pass filter settings, as well as the same fundamental frequencies and individual tone durations as in Experiment 1. Furthermore, sequences for Experiment 2 were constructed using the same arpeggiation technique as described for Experiment 1.

Timbre cues denote different sound sources which can become grouped into streams. Timbres for the tones in Experiment 2 were based on the spectral envelopes of musical instruments. It was important that these timbre cues were maximally different from one another in order to create an optimal situation where multiple sound sources could be perceived. Multi-dimensional scaling solutions for musical instrument timbre (e.g., Grey & Gordon, 1978; Iverson & Krumhansl, 1993) were used to select instrument timbres for this study that would be optimally different in spectral envelope from one another. The timbres for this study were created based on the spectral envelopes of the violin, oboe, clarinet, grand piano, and trumpet. The instrument recordings used to obtain the spectral envelopes for this investigation are available in Ableton Suite 8 (grand piano) as well as the Ableton Essential Instrument Collection Version 2 (violin solo legato, B^b clarinet solo legato, oboe solo legato, and trumpet solo legato). The measurable formants and corresponding bandwidths for these instrument tones are provided in Table 2. As can be seen in this table, the location of spectral peaks is very different across tones. For example the center frequency of the second formant of the clarinet (2644 Hz) is 858 Hz removed from the nearest spectral peak for the adjacent piano tone (1786 Hz) of the sequence.

The natural instrument tones themselves could not be used for this experiment due to the short tone durations within each sequence. The attack phase for some instruments is much longer than the entire tone duration would allow. This would result in little to no remaining energy in the signal. Instead, tones were created using a static spectral envelope based upon the average spectral envelope for each naturally occurring instrument tone. For the sake of simplicity, each artificially created tone will be referred

to by the instrument that it is based upon. For example, the phrase “violin tone” will be used to refer to the artificial tone created using the average spectral envelope of an actual violin tone. The average spectral envelope for each instrument tone was determined by spectral analysis. Relative dB levels were recorded for the average amplitude for each harmonic, which were re-synthesized using Camel Audio’s Alchemy.

Trials for Experiment 2 consisted of 2- to 5-tone sequences presented in the *UpDown* pattern from Experiment 1. The *UpDown* pattern was selected for this task because it creates increased repetitions of the middle-frequency tones, which provides the possibility that it may aid in segregation of the middle-frequency tones. Multiple timbre conditions were used in this experiment. First, to corroborate the results from Experiment 1 with the new set of participants, sawtooth tone sequences were included. This can be seen as Condition A in Figure 7, which depicts the timbre conditions that will be used for Experiment 2. Then next condition used a fixed instrument-derived timbre across all tones. For this condition the spectral envelope of a piano¹ was applied to all of the tones in the sequence. The piano is comparable to the sawtooth because they both cover a full spectrum and contain all integers multiples of the harmonics. Pianos have inharmonic energy due to the physical design of the instrument. This inharmonic energy was not included in the synthesis method. The fixed timbre condition was included to determine if participants are limited to only three streams for 4- and 5-tone sequences when a single timbre is present. The fixed timbre condition for the 5-tone sequence is depicted in Figure 7 as Condition B.

In an attempt to create a situation where more than three streams could be perceived, a multiple timbre condition was created with five different timbres present. This condition is depicted in Figure 7 as Condition C. For this condition a unique spectral envelope shape was assigned to each F_0 in the sequence, permitting up to five different sound sources to potentially be perceived. Each pitch (C2, B2, A[#]3, A4, and G[#]5) was assigned a specific instrument timbre (Clarinet, Piano, Oboe, Trumpet, and Violin respectively) across all sequences in this condition.

Procedure.

For each trial in this task participants were presented a sequence containing one of the timbre patterns and asked to report how many tonal elements they perceived. This task followed the same procedure as the Complex Tone Number Judgment Task in Experiment 1.

Participants received 4 blocks of trials, each containing 72 randomized trials. Each of the conditions was produced at each presentation rate 8 times within the experiment. Each block lasted approximately 12 minutes. Upon completing each block, participants were instructed to take a rest break before the next block of trials started. Stimuli again were delivered, and data were collected, using the *Music Experiment Development System (MEDS 2002-B-1, Kendall, 2002)*.

It was hypothesized that at faster speeds the number of tonal elements reported for the fixed timbre condition and the sawtooth condition would be limited to three streams. Listener's performance for the fixed timbre condition was expected to be similar to that

of the sawtooth condition because the two conditions differ only by the general shape of the spectral envelope across tones. Because the presence of multiple timbres provides strong cues for segregation (e.g., see Bregman, 1990; Bregman *et al.*, 1990; Singh, 1987) it was hypothesized that listeners would be more likely to report more than three streams for sequences with the multiple timbre condition, even as presentation rate increased.

Results & Discussion

The mean numbers of perceived events were calculated for each combination of presentation rate, timbre condition, and number of tones for each participant. Given the absence of an orthogonal design, it was not possible to include all of the conditions within a single analysis. More specifically, the 5-tone sequences were the only sequences to include the sawtooth timbre. As a result, data were submitted to two repeated measures ANOVAs. The first analysis included all manipulated variables except the 5-tone *Mixed* condition, and will be referred to from here on out as the *general* analysis. This consisted of a 4 x 4 x 2 ANOVA with presentation rate (31.25, 62.5, 125, and 250 ms/tone), number of tones within the sequence (2, 3, 4, or 5), and timbre condition (piano timbre only v. different timbres). The second analysis was conducted to permit the inclusion of the sawtooth timbre in the 5-tone sequence, and will henceforth be referred to as the *timbre* analysis. This consisted of a 4 x 3 ANOVA with presentation rate (as specified for the general analysis) and timbre condition (different timbres, piano timbres, and sawtooth waves).

As in Experiment 1, evidence that participants were sensitive to the number of tonal elements was obtained in Experiment 2. Specifically, more events were reported for

sequences that contained more tonal elements. This can be easily seen in Figure 8, which depicts the mean numbers of reported events, along with corresponding standard error bars, as a function of presentation rate, number of tones, and timbre condition. (Data from the 5-tone sawtooth condition is excluded from this figure to avoid confusion, and is depicted with the other 5-tone conditions in Figure 9). This response tendency contributed to a main effect of the number of tones, $F(1.1, 14.0) = 75.68, p < .001$. Pair-wise comparisons of means revealed that every one-tone increase in the length of the sequence pattern produced a significant increase in the number of reported events ($p < .001$). Thus, it is clear that listeners were able, at least under some of the (slower) conditions, to report more than three events.

Inclusion of a different timbre for each tonal element within the sequence was hypothesized to provide further cues for stream segregation, thereby creating a situation in which additional streams could occur. However this was not the case. The mean number of reported events and their corresponding standard error bars for the 5-tone sequences at each of the timbre conditions is depicted in Figure 9. As can be seen in the figure, there is no statistically meaningful difference between the mean numbers of reported events across timbre conditions. There was no main effect for the timbre analysis, $F(2, 26) = .31, p > .05$, or in the general analysis, $F(1, 13) = 3.9, p > .05$. However, there was an interaction between timbre condition and the number of tonal elements, $F(2.75, 35.75) = 3.40, p < .05$. Pair-wise comparisons of means further revealed that the number of reported events for the 3-tone sequence was significantly greater for the piano timbre condition than the all timbre condition ($p < .01$). The lack of

a main effect for timbre suggests that the inclusion of various timbres was not a strong enough cue within sequences to promote additional segregation.

As expected, the average number of reported tonal events tended to decrease as presentation rate increased. As seen in both Figure 8 and Figure 9, the mean number of reported events decreases as presentation rate increases. This pattern of results contributed to the main effect of presentation rate in both the general and timbre analyses, $F(3, 39) = 17.77, p < .001$ and $F(1.29, 16.36) = 27.95, p < .001$ respectively. Pair-wise mean comparisons confirmed that the mean reported number of events was significantly lower for the 31.25 ms condition than for the other three conditions ($p < .05$). Additionally, the number of reported events was significantly lower for the 62.5 ms condition than the 125 and 250 ms conditions ($p < .05$). As can be seen in Figure 8, the number of reported events was more drastically affected by presentation rate as the number of tonal elements within the sequences increased. This tendency contributed to the interaction between the number of tones and presentation rate, $F(9, 117) = 25.49, p < .001$. Pair-wise comparisons of means clarified the nature of this interaction. There was a significant decrease in the mean number of reported events for each increase in presentation rates for the 4- and 5-tone sequences ($p < .05$). The mean number of reported events was significantly less for the 31.25 ms condition than any of the other presentation rate conditions for the 3-tone sequence ($p < .01$). Similar to the findings of Experiment 1, the main effect of presentation rate suggests that participants had an increased difficulty segregating tones as they became shorter in duration and were presented more rapidly within the sequence.

CHAPTER 4

General Discussion

This study initially sought to determine the perceptual limit on the number of concurrent auditory streams that listeners can attend to when presented with rapid sequences of complex tones, and then to assess whether or not this limit can be overcome by including timbre as an additional stream segregation cue. Regarding the initial goal of the investigation, the perceptual limitation of listeners to three auditory streams at fast presentation rates appears to extend to sequences that contain complex tones. This conclusion is supported by several findings from Experiment 1. In the Number Judgment Task listeners reported perceiving no more than three auditory streams for sequences at the 31.25 ms condition, which can be seen in Figure 5. Furthermore, in the Perceived Sequence Structure Task nine of the ten participants showed a pattern of stream segregation that impacted perception of the middle-frequency stream. Eight of these listeners reported a complex middle-frequency stream. Formation of this complex stream was typically accompanied by a perceptual alteration of temporal information. This can be seen in the image choices depicted in Figure 4 that were most frequently selected by participants (Images *B*, *C*, and *E*). These selected structures all depict the middle-frequency tones to either overlap in time with one another or alternatively, reflect shorter perceived tone durations than the remaining (upper- and lower-frequency) tones in the sequence.

Despite the addition of different timbres in Experiment 2, which was expected to aid stream segregation, the average number of reported auditory streams at faster

presentation rates was still limited to approximately three events (i.e., a mean of less than 3.5 reported tonal events). Figure 9 clearly depicts that participants' responses to the 5-tone sequence did not statistically differ between the stimulus conditions that reflected a single timbre across all tones and those consisting of multiple timbres within the sequence. The same relationship was generally seen across the other number-tone conditions depicted in Figure 8, the one exception being the 3-tone sequences at 125 ms. At this presentation rate for the 3-tone sequences, participants reported more tonal events for the piano-timbre sequence than the multiple-timbres sequence. Given that participants' mean number of reported events only differ significantly for the 125 ms 3-tone condition, it is likely that this is anomalous.

Several arguments could be made that represent potential alternative explanations as to why the introduction of timbre variations in Experiment 2 failed to increase the number of streams reported by listeners. For example, it is possible that the various timbres in Experiment 2 did not aid in stream segregation because a fixed amplitude envelope was used for all of the tones. Previous studies suggest that the attack phase of a tone plays a significant role in the perception of its timbre (e.g., Grey, 1977; Grey & Gordon, 1978; Grey & Moorer, 1976). For example, Grey and Moorer (1976) compared original and resynthesized tones with and without the attack phase present. On a two-dimensional scaling solution for the distance estimation, the greatest magnitude of difference was between the removed attack phase condition and all other conditions. Based on these results, Grey and Moorer (1976) described the onset pattern of instrumental tones to be of extreme importance. Additional multidimensional scaling studies, such as, Samson, Zatorre, and Ramsay (1997) have also found that altering the

logarithmic rise time of a tone significantly affects the perception of its timbre, suggesting that the initial portion of a tone plays an important role in the timbre definition.

The lack of a unique attack transient for the tones used in the current investigation could have decreased timbre differences between tones. However, at faster presentation rates, using natural attack transients and decay functions would have resulted in some tones being too brief to allow completion of their attack transient. Such occurrences would leave little to no energy on which to base decisions about instrument recognition. Therefore, the inclusion of individual attack transients for each timbre was not possible for the tone duration used in this study.

Another important characteristic of timbre that was not available to listeners in Experiment 2 was change in the spectral envelope over time, known as spectral flux. Several studies have suggested that spectral flux is one of the most salient physical parameters related to timbre discrimination (e.g., Grey, 1977; McAdams, Beauchamp, & Meneguzzi, 1998; McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995). For example, Grey (1977) presented listeners with pairs of original and resynthesized tones in a discrimination task to determine which dimensions of timbre were most salient. Analysis of the results provided a three-dimensional scaling solution, with one of these dimensions related to the level of spectral fluctuation through time.

Unfortunately, since spectral flux is temporally defined, there was no way to correctly convey natural levels of flux in the short tone durations used in the current investigation. As a result, lack of spectral flux may have reduced discriminability across

the various timbres. On the other hand, by removing spectral flux from tones in the current investigation, each tone's static spectral envelope remained maximally different through the duration of each tone. This aspect of stimulus generation actually should have helped to maximize timbre differences between sequenced tones in Experiment 2. In other words, while it is acknowledged that the loss of spectral flux information represents removal of a critical aspect of natural instrument timbres, there should have remained very salient cues to distinguish the timbres within the experiment.

One might argue that tones in a natural setting would most likely have a slightly longer duration than those that were used in the current investigation. However, this assumption would not take into consideration that tones in a natural setting would come from different sound sources and therefore could be presented almost simultaneously. The resulting overlap of tonal events would make segregation even more difficult for listeners. The use of non-overlapping tones in the current investigation should have exaggerated the differences between tones relative to such naturally-overlapping conditions. This, coupled with perceptually large differences in spectral envelope shape for each tone, should have provided participants with optimal, controlled conditions for additional stream segregation to occur.

It also might be argued that under more natural listening conditions, listeners often have the advantage of sound sources coming from different locations, thereby allowing an additional localization cue to be used. However, this possibility does not seem likely given existing empirical evidence. For instance, Bregman and Steiger (1980) found that listeners' localization of a target tone was influenced by an accompanying tone

when the accompanying tone streamed with the target tone. This suggests that streaming decreases listeners' ability to use localization cues, making it highly unlikely then that localization would aid tone segregation. Therefore, it also is not probable that localization cues would have aided in any additional segregation in the current investigation.

Localization has generally been discussed as a weak cue for discrimination. Even to the point where listeners quite often must determine what a sound source is before they are able to determine where the source originated (e.g., see Yost, 2000).

Generally, the results from Experiment 2 suggest that the addition of timbre cues did not aid listeners in perceiving more than three streams at fast presentation rates. However, the results from a secondary, follow-up task involving laboratory personnel raise the possibility that timbre has an effect on the nature of the perceptual streams that were formed. This secondary task was included to gain additional, albeit anecdotal, evidence about stream formation. Three members of the laboratory were asked to report the way in which they perceived streaming to occur in the 5-tone sequences. Laboratory personnel were selected for this task because they could openly interpret and easily report what they perceived in a consistent manner. These listeners were randomly presented with both the piano-timbre and the multiple-timbre conditions for the 5-tone sequences at each presentation rate and were asked to draw their own depiction of the way in which the tonal elements were perceptually segregating.

All three listeners agreed that only three streams in the sequence could be perceived when event durations were shortened to 62.5 ms or less. In the absence of timbre differences between tones in the sequence, all three listeners reported perceiving

the two outer tones within their own streams as the middle tones formed the middle-frequency stream. However, there was some variation in the way these listeners perceived each of the three streams when the sequence contained tones with different timbres from one another. One listener still reported the middle tones to be streaming together despite timbre differences between tones. However, the other two participants reported perceiving at least one of the middle tones as part of an outer (i.e., higher- or lower-pitch) stream. One of these listeners claimed that one of the middle tones formed a stream with the adjacent lowest-frequency tone, while the other lab member reported that one of the middle tones formed a stream with the adjacent highest-frequency tone.

Individual differences in participant history could have contributed to some of the observed variation in participants' responses. One such individual difference that has been suggested by previous studies is musical training (e.g., Bever & Chiarello, 2009; Huron, 1989; McAdams *et al.*, 1995). Musical experience could aid listeners in their ability to segregate tones, which could potentially contribute to the perceptual report of additional auditory streams.

Range of musical training differed considerably across the two experiments. For Experiment 1, standard deviation in the amount of musical experience between participants was low, ($M = 4.4$ years; $SD = 2.12$; range = 0 to 8 years). However, the standard deviation in the amount of musical experience between participants in Experiment 2 was much higher ($M = 5.62$ years; $SD = 5.14$; range = 0 to 14 years). This pronounced individual difference in the amount of musical training across participants in Experiment 2 is largely due to the inclusion of a few participants with large amounts of

musical training. The inclusion of such listeners permitted a preliminary analysis of whether there is a relationship between extensive training and greater numbers of reported auditory streams within the tone sequences.

Average responses for some of the listeners in Experiment 2 who had the most extensive musical training were plotted against the average responses in the Number Judgment Task. A comparison of these participants' responses is provided in Figure 10. Participant 5 reported a total of 14 years experience across 4 instruments. This participant's mean number of reported tones for the 31.25 and 62.5 ms presentation rates is accurately closer to 5 events. However there is still a decrease in the mean number of reported events from the reported for the 125 and 250 ms presentation rates, suggesting that some form of compression is still occurring due to streaming. Participant 8 reported 11 years of experience across 2 instruments. Unlike Participant 5, the mean number of reported events for the 5-tone sequence for Participant 8 was similar to that of the group mean at each presentation rate.

Participant 11 reported 7 years of musical experience across 4 instruments. During debriefing, this participant mentioned that it was easier to correctly identify the number of tonal elements in the sequences that contained multiple timbres. This participant did report more events for the multiple-timbre condition at the 31.25 ms presentation rate, but reported less events than the single-timbre conditions at the 62.5 ms presentation rate. Overall this participant reported more than five events at each of the presentation rates of the 5-tone sequence, suggesting the possibility of some sort of enhanced ability.

Interestingly, Participant 5 reported more tones for the piano condition than the other two timbre conditions. This participant reported 10 years of training with the piano, which may have aided in identification of the specific timbre as a result of prolonged exposure. However, Participant 8 reported 11 years of piano training and did not exhibit any benefit from this when reporting the number of perceived events in the piano timbre condition.

Although Participants 5 and 11 reported more than five events for the 5-tone sequences, both participants reported perceiving fewer events as presentation rate increased. This suggests that musical training may have aid these participants in reporting the number of events, but only to an extent. Similarly, Huron (1989) found that musicians were slightly more accurate than non-musicians at reporting the correct number of streams at slower presentation rates, yet in his study, all participants were limited to perceiving only three streams at faster presentation rates.

The perceptual limitation in the number of auditory streams found in this study is congruent with the findings of previous studies. For instance, Hall and Schuett (2008) found that listeners were limited to perceiving only three streams at fast presentation rates in sequences consisting of sine tones. However, the results of their study could not fully suggest the existence of this limitation outside of a laboratory setting because sine tones do not occur naturally. The current investigation extends initial results by using complex tones to mimic more naturally occurring events. The observance of a corresponding finding with complex tones suggests that listeners in a natural environment are limited to three auditory streams when vibrational components are presented in rapid succession.

Findings from the current investigation also agree with the findings obtained from more ecologically valid, perceptual studies based upon recordings of natural musical performances. For example, Huron (1989) found that listeners were perceptually limited to three auditory streams when asked to report how many musical voices they perceived within a selected piece of music. Although it is acknowledged that tonal elements in the current investigation were less natural than Huron's recordings, this permitted much greater control over stimulus generation allowing sequences to be designed in a manner that should maximize stream segregation. When taken together, the findings of Huron (1989) and the current investigation appear to indicate that timbre might not be a strong enough segregation cue to promote more than three streams at fast presentation rates. In fact, timbre might be inadequate in promoting more than three streams, regardless of how naturally identifiable or how optimally different it is for tones within a sequence.

Major findings from the current investigation also concur with those from the only other controlled study that directly investigated perceptual organization in rapid sequences of three or more tones under conditions where they might be expected to segregate into different streams. Specifically, when asking participants to listen for temporal irregularities between tones in a sequence, Brochard *et al.* (1999) found that it was harder for listeners to focus on the inner or middle subsequences than the outer subsequences. It was suggested that information about the middle events was harder for listeners to perceive because these events were streaming together while the outer events were segregated from the middle stream. This observation is similar to the obtained pattern of participants' responses for the Perceived Sequence Structure Task of Experiment 1. Eight out of the ten participants reported that the middle tones were

streaming together. A slight difference between studies is that Brochard *et al.* (1999) used pure tones as their stimuli, while Experiment 1 of the current investigation used complex tones. However, stimuli for both studies consisted of fixed timbres for all tonal elements within each sequence. Thus, the tendency to group middle-frequency components into a single stream seems to hold constant for cases where a timbre is shared across events.

The collection of all middle-frequency components into a single stream may not always occur in instances where timbre is varied across events within a stream. Anecdotal reports given by lab personnel in Experiment 2 suggest that the inclusion of timbre differences among tones in a sequence could be altering the way in which the elements are streaming. Therefore it is possible that the tendency for a middle stream to emerge, as suggested by Brochard *et al.* (1999), primarily occurs when only one timbre is present in the sequence.

Although it seems quite apparent that listeners are perceptually limited in the number of streams they can attend when tonal elements are presented rapidly, the streaming literature has generally failed to address this limit. Past studies have suggested that the addition of timbre differences between tones will aid in their segregation (e.g., Bregman & Campbell, 1971; Singh, 1987; Warren *et al.*, 1969). However, these studies were primarily interested in whether or not stream segregation occurred, they did not address the number of perceivable streams present when streaming did occur. The limit in the number of reported streams observed in the current investigation suggests that there may be some natural boundary on the number of perceivable streams, possibly as a function of presentation rate. It is possible that this natural boundary occurs regardless of

how segregation cues are available. Alternatively, the cumulative impact of multiple segregation cues, which are presumably additive, might be reduced as conditions of the sequence become more naturally complex.

When considered collectively, the current investigation and the limited existing literature indicate that (at least for the stimulus conditions that have been evaluated) perception of more than three auditory streams during fast presentation rates is not likely to occur. This was the case despite relying on stimuli that were designed to provide optimal conditions for additional stream segregation. If this represents a general perceptual limitation, then the most fundamental question that is raised concerns how to accurately maintain perceptual organization in situations with multiple sound sources. Probably the most effective way to correctly organize auditory streams would be, if at all possible, to limit the number of sound sources. Although the suggestion is somewhat of a stretch, if a listener who is driving their car wants to correctly perceive important information such as direction from a GPS unit, they would potentially have the best chance of safely doing so by limiting other sound sources, such as choosing to eliminate cell phone use or to turn off the radio.

Obviously, listeners do not always have the luxury of being able to control the number of sound sources present. For instance, soldiers on a battlefield are surrounded by many sound sources at one time, often resulting in the loss of information that is most vital to them. Such scenarios can have dire consequences, thus highlighting the importance and need for future research into ways of conveying information to listeners

without it being lost due to general perceptual constraints on the number of simultaneous auditory streams that can rapidly be perceived within the environment.

A useful direction for future research would be to determine if the presence of the attack transient or spectral flux for tones with slightly longer durations results in aided segregation. If that is the cause, it could be possible that the lack of information from the amplitude envelope due to short tone durations was responsible for the ineffectiveness of timbre as a cue in the current investigation. Additionally, further investigation into how tones are segregating within a sequence when multiple timbres are included could provide useful information into the role timbre differences play. The current investigation only received anecdotal evidence from three listeners, but perhaps with additional participants it could become clearer as to how timbre cues affect the formation of streams within a sequence. A better understanding of timbre's effect on stream formation could help establish which specific timbres are more likely to segregate.

For now, the major findings of the current investigation indicate what appears to be a basic perceptual limit in the number of concurrent auditory streams that can be maintained in sequences at rapid presentation rates. This perceptual limit might not be easily overcome by the inclusion of additional segregation cues. When these results are combined with the suggested avenues of future research with multiple segregation cues, additional clarification should be provided on whether or not this limit truly represents an absolute boundary.

Footnote

¹ Although the violin timbre would be ideal because of its close similarity to the sawtooth (Askenfelt, 1991), limitations in the instrument's range prevent it from naturally playing all of the frequencies in the 5-tone sequence. The piano is physically able to play all of the fundamental frequencies in this study; therefore it was selected for this task.

Table 1

Harmonic Frequencies for the 65 Hz and 124 Hz Fundamental Frequency Tones

| <u>Harmonic</u> | <u>C2</u> | <u>B2</u> |
|-----------------|-----------|-----------|
| | 65 | 124 |
| 1 | 130 | 248 |
| 2 | 195 | 372 |
| 3 | 260 | 496 |
| 4 | 325 | 620 |
| 5 | 390 | 744 |
| 6 | 455 | 868 |
| 7 | 520 | 992 |
| 8 | 585 | 1116 |
| 9 | 650 | 1240 |
| 10 | 715 | 1364 |
| 11 | 780 | 1488 |
| 12 | 845 | 1612 |
| 13 | 910 | 1736 |
| 14 | 975 | 1860 |
| 15 | 1040 | 1984 |

Table 2

Measurable Formants (F) and Corresponding Bandwidths (B) for Each Instrument that Constituted the Different Spectral Envelopes in Experiment 2

| | Clarinet | Piano | Oboe | Trumpet | Violin |
|-----------|-----------------|--------------|-------------|----------------|---------------|
| F1 | 891 | 711 | 1196 | 963 | 1731 |
| F2 | 7644 | 1786 | 2603 | 2306 | 3507 |
| F3 | 5244 | 3623 | 4584 | 3081 | 5675 |
| F4 | | 5892 | | 4444 | 7631 |
| F5 | | | | 5719 | 10630 |
| F6 | | | | | 12977 |
| B1 | 526 | 611 | 325 | 303 | 515 |
| B2 | 696 | 267 | 528 | 283 | 915 |
| B3 | 2275 | 419 | 873 | 532 | 288 |
| B4 | | 2375 | | 666 | 715 |
| B5 | | | | 613 | 1148 |
| B6 | | | | | 715 |

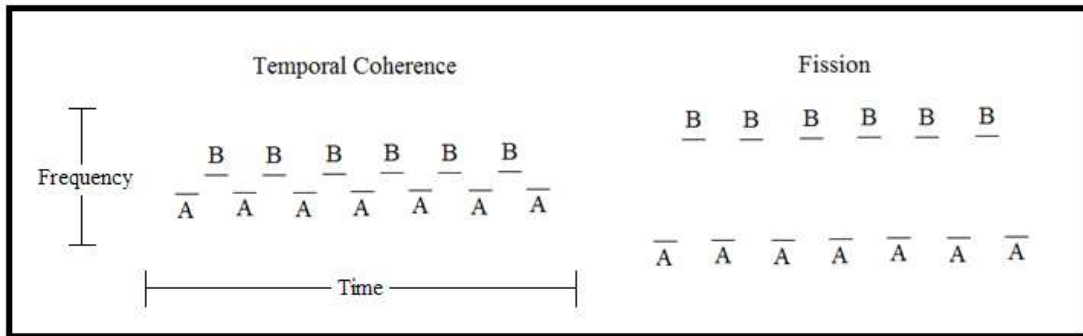


Figure 1. Temporal Coherence and Fission as described by van Noorden (1975).

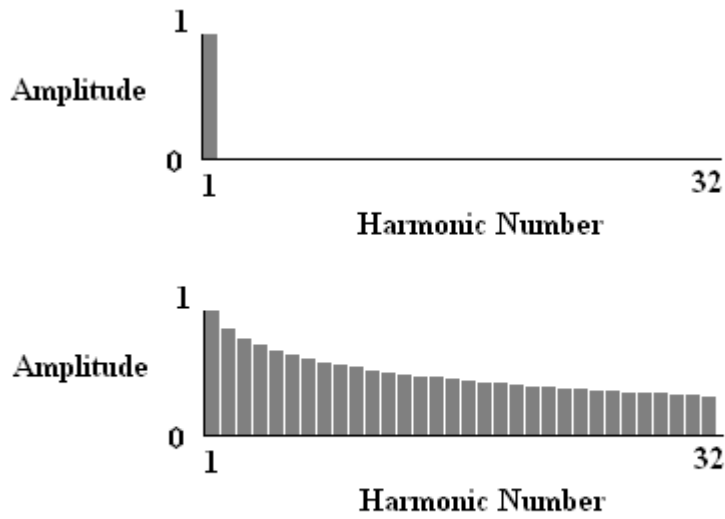


Figure 2. The sine wave (top) and the sawtooth wave (bottom) represented by the amplitude spectra for each.

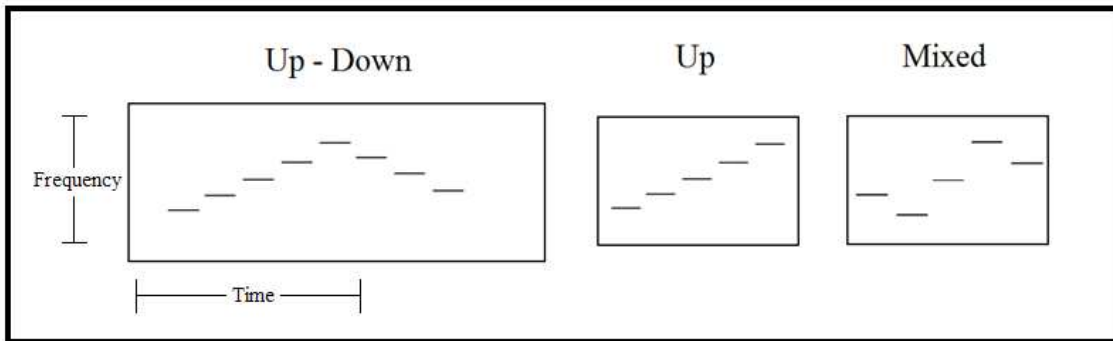


Figure 3. Depiction of the tonal elements within the 5-tone *UpDown*, *Up*, and *Mixed* sequence patterns used in Experiment 1.

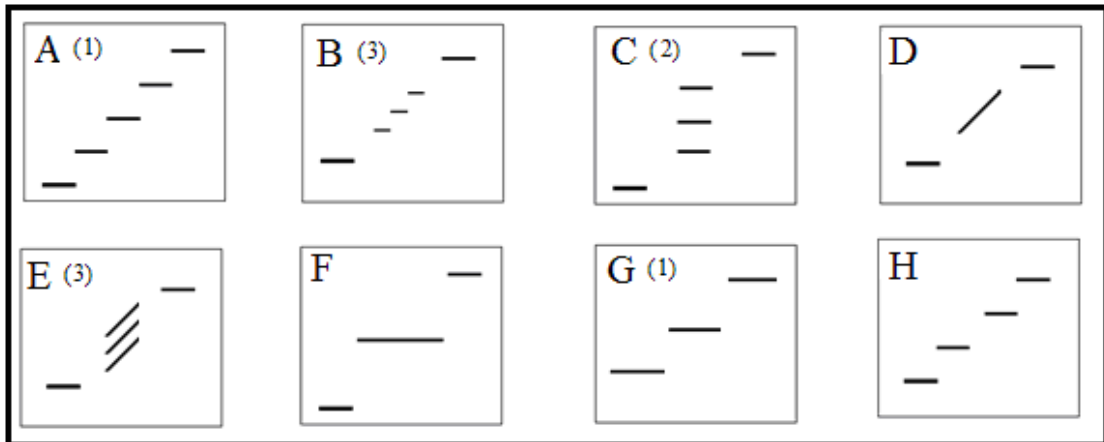


Figure 4. The eight image choices representing the portion of the sequence, as it rises in pitch, presented to listeners to choose from in the Perceived Sequence Structure Task. In these images frequency is represented vertically, and time, horizontally. The number in parentheses next to some of the images represents the number of participants that reported perception of that sequence structure in response to the 62.5 ms 5-tone sequence that was likely to promote segregation.

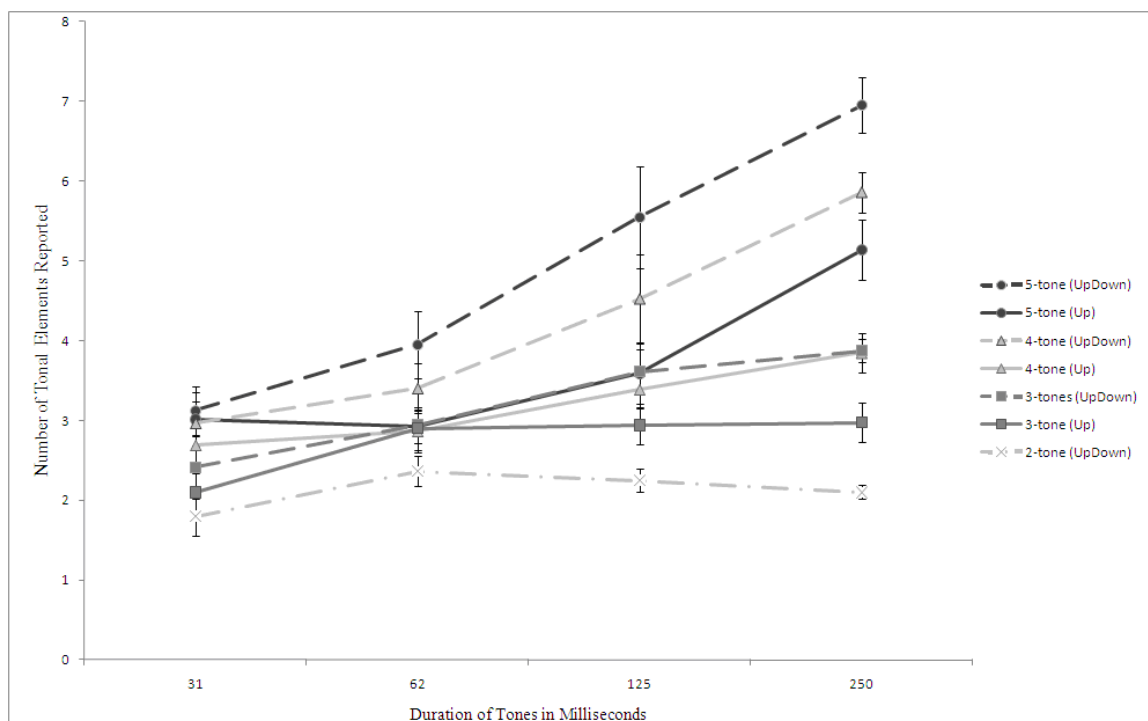


Figure 5. Mean number of reported events (and corresponding standard error bars) as a function of the number of tones (2-5) and pitch direction within the tone sequence (Up and UpDown).

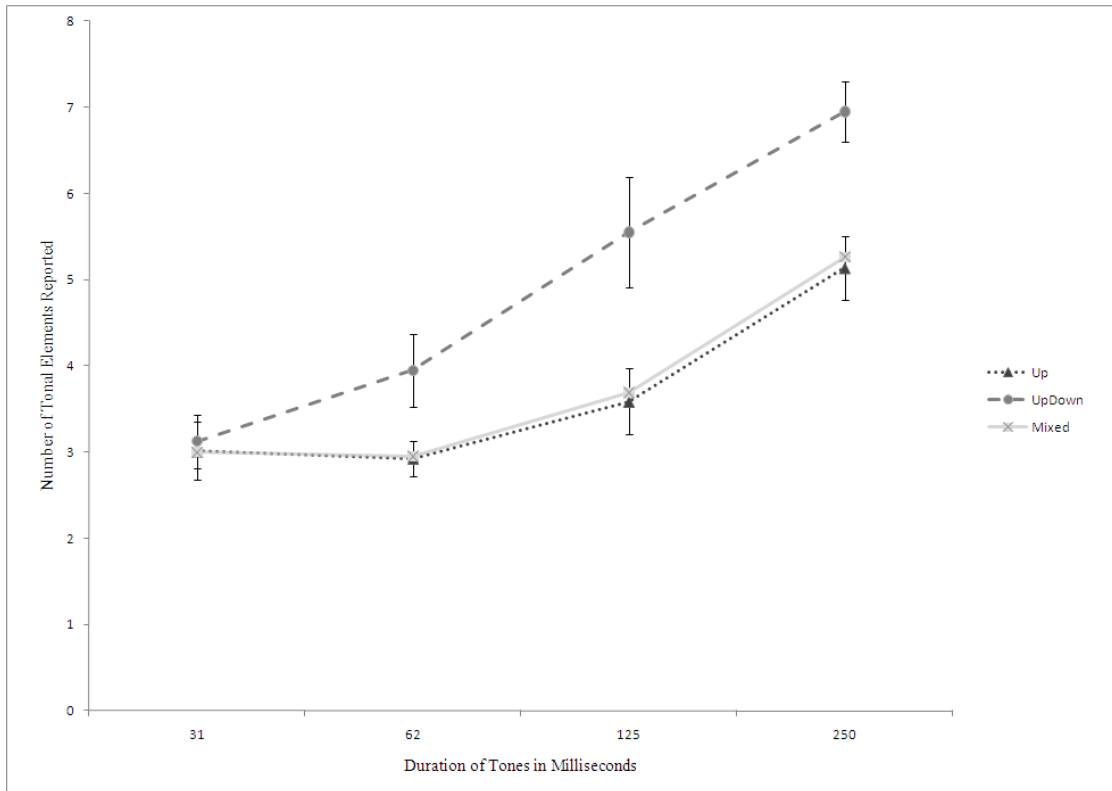


Figure 6. Mean number of reported events (and corresponding standard error bars) as a function of pitch direction (Up, UpDown, and Mixed) for the 5-tone sequences.

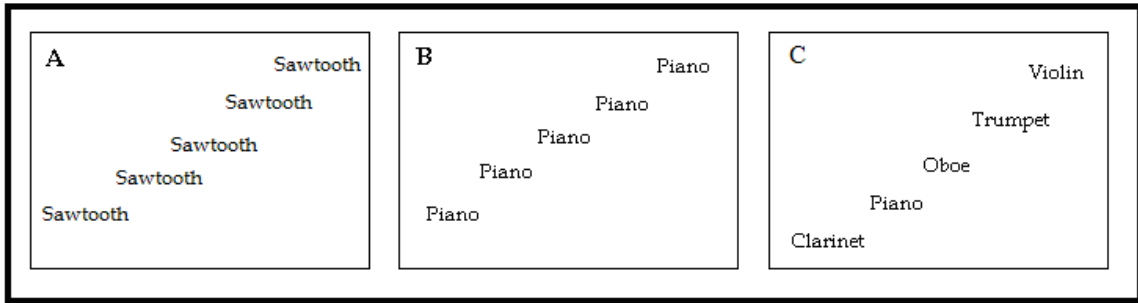


Figure 7. Depiction of the timbre conditions in Experiment 2. In these images frequency is represented vertically, and time, horizontally.

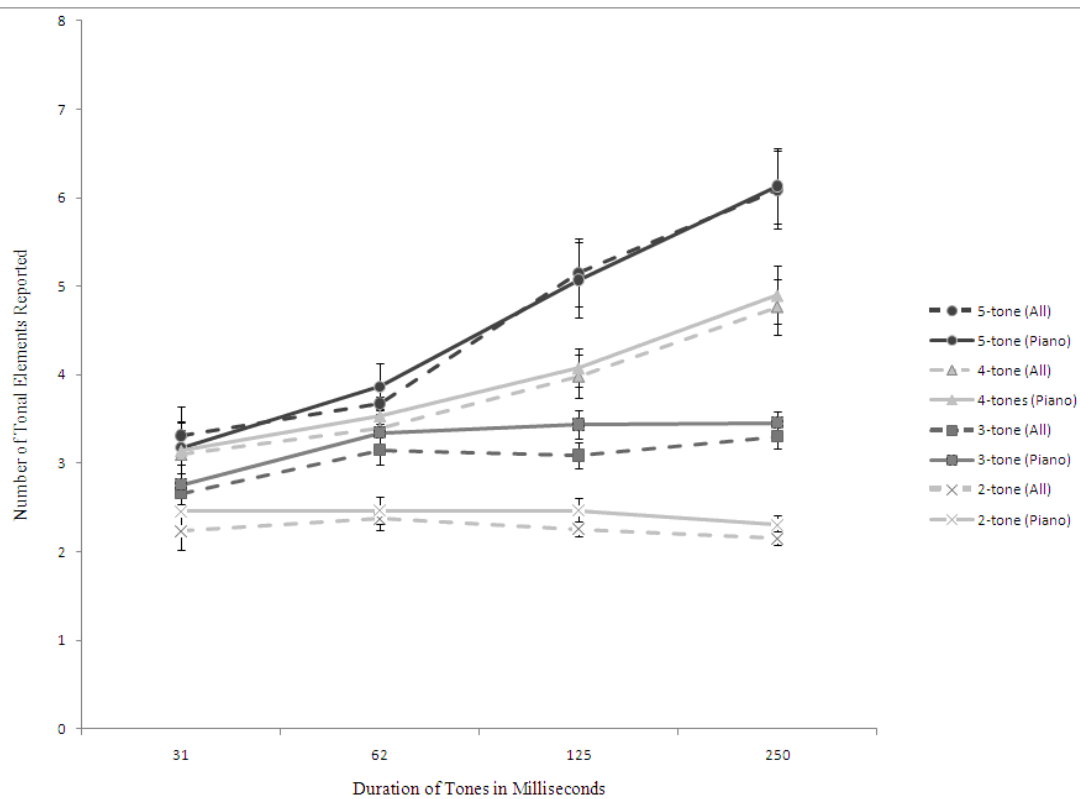


Figure 8. Mean number of reported events (and corresponding error bars) as a function of the number of tones (2-5) and sequence timbre pattern (all different timbres and piano timbre).

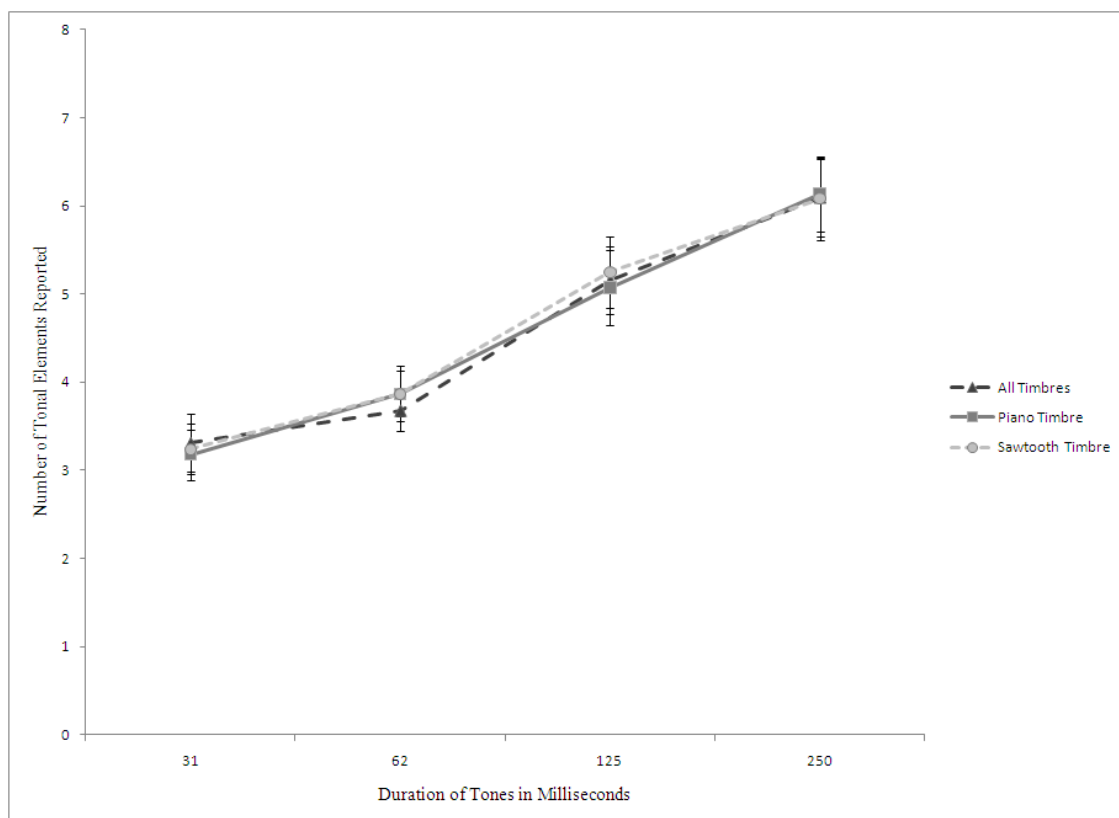


Figure 9. Mean number of reported events (and corresponding standard error bars) as a function of sequence timbre pattern (all tones different timbre, piano timbre, sawtooth tones) for the 5-tone sequence.

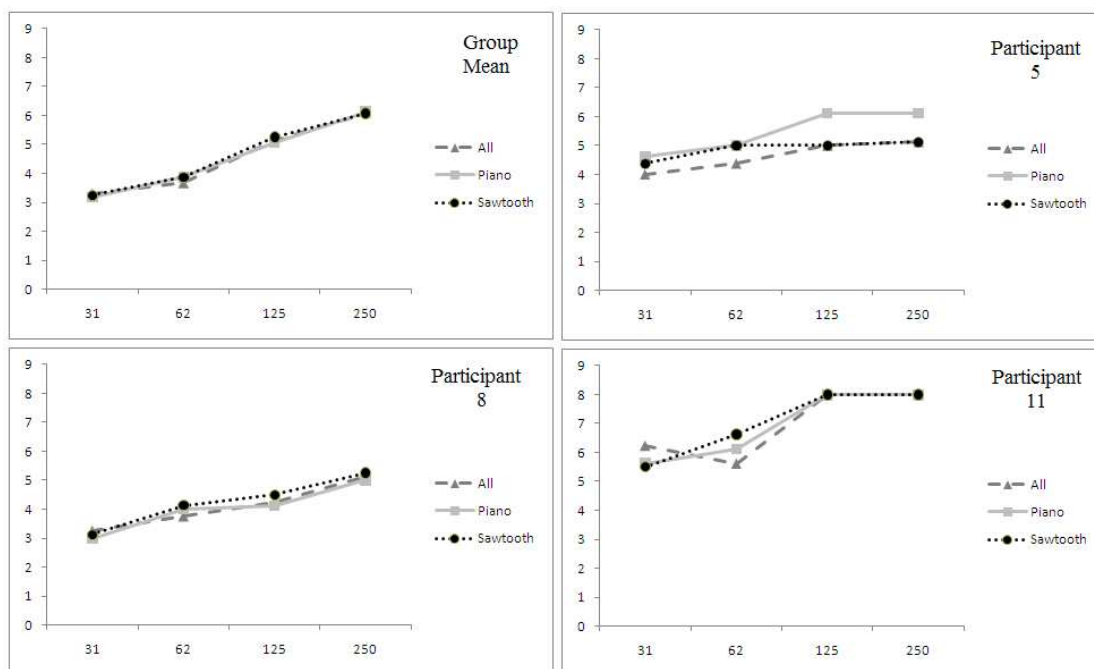


Figure 10. Mean number of reported events for the 5-tone sequences depicted for the group mean and participants with some of the most musical training.

*Appendix A***Musical History Questionnaire**

1. What instrument(s) have you played, and how many years have you studied each instrument?

2. How long have you been playing music? (If no longer a practicing musician, please note the period of time in which you were playing, including estimated dates).

3. Are you a formally trained musician, self-taught, or both (if both, please describe)?
 trained self-taught

4. Do you possess any knowledge of music theory, or are you purely a musician who plays by ear?
 yes (know theory) no (ear musician only)

If you answered *yes*, check all that apply:

- scale patterns chord progressions formal theory instruction
- formal ear training (scale values and intervals)
-
5. Have you received or are you currently receiving instruction in musical performance (if answering *no*, skip to question #8)?
 yes no

 6. Where have you received this instruction (please mark one or more of the following)?
 high school college private instructor literature
 other(specify) _____

7. Over what approximate dates has this instruction taken place?
8. Have you ever been paid in exchange for musical performance (if answering *yes*, proceed with question #9)?
 yes no
9. What were the circumstances under which you were paid for musical performance (e.g., party v. night clubs v. touring, as well as classically trained ensemble v. popular music group)? If you are a professional musician, please estimate the frequency with which you are a paid performer.

References

- Anstis, A., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(3), 257-271.
- Askenfelt, A. (1991). Voices and strings. Cousins or not? In: J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech and brain: Proceedings of an international symposium at the Wenner-Gren Center, Stockholm 5-8 Sept. 1999* (Vol. 59, pp. 243-259) London: Macmillan Press.
- Bever, T. G. & Chiarello, R. J. (2009). Cerebral dominance in musicians and nonmusicians. *Journal of Neuropsychiatry and Clinical Neurosciences*, *21*(1), 94-96.
- Brochard, R., Drake, C., Botte, M., & McAdams, S. (1999). Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1742-1759.
- Bregman, A. S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(3), 380-387.
- Bregman, A. S. (1990). Auditory scene analysis. Cambridge, MA: MIT Press.
- Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, *89*(2), 244-249.

- Bregman, A. S., Liao C., & Levitan R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian Journal of Psychology*, 44(3), 400-413.
- Bregman, A. S., & Steiger, H. (1980). Auditory streaming and vertical localization: Interdependence of “what” and “where” decisions in audition. *Perception & Psychophysics*, 28(6), 539-546.
- Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61(5), 1270-1277.
- Grey, J. M., & Gordon, J. W. (1978). Perception of spectral modifications on orchestral instrument tones. *Computer Music Journal*, 2(1), 24-31.
- Grey, J. M., & Moorer, J. A. (1976). Perceptual evaluations of synthesized musical instrument tones. *Journal of the Acoustical Society of America*, 62, 454-462.
- Hall, M. D. & Schuett, J. H. (2008, November). *Limits on the number of concurrent auditory streams in arpeggiated tone sequences*. Poster presented at the 49th annual meeting of the Psychonomic Society, Chicago, IL.
- Huron, D. (1989). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception*, 6(4), 361-382.
- Iverson, P., & Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Accoustical Society of America*, 95(5), 2595-2603.

- Kendall, R. A. (2002). Music Experiment Development System (Version 2002-B) [Computer software]. Los Angeles, CA: Author, www.ethnomusic.ucla.edu/systematic/Faculty/Kendall/meds.htm.
- McAdams, S., Beauchamp, J. W., & Meneguzzi, S. (1999). Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. *Journal of the Acoustical Society of America*, *105*(2), 882-897.
- McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*(3), 177-192.
- Miller, G. A., & Heise, G. A. (1950). The trill threshold. *The Journal of the Acoustical Society of America*. *22*(5), 637-638.
- Moore, B. C. J. (2003). *An introduction to the psychology of hearing*. San Diego, CA: Elsevier Science.
- Moore, T. J., & Welsh, J. R. (1969). Forward and backward enhancement of sensitivity in the auditory system. *The Journal of the Acoustical Society of America*, *47*(2), 534-539.
- Rose, M. M., & Moore, B. C. J. (2000). Effects of frequency and level on auditory stream segregation. *The Journal of the Acoustical Society of America*, *108*(3), 1209-1214.
- Samson, S., Zatorre, R. J., Ramsay, J. O. (1997). Multidimensional scaling of synthetic musical timbre: Perception of spectral and temporal characteristics. *Canadian Journal of Experimental Psychology*, *51*(4), 307-315.

- Schneider, B., Parker, S., & Upenieks, E. G. (1982). The perceptual basis of judgments of pitch differences and pitch ratios. *Canadian Journal of Psychology*, 36(1), 4-23.
- Serra, X., & Smith, J. (1990). Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 14(4), 12-24.
- Singh, P. G. (1987). Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre? *The Journal of the Acoustical Society of America*, 82(3), 886-899.
- Snyder, J. S., Carter, O. L., Lee, S., Hannon, E. E., & Alain, C. (2008). Effects of context on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, 34(4), 1007-1016.
- Tingen, P. (2007, November). Inside Track: Secrets of the Mix Engineers: JJ Puig. *Sound on Sound*, 23(1), p. 145.
- Van Noorden, L. P. A. S. (1975). Temporal Coherence in the Perception of Tone Sequences. *Unpublished doctoral dissertation*, Eindhoven University of Technology.
- Warren, R. M., Obusek, C. J., Farmer, R. M., Warren, R. P. (1969). Auditory Sequence: Confusion of patterns other than speech or music. *Science*, 164, 586-587.
- Yost, W. A. (2000). *Fundamentals of hearing: An introduction*. San Diego, CA: Academic Press.