Aalborg University
Institute of Electronic Systems
Frederik Bajers Vej 7, 9220 Aalborg
Denmark

Title: Influence of Temporal Attributes in Timbre Streaming

Theme: Masters Thesis

Project period: 1st February - 6th June 2000

Project group: 1066

Members: Felipe A. Otondo

Supervisor: Morten Lydolf

Number of copies printed: 8

Pages: 58

The front page illustration correspond to part of the aural score of Gyorgy Ligeti's electronic piece "Articulation" by Rainer Wehinger.

Abstract

Temporal characteristics of signals as rise time has been studied showing to be a significant cue for the perceptual organization of complex tone sequences [P. G. Singh and A. S. Bregman, J. Acoust. Soc. Am. 102, 1943-1952 (1997)]. The perceptual contributions of spectral synchronicity and temporal envelopes of complex tone sequences were investigated considering variations in the amplitude distribution of four harmonics and three different onset and offset overall characteristics. The perceived differences in the timbre features were measured using the *F0*-tracking method, where the relative pitch of one of two dissimilar complex tones is increased or decreased until the segregation takes place. Results show that the synchrony and temporal envelope features do not have a significant importance in facilitating the segregation of the ABA triplets compared with the AAA monotimbral sequences. The pitch direction of the sequences showed to have a significant influence in streaming the timbral attributes. The musical experience of the participants showed to be significant for the segregation of some of the timbral features studied

Abstracts in other languages

Spanish Abstract

Influencia de atributos temporales en la percepción timbrística de sequencias de sonidos complejos

Las características temporales de señales han sido estudiadas mostrando ser importantes factores para la organización perceptual de secuencias de tonos complejos [P. G. Singh and A. S. Bregman, J. Acoust. Soc. Am. 102, 1943-1952 (1997)]. La contribución perceptual de la sincronicidad espectral y las características envolventes espectrales de secuencias de tonos complejos fueron investigadas. Para tal efecto se consideraron variaciones en las distribuciones de amplitud de cuatro armónicos y tres distintos tipos de ataques y decaimientos de amplitud de la envolvente total. Las diferencias percibidas en las características timbrísticas fueron medidas utilizando el método de seguimiento de frequencia fundamental *F0*, donde la frecuencia relativa de uno de dos tonos complejos dísimiles es variada hasta que la segregación timbristica ocurre. Los resultados muestran que la sincronicidad espectral y las características envolventes totales no tienen una una importancia significante en facilitar la segregación timbristica de sonidos modificados en comparación a aquellos no modificados. La dirección de las variaciones de frecuencia mostró tener una influencia significante en la segregacion de las características del timbre de los tonos complejos. La experiencia musical de los participantes de los experimentos demostró ser un factor significante en la percepción de las características temporales estudiadas.

Danish Abstract

Indflydelsen af tidsmæssige egenskaber i timbre adskillelse

Den tidsmæssige karakteristik af signalers dynamik er blever undersøgt, hvor de viser en signifikant egenskab for organisationen af opfattelsen for komplekse tone sekvenser [P. G. Singh & A. S. Bregman, J. Acoust. Soc. Am. 102, 1943-1952 (1997)]. Bidraget for opfattelsen af spektral synkronitet og tidsmæssige indhyllingskurver af komplekse tone sekvenser er blevet studeret for forandringen i amplitude distributionen af fire harmoniske og tre forskellige onset og offset indhyllingskurve karakteristikker. Den forskellige opfattelse af timbre egenskaberne blev målt ved at bruge F0- tracking metoden, hvor den relative tonehøjde for et af to forskellige komplekse toner er forøget, indtil en adskillelse finder sted. Resultater viser at synkroniteten og tidsmæssige indhyllingskurve egenskaber ikke har nogen signifikant betydning i lette adskillelsen af ABA triplets sammenlignet med AAA monotimbral sekvenser. Tonehøjde retningen af sekvenserne viser at have en signifikant indflydelse i adskillelsen i timbral egenskaber. Den musikalske erfaring af eksperimentets deltagere viste, at være signifikant i adskillelsen for nogle af de undersøgte timbral egenskaber.

Preface

This report corresponds to the documentation of the work done in the third semester of the International Master in Acoustics at Aalborg University. The work has been carried out at the Acoustics Department in the Institute for Electronic Systems at Aalborg University, Denmark, during the period, February 1st to June 6th of the year 2000. The report includes the technical information of the investigation carried out as well as complementary information in the appendices and enclosure. The appendices include the written instructions for the experiments, letters to and from researchers and finally a list of related literature to the topic of timbre perception. In the enclosure the reader will find the codes of the programs used for creating the tones and controlling the experiments. I would like to thank Jan Plogsties, Søren Krarup and Brian Lykkegard for their help with C++ programming and the multiple experiments issues. Also, I would like to thank my colleagues for the help with software topics, Kristian Knudsen for the abstract translation and all the persons that participated in the experiments. Finally, I would like to thank the staff of the acoustic departament, and especially Dorthe Hammershøi and Sofus Birkedal for the friendly cordination of this Masters education.

Felipe Otondo
Aalborg, June 2000

......the cues for timbre depend on context: the duration, intensity, and the frequency of the notes, the set of comparison sounds, the task, and the experience of the subjects all determine the outcomes. At this point, no known acoustic invariants can be said to underlie timbre.

Stephen Handel

Contents

Appendices

Enclosure

Introduction

In this chapter the topics of timbre and timbre perception will be introduced. First, the definition of timbre will be approached from an intuitive point of view in order to clarify the definition that will be used all along this report. As a second issue, the main sound physical characteristics of sound involved in timbre perception will be described.

1.1 What is timbre?

When we reefer to timbre we get the idea of many things, but not really something very precise. A musician would probably relate it to the characteristics that enable him to distinguish one particular musical instrument from another and probably an expert in speech recognition would associate it with the physical characteristics of vowels in a spectrogram. In order to solve the diverse interpretations and confusions with the concept and definition of timbre there have been several attempts to create a precise terminology in the theme. The following brief description will follow the analysis of different definitions of timbre carried out by Letowski (1992). The analysis ends with the definition that has been used as a basis for this study.

The American Standard Association (ASA) (1962), following the original idea of timbre described by Helmholtz (1885), describes timbre as “the attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented as having the same loudness and pitch are dissimilar”. This definition is supplemented by a note that states that “timbre depends primarily upon the spectrum of the stimulus, but it also depends upon the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics of the stimulus.” Timbre, as understood in the previous definition, involves a variety of phenomena but is limited to the comparison of sounds of equal loudness, pitch and duration. An attempt to define timbre, compensating for the rigid approach of the definition of the ASA, was given by Pratt and Doak (1976). They defined timbre as “that attribute of auditory sensation whereby a listener can judge that two [similarly presented] sounds are dissimilar using any criteria other than pitch, loudness or duration”. In a general approach, the last definition states that timbre depends on the spectrum of the stimulus and can be interpreted as the perceived spectrum, or also as the listener’s reaction to the distribution of sound energy along the frequency scale. This reaction involves the spectral envelope and the spectral distribution of sound components. Strictly speaking, the last definition of timbre can be applied only to steady state sounds (static timbre). The timbre of sounds varying in time (dynamic timbre) depends also upon the temporal characteristics of the stimulus. In this study a modified version of the definition of timbre proposed by Letowski (1992) will be used, as described below:

> *“Timbre is that multidimensional attribute of an auditory image in terms of which the listener judges the spectral and temporal character of sound. Timbral differences observed among a group of sounds are meaningful only as long as they refer to the same condition of comparison.”*

1.2 Spectral features involved in timbre perception

The spectral shape, or relative amplitude of each partial, is the classical and probably the most influential cue of the acoustical properties of complex tones in relation to timbre perception. There are different ways to characterize the relative amplitudes of the partials, being the most common to calculate the overall distribution of spectral energy included in the signal obtained by adding the squared amplitudes of each frequency component. One of the important perceptual characteristics of the spectra shape is the fundamental frequency (*F0*). This frequency, corresponding to the first maximum in the spectral shape, allows us to distinguish two complex tones heard together and plays an important roles in the perceptual fusion of harmonics (Moore, 1997). An alternative frequency characterization of a signal can also be obtained by means of the formant frequencies. These frequencies correspond to spectral prominences in spectra and can be obtained by averaging the frequency spectra across the signal range (Handel 1995). Fig. 1 illustrates how a particular kind of regular harmonic source is modified by a filter that has two prominent "pass bands" resonances or "formants" - two hills that reinforce the source frequencies decreasing with them (Slawson, 1982). The filter provides reinforcement to some harmonics and attenuates others, giving to the original spectra the pitch characteristics, as it happens with musical instruments that include mouthpieces (Slawson, 1985).

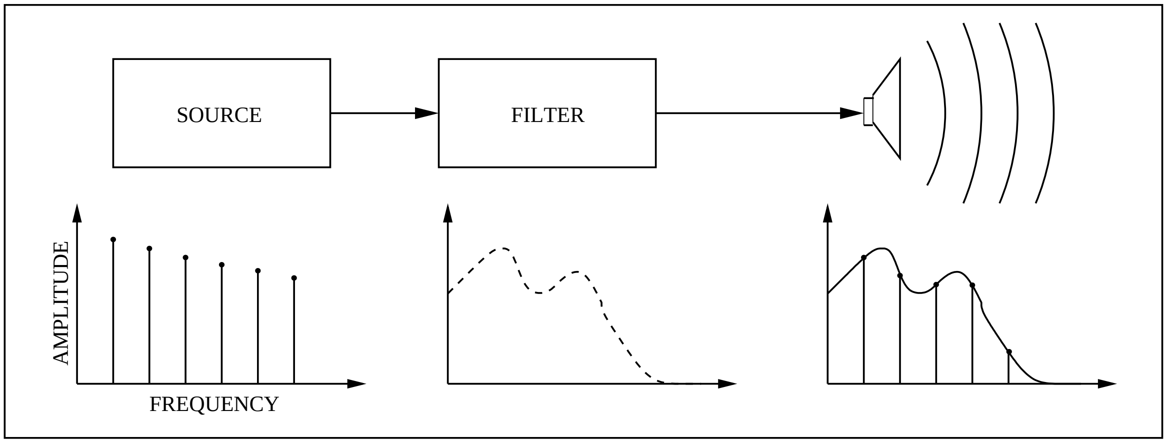


FIG. 1. The Source/Filter Model of Sound Production; the filter modifies the energy coming from the source according to the filter's spectrum envelope.

1.3 Temporal features involved in timbre perception

In non static timbres, timbres that change in the time domain, the temporal acoustical characteristics of a signal are important cues for the identification. Different temporal characteristics have showed to have a significant role in perception of timbre. Onset and offset characteristics of a "string" musical instrument, for example, are related to the method of excitation (plucked or bowed) and variations in the damping of the various vibration modes of the source and filter system mentioned in the previous section (Handel 1995). Researchers, dealing mostly with multidimensional scaling (MDS) studies of timbre of a single note of musical instruments, have used qualitative different descriptions to describe some temporal characteristics involved in timbre perception (Krumhansl, 1989; Winsberg and Carroll, 1989; Grey, 1977). One temporal parameter that has been manipulated in this studies is the overall attack and decay of complex tones. Two tones having the same harmonic structure but different dynamic transitions will be perceived to have different timbres. Another temporal feature that has shown to be important in these studies is the stability of the spectra. This stability, or complexity of the time-varying changes in the spectrum, corresponds to the rate at which the different harmonic intensities fluctuate to each other. If all the harmonics come at the same time, the spectra will be said to be stable or synchronous, and if they come at different rates it will be unstable or asynchronous. Figure 2 shows the different attacks and decays of 8 synthesized harmonics of a trumpet tone (Risset and Mathews, 1969).

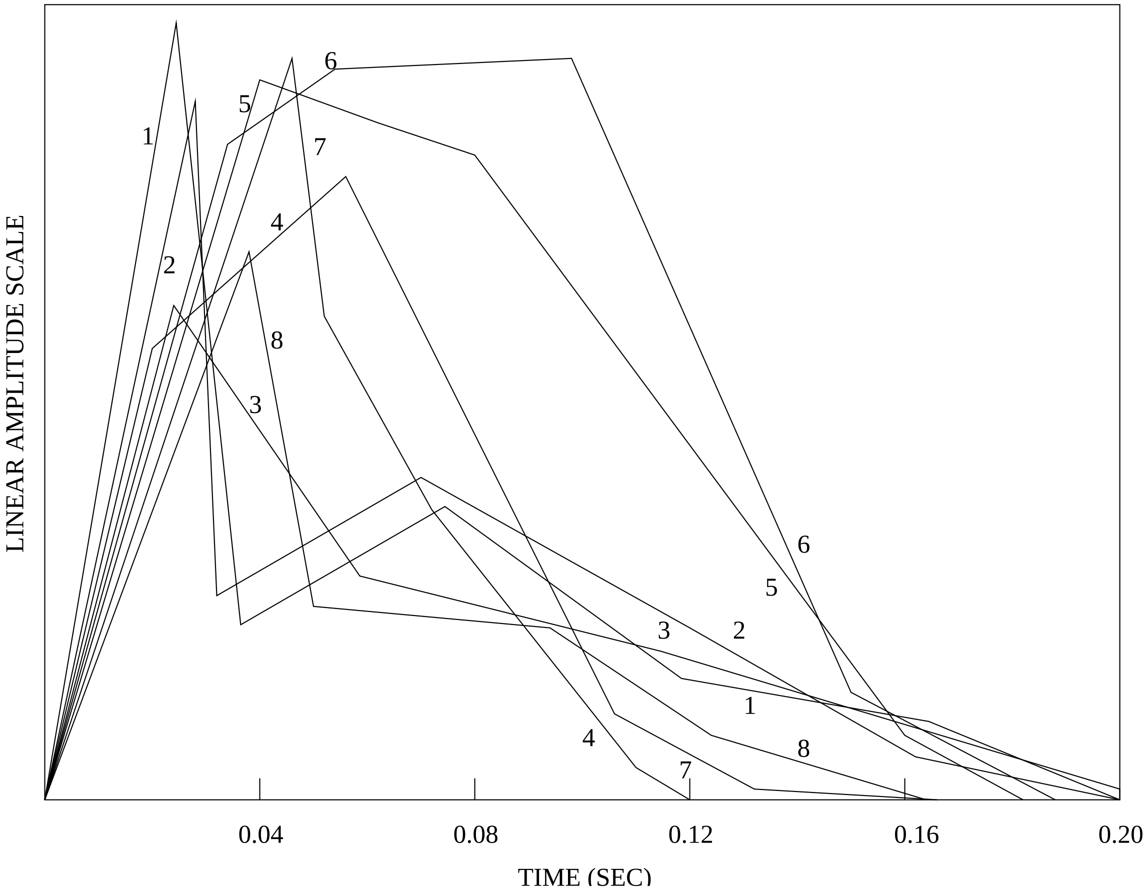


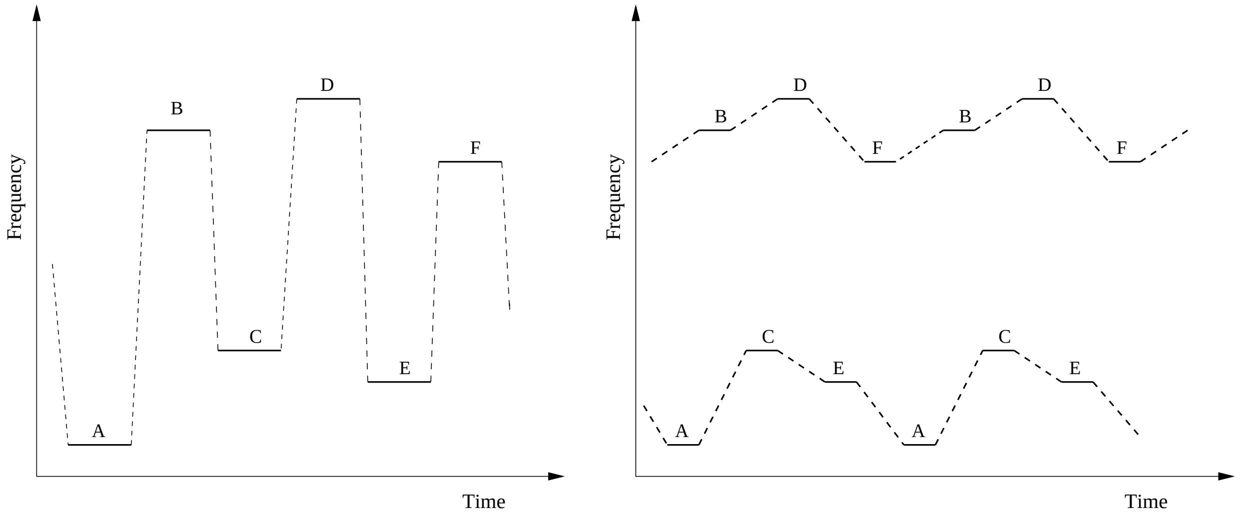
FIG. 2. Line segment functions that approximate the evolution of 8 harmonics of a D4 trumpet tone of 0.2 sec.

2 Stream segregation and purpose of this work

This chapter introduces the concept of stream segregation and some of the research carried out in the field related to the influence of temporal attributes in timbre perception. The inspiring research for the present work by Singh and Bregman will then be described with its goals and results. Finally, the scope and goals of this investigation will be introduced.

2.1 Temporal attributes of timbre involved in stream segregation

Auditory stream segregation can be described as the process by which the auditory system determines whether a sequence of acoustic events result from one, or more than one, “source” (McAdams and Bregman, 1979). If we have a sequence of two tones differing in a small frequency difference played at a high speed, the human auditory system tends to integrate the stream in one melody of fluctuating pitches (one source). When we have a larger frequency separation, the auditory system tends to separate the stream in two independent melodies with different pitches (two sources). In this case the pitch differences cause the tones to segregate (Iverson, 1995). Fig. 3 shows two figures of a repeating sequences of six high and low tones heard as one or two streams, respectively. In the left figure we can distinguish the sequence as one single melody of tones coming from one source, while at the figure to the right we can distinguish two different melodies (McAdams and Bregman, 1979).



ONE STREAM

TWO STREAMS

FIG. 3. A repeating 6-tone sequence of interlaced high and low tones. In figure 3a, the sequence of tones is played at 5 tones/sec., one perceptual stream is heard. In figure 3b, at a tempo of 10 tones/sec., the high tones perceptually segregate from the lower tones forming two streams.

The contribution of acoustical attributes of timbre to streaming has been studied by Iverson (1995) showing that the streaming judgments correspond well to the perceived similarity of timbres of previous MDS studies (Grey, 1975; Wessel, 1979; Krumhansl, 1989). Iverson suggested that the auditory stream segregation relies on the same dynamic and acoustic attributes that determine the similarity judgments of MDS studies. His study showed that, apart from the known spectral effects, changes in the temporal-amplitude envelopes of tones induce a higher segregation when compared with unmodified ones. This results contrast with the previous ones obtained by Hartmann and Johnson (1991) where envelope differences showed to have little, or no, influence in stream segregation. The opposing results of the findings of Hartmann and Johnson and those of Iverson could be explained from the different kind of stimuli used in the two investigations, as proposed by Singh and Bregman (1997).

2.2 The study of Singh and Bregman

Considering the studies of Iverson (1995), Hartmann and Johnson (1991), and other related studies, Singh and Bregman (1997) decided to investigate the influence of different timbre attributes on the perceptual segregation of complex-tone sequences. Their approach was focused on two main features: harmonic content and spectral envelope characteristics. They manipulated the timbral attributes using a variant of the paradigm introduced by van Noorden (1975). This original paradigm used a three element sequences composed of two pure tones created to form a triplet pattern ABA. The three tones are repeated with a period of silence equal to the duration of the B tone inserted between the repetitions. In the "temporal coherence state", or when we don't have any segregation between the two auditory streams, the sequence of triplets is perceived as a single stream resembling a sound of a galloping rhythm. In the "fission" state, or when the segregation takes place, the sequence is perceived as two independent streams of the A and B tones independently. In Fig. 4 it can be observed the galloping sequence used by van Noorden with the "temporal coherence state" (above) and the "fission" state (below). Singh and Bregman used the same method to measure the stream segregation replacing the pure tones with four harmonics complex tones sequences using the displacements in the frequency domain as a quantitative measure for the perceived timbre differences. The results of their study confirm the importance of spectral differences in facilitating stream segregation and show some significant, but minor, importance of the amplitude-envelope features. As one of their conclusions, considering the results proposed Hartmann and Johnson (1991), Iverson (1995), and their own findings, the authors propose as a further studies a more detailed examination of the role of temporal cues in stream segregation using a wider range of envelope differences and temporal envelope modulation differences. Also, they conclude as a complementary observation that the musical experience of some of the subjects in their test could have had some influence in the perception of timbre as shown in previous studies (Pitt, 1994). It is suggested (by Singh and Bregman) a further study of the influence of the subjects musical experience considering a different type of musical experience from the participants.

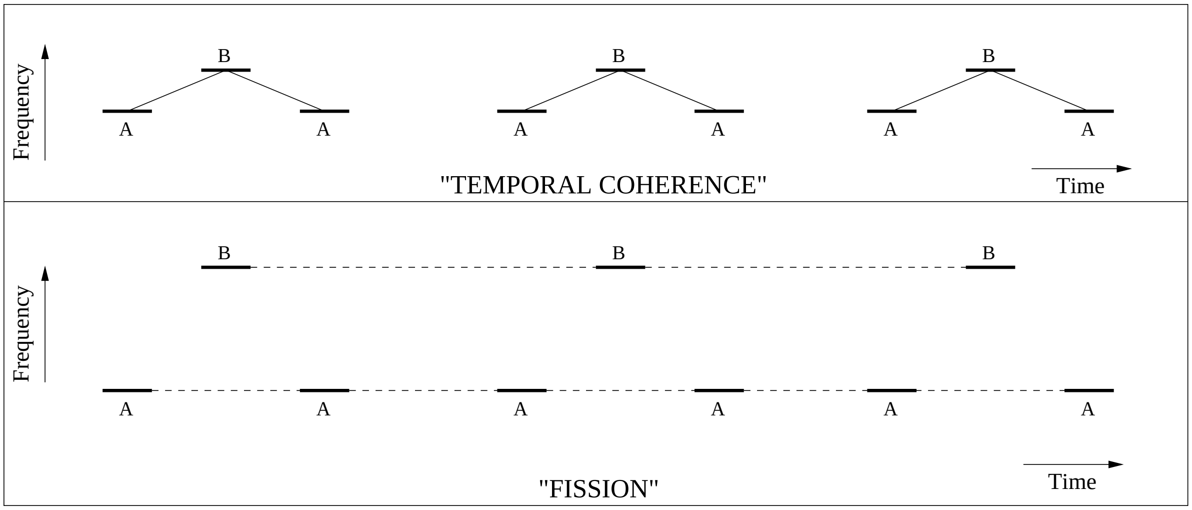


FIG. 4. The “ABA” galloping sequence used by van Noorden (1975). The top panel represents the “temporal coherence” state where the galloping rhythm is heard as one single stream. The bottom panel shows the “fission” state, where the sequence is perceptually segregated into two overlapping sequences of A and B tones, with one perceived to have the double tempo of the other.

2.3 Purpose and scope

Considering the results obtained by Singh and Bregman, it was decided as the main goal of the present work to study selected temporal attributes involved in timbre perception in detail. Following the spirit of the mentioned study, it was decided to use their variant of the van Noorden method considering two different types of temporal features to manipulate. The first temporal attribute considered was a range of three different overall amplitude differences of the complex tones. The second temporal attribute considered to manipulate was the complexity of time-varying changes in the spectrum of the signals. This feature involves the different onset and offset characteristics of different harmonics of the complex tone.

As a secondary goal it was intended to study, if the musical experience of the subjects that participate in the experiments could have some kind of influence in the perception of the timbre attributes studied.

3 Method

In this chapter the method used for the listening experiments will be described. Initially, the type of stimulus used will be explained considering the test method and the features of the synthesized sounds. As a second issue, the equipment used for the experiments will be described considering software, hardware and audio devices used. The procedure followed in the listening experiments will be explicated considering the way the stimulus were presented and how the experience was controlled through the computer. Finally, the general characteristics of the subjects that participated in the experiments will be outlined.

3.1 Stimulus

First of all it is important to define the terminology that will be used in this chapter and in the rest of the report referring the sounds used as a stimulus. One group of ABA sounds will be mentioned in this chapter as a triplet and one group of 24 triplets with the same timbral attributes between each other will be mentioned as one series. A number of series will define a session according to the specifications that will be mentioned subsequently in the next section.

The stimulus sound series used for the listening experiments were constructed considering a variant of the van Noorden galloping ABA sequence employed by Singh and Bregman in their experiments. Complex tones consisting of four harmonics of 256, 512 ,1024 and 2048 Hz were synthesized in phase in order to build each of the sounds series. Each triplet was builded considering a duration of the tones of 100 ms and a period of silence of 10 ms between them. After the third tone was played a gap of silence of 120 ms was included before the next triplet started. This kind of configuration of the triplets implies the audition of a characteristic galloping rhythm that should break into streams whenever the sequence perceptually segregates into independent A and B streams, as depicted in Fig. 4 of the previous chapter. In each of the series of triplets the pitch of the B tone was increasing, or decreasing, depending of the type of ascending or descending series played. The ascending series will imply an increase in pitch in the frequency stepsize while the descending will imply a reduction. With this procedure the pitch was used as a basis of comparison for the segregation of the timbral features. At each triplet repetition the pitch is increased in an eighth of a musical tone, that is a quarter of a semitone or an amount equal to 1.5% of the reference *F0*. This value for the differences in pitch differs from the one of one quarter of a tone used in the study of Singh and Bregman. The reason for this difference in the stimulus was that the results of some pilot experiments, using a quarter of a tone as a unit of comparison, showed large variations in the answers of the subjects in the ascending pitch method. The subjects argued that the segregation took place too fast after the start of the series of triplets and when they indicate that the segregation had taken place it had already gone a period of time. In order to make the transitions slower, and allow the subject to have enough time to tell more accurately when the segregation streaming occurred, an eighth of a tone unit was used in new pilot experiments showing to reduce the response time of the subjects and the variations in the answers for the same series.

The first temporal attribute manipulated for the triplets series was the complexity of the time varying changes in the spectrum. The amplitudes of one or two of the four harmonics of the tone B were manipulated with a different value from the rest of the harmonics of the A and B tones. For the first feature series, the first harmonic was manipulated with a different temporal envelope from the rest of the harmonics of the B tone. The same procedure was done considering the amplitude of the first and the third, the second and the fourth, and the third and the fourth for the rest of the initial five features. The overall temporal envelopes used for the manipulated harmonic, as well as the rest of the complex tone components, were the envelopes of 95 ms of attack and 5 ms of decay (95/5), as well as 95 ms of attack and 5 ms of decay (95/5). The same procedure was used to generate the next five features, but with the values of the amplitudes of tone A and B exchanged. Table 1 illustrates the different types of stimuli used to create the first 10 feature series of triplets corresponding to the complexity of time varying changes in the spectrum. Fig. 5 shows the amplitude in the time domain of a triplet corresponding to feature 2 of table 1.

Number of feature	Amplitude tone A	Amplitude tone B	Harmonics of tone B manipulated	Amplitude harmonics manipulated
1	5/95	5/95	1	95/5
2	5/95	5/95	1 - 2	95/5
3	5/95	5/95	1 - 3	95/5
4	5/95	5/95	2 - 4	95/5
5	5/95	5/95	3 - 4	95/5
6	95/5	95/5	1	5/95
7	95/5	95/5	1 - 2	5/95
8	95/5	95/5	1 - 3	5/95
9	95/5	95/5	2 - 4	5/95
10	95/5	95/5	3 - 4	5/95

Table 1. Spectrum synchrony characteristics used to construct the 10 first features series for the experiments.

The second temporal attribute manipulated for the triplets series was the overall amplitude of the tones. Three different amplitude characteristics were considered consisting of three values for the attack and decay of the complex sounds. The values used were: an attack of 5 ms and a decay of 95 ms (5/95); an attack of 50 ms and a decay of 50 ms (50/50); and an attack of 95 ms and a decay of 5 ms (95/5). Table 2 shows the three main envelope characteristics with the 9 possible subgroup combinations, which correspond to the last 9 of the 19 subgroups of features used for the experiments. Mono timbral sequences (AAA) were also included as a basis of comparison for both temporal attributes manipulated. Fig. 6 shows a triplet corresponding to feature 19 of table 2.

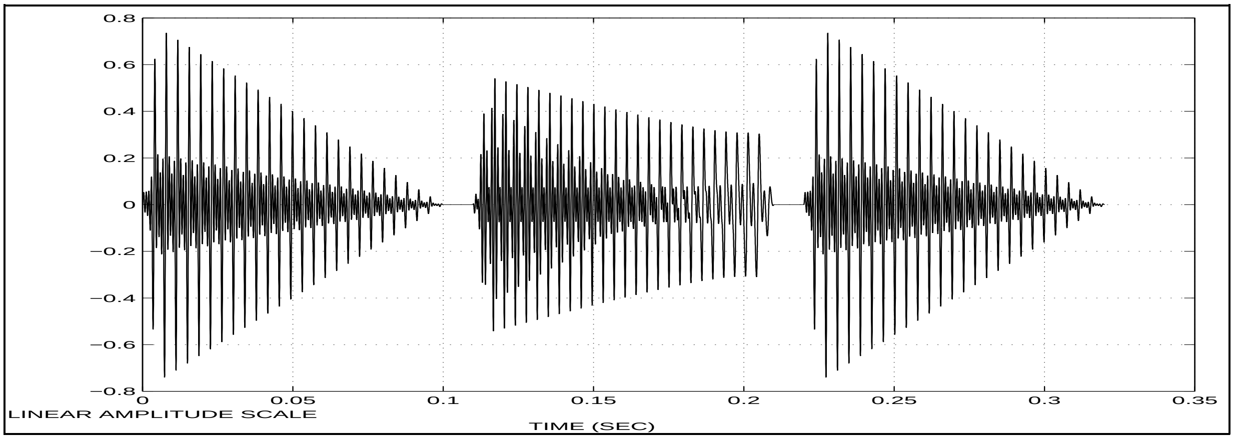


FIG. 5. A triplet of sounds ABA used for the experiments. In this triplet the amplitude of the first and the second harmonics of the B tone have been manipulated with an envelope of (95/5), according to feature 2 of table 1. The first and the last tones have the same amplitude (5/95).

Number of feature	Amplitude tone A	Amplitude tone B
11	5/95	5/95
12	5/95	50/50
13	5/95	95/5
14	50/50	50/50
15	50/50	5/95
16	50/50	95/5
17	95/5	95/5
18	95/5	5/95
19	95/5	50/50

Table 2. The temporal envelope characteristics for each of the last 9 triplets sequences of all the features manipulated.

For each of the feature series of triplets the frequency range considered for the pitch variations of the van Noorden method covered approximately half an octave. This range started with a value of fundamental frequency (*F0*) of 262 Hz and ended in a value approximately of 365 Hz. This range was chosen in order to fit the characteristics of the frequency units of one eighth of a tone. All the resulting triplets series were compensated for their overall level in order to compensate the amplitude differences in the reproduction process through headphones.

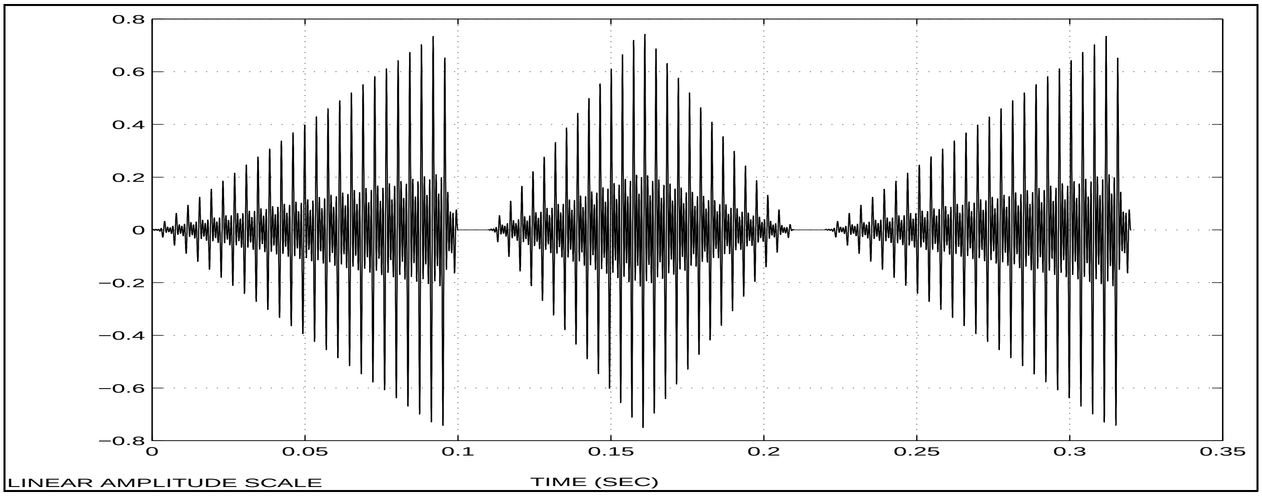


FIG. 6. A triplet of sounds ABA where the spectral characteristics of the B tone have been manipulated. The amplitude of the tones A and B correspond to the ones of feature 19 of table 2.

3.2 Apparatus

The sounds used to build the sequences were synthesized digitally using the software *Matlab*. The code of one of the programs used for the synthesis can be seen in the Enclosure of this report. The stimulus presentation and the data collection were controlled with a *Hawlett Packard* pentium computer, 75 MHz, fitted with a *Vortex 1* sound card set at a sampling frequency of 48 kHz and 16-bit resolution. The levels of the presentations were controlled with a *Behringer* headphone amplifier model H1903 and verified with a B&K artificial ear type 4153 fitted with a B&K condenser microphone type 4134 with preamplifier 2619 connected to a B&K analyzer type 2133. The listeners were seated in a double wall audiometry room with a background noise level of -10 dB referred to the human sound threshold, measured in octave bands. The subjects received the stimulus via *Sennheiser* headphones model HDA 200 at an overall sound pressure level of approximately 70 dB.

3.3 Procedure

The series of triplets were presented to the listeners using an interactive procedure routine program created in C++ language. The program’s code used for the presentation of the triplets series and achievement of the results can be seen in the Enclosure. For each series of sounds, the subjects were presented with a repeated sequence of triplets which increased in pitch in an eighth of a tone steps. The triplet series were played through the computer until they were perceptually segregated by the listeners into a galloping rhythm or two different streams, depending on the ascending or descending pitch disposition of the series. If the sequence was ascending in pitch, the subjects were asked to indicate when the constant pitch galloping rhythm stopped and two different streams were heard, as shown in Fig. 4 of the previous chapter. If the sequence was descending in pitch, the subjects were asked to indicate when the two different streams transform into a constant pitch galloping rhythm. The series of triplets were played by the computer until the listener pressed a push-button which lead to a gap of 2 seconds of silence followed by the next series of triplets. The number of the last triplet played was then achieved in a *Matlab* file and a text file, respectively. The control program allowed the controller of the experiment to monitor the inclusion of the different sound files corresponding to the different timbral features and hear the sounds played during the experiments. For each triplet of the 19 features series a sound file was played in the ascending or descending pitch disposition. The order of the series of triplets sequences was controlled by the program such that the order in which the a series was different for each subject, and also the first triplet in each session different between each of the sessions of the same subject. At each of the three sessions long and short series of descending and ascending triplets were played to the listeners considering short pauses between them. The order was the same for all the subjects and consisted of one short and one long descending series for the first session, one short and one long ascending series for the second session, and one short ascending and one short descending for the final session. This order was chosen to be fixed to facilitate the performance of the subjects and was structured according to the results obtained in the pilot experiments.

Summarizing, the order of the series in the three sessions was:

(1) short descending and long descending series;

(2) short ascending and long ascending series; and

(3) short ascending and short descending series.

Each of the short series included the 19 features once, while the long series of repetitions included the 19 features twice. After each of the sessions a pause of 10 minutes was held where the subjects could take a break and drink coffee or juice in a resting room. All and all, each of the series of the 19 features were played to the listeners four times for each ascending and descending method.

3.4 Subjects

Twelve subjects, three females and nine males, participated in the experiments with ages between 22 and 32 years. Three of subjects had participated in listening experiments before. Most were students from the electronic systems and music departments at Aalborg University, and were chosen according to their musical experience. Six of the subjects were piano players with a formal musical education. The remaining six subjects did not have previous musical education and do not play any musical instrument. To each subject an audiometric test was performed after the listening tests showing that the hearing loss of the subjects was less than 15 dB at octave frequencies from 125 Hz to 8 kHz. All subjects were given written and oral instructions before the experiments, as well as an auditory demonstration of an ascending and descending sequence of repetitions of triplets. The written instructions are included in Appendix A of this report. The initial descending and ascending short series of the first and second sessions were used as a trial for the subjects and were not included as part of the results. This initial series were repeated in case it was detected the subjects had not understood correctly the task of the experiments. The results corresponding to the first two subjects were not used for the analysis due to a synchronization problem that was detected in the reproduction process by the computer.

4 Results and discussion

In this chapter the results from the listening experiments will be presented and discussed. The first section will include the results from the temporal attributes studied, followed by the results from planned comparisons. In the second section a general discussion will be included where the results will be analyzed and compared with previous studies. It must be noted that in all the analysis of the results and the planned comparisons a value of 0.05 was considered as a reference for the probability under the null hypothesis of obtaining a result equal in the one-way ANOVA comparisons.

4.1 Results

4.1.1 Synchrony features

The results obtained for the first five synchrony features for the ascending and descending *F0* trials can be seen in Fig. 8. The ordinate gives the mean *F0* differences for 10 subjects required for stream segregation in eighth-tone steps calculated as an average for the 10 subjects. The abscissa shows the first five synchrony features as defined in tables 1 and 2. The effect of the first five synchrony features did not show to be significant for the ascending and descending method compared with the reference (5/95). The descending values are lower than the ascending values for all the features.

Figure 9 shows the results obtained for the second five synchrony features for the ascending and descending *F0* trials. The ordinate and abscissa are the same as in Fig. 8. The results for the ascending and descending method show that the second five synchrony features' mean crossover points values do not show significant different with respect to the reference (95/5). The descending values are lower than the ascending values for all the features.

The mean *F0* crossover points for the synchrony features averaged for the order in which the A and B tones were played is shown in Fig.10, for the ascending and descending method. The ordinate and abscissa are the same as in the previous figures. The effect of the averaged features did not show to be significant for the ascending and descending method compared to the references. The descending values are lower than the ascending values for all the features showing differences that fluctuate from small values (0.7 of a crossover point approx.) to large values (2 crossover points approx.).

Figure 11 shows the results obtained for the synchrony features and references averaged for the ascending and descending *F0* trials and the order of the A and B tones. The ordinate and abscissa are the same as in the previous figures. A planned comparison between the different features and the references can be seen in table 3. These results show that the differences in the segregation are not significant for any of the features compared with the respective references.

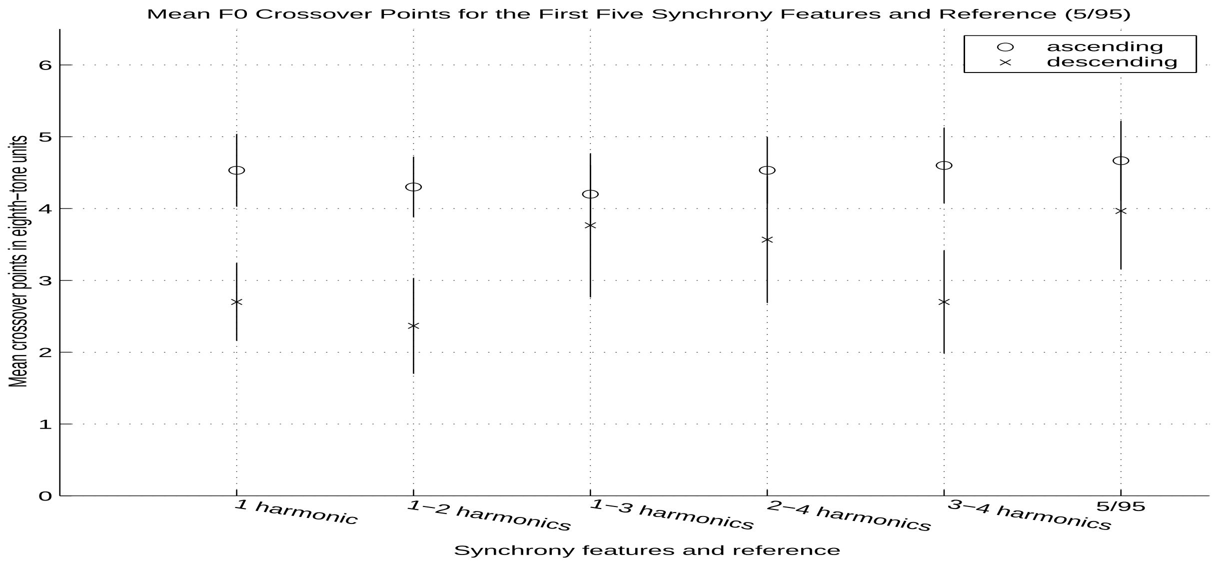


FIG. 8. Mean crossover points (in eight-tone steps) obtained for ten subjects for the first five synchrony features and reference (5/95) shown along the abscissa. Empty symbols correspond to crossover points for ascending *F0* trials. Cross symbols represent the descending *F0* trials. Error bars correspond to the standard error of the data considered.

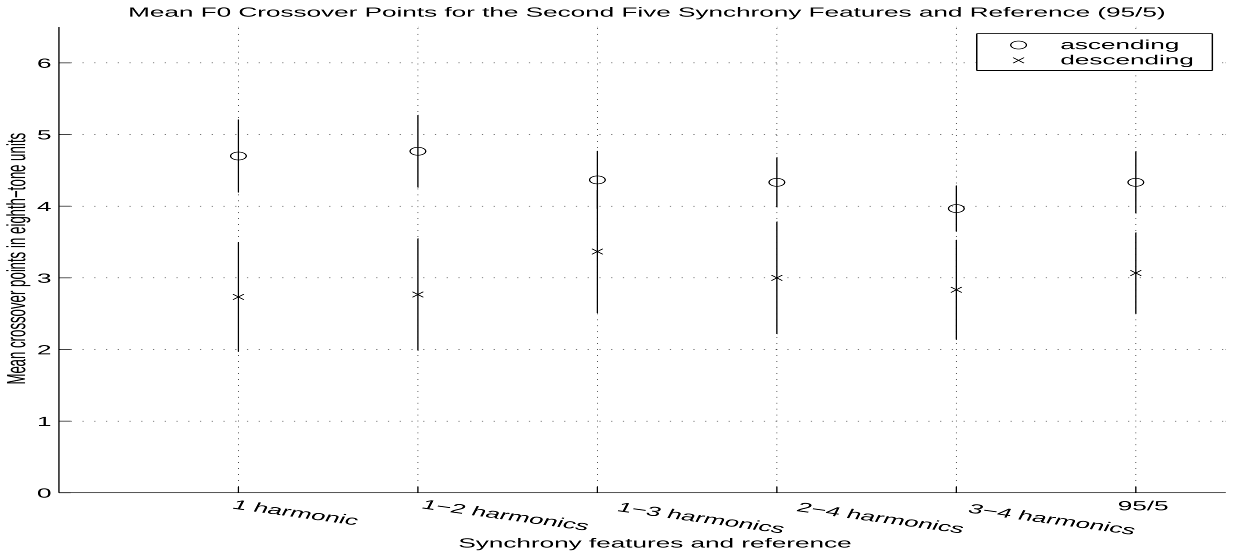


FIG. 9. Mean crossover points (in eight-tone steps) obtained for ten subjects for the second five synchrony features and reference (95/5) shown along the abscissa. Empty symbols correspond to crossover points for ascending *F0* trials. Cross symbols represent the descending *F0* trials. Error bars correspond to the standard error of the data considered.

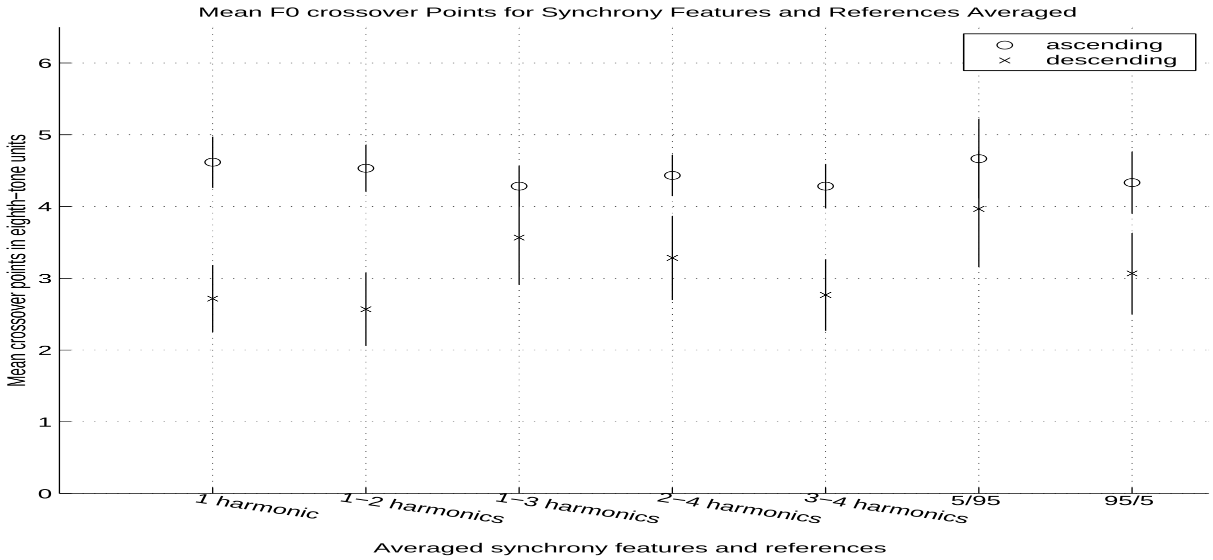


FIG. 10. Mean crossover points (in eight-tone steps) obtained for ten subjects for the synchrony features and references averaged for the order in which the A and B tones were played. Empty symbols correspond to crossover points for ascending *F0* trials. Cross symbols represent the descending *F0* trials. Error bars correspond to the standard error of the data considered.

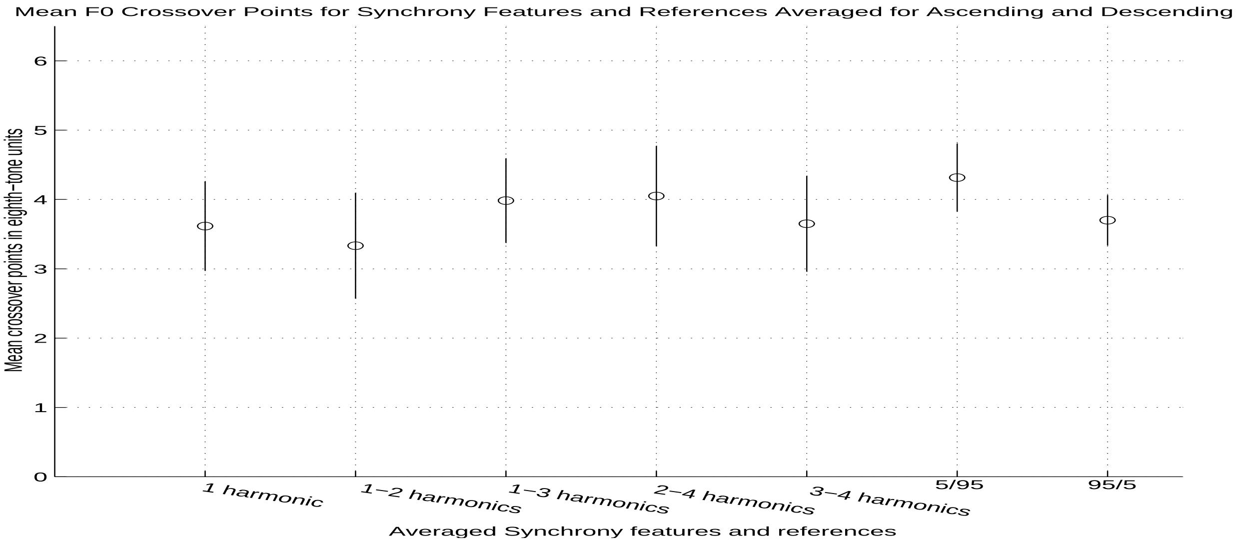


FIG. 11. Mean crossover points (in eight-tone steps) obtained for ten subjects for the synchrony features and references averaged for the order in which the A and B tones were played and the ascending and descending method. Error bars correspond to the standard error of the data considered

Features Comparisons	**p values**
mean (1 & 6) vs. 11	0.922
mean (2 & 7) vs. 11	0.929
mean (3 & 8) vs. 11	0.919
mean (4 & 9) vs. 11	0.927
mean (5 & 10) vs. 11	0.638
mean (1 & 6) vs. 17	0.633
mean (2 & 7) vs. 17	0.795
mean (3 & 8) vs. 17	0.767
mean (4 & 9) vs. 17	0.788
mean (5 & 10) vs. 17	0.411

Table 3. Results of planned comparison for different averages of synchrony features and references. The first column shows the features averaged and the reference for which the one-way ANOVA was done. The second column shows the values for the null hypothesis that the means of the columns are equals. The numbers of each feature correspond to the ones specified in table 1 and 2.

4.1.2 Temporal envelope features

The results obtained for the temporal envelope features for the ascending and descending *F0* trials are shown in Fig. 12. The ordinate gives the mean *F0* difference for 10 subjects required for the stream segregation in eighth-tone steps. The abscissa shows the corresponding temporal envelope features as defined in table 2. The effect of the different temporal envelope features did not show to be significant for the ascending and descending method compared with the references. The descending method values are lower than the descending for all the features, with differences that fluctuate from small values (0.5 of a unit approx.) to large values (2 units approx.).

Figure 13 shows the mean *F0* crossover points for temporal features and references averaged. The averages were considered for the order of the A and B tones for features 12 and 15, 13 and 18, and 16 and 19, respectively. The reference corresponds to the features where the tones A and B had the same temporal envelope in the triplet, that is features 11, 14 and 17 from table 2. The ordinate is the same as in Fig. 12 and the abscissa includes the values for the average in the first three positions and following the references. The results for the ascending and descending method show that the averaged temporal envelope features are not significantly different compared with the references. The descending method values are lower than the ascending method values for all the averaged features and references with fluctuations in the difference between them (from 0.5 to 2 units approx.).

Figure 14 shows the results obtained for the average for the ascending and descending method and the order of the A and B tones for the temporal envelope features. The ordinate and abscissa are the same as in the previous figure. A planned comparison between the three features averaged for the order of the A and B tones in the triplet and the references can be seen in table 4. The comparison between the references is also included. The results show that for all the features compared with the references the results are not significantly different. The same can be stated about the streaming between the references.

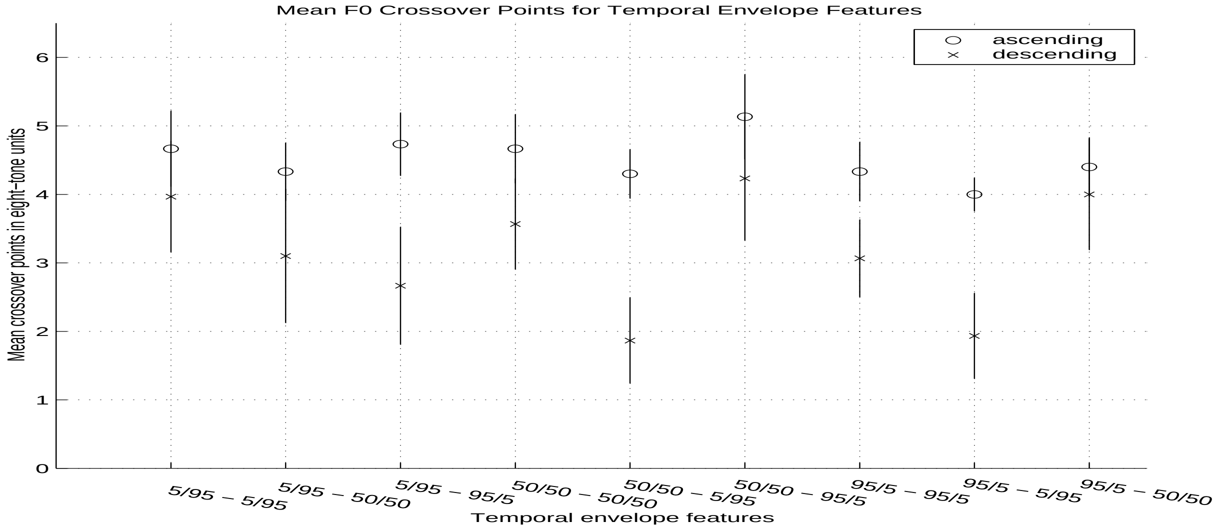


FIG. 12. Mean crossover points (in eight-tone steps) obtained for ten subjects for nine envelope temporal envelope features. Empty symbols correspond to crossover points for ascending *F0* trials. Cross symbols represent the descending *F0* trials. Error bars correspond to the standard error of the data considered.

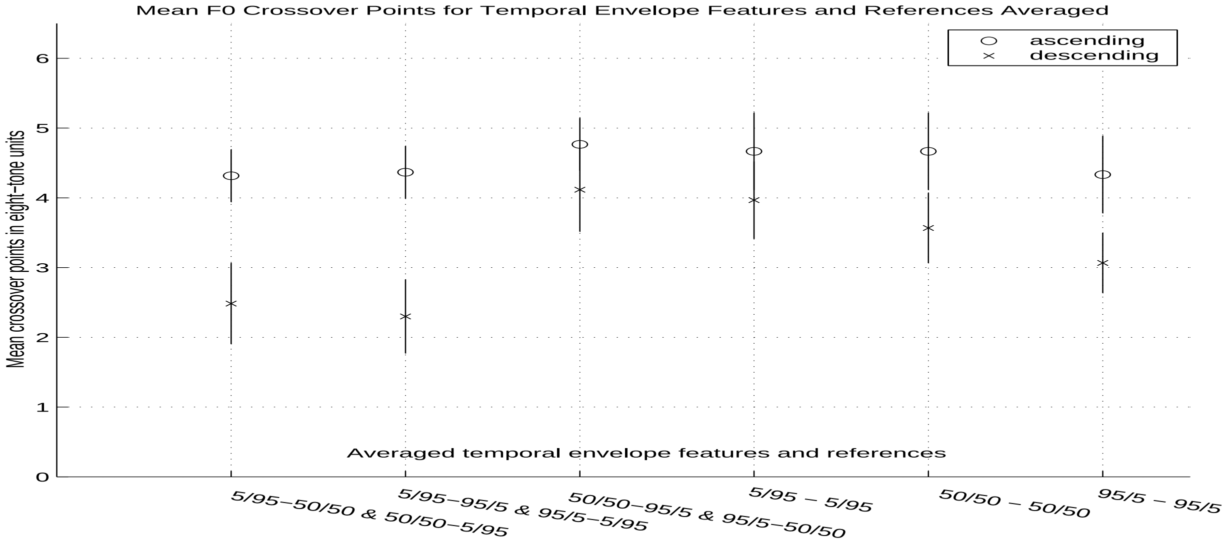


FIG. 13. Mean crossover points (in eight-tone steps) obtained for ten subjects for the temporal envelope features averaged for the order of tones A and B and references. Empty symbols correspond to crossover points for ascending *F0* trials. Cross symbols represent the descending *F0* trials. Error bars correspond to the standard error of the data considered.

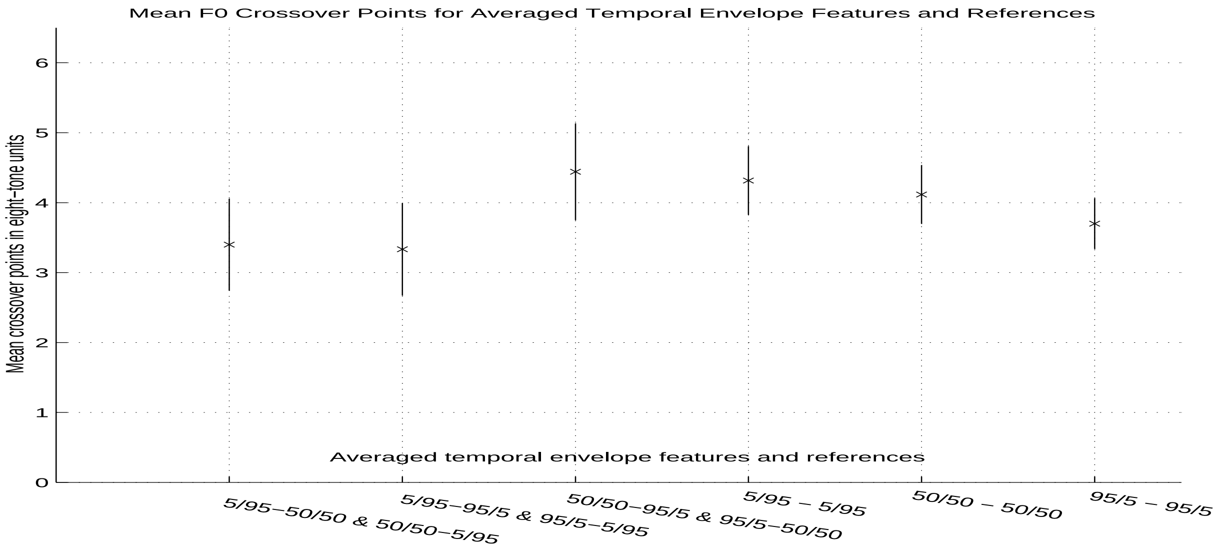


FIG. 14. Mean crossover points (in eight-tone steps) obtained for ten subjects for the temporal envelope features averaged for the order of tones A and B and for the ascending and descending *F0* trials. Error bars correspond to the standard error of the data considered.

Features Comparisons	p values
mean (12 & 15) vs. 11	0.479
mean (13 & 18) vs. 11	1
mean (16 & 19) vs. 11	0.695
mean (12 & 15) vs. 14	0.415
mean (13 & 18) vs. 14	0.856
mean (16 & 19) vs. 14	0.859
mean (12 & 15) vs. 17	0.397
mean (13 & 18) vs. 17	0.851
mean (16 & 19) vs. 17	0.854
5 vs. 6	1
4 vs. 6	0.824
4 vs. 5	0.833

Table 4. Results of planned comparison for different temporal envelope features and references. The first column shows the features averaged and the reference for which the one-way ANOVA was done. The second column shows the values for the null hypothesis that the means of the columns are equals. The numbers of each feature correspond to the ones specified in table 2.

4.1.3 Influence of the order of timbres for A or B tones

Table 5 shows the results of the one-way ANOVA with the two types of order of the A and B tones of the triplets. For each of the streaming conditions the comparisons were done with the values obtained with one of the order of A and B tones. The first five features correspond to the synchrony features and the last three to the temporal envelope features. Results show that for all the features studied the order of the A and B tones did not had a significant effect on the timbral segregation.

Features Comparisons	p value
1 vs. 6	0.816
2 vs. 7	0.481
3 vs. 8	0.771
4 vs. 9	0.731
5 vs. 10	0.308
12 vs. 15	0.952
13 vs. 18	0.164
16 vs. 19	0.336

Table 5. Results of planned comparison for the order of the A and B tones in the triplets. The first column shows the features compared in a one-way ANOVA . The second column shows the values for the null hypothesis that the means of the features compared are equals. The numbers of each feature correspond to the ones specified in table 2.

4.1.4 Influence of direction of *F0* change

Table 6 shows the results of the one-way ANOVA analysis for the influence of the ascending and descending method in the streaming of the 19 features studied. The first ten features correspond to the synchrony features and the last nine to the temporal envelope features. The results show that the segregation is significantly different for the ascending and descending methods.

4.1.5 Influence of the musical experience of subjects

The results of the one-way ANOVA for the influence of the musical experience of eight of the subjects (4 musicians and 4 nonmusicians) is shown in tables 7 and 8. The results are shown for all the features studied where the first ten conditions correspond to the synchrony features and the last nine to the temporal envelope features. The results for the ascending method of table 7 show that for most of the features (11 of 19) the results of the musicians and nonmusicians timbral segregation are significantly different. The descending method’s results of table 8 show that for most of the features (16 of 19) the results of the musicians and nonmusicians timbral segregation were not significantly different.

Features Comparisons	**p value**
1	0.016
2	0.000
3	0.006
4	0.004
5	0.000
6	0.000
7	0.000
9	0.000
10	0.000
11	0.000
12	0.000
13	0.000
14	0.000
15	0.000
16	0.000
17	0.000
18	0.000
19	0.000

Table 6. Results of planned comparison for ascending and descending method. The first column shows the features compared in a one-way ANOVA of the ascending results and descending results. The second column shows the values for the null hypothesis that the means of the features compared are equals. The values stated as 0.000 are p values smaller than 5 in 10000 ($p<0.0005$).

Features Comparisons	p value
1	0.019
2	0.137
3	0.076
4	0.012
5	0.027
6	0.011
7	0.053
8	0.090
9	0.029
10	0.004
11	0.002
12	0.019
13	0.009
14	0.003
15	0.064
16	0.003
17	0.015
18	0.312
19	0.054

Table 7. Results of planned comparison for the average of the streaming of musicians and nonmusicians for the 19 features using the ascending method. The first column shows the features compared in a one-way ANOVA. The second column shows the values for the null hypothesis that the means of the features compared are equals.

Features Comparisons	**p value**
1	0.426
2	0.912
3	0.915
4	0.231
5	0.803
6	0.245
7	0.162
8	0.192
9	0.759
10	0.958
11	0.045
12	0.040
13	0.307
14	0.092
15	0.545
16	0.186
17	0.034
18	0.214
19	0.103

Table 8. Results of planned comparison for the average of the streaming of musicians and nonmusicians for the 19 features using the descending method. The first column shows the features compared in a one-way ANOVA. The second column shows the values for the null hypothesis that the means of the compared features are equals.

4.2 Discussion

The results of the experiments do not show an importance of the synchrony of the spectrum features in facilitating stream segregation. These findings are not in concordance with the ones of MDS studies, where the synchrony the time varying change of the spectrum showed to be important for the discrimination between modified and non-modified musical instruments sounds (Krumhansl, 1989; Grey, 1977). A possible explanation for the different results could rely on the very different kind of stimulus used in both kinds of studies. The mentioned MDS studies considered similarity experiments with synthesized musical instruments tones which contain a wide range of harmonic components emulating the natural ones. The influence of these spectral characteristics has proven to affect the perceptual segregation of the timbral characteristics to a higher extent in studies of static-spectra tones (Plomp, 1970; Wessel, 1979).

The results of the experiments do not exhibit an importance of the temporal envelope features in facilitating stream segregation. These results do not corroborate the ones of Singh and Bregman (1997) where two temporal envelope features (95/5 and 5/95) showed to have a significant importance in facilitating the segregation of the ABA triplets compared with the AAA monotimbral sequences. Figure 15 shows the results of the investigation of Singh and Bregman plotted with the ones of the present investigation. The stepsize has been adapted to the one of the present study by multiplying the quarter-tones units by two in order to obtain eighth-tone units. The ordinate gives the mean the mean *F0* difference between tones required for stream segregation in eighth-tone steps. The abscissa shows the monotimbral references and the temporal envelope features averaged for the order of the A and B tones and the ascending and descending method. The circles correspond to the results of Singh and Bregman and the crosses to the present investigation. As it is seen the results of both investigations are in different ranges and the differences from the references to the temporal envelope features differ. The reason for the discrepancies in the range could be explained by the influence of the modifications introduced in the method with the change in the frequency stepsize. Comparing the results with the ones obtained in previous studies to the ones of Singh and Bregman, the findings of Hartmann and Johnson (1991) are confirmed in opposition to the ones of Iverson (1991). This supports the thesis that the kind stimulus could make the difference for the streaming in both studies. Iverson used synthesized musical tones while Hartmann and Johnson used harmonic tones, as in the present investigation. As discussed for the synchrony of the spectrum streaming results, the characteristics of the synthesized tones could have implied a higher segregation due to their wide frequency range that includes higher order components that the harmonic tones do not have.

The results of the planned comparisons for the order of the timbres in tones A and B showed no significant importance in the timbral segregation of all the features studied in agree with the findings of Singh and Bregman. Further investigation could include multiple comparison between the features and more dynamic characteristics.

The planned comparisons for the influence of the ascending and descending method in stream segregation showed a highly significant difference for the crossover points for each of the directions of *F0* change. This result also corroborates the results of Singh and Bregman where the ascending and descending method showed to contribute to highly significant differences in the crossover points for the monotimbral sequences and the spectral envelope features. In this case no distinction could be found that could reveal that the *F0* direction only made a difference for envelopes with sharper rise times as found by Singh and Bregman. Further investigations should include a detailed study of the importance of the *F0* direction in the response time of the subjects with very different kinds of timbre attributes. This features should include some contrasting timbre characteristics that will be easy and hard to segregate for the listener. It can be speculated that features that have more subtle perceptual characteristics, as the temporal attributes of this study, will be more dependent of the direction of *F0* change than features that are known to have a high segregation, as the pure spectral characteristics.

Planned comparisons between the results of musicians and non musicians showed that for some features the segregation was significantly different between the groups when using the ascending method (11 of 19). Using the descending method only for few features the segregation was significantly different (2 of 19). It could be hypothesized that the musicians were faster to familiarize with the tones when presented in the ascending method and probably segregate faster. The results of these planned comparisons agree with the ones of Pitt (1994) that showed some distinction for the perception of timbre and pitch between musicians and nonmusicians. It is not possible with the present study to tell which of the two groups was more sensible for timbre streaming. In the present study all the musicians were selected according to their type of musical experience with one particular musical instrument (all were piano players). Further studies could include subjects with other type of musical experience.

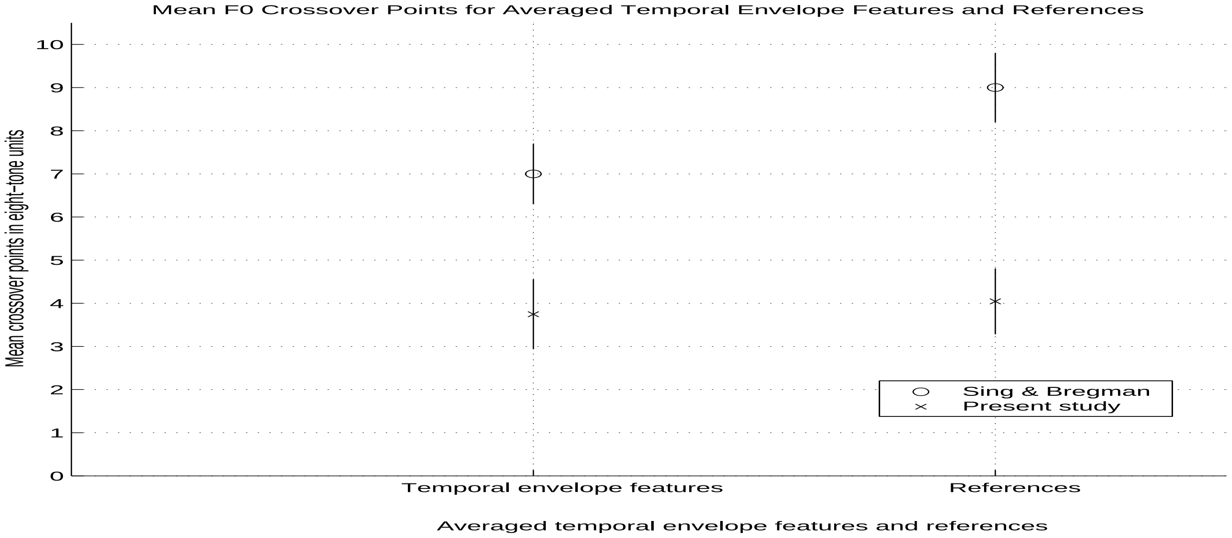


FIG. 15. Mean crossover points (in eight-tone steps) of the work of Singh and Bregman and the present investigation obtained for ten subjects for the temporal features averaged for the order of tones A and B and for the ascending and descending *F0* trials. The empty symbols correspond to the results of Singh and Bregman and the cross symbols to the present work. Error bars correspond to the standard error of the data considered.

5 Conclusion

The influence in stream segregation of two different types of temporal attributes of complex tones has been evaluated in listening experiments with twelve subjects. The temporal features considered were the time-varying changes in the spectrum and three different temporal envelope characteristics. The method used to measure the timbral streaming was the F0-tracking method which enabled measuring the segregation of the differences using the frequency space as a unit of comparison.

The effect of the studied temporal features in the segregation did not showed to be significant when comparing the timbre modified triplets with the non modified ones. The results of previous comparison MDS studies were not confirmed in the streaming of the synchrony of the spectrum features. The different stimulus of both studies could explain the divergences in the findings considering the high frequency components included in the synthesized musical tones of the MDS studies. The results of the segregation of the three different envelope features confirmed the findings of Hartmann and Johnson (1991), which showed that little or no stream segregation is observed when manipulating differences in amplitude envelope shape. These findings also support the hypothesis that synthesized musical instruments tones segregate more than harmonic tones. The effect of the stepsize in the experiments could have implied a lower segregation of the temporal attributes. It remains to be investigated if the the same experiments with a quarter-tone stepsize would give similar results.

The *F0*-tracking method used for this study enabled measuring the the influence of the different timbre attributes on perceptual segregation using crossover points as a common denominator. The method seems to suit very well to measure differences in pure spectral changes (as shown by Singh and Bregman), but does not seem to fit so well to measure more subtle differences as the ones used in the present study. The quality of this method, to measure differences in timbral attributes that are not easily distinguished by the auditory system, remains to be studied in detail. There also remains to be studied the influence of other frequency stepsizes and other method's characteristics as tempo of the sequences and frequency range of the stimulus.

The musical experience of the participants of the experiments showed to be a factor of significance in the streaming of some of the temporal features. The direction of the pitch differences showed to be an important cue for the segregation increasing the differences when the sequences were presented in ascending order. A possible explanation for this tendency could rely on the fact that musicians are faster to determine timbral changes than non musicians. Further studies should emphasize these presentation differences and study if the type of musical experience of the participants affects timbre streaming.

Bibliography

American National Standards Institute. (**1960**). *USA Standard Acoustical Terminology*, S1.1-1960.

Handel, S. (**1995**). *Timbre perception and auditory object identification, in Hearing, edited by B. C. J. Moore* (Academic Press, Florida).

Hartmann, W. M. and Johnson, D.(**1991**). “Stream segregation and peripheral channeling, ” Mus. Perc. **9**, 155-184.

Helmholtz, H. L. F. (**1885**). *On the sensation of tone* (London: Longmans & Co).

Krumhansl, C. L. (**1989**). "Why is musical timbre so hard to understand?,” in *Structure and perception of electroacoustic sounds and music*, edited by S. Nielzen and O. Olson (Excerpta Medica 846, Amsterdam), 43-53.

Grey, J. M. (**1975**). "Multidimensional perceptual scaling of musical timbres," J. Acoust. Soc. Am. **61**, 1270-1277.

Iverson, P. (**1995**). “Auditory stream segregation by musical timbre: Effects of static and dynamic acoustic attributes,” J. Exp. Psychol.: Hum. Perc. Perf. **21**, 751-763.

Letowski, T. (**1992**) “Timbre, tone color, and sound quality: concepts and definitions,” Archives of Acoustics, **17**, 1, 17-30.

Moore, B. C. J. (**1997**). *An Introduction to the psychology of hearing* (Academic Press).

van Noorden, L. P. A. S. (**1975**). “Temporal coherence in the perception of tone sequences,” Ph.D. dissertation, IPO, The Netherlands

Pitt, M.A.(**1994**). "Perception of pitch and timbre by musically trained and untrained listeners," J. Exp. Psychol.: Hum. Perc. Perf. **20**, 976-986.

Plomp, R.(**1970**). “Timbre as a multidimensional attribute of complex tones,”in *Frequency analysis and periodicity and detection in hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), 397-414.

Pratt, R., and Doak, P. (**1976**). “A subjective rating scale for timbre,” J. Sound. Vib. **45**, 317-328.

Risset, J. C. and Mathews, M. V. (**1969**). ”Analysis of musical-instrument tones,” Physics Today. **22**, 23-30.

Singh, P.G., and Bregman, A.S. (**1997**). "The influence of different timbre attributes on the perceptual segregation of complex-tone sequences," J. Acoust. Soc. Am. **102**, 1943-1952.

Slawson, A.W. (**1982**). “The musical control of sound color ,” Canad. Univ. Mus. Rev. **3**, 67-79.

Slawson, A. W. (**1985**). *Sound Color* (Berkeley: University of California Press).

Wessel, D. L. (**1979**). "Timbre space as a musical control structure," Comput. Music. J. **3**, 45-52.

Winsberg, S., and Carroll, J. D. (**1989**). “A quasi-nonmetric method for multidimensional scaling via the extended Euclidean model,” Psychometrica, **54**, 217-229.

Appendix A

Instructions for the listening experiments

The following are the written instructions given to the participants of the experiments. The instructions for the first and second experiment were handled separately to the sujects before each experience.

A.1 General Instructions

The listening experiments in which you will participate are part of an investigation of human sound perception. Initially you will hear a group of three sounds that will be repeated several times. These sounds will be presented one after each other. Before the three sounds are repeated there will be a short pause of silence. The first and the third sound in each repetition will always be the same and the only sound that will change will be the second sound. This sound will be different in some repetitions and in others it will be the same. When the second sound will be the same, the three sounds will be heard as a constant galloping rhythm with no difference at all. When the second sound will be different, no constant galloping rhythm will be heard, but a different sound in the second place.

If you have any questions, or you will wish to have some more training, please ask.

In case you have make a mistake you must tell the operator at once. It is very easy to repeat the experiments and there is no problem in making as many repetitions as you like.

Thanks for your participation.

A.2 Instructions for the first experiment

The repetitions in this experiment will start with a very different second sound that each time will be getting more and more similar to the first and the third sounds. Your task in this experiment will be to hear the repetitions of three sounds and press the button whenever the three sounds are heard as a constant galloping rhythm with no perceived difference between the three sounds.

A.3 Instructions for the second experiment

The repetitions in this experiment will start with a very similar second sound that each time will be getting more and more different to the first and the third sounds. Your task in the experiment will be to hear the repetitions of three sounds and press the button whenever the constant galloping rhythm is replaced by three sounds that have one different sound in the second position.

Appendix B

Letters to and from researchers

The following are some electronic mails that were send and received from authors dealing with the topic of timbre streaming. Only the mail sent to Albert Bregman is shown due to the fact that the others were very similar, only changing in small details.

Electronic mail sent to Albert Bregman

Mr. Bregman, I am a student of the last semester of the master in acoustics at Aalborg University in the north of Denmark (www.acoustics.dk) and I am making my Master thesis in the topic of timbre perception. I have been reading some of the literature of the field in the last months and I became very interested in the topic of stream segregation of complex tones and particularly in your publication with Punita Singh of the timbre characteristics involved in the perceptual segregation of complex tones. [P. G. Singh and A. S. Bregman, J. Acoust. Soc. Am. 102, 1943-1952 (1997)]. My idea would be to make an experiment using the same method that you used in the experiments (F0-tracking method) but trying to emphasize the perceptual contributions of temporal characteristics of complex-tone sequences (as it is proposed for further developments in the theme at the end of the cited publication). The initial idea is to make experiments using a set of equal complex sounds with different attacks times (and probably decays times) in order to determine the the role of temporal cues in stream segregation. I would like to ask you your opinion about the goal of my project and if possible some feedback about the possible features of the temporal cues that could be used in the experiment. I have also mailed Punita Singh in order to get some feedback from her. Finally, I would like to tell you that it has been very encouraging for me as a student to read your publications and your book Auditory Scene Analysis.

Thanks.

Felipe

Electronic mail from Albert Bregman

Dear Felipe,

It is interesting that you are doing your research on timbre perception. It was a good idea to contact Punita Singh for comments.The rate of onset and of offsets are two obvious things to vary in an experiment. Another possibility for manipulating temporal properties might be the complexity of time-varying changes in the spectrum. This property has come out of some multi-dimensional scaling studies as an important feature whereby listeners discriminate different musical instruments. For example, in one tone the harmonics might come on at different rates from each other, or might fluctuate in intensity independently of each other, whereas in a second tone all the harmonics would come on at the same time and maintain a constant intensity (or fluctuate synchronously with one another). In other words, tones could be different in the "stability" of their spectra, especially at the moment of onset.

Best of luck,

Al Bregman

Electronic mail from Stephen McAdams

> I would like to ask you your opinion
> about the goal of my project

I think you should look at a paper in Music Perception by Hartmann & Johnson who found that temporal factors had no effect, and another in Cognitive Psychology by Iverson who found that they did to some extent.

Stephen McAdams

Electronic mail from Paul Iverson

Felipe, thanks for the note. I've often thought of doing a study like you propose, but I have been focusing more on speech perception for the last several years. I'll be interested in what you find.

My main suggestion is that you should try to look at a wide range of temporal variation. Hartmann and Johnson looked at a temporal contrast that they thought would be maximally different - a tone with a short attack and a long decay and the mirror image tone with a long attack and a short decay. However, I have a suspicion that this difference may have had only a marginal difference on stream segregation. My guess is that one of their tones would segregate more with a tone that had long transitions on both sides. I'm also guessing that spectral flux interacts with amplitude variation in an important way, so that amplitude differences will be less likely to cause stream segregation if spectral variation is absent.

These are only speculations on my part, but my general point is that no one really knows what aspects of temporal variation are important for stream segregation, so you run the risk of missing some interesting phenomena if you make too many apriori assumptions about what stimulus dimensions are most important. Of course, there is also the risk of producing an uninterpretable mess if you include too much.

I'm moving to take a faculty position at University College London in the fall and I understand that Aalborg is not right next door to London, but feel free to stop by my futureUCL office if you are in that area.

Best,

Paul

Appendix C

Some literature related to Timbre Perception

The following is a list of selected literature, by topic, related to timbre perception that was reviewed during the development of the investigation. Some references appear more than once due to the fact that they involve more than one topic.

Auditory Stream Segregation and Timbre

Bregman, A. S.(1978). "Auditory Streaming and the building of Timbre," Canad. J. Psychol. 32, 19-31.

Bregman, A. S. (1990). “Auditory Scene Analysis: The Perceptual Organization of Sound” (MIT, Cambridge, MA). (Chapter 2, 92-130)

Hartmann, W. M. and Johnson, D. (1991). "Stream segregation and peripheral channeling," Mus. Perc. 9, 155-184.

Iverson, P.(1995). "Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes," J. Exp. Psychol: Hum. Percept. Perf. 21, 751-763.

McAdams, S.E. and Bregman, A. S. (1979). "Hearing Musical Streams," Comput. Mus. J. 3, 26-43.

Moore, B. C. J. (1997). An Introduction to the psychology of hearing (Academic Press). (Chapter 7, 249-270).

Singh, P. G. and Bregman, A. S. (1997). " The influence of different timbre attributes on the perceptual segregation of complex-tone sequences," J. Acoust. Soc. Am. 102, 1943-1952.

Timbre and Pitch interaction

Balzano, G. J. (1986). "Whay are musical Pitch and Timbre?," Music Perception. 3(3), 297-314.

Singh, P. G. (1987). "Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre?," J. Acoust. Soc. Am. 82, 886-899.

Singh, P. G. and Hirsh, I .J.(1992). " Influence of spectral locus an F0 change on the pitch and timbre of complex tones," J. Acoust. Soc. Am. 92, 2650-2661.

Moore, B. C. J. (1997). An Introduction to the psychology of hearing (Academic Press).
(Chapter 5)

Temporal attributes of signals involved in Timbre Perception

Patterson, R. D. (1994b). " The sound of a sinusoid: Time-intervals models ," J. Acoust. Soc. Am. 96, 1419-1428.

Influence of phase in Timbre Perception

Plomp, R. and Steeneken, J. M. (1969). "Effect of phase on the timbre of complex tones ," J. Acoust. Soc. Am. 46, 409-421.

Influence of spectral attributes of signals in Timbre Perception

Patterson, R. D. (1994a). "The sound of a sinusoid: Spectral models ," J. Acoust. Soc. Am.96, 1409-1418.

Timbre Perception and Musical Instruments

Saldanha, E. L., and Corso, J .F. (1964). " Timbre Cues and the Identification of musical instruments," J. Acoust. Soc. Am. 36, 2021-2026.

Slawson, A.W. (1985). “Sound Color” (Berkeley: University of California Press) (Chapter 2)

Timbre Synthesis

Risset, J. C., and Mathews, M. V. (1969). "Analysis of musical-instrument tones," Physics Today 22, (2), 23-30.

Wessel, D. L. (1979). "Timbre space as a musical control structure," Comput. Music. J.3, 45-52.

Speech and timbre perception

Assman, P. F. (1987). "Perception of front vowels: The role of harmonics in the first formant region," J. Acoust. Soc. Am. 81, 520-534.

Darwin, C. J. and Gardner, R. B. (1986). "Mistuning a harmonic of a vowel: Grouping and phase effect on vowel quality," J. Acoust. Soc. Am. 79, 838-845.

Timbre characterization

Bailey, N. J. and Cooper, D. (2000). " Sculptor: Exploring Timbral Spaces in Real Time," J. Audio. Eng. Soc. 48, 174-180.

Multidimensional Scaling of Timbre

Grey, J. M. (1975). "Multidimensional perceptual scaling of musical timbres," J. Acoust. Soc. Am. 61, 1270-1277.

Handel, S. (**1995**). *Timbre perception and auditory object identification, in Hearing, edited by B. C. J. Moore* (Academic Press, Florida).

Krumhansl, C. L.(1989). "Why is musical timbre so hard to understand?,"Structure and perception of electroacoustic sounds and music", Amsterdam:Elsevier (Excerpta Medica 846), edited by S.Nielzen and O.Olson , 43-53.

Shepard,R. N. (1962b). "The analysis of proximities: Multidimensional scaling with unknown distance function ," Psychometrika, 27, 219-246.

Wedin, L. and Goude, G. (1972). "Dimensional analysis of the perception of instrumental timbre," Scand. J. Psych.,13, 228-240.

Wessel, D.L. (1979). "Timbre space as a musical control structure," Comput. Music. J.3, 45-52.

Winsberg, S., and Carroll, J.D. (**1989**). “A quasi-nonmetric method for multidimensional scaling via the extended Euclidean model,” Psychometrica, **54**, 217-229.

Timbre Perception and Music

Bregman, A.S. (1990). “Auditory Scene Analysis: The Perceptual Organization of Sound” (MIT, Cambridge, MA). (Chapter 5 ,481-490)

Handel, S. (**1995**). *Timbre perception and auditory object identification, in Hearing, edited by B. C. J. Moore* (Academic Press, Florida).

McAdams, S. and K. Saariaho (1985). "Qualities and Functions of Musical Timbre," Comput. Mus. Proced. 367-374.

Pitt, M.A.(1994). "Perception of Pitch and Timbre by Musically Trained and Untrained Listeners," J. Exp. Psychol: Hum. Percept. Perf. 20, 976-986.

Slawson, A.W. (1982). "The musical control of sound color" , Canad. Univ. Mus. Rev. 3,67-79.

Timbre perception and musical composition

McAdams, S. and K. Saariaho (1985). "Qualities and Functions of Musical Timbre," Comput. Mus. Proced. 367-374.

Slawson, A. W. (1985). “Sound Color” (Berkeley: University of California Press) (Chapters 1, 6 & 7)

Timbre definition and concepts

Balzano, G. J. (1986). " What are musical Pitch and Timbre?, " Music Perception. 3(3), 297-314.

Handel, S. (**1995**). *Timbre perception and auditory object identification, in Hearing, edited by B. C. J. Moore* (Academic Press, Florida).

Letowski, T.(1992)."Timbre, tone color, and sound quality: concepts and definitions," Archives of Acoustics, 17, 1 , 17-30.

Matlab program to synthesize the tones

The following is an excerpt of the Matlab program was used to synthesize the complex tone triplets. In this part of the original program only the creation of the second synchrony feature complex tones has been included.

```
%% Program that creates the second synchrony feature wav file
%% for the experiments  with headphone correction
%% first the 10 synchrony features and after the 10 temporal features
%% Program that creates 3 complex tones  A - B - A  %%%%%%%%%%%
%% makes the graph of one triplet for each feature
%% temporal envelope characteristics
%% win1=(95 / 5)
%% win2=(50 / 50)
%% win3=(5 / 95)

clear all
close all

%%%%% Corrections headphones for the different harmonics

C2 = 1.3;
C3 = 1.7179;
C4 = 1.9498;

%%%%%% Constants

fs=44100;                     %sampling frequency
C=((2)^(1/48));               %constant of the increase in pitch
                             %of one eighth of a tone
n1=100/1000;                  %beginning of the first period of time
t1 = 0+1/fs:1/fs:n1;          %period of time of the first component

%%%%%%%%%% calculations of the first 4 harmonics of the complex tone A

x1=sin(262*2*pi*t1);                %first harmonic tone
x2=sin(2*262*2*pi*t1);              %second harmonic tone A
x3=sin(3*262*2*pi*t1);              %third harmonic tone A
x4=sin(4*262*2*pi*t1);              %fourth harmonic tone A
x= x1 + C2*x2 + C3*x3 +C4*x4;         %complex sound A with the 4 harmonics

%  building the temporal envelope characteristics win1 : 5/95 (onset =5 ms, offset = 95 ms)

numsamples1=length(x(1:round(fs*5/1000)));    %number of samples from 0 to 5 ms in the first tone
win01= 0+1/numsamples1:1/numsamples1:1;       %onset part of the envelope (5 ms)
numsamples2=length(x(round(5/1000*fs)+1:round(0.1*fs))) ; %number of samples from 5ms to 100ms
win12=-1+1/numsamples2:1/numsamples2:0;       %offset part of the envelope (95 ms)
win1= [win01 -win12];
%x= win1.*x;                          %assuming win1 as a temporal envelope for A (50/50)
```

```
%%%building the temporal envelope characteristics win2 : 50/50 (onset =50ms, offset = 50ms)

numsamples1=length(x(1:round(fs*50/1000)));     %number of samples from 0 to 50ms in the first tone
win01= 0+1/numsamples1:1/numsamples1:1;            %onset part of the envelope (50 ms)
numsamples2=length(x(round(50?1000*fs)+1:round(0.1*fs))) ; %number of samples from 50ms to
                                                                                                    100ms
%%win12=-1+1/numsamples2:1/numsamples2:0;          %offset part of the envelope (50 ms)

win2= [win01 -win12];

%x= win2.*x;                                  %assuming win2 as a temporal envelope for A (50?50)
%building the temporal envelope characteristics for the toneswin3 : 95/5 (onset =95ms, offset = 5ms)

numsamples1=length(x(1:round(fs*95/1000)));       %number of samples from 0 to 95ms in the first tone
win01= 0+1/numsamples1:1/numsamples1:1;            %onset part of the envelope (95 ms)
numsamples2=length(x(round(95/1000*fs)+1:round(0.1*fs))) ;  %number of samples from 95ms to 100ms
win12=-1+1/numsamples2:1/numsamples2:0;                   %offset part of the envelope (5 ms)
win3= [win01 -win12];


%%%%%%%%%% inclusion of the period of silence after tone A ( 10 ms)

n2=110/1000;
t2= n1+1/fs:1/fs:n2;                  %%the silence gap starts in the first sample after 100 ms
u=0*sin(262*2*pi*t2);

%%%%%%%%%% inclusion of the period of silence after tone B (10 ms)

n3=210/1000;                          %end of the B tone (110 ms)
t3= n2+1/fs:1/fs:n3;

n4=220/1000 ;                         %period of silence of 10 ms (gap)
t4= n3+1/fs:1/fs:n4;
q=0*sin(262*2*pi*t4);

%%%%%%%%%% inclusion of the period of silence after tone A (120 ms)

n5=320/1000;
n6=440/1000;                          %final period of time of 120 ms (gap)
t6=n5+1/fs:1/fs:n6;
v=0*sin(262*2*pi*t6);

%%%%%%%%%% Final time vector

t=0:1/fs:n6+1/fs;
```

```
%%%%%%% Creating the synchrony characteristic %%%

%%% SYNCHRONY FEATURE #2 : 5/95 ,1-2 95/5

n=2  %number of synchrony feature
win=win1;
win_a=win1;      %%envelope TONE A

T=[ ];

for N =  0 : 23,       % loop used to create the 24 tones
  Cmod= C ^ N;         % the new value of C modified (1/8 tone higher)
  n3=210?1000;         % beginning of the tone B (110 ms)
  y1= win3.*sin(262*Cmod*2*pi*t3) ;         %first harmonic tone B
  y2= win3.*sin(2*262*Cmod*2*pi*t3);        %second harmonic tone B
  y3= win.*sin(3*262*Cmod*2*pi*t3);         %third harmonic tone B
  y4= win.*sin(4*262*Cmod*2*pi*t3);         %fourth harmonic tone B
  W = y1+C2*y2+C3*y3+C4*y4;                        %tone B with the 4 harmonics
  r=[];
  s= win_a.*x;            % assigning the amplitude characteristics to tone A
  r = [ s u W u s];       %vector with the new value of pitch of the B tone
  T= [T;r];               %vector with the 24 values of pitch
end

T = T*0.15;               %%compensation for amplitude saturation


%%%%  General string labels for the wav files %%%%

D2=['ex_01_00','ex_01_01','ex_01_02','ex_01_03','ex_01_04','ex_01_05','ex_01_06','ex_01_07','ex_01_0
8','ex_01_09','ex_01_10','ex_01_11','ex_01_12','ex_01_13','ex_01_14','ex_01_15','ex_01_16','ex_01_17','
ex_01_18','ex_01_19','ex_01_20','ex_01_21','ex_01_22', 'ex_01_23'];

for e = 1:8:length (D2)   % loop to assign the label to the wav file
   D2(e + 4) = char(48+n);% corresponding to the second synchrony feature
end

J=1;
for I= 1 : 24,  % creating wav files with the corresponding labels
  wavwrite(T(I,1:1:length(T)),44100,D2(J:1:J+7));
  P= D2(J:1:J+7)   %prints to the screen the value of the wav file created
  J=J+8;
end

figure(n)
plot(T(4,1:length(T))); % plots the wav file to the screen
```

C++ program created for the listening experiments

The following is part of the code created to control the listening experiments. The following excerpt corresponds the first session of the experiments. The other sessions were not included due to space reasons.

```
// C++ program created to control the listening experiments ///

// This code only includes the routines for session 1
// The other 7 sessions follow the same procedure with the only
// difference in the order in which the triplets are played (ascending and descending)

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

// the next lines correspond to the initialization of the press-button through the parallel port of the computer

BOOL bButtonPressed=FALSE;

DWORD WINAPI check_button_thread(LPVOID){
   while (1) { // while program running - always
      //if (_inp(0x3F8)&1) { // Means that bit 1 is set on the parallel port = button pressed
      //   bButtonPressed=TRUE;
            if (_inp(0x379)&32) { // Means that bit 5 is set on the parallel port = button pressed
         bButtonPressed=TRUE;
            }
   }
   return 0;
}

main()
{
 int i,j,counter=0; // initialization of variables
   int last_triplet=0;
   HANDLE hThread;
   DWORD dwThreadID;
   char File[81];
   hThread = CreateThread(NULL,0,check_button_thread,0,0,&dwThreadID);
   char filename[15];
   char filename_matlab[25];
   FILE *fp;
   FILE *fp2;
   int session,subject;
   int first_feature=1;
   int noFeatures=19;
   int *result=new int[(2*noFeatures)+1];   //defining result vector with length

  _outp(0x378,2);    // brings up bit #2
```

```
// Here the experiment routine starts

printf("Enter number of subject:");
scanf("%02d",&subject);  // input number of session
printf("\n");           //
jumps to the next line
printf("Enter number of session:");
scanf("%d",&session);  // input number of session

if ( subject <11 )  //  making the first features different for each the first ten subjects
{first_feature = (2*subject - 1 );}

if ( subject > 10 )  // making the first features different for the second ten subjects
{
 if (subject <21)
        first_feature = (2*(subject-10) - 1 );}

 ///   SESSION #1

if (session == 1) //routine first session
{

printf("DESCENDING  # 1 SUBJECT %02d\n",subject);//prints to the screen the session and the number
of the subject
int *result=new int[(noFeatures)+1];

for (j = first_feature; j<noFeatures+1; ++j ) //loop that plays the series of triplets corresponding
                                              //to the features from the first feature for each subject until
                                              //the last one (the rest of the features (before first feature)
                                             // are played a loop afterwards
{
     Sleep(2000);  // waits 2 seconds before it starts each series
     bButtonPressed=FALSE;  // assumes that the button has not been pressed

                for (i = 23; i>-1; --i )    // i = number of triplets in a sequence  ( M = 24 )
         {
                        sprintf(File, "ex_%02d_%02d.wav",j,i);

//assigns the wav file to the variable File
                        PlaySound(File,NULL,SND_FILENAME);   //plays the file
                        printf("%s\n",File);                 // prints the file played to the screen








if (bButtonPressed) {     // if loop that checks if the butoon has been pressed
                         result[j]=i;  // stores the number of the last triplet in a vector of results
                         printf("%d %d\n",j,result[j]); // prints the last triplet to the screen
```

```
                        while (_inp(0x379)&32) {   // this while loop waits that the button is released
                       _outp(0x378,0);   // resets bit #2
                                 Sleep(1);   // time to wait that it resets
                                 _outp(0x378,2);  // brings it up again
                                                                     }
                         bButtonPressed=FALSE;
                         break;
               }

        Sleep (120); // gap of silence after the last tone A
        }
}


  for (j = 1; j<first_feature; ++j ) // j = number of sequences in a session (N = 20 )
{
                                   // loop that plays the rest of the features

                 Sleep(2000);        // Time interval before the new set of sequences starts
           bButtonPressed=FALSE;
               for (i = 23; i>-1; --i )
        {
                       sprintf(File, "ex_%02d_%02d.wav",j,i);
                       PlaySound(File,NULL,SND_FILENAME);   //plays the file
                       printf("%s\n",File);                  // copies to the screen

                       if (bButtonPressed) {
         result[j]=i;
         printf("%d %d\n",j,result[j]);

                         while (_inp(0x379)&32) {   // this while loop waits that the button is released

                         _outp(0x378,0);   // resets bit #2
                                 Sleep(1);   // time to wait that it resets
                                 _outp(0x378,2);  // brings it up again
                                                                     }
                         bButtonPressed=FALSE;
                         break;
               }
        Sleep (120);
        }
}


sprintf(filename,"subject%02d0%d.txt",subject,session);    // creating the text file string of the file to write
the results
sprintf(filename_matlab,"subject%02d0%d.m",subject,session); // creating the matlab file string of the file
to write the results

fp2=fopen(filename_matlab,"w");
// putting the results in a column vector
```

```
fp=fopen(filename,"w");           // putting the results in a column vector


  for (i=1;i<(noFeatures+1);++i)
                    {
                    fprintf(fp,"%02d\n",result[i]);  // assigning the results in the vector
                    fprintf(fp2,"featureS%02d(1,%02d) = %02d ;\n",subject,i,result[i]);
                    }
fclose(fp);
fclose(fp2);
}   // end if session = 1

TerminateThread(hThread,0); // Stops the background thread

printf("\n");
printf("End session %02d of subject %02d\n",session,subject); //print the session ended and the corresponding subject number
delete result;
return 0;
}
```