1984

Dimensional tradeoffs in the perception of complex tone sequences

Punita Gurpreet Singh

WASHINGTON UNIVERSITY

Department of Psychology

DIMENSIONAL TRADEOFFS IN THE PERCEPTION OF COMPLEX TONE SEQUENCES

by

Punita Gurpreet Singh

A thesis presented to the
Graduate School of Arts and Sciences
of Washington University
in partial fulfilment of the
requirements for the
degree of Master of Arts

August, 1984

Saint Louis, Missouri

Sound upon sound upon sound compound
many are the ways in which they are bound
mutually dependent on other sounds they surround
with frequency and time being dimensions confound.

The interaction that goes on, the "give" and the "take"
the creation of tones and the patterns they make,
make "grouping" necessary, for efficiency's sake.
So order is achieved, and yet "order" is at stake ...

"What came first ?" __ it becomes difficult to tell,
was it the tone or the hiss or the buzz or the yell ?
But the sequence comprises subpatterns as well,
where the members of a category together do dwell

Using changes along multiple dimensions as cues,
such as the pitch range, the tempo and timbral hues,
ascertain what is to breakup and what is to fuse ...
what we focus on then, is for our attention to choose.

To hear, to listen, to perceive, to appreciate,
how easily we do it, an ability innate,
of a system that copes with changes of state
enables us to synthesize, to analyze or segregate.

It happens so naturally, yet how difficult it seems,
from a motley input, a melody it redeems
and efficiently sifts, and channels into "streams"
the sounds that once composed a composer's dreams ...

ACKNOWLEDGEMENTS

The completion of this thesis represents the culmination of a time when I have had the unique opportunity of working in close contact with people and materials rooted in the various disciplines allied with the study of sound, and music in particular. Interaction with musicians, psychologists, physicists, engineers, architects, and people in the auditory sciences has enabled me to have a more informed view about the various processes that go into the creation, perception, transmission, and reception of music.

For this valuable education, I am deeply indebted to all who have helped me learn, and who have watched over me with guiding eyes as I have tried to fulfil my urge to "acquire knowledge". In so doing, I have had to undergo the inevitable experience of "growing up" and learning to stand on my own feet with some degree of confidence. Though such independence was often not desired or appreciated in the "tougher" moments, I feel stronger for having had to cope with it, and am grateful to those who helped the transition to be made as painlessly as possible.

While the extent of my gratitude and respect goes much further than can be expressed here, I would like to thank on these few pages of this document, the place and the people that helped make it possible.

The experiments were carried out in the Psychoacoustics Laboratory of the Research Department at the Central Institute for the Deaf. I am indebted to the Institute for providing me access to these facilities and for furnishing financial support for this research from a grant, NS-03856, awarded to the Institute by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS).

To Dr. Ira Hirsh go my eternal "thanks" for the help he has given me in various ways; as counselor, scientist, father, and friend. He has been instrumental in allowing me to pursue such a versatile course of study, and has provided the restraint necessary to prevent such a venture from becoming a non-specific "Jack of all trades" type pursuit. I am also grateful to him for his immense patience in dealing with my occasional childish outbursts, and misuse of prepositions !

I am grateful also to Dr. Judy Lauter who helped me get initiated in the lab, listened to my ideas, and has provided advice, criticism and encouragement from time to time.

Dr. Bob Lutfi, and Dr. Bob Gilkey have also listened to ideas regarding the type of phenomena I wanted to investigate, and have offered their opinion, suggestions and help in execution. I am grateful to them for their interest, and for the enlightening conversations (and arguments) that I have had with them. Thanks are also due to the "saviors" in all technical crises; Arnie Heidebreder, Joe Sharp and Dr. Engebretson, and to Derek Morgan and Sally Goldwasser for providing some of the software used in the experiment.

Various others have played the role of morale boosters and cynics, and have in one way or another helped this thesis arrive. Amongst them, Ken, Debby, Stacey, Jan, Dr. Arthur Niemoeller, Dr. Don Ronken, Mary Sicking, Dr. Mari Jones, Rich O'Donnell, and my listeners Dan, Michael, Jude, Pete, and Wendy.

Another acknowledgement of never-ending "thanks" goes to my brother Ina who has shared my "agony and ecstacy" in using computers, and acted beyond the call of fraternal duty in "looking after" me, by offering without any reservations, his home, car, computer, shoulder, and good humor.

Finally, the people because of whom I am here, and for whom I will always be there ; my friend Madhav, and our parents...

Abstract

DIMENSIONAL TRADEOFFS IN THE PERCEPTION OF COMPLEX TONE SEQUENCES

by Punita Gurpreet Singh

Chairman: Ira J. Hirsh.

Given a rapidly occurring sequence comprising sounds that differ from each other along some dimension, the presented sequence of "different" sounds is often found to be perceptually grouped into sub-sequences or "streams" where such differences are diminished. This study demonstrates the use of differences in "timbre" as providing the factor necessary to initiate this segregation into sub-groups.

The "timbral" attribute of four complex tones constituting a sequence has been put in "competition" with their "pitch" attribute, in order to compare the tradeoffs between the two cues that take place in dominating the grouped percept. This was done by constructing sequences of the type:

T2P1 TmP1 T2Pn TmPn

in which the first pair of tones was assigned the same pitch P1, but the tones were made to differ in their timbre. Similarly, the second pair of tones had the same pitch Pn, but differed in their timbre in the same way that the first pair did. The "pitch" and "timbre" attributes of the four-component complex tones synthesized, were both based on appropriate manipulation of the frequency dimension, the "absolute" position of the spectrum along the frequency axis, being varied to provide the timbral "distances" between sounds, and the "relative" spacing or "width" between harmonics of the spectrum, being varied to provide differences in "pitch" between sounds.

Forty-nine sequences of this type were presented to six listeners, who indicated their perceived grouping of the pattern, as being one based on pitch, one based on timbral segregation, or an ambiguous grouping that was not dominated by either cue. The results validate the hypothesis that timbre can serve as a cue to segregate sequences, and indicate that the notion of "pitch" as the dominant factor causing stream segregation, needs to be revised, and viewed in terms of "frequency", since the different timbres juxtaposed in this study always shared the same pitch, but differed in the frequency region of the harmonics used in the spectrum. Large jumps in spectral region of the neighboring tones seem to encourage the "streaming" together of the alternate tones that are nearer in frequency region, and designed to be of the same "timbre", thereby disrupting the perception of temporal order of the sequence tones.

It is thus suggested that the stream segregation phenomenon may be based on proximity in "frequency" region of the sequence events, rather than on the basis of their "pitch" per se, as has often been assumed. For pure tone sequences, this distinction is trivial, and grouping by "pitch" is equivalent to the grouping by "frequency". For sequences of complex tones however, it appears that it is the difference in frequency region of their spectra, that determines grouping, such differences affecting either the "timbre" of the tones, or the "pitch", or both, with consequent tradeoffs between these attributes taking place as the spectra of the complex tones vary for different sequences.

LIST OF FIGURES

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

LIST OF TABLES

TABLE OF CONTENTS

TABLE OF CONTENTS

(Continued)

TABLE OF CONTENTS

(Continued)

1. OVERTURE

As inhabitants of a dynamic universe characterized by "movement" at every level of observation, we are constantly in the midst of ongoing phenomena that imply changes in time and space. Changes in time are irreversible. Time is a unidirectional vector quantity that cannot be held back, and must progress onward. One can however perceive a type of constancy by the process of repetition. Things that repeat regularly in time acquire a "pattern" that may then be thought of as a constant feature of a succession of changes. The changes themselves are of course transitory, but their repetition in terms of their magnitudes and juxtaposition with other changes can render them invariant at a "relative" level. The rate of repetition or "periodicity" may be of different types, such as one way cyclical motion that characterizes revolution and rotation, or the "to and fro" oscillatory behavior exhibited by the vibrations that constitute sound and light.

Our constant interaction with such changes in time and their causes, breeds familiarity and the ability to identify and discriminate between them. This ability, in conjunction with the remarkable perceptual and motor mechanisms that we are endowed with, enables us to exploit the world of vibrations such as sound and allows us to "organize" them into meaningful patterns, to create systems of communication and expression such as speech and music.

Since these systems involve the organization and control of sounds, any sound may be a potential ingredient. However, anatomical and physiological limitations restrict the range of possible speech sounds to a rather narrow range that does not unduly tax the auditory system and thus makes for an efficient communication medium.

Similarly, in keeping with the requirements of a primarily aesthetic, artistic activity, the palette of available sounds that are typically used in music is wide and many hued. However, unlike the visual arts like painting, music unfolds its form on a canvas made of time. Akin to the space dimension in vision, time is the embedding dimension in audition. The elements of sound that occur in a small enough "time window" to be perceived as simultaneous, comprise an "event", and the stringing together of these events forms the lines of music, that are etched out, reiterated, interconnected, embellished, varied and finally brought to a meaningful end, all in the very presence of the beholder.

Organization of these events in the frequency and time domains with appropriate variations of intensity results in the formation of "ordered" auditory patterns that form the melodic, harmonic and rhythmic backbone of music. Superposition of such patterns further results in the stratified temporal superstructure that we refer to as music.

Musicians over the years have manipulated sounds in various ways and devised elaborate systems of composition rules and grouping schemes in an attempt to organize sound into aesthetically as well as intellectually pleasing structures that we have been conditioned to perceive as "musical".

Whether our response to such sounds is indeed "conditioned" by exposure and learning, or whether there is some inherent physical feature of the sound that appeals to us psychologically or physiologically is still a matter of debate and psychoacoustical studies dealing with the issue of musical perception aim to determine the nature of the relationship between the physical features of sound patterns, and the features of the patterns as perceived by the listener.

1.1 Grouping and Organization in Music

Such exploitation of the world of sound and the profound reactions evoked by such sounds necessarily implies that we must have a sensory-perceptual system that is capable of processing such a complex input, and that is able to assign the various attributes of the sounds to their rightful possessors within the input mixture.

In order to achieve this seemingly hopeless task of "grouping" and "assignment", the perceptual system would need to be "referential" or context dependent, to allow for appropriate comparisons of attributes to be made among the received sounds. Since the sounds typically arrive in "hordes", and at the same time as well as in succession, the system would need to be capable of "tracking" the auditory input over time, as well as in frequency.

Since these events take place in time, music places a heavy responsibility on memory to store the features of the events and of the larger unit they comprise, so that what is to come may be compared with what has been, thus allowing for the establishment of linkages between events, or between the "groups" of events comprising a "figure" or a larger unit. Davies (1978), commenting on the transitory nature of auditory events states that " music comprises nothing more than a series of events of the present that come together to form something which is always an event of the past, the events themselves having occurred and vanished. We can only make judgments about music by comparing recent memories with less recent memories, all concerning events which have taken place in the past ".

However, the fact that music is generally perceived in terms of motifs, melodic phrases, rhythmic groups, movements, and other such units, rather than a mere collection of isolated acoustical events seems to imply that the attributes of the events themselves may not be as salient in perception as the attributes of the larger "chunk" they form (after Miller, 1962) , which in fact becomes a unit unto itself, having an overall shape or form. It thus seems that the listening of music is not a passive process of mere reception, but rather one of active construction that serves to sort out the "motley input" into perceptually unified groups. Indeed, Neisser (1967) has written that "..seeing, hearing and remembering are all acts of construction, which may make more or less use of the stimulus information depending on the circumstances". The inclusion of "circumstances" as a variable that may effect the percept serves to act as a word of caution to researchers to hold back from making general inferences about the kind of criteria that determine grouping, since the statement indicates that the strategy used in the construction of the percept may in fact vary from situation to situation . Thus it seems that there may not be any universal rules that are always followed in organizing a complex auditory input. Rather, there may be a whole set of context dependent hierarchical rules that offer alternative organizing schemes, out of which the most salient one is followed , depending on the circumstances.

A similar conclusion was reached by Stephen Handel (1974) on studying the perception of melodic and rhythmic patterns. He concludes that the perception of complex patterns is a function of all the component patterns that exist within, thus enabling a multitude of possible organizations with no mode of organizing being more fundamental than another. Further, he feels that no complete set of principles for organization of patterns may ever be found. Be that as it may, the contention in this thesis is that grouping rules do in fact differ in their "salience" along some perceptual dimension and vie with each other for dictating the organization format. There may thus be a trading off of "power" between rules with ultimately the strongest one, as is the case in evolution, surviving.

Musical systems such as scales, chords and time signatures are created essentially to facilitate the formation of such linkages, by establishing a referential framework using which the incoming events may be appropriately ordered. Music theorist Graham George (1970) aptly sums up this aspect of musical form in his supposition that " A work of art is a meaningful unit as distinct from a mere succession of events, and that the comprehension of a series of musical procedures that gather the events into a meaningful unit requires a unifying mental act. " The statement echoes an earlier remark made by Bingham (1910) who writes, " The unity, then, which marks the difference between a mere succession of discrete tonal stimuli and a melody, arises not from the tones themselves; it is contributed by act of listener."

1.2 Gestalt Principles

This perceptual unification of separate pieces to form patterns and configurations may well be brought about by following of the grouping principles proposed by the Gestalt psychologists (Von Ehrenfels, 1890, Wertheimer, 1923/1938, Koffka, 1935). These principles operate along a certain specific dimension or attribute of the stimulus, using criteria such as proximity, similarity, closure, good continuation, covariance and "belongingness" to form coherent structures out of an otherwise confusing, complex stimulus. Since many alternative ways of organization may exist, Koffka reasoned that "the psychological organization will always be as good as the prevailing conditions allow" with the most stable organization emerging as the dominant percept (The law of Pragnanz or "best figure"). The emergence of a melody from a collection of physically disconnected sounds may thus be analogous to the emergence of the "figure" from the "ground" as spoken of by Vernon (1937) with regard to visual objects and patterns.

Melody in music is viewed by some to be the gestalt equivalent of shape in vision as it can withstand some types of transformations such as pitch transposition and tempo changes (within limits), the former being an operation in the frequency domain that preserves the frequency intervals (i.e "relative" pitch distances) between tones, while shifting them to different frequency regions, and the latter being a similar operation in the time domain, that preserves the temporal relationships (i.e "relative" durations) between tones, while changing the actual time values so that the melody as a "whole", becomes faster or slower.

However in the face of other changes such as retrogression (i.e temporal order inversion), pitch interval inversion, retrogade inversion (i.e simultaneous temporal order and pitch interval inversion) and other contracting or stretching forces acting in the auditory space defined by frequency and time, the melodic object loses its identity and undergoes perceptual breakdown followed by a process of reorganization based on another grouping scheme in keeping with the changed environment. (Recognition of such melodic transformations has been described by Dowling (1972)).

The nature of the melody in terms of its pitch relationships, timbral assignment (orchestration), rhythmic features and overall tempo seems to be an important determinant of its eventual perceptual organization, with discontinuities such as abrupt or large changes along dimensions, like increases in tempo (equivalent to reducing the inter event time intervals), and in frequency intervals between events encouraging reorganization into groups where such inconsistency is absent.

The extent of allowable change beyond which a melody gets recognizably deformed varies with the physical features of the pattern and depends on the dictates of the musical system (such as tonality) imposed on the structure. The cohering architectonic forces in music that "hold the notes together" are probably derived from perceptual experience in dealing with auditory patterns, leading to the establishment of composition rules such as the use of small pitch steps between adjacent notes, or the use of legato (smooth changes between notes) to get the effect of continuity within a phrase. Conversely, silent gaps (rests), and accents , brought about by increase in loudness, a large pitch jump or by the use of staccato (i.e abrupt execution of notes) or other such discontinuous changes are commonly used to demarcate the ending of one phrase and the beginning of another.

1.3 Frequency -Time Relations

While the auditory system is extremely good at detecting minute changes in frequency and time intervals, it seems to overlook or "forgive" much larger changes that occur in music,to some degree ascertained by the musical context in which they occur. As Davies (1978) points out ; "one changed note does not a new melody make". Indeed, notes are often changed intentionally in music in the act of improvisation or "variation on a theme", such changes being noticed, yet not construed as having changed the melody enough for it to be called "different". Thus it appears that the "difference limens" for pitch discrimination in a melodic context are markedly different compared with the acute frequency resolving power of the ear. This difference in perception may well be linked to the different roles played by the peripheral and central nervous systems in the analysis of incoming sounds on the one hand, and perceptual synthesis of the analyzed material based on higher order processes such as attention, learning and memory on the other

According to Dowling (1982), changes along a hierarchy of melodic features such as pitch, contour, tonality, and interval size determine the invariant properties of melodies, the ability to use these features for discrimination varying across the different developmental stages from infancy to adulthood. Other studies by Cuddy et al. (1976, 1981), and Dowling and Fujitani (1971), indicate that in-key pitch changes that retain the basic contour of a melody are harder to identify than changes that violate the contour and tonality. This difference in the ability to discriminate between distorted versions of a melody could thus arise due to the process of learning that automatically takes place on repeatedly encountering music composed within the framework of a particular tonal system, which enhances the information processing ability of the listener.

Another factor that affects melody recognition has been observed by Jones (1983) to be the temporal structure of the melody. In presenting a standard melody to be compared with a comparison melody differing only in the pitch of a single element, she observed that the rhythmic structure (i.e the ordering of relative durations and silent gaps between elements) effected the ability of the listener to discriminate between the patterns. Since "rhythm" implies the patterning of time into intervals demarcated by accents, the listener begins to anticipate the "beat" or the pulse of the rhythm, specially attending to the "salient" points within the pattern, as determined by the accent structure. Pitch changes occurring at these focal points are generally better discriminated than those occurring elsewhere within the pattern. On changing the rhythm across trials, the temporal features of the pattern get restructured, leading to disruption of the listener's anticipation pattern which is also correspondingly changed. Listeners who were subjected to changing rhythms across trials did poorly on the single element pitch change melody discrimination task. Jones suggests that temporal patterning guides attending, with even subtle pitch changes being discriminated better when the point of change is anticipated. Thus it appears that it is not just "what" happens, but also "when" it happens that guides the discrimination task described above, a finding similar to that obtained by Watson et al.(1975, 1976) in their study on the discrimination of word length tonal patterns. They observed that frequency resolution for individual components of a ten tone pattern was severely degraded toward the beginning of the pattern, but approached normal levels toward the end, a phenomenon they termed "the recency effect". They later attributed this position based degradation of resolving power to the "level of trial-to-trial stimulus uncertainty" between the various psychophysical conditions they employed. When stimulus uncertainty was reduced to its psychophysical minimum, so that the trained subjects now knew "when" in the complex sequence they should be listening, the frequency resolution went up to a level only slightly less than that for isolated tones, regardless of the position of the tone within the sequence.

Although the test materials and procedures used by Jones and Watson et al. were very different, both studies conclude that the uncertainty in the temporal structure of a sequence is in some way important to the perception of non-temporal structure. Jones feels that this happens because the listener is unable to use appropriate rhythmic invariants to guide responding, when given novel or ambiguous rhythms. Watson et al. imply the same reasoning by providing a constant (invariant) spectral and temporal context across trials, thereby minimizing stimulus uncertainty in their procedure , resulting in improved discrimination performance.

A subsequent study by Espinoza-Varas (1983), further suggests that listeners integrate spectral and temporal cues in the discrimination of sound sequences. Using sinusoidal tone pairs as a stimulus lacking a linguistic correlate, Espinoza-Varas found that listeners could integrate changes in frequency and duration of the component tones in discriminating between pairs. Both subthreshold and suprathreshold values of change along the two dimensions were found to yield integration, independent of the order of presentation of the cues. Varas thus proposes that "the ability to integrate spectral and temporal cues is within the repertoire of auditory processing capabilities."

This interaction between the frequency and time dimensions that frame the space in which auditory patterns are constructed, is further manifested in a study investigating the figural properties of auditory patterns carried out by Divenyi and Hirsh (1978). They found that tonal patterns remained perceptually invariant on transposition, but noted that on distortion of the frequency intervals between the component tones of the pattern, identification begins to deteriorate monotonically with the extent of interval change. The deterioration in identification is manifested by a confusion in the temporal order of the tones, the latter being preserved over intertone frequency intervals of about a half-octave, but suddenly getting disrupted when the frequency range is further increased. Since the actual physical order of the tones remains identical during such distortion, this confusion reflects a purely perceptual phenomenon that comes into play when faced with large jumps in frequency or some other inconsistency in the physical structure of the pattern that defies the Gestalt principle of good continuation.

Thus here, a difference along the frequencny dimension resulted in confusion along the time dimension. As it turns out , the frequency interval mentioned above , where a sudden deterioration in performance was noted, is close to the limits of frequency beyond which previously connected tones now appear to be disconnected and perceptually isolated into their own auditory "streams".

This phenomenon, variously labelled as "the trill threshold" (Miller and Heise, 1950), "rhythmic fission " (Dowling, 1968, Van Noorden, 1971) and , " auditory stream segregation " (Bregman and Campbell, 1971) refers to the break up of a rapidly occurring sequence of different sounds into subsequences, reflecting the use of grouping principles that serve to organize the incoming auditory input into categories based on some similarity between the elements. A loss of veridical serial order perception thus results, with the original set of sequentially ordered "different" sounds being erroneously perceived as co-occurring simultaneous subsets of "similar" sounds. "Different" and "similar" here refer to magnitudes of changes along the primary acoustic dimensions such as frequency, intensity and timbre, and also to the spatial attributes of the sounds, such as the apparent locations of their sources.

This phenomenon is perhaps the most convincing example of the organizational capabilities of the auditory nervous system , since the groups formed are purely perceptual and depend on the changes made along the primary physical dimensions of the component sounds of the pattern, the percept changing in a nonrandom fashion with the changes in physical pattern structure.

The perceptual system seems to use such cues in accomplishing the task of sorting out a typically heterogeneous input into meaningful subunits. The different elements appear to be compared and related on the basis of these dimensional attributes, and the proximal elements filed into their own relatively homogeneous "stream".

Since relations along many different dimensions may serve to be "the" organizing factor that is used in segregation, the system often needs to "weigh" the alternative competing options and decide to organize according to the most "powerful" one. We are rather adept at this task of auditory organization, although competition amongst alternative ways of organizing often lead to differences in perception.

As a consequence of this perceptual reorganization of the input pattern, the time relationships between all the component events may be obscure, but temporal order of the events comprising a subsequence is preserved. The events within a subsequence are members of one perceptual group, with the sequence as a whole thus comprising members across perceptual boundaries. Thus although the ability to serially organize the entire input sequence gets worse, the ability to discriminate between subpatterns gets better.

However, the changes occurring along any of the available dimensions seem to be implicitly dependent on the time dimension, since such perceptual segregation phenomena only seem to come into play at fast presentation rates of the order of eight events per second and higher (according to Jones, 1978). Jones (1976) maintains that that "there is a lawful relationship between the degree of physical change within a sound pattern and the durations of sounds changed, such that as the magnitude of the physical change increases relative to a constant time duration, auditory streaming will occur. In general, it appears that disruptions in the lawfulness of relationships of these several kinds lead to confusions in temporal order of the sounds involved."

This lapse in the veridical perception of patterns may therefore paradoxically be viewed as "confusion" or perceptual breakdown, or as an efficient alternate way of organizing, and therefore as an accomplishment.

A clear illustration of this fact is evident in Dowling´s (1973) experiment on the perception of interleaved melodies. His stimulus consisted of two melodies whose elements were temporally interleaved. The interval sizes between the notes of each melody were thus destroyed by the inserted "foreign" pitches. He found that transposing either melody so that its notes got pulled out into a frequency region different from that of the other melody, resulted in restoring the individual identities of the melodies, enabling their correct detection. For fast rates of presentation (8 to 10 tones/sec), smaller pitch separations were needed for the melodies to split apart into their own coherent auditory streams.

In another study on melody recognition, Deutsch (1972), did the reverse of the situation just described, in that she distorted interval sizes between notes of a melody by distributing them across octaves. Such scrambling resulted in the lack of recognition of even familiar melodies, probably because the notes of the melody were now in their own perceptual streams formed on the basis of pitch distance, resulting in the loss of integration between consecutive elements that resided in different streams.

The streaming phenomenon thus might account for our success or failure in recognizing melodies as a function of the dimensional relations employed within their structure. Melodies that have events widely separated in time can seemingly withstand larger changes in frequency as well, as compared to a fast melody that tends to break up in the face of the same frequency jumps. All melodies, or all auditory patterns in general are thus not equally resistant to deformation. Some patterns are sturdier than others ...

Indeed, this phenomenon appears to be the outcome of higher order organizational ability of the auditory nervous system that comes into play to sort out the auditory environment into cogent perceptual groups.

A more complete understanding of the structural properties of the patterns that cause this to happen might shed light on some very fundamental issues in the perception of music, and pattern perception in general, such as : why complex tones are perceived as cohesive units while complex tones produced by different sources yielding different timbres are kept apart perceptually , and if this is indeed a "hard and fast" property of a sound or a context dependent attribute that might undergo perceptual reorganization when temporally juxtaposed with other sounds. Exploration of such disorganization or reorganization of sounds brought about by juxtaposition, in terms of the physical features of the sounds, may further reveal the role of "context" in auditory perception.

Being the epitome of a "grouping phenomenon", studies on stream segregation may well be applicable on a larger time scale, in explaining the circular way in which rhythm is used in music to "measure" time and provide the basic chassis or framework, referred to which changes along other dimensions can be correlated and grouped.

It is in the quest of answers to questions such as the ones raised above, that this research is undertaken. In particular, the study presented here aims to provide empirical evidence to support the speculation that the features of sounds in isolation may markedly differ from those that result when sounds are concatenated in time, due to the existence of "choices" that are made available to the auditory system by virtue of the very nature of the "pattern" itself.

The specific details of this query are outlined in the chapter that follows the next one dealing with past research in this area of pattern perception.

2. RECAPITULATION

" ... no sound is an island ..."

2.1 Sounds in Sequence

Due to the ubiquitous frictional forces in nature, that tend to "damp" the oscillatory motion of sound sources, one rarely gets to hear the so called "steady state" sounds that have been the subject of such avid investigation in the laboratory. Rather, sound sources typically tend to come on and off, their activation being dependent on the supply of energy provided by a natural driving force such as wind, or by calculated control over some generating source such as electricity, or by mechanical excitation as in the bowing or plucking of a string, the striking of a drumhead, and in the controlled exhalation of air resulting in the structured vibrations representing speech and song. The duration for which different sound sources remain activated varies, as does the quality of the sound produced, and its pitch and loudness.

The vast number of sound sources that exist and their sporadic or repetitive production of sound makes for quite an active and multifaceted auditory environment the features of which are communicated to us via a "net" vibrational pattern representative of the various concurrent vibrations present simultaneously in the medium. This complicated pattern traverses the transmission medium and finally arrives at the ear of a listener, the portal to the auditory nervous system, where it gives itself up for decoding, thereby conveying information about the environment, such information being vital for our very survival. Indeed, "hearing" is here viewed to be the primary sense, responsible for the communication between organism and environment, for the major part. The auditory system is extremely sensitive indeed, and "unlike the visual or the tactual world, the world of sound cannot be "locked out" of the domain of our hearing by the simple shutting of an eyelid, or by physical isolation." (Singh, 1983).

In order to be an effective interpreter of the environment, the auditory system must carefully monitor the complicated mixture of sounds and determine some "method in the madness", so that the sources of the sounds, the order of the sounds, and other relationships between them may be ascertained, allowing for some meaningful structure to be assigned to an otherwise seemingly chaotic "mess" of intertwined sounds. The relationships perceived between the various sounds and their components may reflect the blatant physical relations between them, resulting in so called "veridical" perception, or they may transcend the obvious, and reflect the subtle relations that might exist between the components of the sounds at the "microsonic" level. The relationships thus perceived may therefore seem illusory at a superficial level, while reflecting the "truth" at some other level.

Leaving aside the issue of unitary events for the moment,the present review covers the perceptual aspects of "sequences" of sounds. Such patterns comprise separate events occurring in succession. Each "event" itself may also be a grouping of synchronous or asynchronous sounds, that are nevertheless perceived as a single unit, because they occur in a small enough time window to cause the perceptual system to temporally "smear" them into a single fused percept, such as a sound of a peculiar "texture", or even a particular "pitch", determined by the rate at which the "elementary" sound occurs. (A phenomenon described as "periodicity pitch" or "repetition pitch" (reviewed in Plomp 1976, p.138)).

When presented a serially ordered pattern comprising such sounds, a listener is faced with the task of not only observing the multidimensional attributes such as pitch, timbre or intensity of the single events themselves, but also the task oftracking down the changes that occur along these dimensions from event to event over time.

2.2 Perception of Order in Sequences of Sounds

The perception of sound patterns that comprise speech and music requires a listener to be able to discriminate between the events on the basis of features such as pitch or timbre, and also to be able to order the events to arrive at conclusions about the nature of the pattern, such as "an ascending melodic pattern", or the distinction between words containing the same phonemes, but grouped in a different order as is the case in "fits" and "fist". In order to perceive these words as different, a listener "must not only be able to discriminate between the two speech sounds (/s/ and /t/) themselves, but further must be able to perceive their order of occurrence." (Hirsh, 1959).

The rate at which these events occur in time, as inferred from the inter event time intervals separating them in the sequence, seems to be an important determinant of how these events are to be grouped.

Based on the range of time intervals used, Hirsh (1974) proposes that different mechanisms may be employed in the perception of temporal order within auditory sequences. According to him, there appear to be three distinct categories differentiating patterns by virtue of the temporal relations between them. For onset to onset intervals upto 20 msec, he suggests that order discrimination of the events of short sequences is subserved by any cue that provides a qualitative difference, since such sequences are so rapid as to be perceived as single entities, with any change being noticed as a change in the overall quality of the compound sound. For the range from 20 to 100 msec, sequences are often perceived as gestalten or "figures", such that the events constituting the figure are related due to some common property. Changes in stimulus order in such sequences results in the perception of a change in the figural properties of the sequence, facilitating discrimination, even though the precise change in order may not be identified by the listener. For larger time differences beyond 100 msec, an item by item analysis of the pattern is suggested, allowing for the correct perception of order.

While these three categories suggested by Hirsh may infact exist, there seems to be a lot of variability in the demarcation of the time values separating them, with the actual values being dependent on the number and the type of sounds used, the task set, the number of times the sequence is presented, and the experience of the listeners.

For pairs of sounds presented one at a time for identification of order, Hirsh (1959) found that order could be reported correctly for time intervals of the order of 15 to 20 msec separating the sounds. Using a variety of sounds such as tones of different frequency, clicks (pulse tones), and noises of different bandwidths, Hirsh observed that the detection of order seemed to require a separation time of about 20 msec independent of the kind of sounds used. While acknowledging that temporal features of the individual sounds themselves, such as rise-time and duration, effected the judgement to some degree, Hirsh goes on to ignore these factors and concludes that 20 msec seems to be an adequate length of time between sounds to be able to say "which came first".

Extending such experimentation with pairs of events to cases when sounds were presented to different ears in succession, or when stimuli crossing the bounds of audition to include vision and touch were presented for judgements of order, Hirsh and Sherrick (1961) reported the same value of 20 msec as being enough to bring about (75 %) correct judgements of order. They thus concluded that a time interval of about 20 msec seemed to be some sort of a fundamental limit for the perception of order, independent of the modality employed.

Hirsh and Sherrick also suggest that temporal "resolving power" has two components; one involving the perception of successivity of events, which is a necessary but not a sufficient condition for the other, involving the perception of order of the events. Thus while a time interval as short as 1 or 2 msec (Green,1969) seems to be enough to enable a listener to judge that there are two sounds as opposed to one, an interval greater by an order of magnitude seems to be needed to then be able to judge which of the two detected sounds preceded the other. The former interval thus reflects the "temporal grain" of the auditory system (Hirsh, 1975), while the latter interval seems to suggest "that the judgement of order requires other mechanisms than those associated with the peripheral auditory system". (Hirsh, 1959).

By the above account, it appears that the auditory system is extremely sensitive to temporal manipulation of sound, as indeed it should be, in order to keep track of the "fast moving multidimensional arrays" that constitute speech and music (Jones, 1976).

On one hand, it appears that listeners can discriminate between sequences on the basis of the order of their component events down till values of only one or two milliseconds for the time interval between them. Such discrimination however seems to be based primarily on the judgment of qualitative differences of the net "sequence", rather than on the individual identification and subsequent ordering of the events themselves. In order to actually report the order of the events, it seems that this limit for the time interval between the onsets of the sounds has to be increased to about the 20 millisecond value reported by Hirsh et al. In contrast to these results reflecting a fine degree of temporal resolution, are the findings of Warren and associates, in their research on the ordering of recycled sequences comprising different types of events. In particular, the finding that the heterogeneous sequence employed by Warren, Obusek, Farmer and Warren (1969) required between 200 and 700 msec to elicit correct ordering seems to indicate some sort of "breakdown" in the temporal resolving power of the system.

A variety of explanations have been offered to account for these paradoxical results. Some researchers (e.g Warren et al.) attribute the supposed "failure" in perception to factors related to the listener's responding ability. Others (e.g. Norman,1967) attribute this breakdown to the limited capacity of the processing mechanisms involved. Still another school of investigators (e.g. Bregman et al.) feel that the "breakdown" in order perception actually reflects "breakup" of the sequence into subsequences based on similarities between the physical features of the sounds. Accurate order perception is assumed to prevail within the subsequences, although the original sequence is incorrectly judged.

This latter group of investigators support their argument by a wide range of experimental results that indicate that the perception of order is indeed effected by the physical properties of the sequence events themselves. In particular, their results imply that it is not just the "rate" of presentation of a sequence as stated by Hirsh (1974), but also the frequency of sequence events relative to each other. that determines how the order will be construed. Furthermore, it appears that it is the inherently interdependent relationship between time values and frequency values, that ultimately effects the perception of order of the events of a sequence of sounds.

2.3 The Effect of Frequency Differences between Events

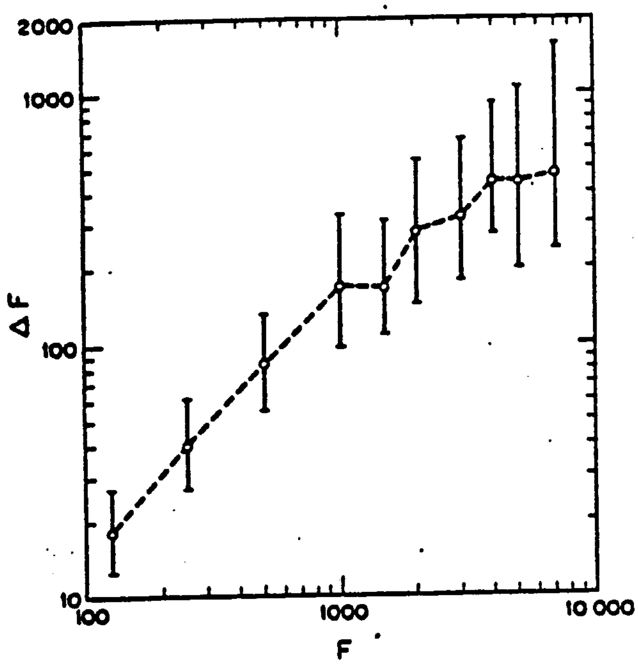
In relating incidents of temporal confusions, to the existence of "limited capacity processors", Norman presents some interesting observations regarding the dependence of temporal ambiguity and consequent misjudgment on factors other than the time relationships between the events. He notes that "a sequence of single auditory notes is perceived as two different patterns with no obvious time correlations between them, if the notes are drawn from two different samples of frequencies separated by a sufficient frequency difference."

In the investigation of parameters involved in making two such perceptual sequences out of one physical sequence, Norman observed that the position of a probe tone inserted in the time interval between two alternating tones of high and low frequency respectively, could be determined if the frequency of the probe was inbetween the frequencies of the sequence tones. However, if its frequency was either much higher or lower than the frequency of the sequence tones, the probe tone could not be temporally localized with accuracy. This experiment seems to indicate that the frequency relationships between the tones is a crucial determinant of their temporal order, capable of disrupting the fine time resolving ability of the auditory nervous system, depending on the particular circumstances. Norman also observed that the splitting effect was not possible with a single presentation of two tones. Rather, some additional tones were needed before or after the stimulus pair, and the splitting effect was best observed when the entire test sequence was repeated over and over again, without very large gaps.

These observations made by Norman echoed the results of an earlier experiment carried out by Miller and Heise (1950), on the effect of the frequency difference between two tones, in the perception of the alternating sequence formed by their repetition.

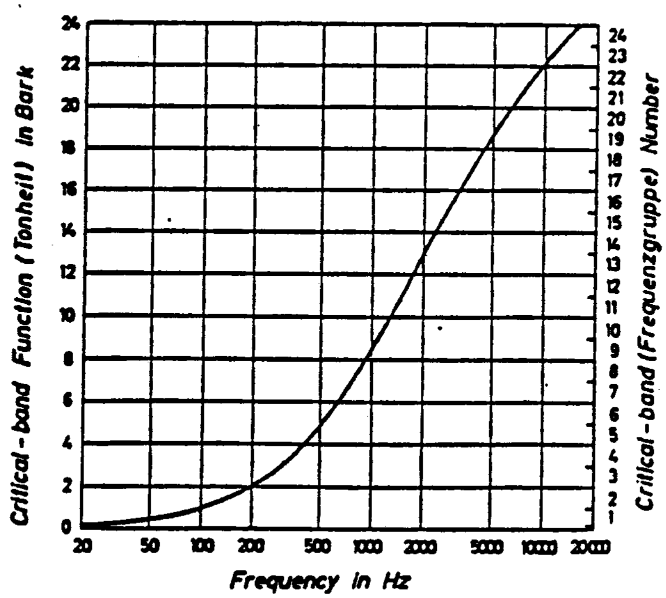
Miller and Heise observed that for a certain value of the frequency difference, the perception of this repeated two tone sequence ABABABA... , underwent a sudden change from being a "trill", reflecting the continous up and down movement of pitch,to being perceived as the co-occurrence of two unrelated, interrupted tones , A.A.A.A... and B.B.B.B... They labelled this critical frequency difference "the trill threshold", and determined its value to be about 15 % for pairs of tones located in various frequency regions. This proportionality between the frequencies of the two tones, resembles the proportional relationship between the critical bandwidth and the particular frequency around which it is determined. Figure 2.1 shows the trill threshold function in comparison with the critical band function suggested by Zwicker (1961).

The similarity thus led Norman to speculate that the ordering mechanism may not be able to keep track of a tone that lies beyond the critical band of the preceding tone. Indeed, the observations of Miller and Heise do seem to encourage the viewpoint of a perceptual breakdown of a mechanism that normally follows the true order of sequential events. This breakdown however occurs only for particular values of frequency difference, between tones presented at a very rapid rate.



Size of trill threshold as a function of frequency.
Length of vertical bars indicates interquartile range.

Figure 2.1 Comparison of the Trill Threshold function (Miller & Heise, 1950) with the Critical Band function (Zwicker, 1961).



Relation between the critical-band function and the frequency. The ordinate scale at the right side gives the critical-band numbers of the present proposal.

Dowling (1968), used the term "rhythmic fission" to describe this phenomenon. For a sequence of tones repeating rapidly at the rate of 10 tones per second, (the same rate used by Miller and Heise), he claimed that if the alternate tones were separated by small frequency differences, while successive tones were separated by large frequency differences, then two simultaneous melodic patterns could be heard, formed by the bunching up of the alternate tones. The original sequence thus undergoes "rhythmic fission". Dowling further claimed that such fission could be brought about by other differences between tones, such as intensity differences and differences in "place", brought about by stereo separation. He stated that both the soft as well as the loud patterns thus emerging, could be attended to as the "figure", while with position cues a melody could be made distinct from a melodic or random background, the latter requiring a greater degree of stereo separation.

Another interesting claim made by Dowling regarded the ability of an observer to ferret out a particular target melody, embedded in a jumbled sequence of tones of the same intensity, the alternating embedding tones being randomly selected from the same frequency range. This active piecing together of a melody from notes hidden amongst other notes, indicates that "attention" may play a vital role in enabling an observer to form groups based on various schemes. According to Dowling, no observer reported hearing any melody after being misinformed about which melody to listen for, thus indicating that this extraction process was not illusory in nature. Rather, it seemed to operate on the actual information present in the stimulus.

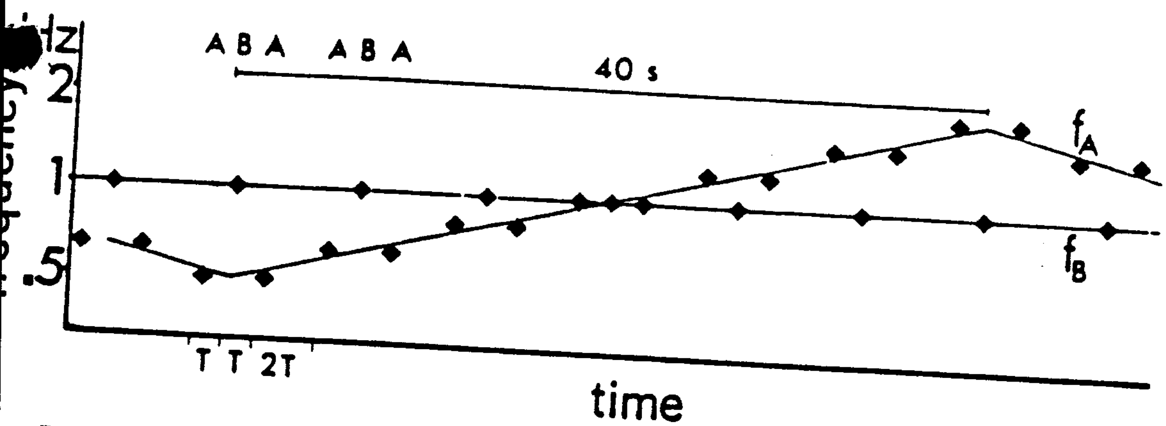
These findings on the perceptual breakup of tonal sequences by Miller, Heise, Norman, and Dowling, although implicitly understood and exploited by musicians at a non-analytic, practical level (discussed in Ortmann, 1926), received little experimental attention till around 1971, from when on a number of investigators (primarily van Noorden and Bregman et al.) decided to undertake research on the subject of temporal coherence of sound sequences in considerable detail.

2.4 The Effect of Rate of Presentation (Tempo)

Since it had long been known that the fission phenomenon only came into play at "rapid" rates of presentation, Leon van Noorden (1971) attempted to quantify this dependence on time, by varying the tone rate in presenting a two tone sequence similar to that used by Miller and Heise.

In his report on "Rhythmic Fission as a Function of Tone Rate", van Noorden (1971) lays stress on the influence of the "attentional set" of a listener in the perception of a tone sequence, and points out that a tone sequence may be listened to in two paradoxical ways, depending on the attentional motivation of the listener. A listener may "try" to listen to a sequence as a single coherent pattern, referred to as "fusion", or "temporal coherence" by him, or a listener might choose to try and hear the sequence as being separated into subsequences, the original pattern undergoing "fission" . The "trill threshold" studied by Miller and Heise thus demarcates the boundary between the fusion and fission states of the tone sequence (ABABAB...) used by them.

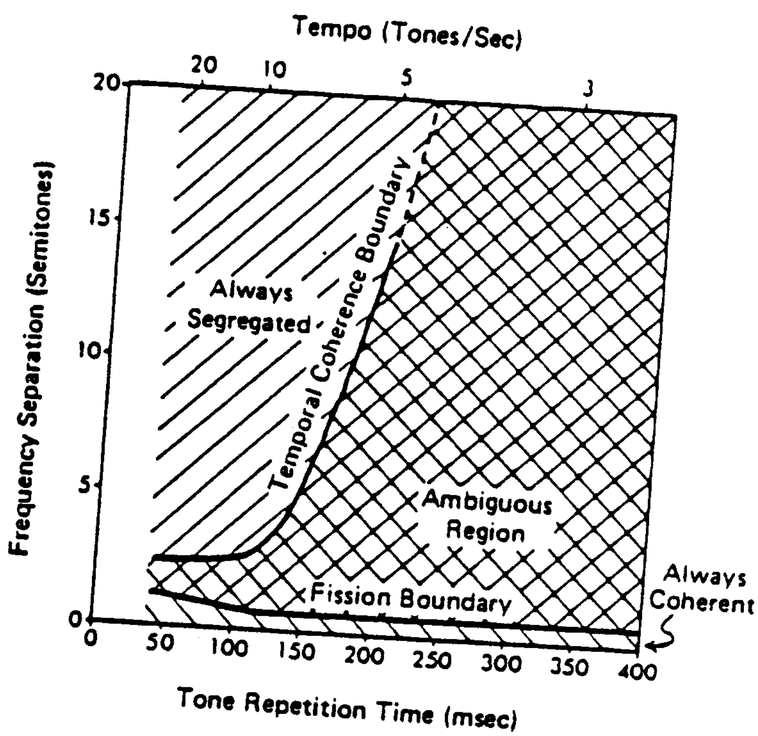
To determine the changeover from the fused coherent state to the fission state, van Noorden employed a rather clever technique, exploiting the rhythmic relation between the tones of the sequence, brought about by the omission of every second "B", from the sequence ABABAB... . In the fused state, this would result in the creation of a "galloping" triplet rhythm for the sequence ABA.ABA.ABA... , while upon undergoing fission, the subsequence A.A.A... would appear to be twice as fast as the subsequence B..B..B... as is shown in figure 2.2.



The frequency sweep of the tones A.

Figure 2.2 a). Rhythmic layout of the sequence used by van Noorden (1971), showing the changeover from the fusion to fission state.

b). The Boundaries of Temporal Coherence and Fission, defining three perceptual regions in the stream relationship between two alternating tones, as a function of both the tempo, and the frequency separation of the tones. (From McAdams & Bregman, 1979).



The creation of the rhythmic "cue", created due to the peculiar relationship between the onsets of the tones made it possible to determine the "switchover point" when the coherent sequence with the galloping rhythm broke up into the two subsequences in the fission state.

Since the fission of the sequence is dependent on the frequency relationship between the two tones, as well as on tone rate, certain parameters of the sequence need to be fixed, to avoid confounding of variables. Fixing tone duration at 50 msec, and the frequency of tone B (Fb) at 1000 Hz, van Noorden provided the listener control over the frequency of tone A (Fa). Thus for different values of the "inter event (time) interval" (IEI), a listener could sweep log Fa linearly in time. The tone rate, determined by the onset to onset interval, or the "IEI", was varied from 50 to 200 msec, corresponding to silent intervals of 0 to 150 msec between tones, incremented in steps of 10 msec.

To determine the "temporal coherence" boundary, van Noorden instructed his listeners to try to listen to the galloping rhythm of the presented sequence, as long as possible, while sweeping the frequency of tone A up and down. The moment the gallop rhythm was imperceptible, and the sequence seemed to break up, the frequency difference (Fa- Fb) was recorded. This was repeated for all interval durations presented in a random order. Similar measurements were made in the determination of the "fission" boundary, by instructing the listeners to try and hear the tones A separated from the tones B, as long as possible, till arriving at a difference in frequency at the particular tone rate being tested, that resulted in the tones cohering into the fused triplet rhythm again.

Using this "sweep" technique for varying the frequency difference between tones, van Noorden differentiated three regions in the frequency-time space for each subject, as is illustrated. For appropriate differences in frequency at the different rates used, he demarcated a region in which fission would be likely to take place. For suitably small frequency differences, a coherence (fusion) boundary was indicated, that seemed to be independent of the onset to onset duration range from 60 to 150 msec. For frequency differences at slower rates, a region was indicated, wherein fusion or fission could take place, depending on the "set" of the listener.

These results support the hypothesis that changes occurring in the time and frequency domains are inherently interactive. As summarised by McAdams and Bregman (1979), "it appears that there is an essentially inverse and strictly interdependent relationship between tempo and frequency relationships among tones: the faster the tones follow one another, the smaller the frequency separation at which they segregate...". "Conversely, the greater the frequency separation, the slower the tempo at which segregation occurs".

These observations are in agreement with those of practising musicians and composers, that the higher the rate of presentation, the less the frequency difference required to split a sequence into separate melodic lines. This phenomenon is exploited to advantage by composers in the perceptual creation of two or more melodies from a single sequence produced by a single source. The relationships between the notes can be cleverly manipulated to form sub-melodies of a desired format, giving the illusion of polyphony. Music of the Baroque era, particularly that of Bach and Telemann, and lately of the group of composers referred to as "the serialists", exemplifies this principle, showing ways of grouping sounds that change as the features of the sounds change.

2.5 Pattern Perception and the Formation of "Streams"

In the course of investigating organizational processes in the perception of rapid sequences of sounds, Bregman and Campbell (1971) encountered the fission phenomenon described above. In their description of the phenomenon, "a single rapid sequence of tones seems to break up perceptually into two or more parallel sequences, as if two or more different instruments, each restricted to a certain class of sounds, or range of frequencies, were playing different but interwoven parts". They labelled this phenomenon "primary auditory stream formation", later abbreviated simply to "stream segregation" or "streaming". They defined a "stream" to be "a sequence of auditory events whose elements are related perceptually to one another, the stream being segregated perceptually from other co-occurring auditory events". Subsequent definitions of a "stream" (Dannenbring and Bregman, 1976, McAdams and Bregman, 1979) define it to be a psychological organization that mentally represents some sequence of acoustic events as emanating from one location, considered to be equivalent to a physical "source". A perceptually organized sequence of this type is assumed to display a certain internal consistency or continuity that allows that sequence to be interpreted as a "whole". Stream segregation is viewed by these authors to be a parsing mechanism, that enables the auditory system to decompose auditory inputs into those that arise from separate sources. Such decomposition thus results in the creation of substreams of sound whenever sounds of different types occur rapidly in a mixed sequence (Bregman and Dannenbring, 1973).

Since the previously referred to experiment on temporal order perception performed by Warren, Obusek, Farmer and Warren (1969), employed a "mixed" repeating sequence containing a hiss, a buzz, a high tone and a low tone, Bregman and Campbell (1971) proposed that the failure in accurate perception of order at fast rates (less than 200 msec per event), might be related to the phenomenon of stream formation, that was resulting in the isolation of the "different" events into their own perceptual streams. Lack of interaction between the streams was assumed to result in temporal ambiguity, and consequent confusion of order. Bregman and Campbell thus decided to test this hypothesis by performing an experiment analogous to Warren et al.'s, but using sounds whose similarities could be known in advance, namely six pure tones, three of which were chosen from a high frequency range (2500, 2000, and 1600 Hz), and three from a low range (550, 430, and 350 Hz). Each of the tones was 100 msec in duration, resulting in the same "rate" of presentation of the sequence, as was used by Miller and Heise, i.e. 10 tones/sec.

In accord with their expectations, Bregman and Campbell observed that in the identification of order, their subjects invariably grouped the high tones together into one stream, and the low tones into another stream. For different arrangements of the six tones in the presented sequence, the preservation of order relationships between the tones of the twenty possible "triplet" subsequences, was looked for, and it was found that the subjects were consistently superior in their judgement of order for "within stream" triplets containing tones from the same frequency range, than they were in ordering of tones forming "across stream" triplets. The subjects had difficulty in relating the events belonging to different streams, yet were able to relate the events belonging to the same stream quite accurately.

To further verify their hypothesis that subjects would be unable to relate elements across streams, Bregman and Campbell included a discrimination task in their experiment. This required listeners to judge whether or not three tones presented in a "standard" sequence, occurred in the same order and temporal spacing in another sequence presented for "comparison". The standard sequences used, were designed to contain either within-stream triplets, comprising three tones of the same frequency range, or across-stream triplets, containing two tones from one frequency range, and one from the other range. The tones used were the same as used in the identification task, their durations, and the durations of the three silent gaps in this case being 100 msec. each. The comparison sequence contained six tones, the silent gaps being filled by the remaining tones not used in the standard sequence. The three tones present in the standard, were thus always presented in the comparison sequence as well, but not necessarily in the same order. Six different conditions varying in the frequency and temporal juxtaposition of the tones were tested.

It is amusing to note that Bregman and Campbell took special care to ensure that the tones within the streams would be arranged such that no rhythmic cue would facilitate the detection of a change in the comparison pattern brought about by streaming, while van Noorden especially relied on this cue, using it to advantage in determining the occurrence of streaming (fission).

On analysis of their data, Bregman and Campbell concluded that "at the rates used, there is essentially no ability to relate material from different streams", although "the speeds involved were not too high for accurate order judgements, provided that the comparison restricted itself to elements of a single stream". Since different temporal arrangements of tones had been used for generality, the authors imply that it is not the shifting of attention from event to event that is a time limited process causing confusion of order due to lack of sufficient "processing time", as suggested by some authors (e.g. Stroud, 1955, Estes, 1972, Massaro, 1972). Rather, it seems that it is the "nature" of the events themselves, determining the perceptual grouping into streams, that results in the time dependent process of shifting attention from stream to stream.

The dependence of order perception ofthe events of a pattern on the physical features of the sounds, is further corroborated by these results of Bregman and Campbell. As Mari Riess Jones (1976) later pointed out in a theory on the role of "time" in perception, it appears that "the magnitude of pitch change, relative to time change within a specified sequence of pitch relations, is crucial in determining which pattern contexts will lose serial order via streaming, and which will not". Large time intervals between events seem to be more capable of handling large changes in frequency, while short intervals can handle only correspondingly small differences in frequency, thereby losing their resilience, and breaking up the sequence, when subjected to large frequency jumps.

2.6 The Role of Context

From the work on streaming reviewed so far, it appears that "differences" in frequency, rather than absolute values of frequency, are what determine the coherence or breakup of auditory patterns at different rates of presentation. However, it appears that even the differences in frequency are not sacred in and of themselves. A given frequency difference between two tones causing streaming in one pattern context may cause the tones to cohere given another context. The physical features of the entire pattern seem to determine the context of the relative differences between the events. The relations between the successive and simultaneous sounds all interact to finally determine if the pattern will undergo restructuring or not. A subsequent change of circumstances ensuing from changed features of the sounds, would further modify the relative salience of pattern events, leading to corresponding reorganization.

Evidence for the occurrence of such a context sensitive organizational process was provided by quite a dramatic experiment performed by Bregman and Rudnicky (1975), in their exploration of the relationship between stream segregation and pattern recognition.

On the basis of earlier work (Bregman et al. 1971, 1973), these investigators were of the opinion that a listener is able to attend to only one stream at a time. This selectivity in attention was assumed because of the manifested inability of listeners in identifying sequences distributed across streams, and lack of correct temporal ordering between elements of the different streams. In attending to only one stream at a time, the listener was assumed to reject the elements of the other stream, so that there was no interaction between the streams.

To verify if this was indeed the case, and the listener was oblivious to the features of the elements of the unattended stream, Bregman and Rudnicky manipulated the physical attributes of the so called "background stream", and observed the effects of this manipulation on the "target stream".

The target stream was created by using small differences in frequency between four tones constituting the sequence XABX or XBAX (A = 2200 Hz, B = 2400 Hz, and X = 1460 Hz). The tones A and B were thus embedded in a stream that contained "distracting tones" (X) flanking the pair on either side as is shown in figure 2.3.

Discrimination of these two sequences would demand a listener to basically judge the order of the tones A and B. At the 66 msec inter-event interval used, this task ought to be fairly easy, considering such ordering was shown to be accomplished easily at values of IEI as short as 20 msec and less (reviewed by Hirsh, 1974, Warren, 1982). However, the inclusion of the tones X before and after the to-be-judged pair, results in the removal of onset and termination cues, that usually aid the identification of order (Warren,1970).

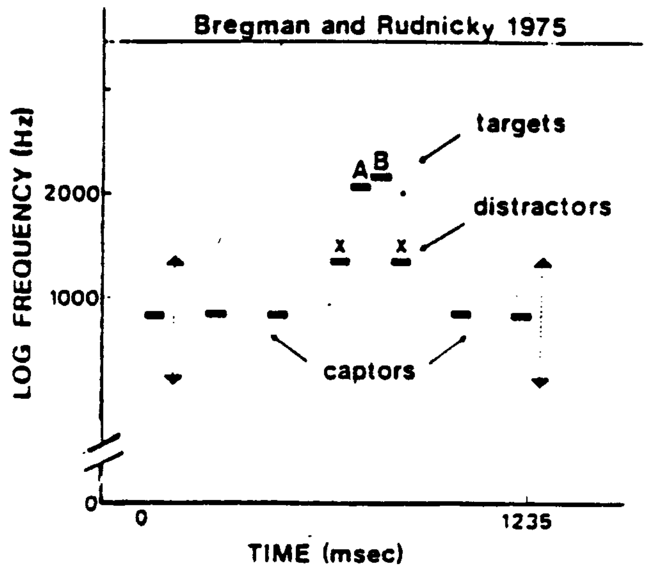


Figure 2.3 The stimulus used by Bregman and Rudnicky (1975), showing "captor" tones (C) of variable frequency that are used to capture the "distractor" tones (X), leaving the "target" tones (AB) free for unhindered judgement of order.

Due to the lack of such cues, it therefore becomes difficult to compare the order of A and B when they are a part of the stream XABX (or XBAX), with the case when they are presented as a "standard" pair not belonging to any stream. The frequency difference between the tones of the target pair and the flanking tones is apparently small enough to keep the four tones glued together within a single stream. The presence of the "distractor" tones thus, indeed seems to distract the listener into confusing temporal order of the target pair.

To see if this was indeed the case, and the frequency difference would always yield streaming between the four tones, Bregman and Rudnicky introduced another set of "captor" tones "C" preceding and succeeding the sequence XABX. By varying the frequency of the tones C (and thus the frequency difference between C and X), they attempted to "capture" the tones X from the target sequence by forming a second stream CCCXXCC. If this could be accomplished, the distracting tones would be stripped away, leaving A and B in a separate two tone sequence whose order could be judged with greater ease. The success of this scheme would depend on the "winner" of the competition between the magnitudes of the frequency intervals (Fa-Fx or Fb-Fx), and (Fx-Fc). The tones X would be assigned to whichever stream was near enough in frequency to capture them.

Using three conditions for the frequency of the captors (590, 1030, and 1460 Hz), and a condition with no captor tones, Bregman and Rudnicky found that it was indeed possible to capture the distracting tones into a separate stream with the captor tones, which could be ignored due to the selective attending entailed by the stream segregation process. The tones A and B were left in a stream of their own, that could then be attended to undisturbed, facilitating order discrimination.

The results of the experiment revealed that order discrimination was hardest for the no captor condition and the far frequency (590 Hz) condition. It was easiest when the captors and distractors were the same frequency, and intermediate in difficulty when the captors were near the frequency region of the XABX pattern.

On the basis of these results, Bregman and Rudnicky propose that "listeners have an adaptive rejection process which alters its frequency range over time". This process apparently starts off with a wide frequency acceptance range, which is subsequently narrowed over time if no alternative frequency events are encountered. Events differing widely in frequency which are encountered after the acceptance range has been narrowed are thus rejected. Conversely, if events not differing too greatly in frequency are encountered before the tuning process has set in, then these events will be accepted within a single "stream".

This plausible but poorly argued proposal made by these authors thus hypothesizes the existence of a filter like mechanism, the pass band of which can change over time. If the filter tunes in to a band around the captor frequency region, it would pass the captors and near frequency events like the distractors unhindered, but would blockade the pair AB from passing through, leaving them sifted from the rest of the pattern. The modification in acceptance region exhibited in this experiment due to the context provided by the values of the frequency differences used, supports the notion that "context" is what determines the "power" of the organizing principle based on pitch proximity, rather than the actual values of the frequencies used.

However, surveying the type of stimulus used by these authors, and the arguments consequently put forth by them, brings to light a curious array of conflicting claims made by them.

To begin with, the stimulus used comprised single presentations of the test sequences, and not the recycling sequences typically employed in streaming studies. It is possible that had the test sequence been repeated, the distractor tones could have formed their own stream separated from the AB stream. The authors however continue to refer to this phenomenon as being one of stream segregation, even though they do not provide enough "similar" types of events in the no captor condition to allow for grouping by similarity (i.e streaming) to take place. The captor condition on the other hand provides three captor tones before and two after the test sequence, such repetition facilitating the formation of streams based on the availability of enough similar and dissimilar events to be compared and then grouped.

Another peculiar feature of the stimulus used is the difference in inter-event-interval between the tones comprising the sequence CCCXABXCC. Although all the tones used here had the same duration (65 msec), the silent gaps between the events were not equal. Indeed, there was no gap at all between the events of the sequence XABX, while the onsets between the tones C and X were maintained at a fixed value of 130 msec. This assignment of silent durations between events makes for a peculiar rhythmic configuration, as is illustrated. It therefore turns out that instead of eliminating onset and termination cues to prevent ease of order identification as they presumed to have done by the arrangement of the distractor tones, these authors unwittingly provided a rhythmic cue that was confounded with the occurrence of the AB pair.

This rhythmic difference itself could result in isolating the AB pair as a separate "fast" sequence, as opposed to the remaining slow sequence. The ease in discriminating order in the captor condition is thus confounded with the existence of both frequency and time cues. The rhythmic cue provided in this sequence is illustrated in figure 2.4.

Jones, Kidd, and Wetzel (1981) investigated this rhythmic confounding further, replicating the experiment of Bregman and Rudnicky for a variety of rhythmic contexts, that resulted in isolating the pair AB on a separate two tone rhythm, or the sequence XABX on a separate four-tone" rhythm, or maintained the whole sequence on an isochronous rhythm, created by providing equal onset to onset intervals between all tones.

The two-tone rhythm that provided the target tones with a different tempo than the one shared by the captor tones and the flanking tones, yielded the best performance. The worst performance proved to result from the four-tone rhythm, that tied the target pair and the flanking tones together in a rhythm that was out of context with that of the captor tones, and therefore was difficult to segregate to yield accurate judgements of order of A and B. Strangely enough, no effect of the captor pitch distance was found. The interaction of pitch distance with rhythm was however significant, reflecting best performance when captor

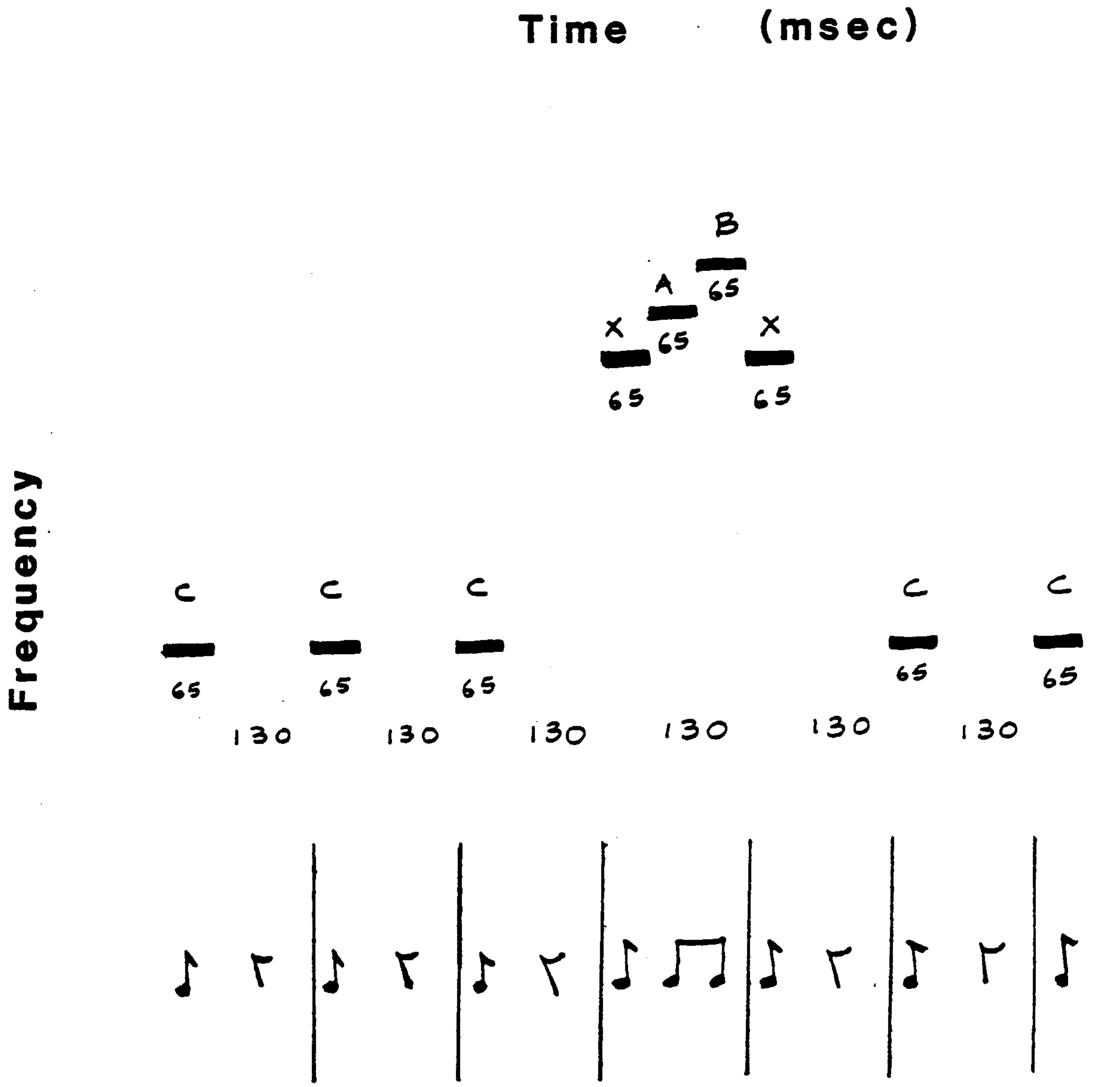


Figure 2.4 Portrayal of the rhythmic cue unwittingly provided in the stimulus used by Bregman and Rudnicky (1975), because of different values of the inter-event-interval used between the different sequence events.

tones had the same frequency as the flanking tones, and also shared their rhythm, thereby differentiating the pair AB in terms of differences in both frequency and rhythm. The sequence layout is illustrated in figure 2.5.

Jones et al. also tested the hypothesis that the rejection process undergoes tuning over time, by increasing the number of captor tones inaugurating the stimulus sequence. They found that the lengthened captor string strengthened initial periodicity, but did not facilitate performance per se. Rather, perfomance continued to be confounded with the rhythmic cue.

The results thus support a rhythm sensitive aspect of the attention mechanism, rather than the wandering filter-like mechanism proposed by Bregman et al. The rhythmic tuning appears to manifest itself by enhancing the listener's ability to temporally predict events within patterns, such as changes in frequency across events. The magnitude of a captor effect based on frequency is thus seen to increase as temporal anticipation becomes stronger.

The conclusions drawn from this study prove to be extremely interesting indeed, as they imply that the temporal features of the events serve as powerful contextual cues in organizing patterns, along with other cues such as pitch distance. The data suggest that

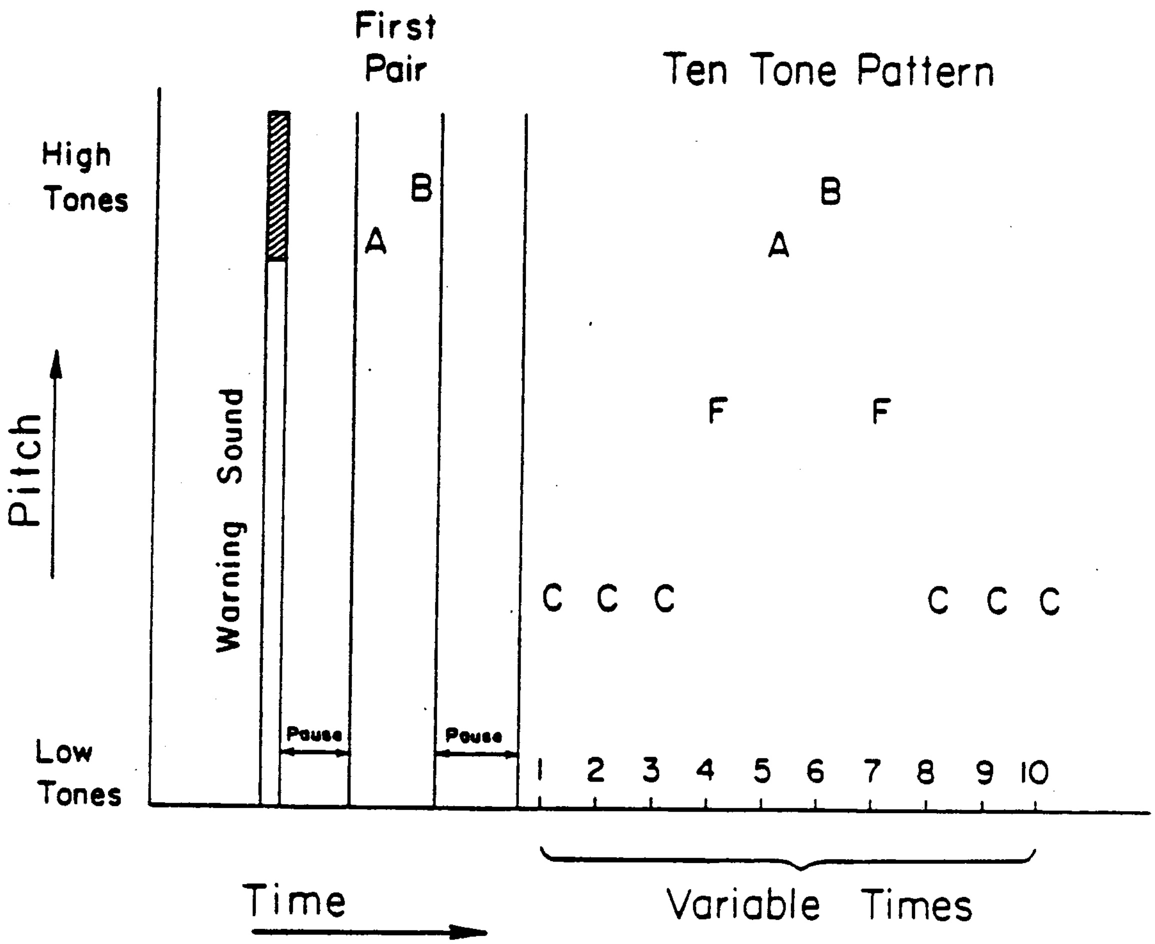


Figure 2.5 The stimulus used by Jones, Kidd & Wetzel (1981) in further investigation of the effect of rhythmic layout of the sequence used by Bregman & Rudnicky (1975) in eliciting "streaming". The inter-event-intervals were varied in order to provide different rhythms for the sequence.

rhythm plays a role in frequency based stream segregation, the time relations between events interacting with the frequency relations to provide cues for segregation.

In the light of this study, it appears that conclusions regarding organizing principles in phenomena such as stream segregation that are drawn from studies using isochronous rhythms, should not be generalized blindly. Rather, it should be acknowledged that such results may hold only for the case of the particular rhythmic context used, and may change radically under different conditions of temporal organization.

Additional evidence for segregation phenomena on the basis of frequency differences has been provided in many studies. The role of temporal factors in bringing about segregation however, remains a less well explored area. A study on temporal order perception carried out by Divenyi and Hirsh (1975) around the same time as Bregman and Rudnicky did however report the effect of both frequency and duration of an irrelevant fourth tone in obscuring the order of a preceding three tone sequence.

In eliciting judgements of order for three contiguous 20 msec tones (891, 1000, and 1118 Hz), they found that the addition of the extra tone at the end of the three note pattern, resulted in degraded judgements of order compared to the case when the pattern was presented in isolation (Divenyi and Hirsh, 1974). This perceptual erasure of a preceding pattern by immediately following irrelevant material was termed "blanking" by these authors, a term they borrowed from studies on similar backward-inteference phenomena in vision ("Type B masking": Kolers, 1962).

The effect of the fourth tone in the blanking of the three-tone pattern was investigated as a function of both the frequency and duration of the added tone. It was found that the fourth tone succeeded in "rubbing out" the image of the target pattern to the greatest degree, when it was identical in frequency to the first tone, or when its frequency lay between 1/6 to 1/3 octaves above the highest pattern frequency. The blanking was also found to be strongest when the duration of the fourth tone was equal to that of the pattern components. The blanking could be made least effective if the added tone was blatantly different from the pattern tones, in terms of its frequency and duration. The effect was also found to be attenuated if the added tone served as a logical extension of the three-tone pattern, forming a regularly ascending or descending set of frequencies.

The obvious similarity of these results to the work on streaming, led these authors to make statements uncannily reminiscent of those made by the streaming researchers, yet apparently in ignorance of the streaming literature.

With no mention of the streaming phenomena studied prior to the time, these authors go on to make identical claims about the importance of the physical features such as frequency and duration of the tones, in determining the context that the added tone acquires. They report that when the frequency of the added tone coincided, or was sufficiently close to the frequency of the pattern tones, "two partially overlapping patterns" were created, which were "likely to compete for the subject's response". The overlap could be reduced by differentiation of the pattern tones and the added tone, by using differences in time and frequency values, that apparently resulted in the two patterns referred to above being separated and organized as two "Gestalts", comprising the three tone pattern as one entity, and the incongruous tone as a separate entity. The perceptual separation thus achieved resulted in restoration of ease of identification of order for the tones comprising the three-tone pattern.

A later experiment carried out by these same investigators (Divenyi and Hirsh, 1978) also indicated that a three tone "figural" pattern embedded sequentially among other irrelevant tones could be separated out from the "background" by providing both "vertical" cues, based on spectral differences between the pattern tones and the background tones, and also by "horizontal" cues in the time domain, such as placing the pattern at the very beginning or even better, at the very end of the sequence. The temporal position and the spectral composition both could therefore be used as separation tools to highlight the camouflaged pattern.

This type of induced separation (or "destreaming") of a confusing sequence leading to the revelation of a target pattern hidden within it, is thus precisely what Bregman and Rudnicky achieved in their experiment cited above. Although their main aim was to show the dramatic effect of context in assigning a new meaning to a frequency interval, they succeeded in achieving the desired segregation on the basis of both frquency and time cues, provided by the strange rhythmic layout of their sequence.

Another experiment carried out by Bregman (1978 a), further stressed the role of context in bringing about the splitting of a sequence into subgroups. Bregman also demonstrated that alternative potential groupings compete to incorporate sequence events into their own streamed structure. An event belonging to one stream may therefore defect to another, if the events of the other stream share some similar feature with the defecting event, thereby luring it to join them.

Using a successive pair of tones A and B separated by a fixed frequency difference (varied across conditions), and a fixed temporal interval (65 msec), Bregman showed that the pair would either be integrated into the same stream, or segregated into separate streams, depending on the context of other tones X and Y, following A and B in a repeating four tone sequence ABXY, as shown in figure 2.6.

The frequencies of the tones X and Y were varied so that in some cases they were too far from those of A and B to have any influence on them. However, if X and Y are shifted to another frequency range that includes the tones A and B, segregation of the original sequence into streams based on frequency proximity can be brought about, giving the percept of two sequences A-X and B-Y, instead of the original A-B X-Y sequence. The "uneven" doublet rhythm of the original sequence correspondingly changes to an "even" single beat rhythm on undergoing fission.

The frequency range within which the perceptual system groups tones on the basis of frequency, is thus not constant. Rather, it varies depending on the values of all the tones or events making up the pattern. Thus it would appear that the perceptual system like the systems of forces and currents in the physical world, chooses to

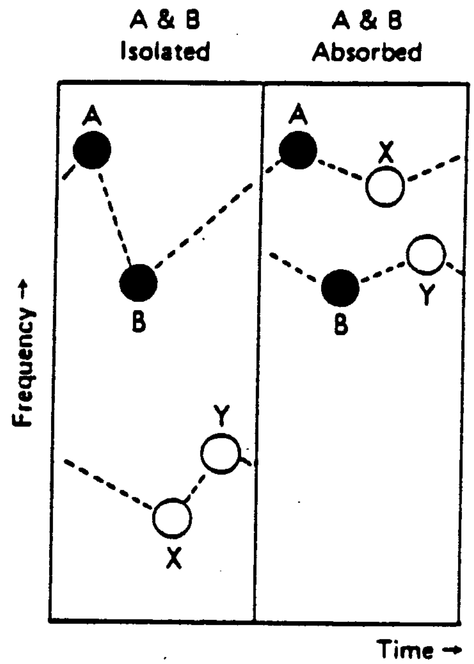


Figure 2.6 Sequence used by Bregman (1978) illustrating the effect of frequency "context" as opposed to frequency difference, in stream formation. The stream "AB" in the first part, is seen to breakup into the streams "AX" and "BY", as tones X and Y are moved into frequency regions nearer those of A and B. The frequency difference between A and B remains the same in both cases. (From McAdams & Bregman, 1979).

travel the path of "least resistance", manifested by the smallest difference in frequency (in the absence of other organizing cues along other dimensions).

Using seven different conditions for frequency differences between the four tones, Bregman observed the most interesting phenomenon that tones separated by frequency differences as great as 10.2 semitones remained coherent within a single stream, although the predictions made by other investigators (Miller and Heise,1950, van Noorden, 1975) would have expected frequency differences between 2.4 and 6 semitones for pure tones alternating at about 10 tones/sec to split up perceptually into separate streams.

This difference in results arises precisely because of the fact that an alternating sequence of two tones was not used. Such repetition would have facilitated streaming of tones with their own repeated counterparts of identical frequency ! The use of the four tone sequence in Bregman´s experiment provided the additional context of two extra tones, thus enabling greater room for comparison and consequent grouping.

Thus the "content" and "context" provided by the events composing a sequence both influence the formation of perceptual groups within the sequence. The grouping usually follows some principle such as nearness in frequency, temporal proximity, or some other rule , all of which are specific to the circumstantial variables such as rate of presentation, the existence and features of surrounding tones, and other non-quantifiable variables such as the attentional set of the listener. Given similar values for the variables under different conditions of presentation may result in an entirely changed percept. Sequential processing thus appears to be a matter of following well defined rigid rules under circumstances that cannot be so precisely defined because of the multitude of variables involved, and their complicated interaction.

Given the apparently complex nature of this phenomenon, leads one to wonder how we are so adept at this task of organizing sounds routinely into the patterns of music and speech. Given the role of context, it seems that there may well be something about the very patterns themselves that we unconsciously create, in order to facilitate organization. It therefore seems to be in order, to consider some of the features of the sounds of speech and music, and look for clues as to what are the special attributes that serve to connect these sounds to form patterns, and to examine the larger structural features of the patterns such as the frequency trajectory representing the intonation pattern or melodic contour.

2.7 Factors Connecting Events in Auditory Patterns

The definite effect of rate of presentation on the fission of sequences as demonstrated in the above described research, readdresses the question as to how and why we are able to preserve the order of rapidly occurring speech sounds, even when such sounds typically occur as fast as 30 phonemes per second (Orr, Friedman and Williams, 1965). In the light of the evidence presented thus far, it would seem that these fast paced speech sequences ought to break up into streams based on some similarity criterion between the different sounds, yet listeners are apparently able to make accurate judgements of temporal order at these speeds. However, there seems to be some discrepancy between the accuracy with which people identify sounds in natural speech, in synthetic speech, and in isolation (reviewed by Jones, 1976).

The experiments performed by Warren in particular, presented the puzzling results (reviewed in Warren, 1982), that though listeners were unable to identify the order of the four unrelated non-speech sounds until their duration was increased to 700 msec, they were surprisingly accurate in their judgement of order for a sequence containing four spoken digits at durations of only 200 msec each. Indeed, it seems paradoxical that correct identification was accomplished with ease by all listeners, even though each of the spoken digits was a "word", entailing the correct ordering of the phoneme sequence within the word itself. Other experiments (by Thomas, Hill,Carroll, and Garcia, 1970, Thomas, Cetti, and Chase, 1971, and Dorman, Cutting, and Raphael 1975, among others) dealing with the perception of order of recycled phoneme sequences, reveal that synthetic vowel sequences, or sequences of abrupt segments of steady-state vowels, need upto about 200 msec intervals, to be correctly ordered, while only 30 msec suffices for identification of vowels in isolation, or in the course of natural speech. The insertion of silent gaps between the vowels, is observed to be an aid in the naming of order (Cole and Scott, 1973), while the insertion of normal articulatory transitions linking the successive events in the sequence, further facilitates the identification of order.

It thus appears that the splitting of a sequence of sounds, is dependent on the "transition" from sound to sound, as well as on the similarity of features of the sounds. Since the transitions between the sounds of speech are not instantaneous or abrupt, streaming tends to get suppressed, resulting in the preservation of order. It is also believed (Jones, 1976) that the syllabic context in speech facilitates the retention of order, due to the transitional rate of change of certain formant frequencies between adjacent phonemes in the syllable. Rather than assuming two separate perceptual systems for the processing of speech and music, it seems logical to believe as Jones does, that there is something in the nature of the "pattern" itself that determines perceptual phenomena such as streaming, brought about by the ability of people to detect and use relative, rather than absolute features between events, such as the relationship between pitch changes and time changes within a pattern, to structure the environment.

With the aim of testing the hypothesis that "the continuity of the acoustic properties of successive moments of sound might be a factor tying the successive sounds of speech into a single stream", Bregman and Dannenbring (1973) carried out two experiments to determine if streaming of a sequence containing high and low tones, could be suppressed by providing a frequency glide ("glissando") , between successive tones.

In the first experiment, they required subjects to discriminate order of the tones, when presented with a standard and comparison sequence of tones, each sequence comprising two high tones (H1 = 2000 Hz, H2 = 1600 Hz.), interspersed between two low tones (L1 = 614 Hz, and L2 = 400 Hz). In a second experiment, the subjects were required to judge if a single presented sequence, did or did not segregate into streams.

The patterns used were made to vary according to two conditions: in the type of transition between tones, and in the length of steady-state time of the tones. The different durations used thus resulted in different rates of presentation of the sequence.

The transition between successive tones could be of three types (as illustrated in figure 2.7). In the "ramped" condition, frequency and amplitude were made to vary linearly from one tone to the next, resulting in a smooth glissando stretching over the 40 msec break between the steady-state segments of the tones. In the "semi-ramped" condition, a frequency change in the direction of the succeeding tone accompanied a 45 dB fall or rise in amplitude over 10 msec, at the end and beginning respectively, of the steady-state portions of the sequence tones, thus leaving a 20 msec silent gap between the tones. In the "discrete" condition, the frequency change between tones was abrupt and disconnected, though a 45 dB drop or rise in amplitude over 10 msec succeeded and preceded the steady portion of the tones as before, with a 20 msec silent gap in between. Three different durations ; 100, 150, and 225 msec, were used for the steady-state portions of the tones.

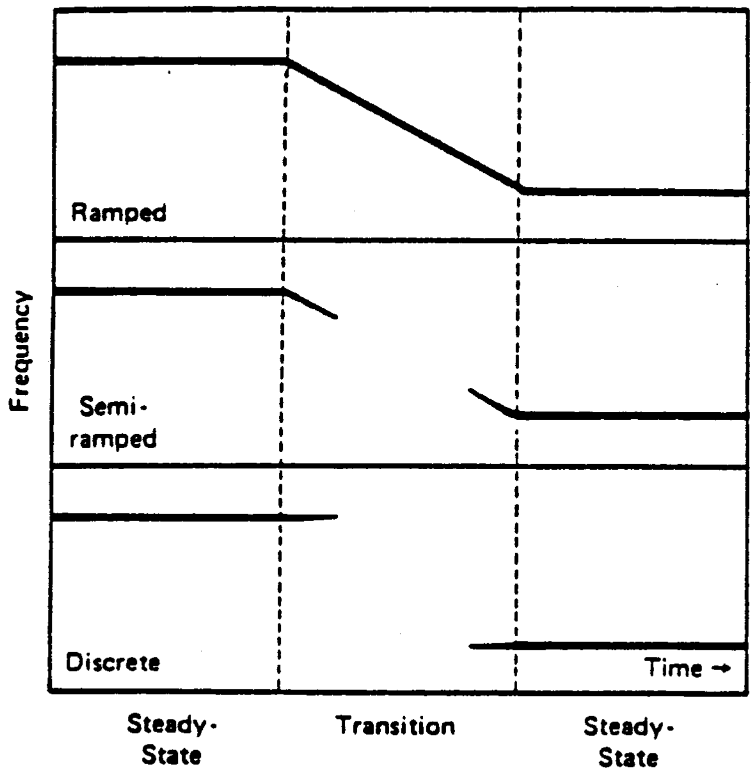


Figure 2.7 The three types of frequency transitions used by Bregman & Dannenbring (1973), in determining the role of such transitions in inhibiting the formation of streams. In the "ramped" condition, the tones are completely connected by a frequency glissando. In the "semi-ramped" condition, partial glides are provided, while in the third condition, the frequency changes abruptly between the tones. (From McAdams & Bregman, 1979).

In a manner quite typical of Bregman et al.'s reporting style, the description of the stimuli is incomplete. The extent of amplitude change for the ramped condition is not mentioned, and the effect of the rhythmic nature of the sequences described, is left open to speculation. In implying that the transition region between the tones is kept fixed at 40 msec, while tonal durations are varied, it would appear that a difference in relative durations of the transition and steady-state portions would result in the creation of different "rhythms" from condition to condition, the existence and effect of which is not mentioned.

Bregman and Dannenbring obtained results that showed that judgements of order were superior for the ramped and semi-ramped conditions, than for the discrete condition at each of the steady-state durations used. Performance was also found to improve as the durations of the tones were increased. From the judgements on "streaming" obtained in the second experiment, it appeared that the discrete sequences were much more prone to splitting, than were the ramped sequences, while the semiramped sequences showed an intermediate degree of segregation. Increase in steady state time of the tones, resulting in slower rates of presentation, was found to inhibit streaming, in keeping with the results of van Noorden described above.

These results obtained by Bregman and Dannenbring, led them to conclude that correct judgements of order were dependent on the occurrence of stream segregation, the likelihood of which increases, when subsets of sounds occupy different frequency regions, when the presentation rate is high, and when the transitions in frequency are discrete.

The use of glides and partial glides in this experiment, shown to have inhibited segregation, lends support to the belief that the sounds of speech are preserved in correct order even at fast rates, because of the existence of transitional acoustic cues between the phonemes. The semi-ramped condition employed in this experiment in particular, appears to be very similar to the spectrograph patterns of syllables studied by Liberman, Cooper, Shankweiler, and Studert-Kennedy (1967). These authors argue that a motor-speech decoder is responsible for the ordering of speech sounds, but it seems to be more reasonable in view of the results just described , to conclude, in agreement with Bregman et al, that "the partial gliding or "pointing" in the acoustic components of speech may be a strong factor holding it together as a unified stream".

Another factor that could be responsible for the preservation of order in speech , is the fact that speech sounds typically comprise sounds of different "types", such as vowels, fricatives, plosives etc. It is possible that these sounds do not recur at fast enough time intervals with respect to each other, to enable subsets of similar sounds to group into streams. The experiments of Cole and Scott (1974), on the recycling of single consonant-vowel syllables indicate that this may be the case. On recycling a single consonant vowel sequence, so that similar speech units were repeating rapidly, they found that it was possible to split up the syllable into separate streams containing the consonant and the vowel. Another finding by these investigators (described by Bregman et al. 1973), was that the glided formant transition between a consonant and a vowel in a syllable, often remained united with the vowel, resisting the effect of stream segregation, while the consonant detatched itself to form a separate stream. A syllable such as "sha" was thus apparently found to be perceived as the syllable "da", due to the retention of the glide before the /a/.

From the review of the work on streaming presented so far, it seems that differences in frequency and the type of transition between events, both contribute in encouraging or discouraging stream segregation from taking place.

In extending the concept of "glide" to apply to melodic "contour" Heise and Miller (1951) studied longer auditory patterns of different "shape" (contour), and degree of "steepness", determined by the magnitude of the ratio of frequencies of successive tones. Using repeating sequences of eleven tones arranged in different orders, such as linearly ascending, descending, V- shape, and inverted V-shape arrangements of the tones separated by different frequency intervals, Heise and Miller attempted to determine the frequency separation required between the center tone and the rest of the pattern, in order to cause it to "just" dissociate from the rest of the pattern tones. Such dissociation causes the center tone to be isolated as a perceptual "pop" distinctly separated from the rest of the sequence. As the patterns were made steeper, i.e the frequency intervals between successive tones within a pattern were made larger, the amount of frequency change required to produce segregation was also found to increase. In addition, it was found that a smaller frequency difference was required for the center tone of the V-shaped patterns to dissociate, rather than for the linear patterns. This trend seems to exhibit a tendency to group the sequence following the Gestalt law of "good continuation", since the linear patterns are more predictable, and therefore more resistant to cleavage, than are the V- shaped patterns that abruptly change direction.

From the work of Heise and Miller,it would appear that these criteria are indeed valid on a larger scale, when referred to a "melody" comprising a series of notes separated by different melodic intervals and having an overall melodic contour or "trajectory". As summarised by McAdams and Bregman (1979), the basic rule seems to be that "large jumps and sudden changes in the direction of a melody produce discontinuity", with the miscreant tones often being removed from the melodic continuity of the rest of the pattern, into their own streamed pattern, that may be considered to be "separate", and emanating from a different source.

The dependence of streaming on melodic contour as implied by Heise and Miller, was further investigated by van Noorden (1974), in his study on "Temporal Coherence in Random Tone Sequences". In an attempt to understand the role of "attention" in the perception of temporal coherence of tone sequences, van Noorden compared the "temporal coherence boundary" for listeners presented with predictable alternating tone sequences, and sequences built from tones separated by randomly chosen tone intervals. His rationale in doing this was the speculation that if attention played an "active" role in the following of tone sequences, then an observer ought to be better at following familiar sequences. If however, the perception of sequences is an automatic process, then it should not make a difference if the sequence is known or not, and the temporal coherence boundary should depend only on the magnitude of the frequency differences.

This indeed seemed to be the case, according to the observations made by van Noorden. Previous knowledge of what tones might follow, did not appear to have any effect on the temporal coherence boundary of the observers, as the results were very similar for the case of both "predictable" alternating patterns, and the random patterns used. The results for the latter sequences showed the expected dependence of fission on the magnitude of the frequency intervals used. The subjects were seen to set longer times for the tone repetition time (T), to bring about the coherence of sequences with larger frequency intervals, than in sequences with small intervals. Sequences containing intervals of three semitones or less, were found to remain coherent to the greatest degree, for increases in the tempo, corresponding to the smallest values of T (the inter event interval), adjusted by the observers .

For two-tone sequences, van Noorden observed that "there is a limit to the rate of the jump in pitch at which temporal coherence can be heard". If the pitch changed too fast from tone to tone, the percept of coherence changed to the impression of simultaneity of the two tones. Small tone intervals and small tone repetition times were found to encourage the fusion of the two tone bursts into a single tone burst, with an ambiguous "in-between" pitch. The tones thus appeared to be joined to each other, gradually losing this connection and becoming more apart with an increase in the inter tone time interval.

This observation corroborates those made by Bregman and Dannenbring (1973), and Nabelek, Nabelek, and Hirsh (1970, 1973), that fusion between tones can be brought about with small values of frequency change and duration, and by insertion of frequency glides or transitions between tones. For pairs of tones, these conditions encourage a fused "pitch" percept, while in longer sequences, fusion between tones following these criteria is encouraged, the fused "stream" in this case, being segregated from other streams containing their own sets of proximal elements.

In the same report (discussed above), van Noorden also describes his experiments on the temporal coherence of three tone sequences. Analogous to the longer patterns used by Heise and Miller (1951), van Noorden studied two possible trajectories formed by the arrangement of three tones. In "linear" sequences, the tonal intervals succeeded each other in the same direction, while in "angular" sequences, the two melodic intervals differed in the direction of frequency change (up or down). The melodic contour thus changed direction at the middle tone, with the first and third tones being nearer to each other in frequency than to the middle tone. As expected, the angular sequences lost their temporal coherence more readily than the linear sequences. The latter were able to retain their coherence at faster tempi, while the former underwent fission.

2.8 Attentional Tracking of Sequence Events

On the basis of the results reported above, van Noorden puts forward the suggestion that the attention of the observer may be considered to be analogous to a filter, that can "tune in" to tones of different frequency. In order to hear temporal coherence, a listener could adapt such a mechanism to keep track of incoming tones, by either increasing the bandwidth of the filter, or by letting the tuning frequency of the filter follow the frequency of the tones.

The assumption of the wideband filter is however immediately rejected, because the temporal coherence boundary, although dependent on the frequency difference between tones, is also dependent on the time interval between them. The time factor is not accountable in accepting the theory of a variable bandwidth filter. A wide band would also not be as selective in accepting or rejecting tones of different frequency lying in a fairly narrow range, as the streaming mechanism appears to be.

If instead the hypothesis of a filter with variable tuning frequency is accepted as true, then a lot of the results obtained from experiments on temporal coherence can be accounted for. In such a case, the filter would change its tuning frequency in concurrence with the frequency of the input tone. If this shifting of frequency is assumed to take place with a limited velocity, then the dependence of streaming on the rate of presentation can also be explained. Longer inter-tone time intervals would allow the filter to change its tuning frequency over a larger extent, thus facilitating temporal coherence over wider frequency ranges, for slower sequences.

Bregman and Dannenbring (1973) also seem to have felt this to be a plausible explanation of the nature of the coding mechanism responsible for sequential processing. In describing the properties of such a mechanism, they claim it to be a "predictive tracking device", that "modifies its criteria for inclusion of an input into a stream, as a function of very recent properties of the stream". This "predictive" tracking enables the incorporation of new input into the stream more easily if it is preceded by some sort of introductory evidence such as a transition in frequency, or some other "pointer" in the direction of the change.

The conclusions drawn by these investigators, based on the results of their experimental endeavours, provide reason to believe that the auditory system may indeed routinely use Gestalt principles such as the laws of "good continuation" and "common fate", in grouping events into coherent sequences.

In holding the streaming phenomenon as being responsible for the confusion in ordering different types of elements in sequences, it was implied by Bregman èt al. (1971,1973), that judgements of order tended to be more accurate for "within stream" sequences, than for those that trespassed the stream boundaries. It was suggested that this inability to relate elements across streams is a consequence of the time dependent shifting of attention from stream to stream. To explore further the assumption that "a central switching process is operative during the sequential tracking of tonal stimuli", and to test the hypothesis that tones widely separated in frequency were processed in separate communication channels with little interaction between them, Fitzgibbons, Pollatsek, and Thomas (1974), conducted experiments to compare the ability of listeners to detect temporal gaps within and between the perceptual tonal groups, formed from a sequence comprising two high and two low tones with a brief silent gap inserted in either of three positions within the sequence.

The results obtained by Fitzgibbons et al. displayed near perfect resolution within both high (H) and low (L) tone groups. Gap detection between the high and low frequency groups was significantly worse. In this case, the temporal gap occurred in conjunction with the jump in frequency, the net effect of which was to make discrimination of HHLL from HH(gap)LL a very difficult task. It thus appeared that the rapid jump in frequency between the "groups" of the standard HHLL sequence, gave the impression of temporal discontinuity between tones, even though the sequence was physically continuous. This perceived temporal discontinuity made discrimination difficult from patterns that did indeed have a genuine discontinuity (gap) in the same position demarcating different frequency regions, resulting in the poorer scores for across stream gap detection.

This experiment reporting the loss of temporal perception acuity due to the frequency structure of the sequence, is almost an exact replication of the work of van Noorden (1971 b), on the discrimination of time intervals bounded by tones of different frequencies.

Using a six-tone sequence with a frequency change occurring between the third and fourth tones (from Fa to Fb), van Noorden provided his subjects control over the magnitude of the time interval between these tones. Using a method of adjustment based on discrimination of time intervals, his subjects were required to assign a value to this interval, that would perceptually seem to be equal to the time intervals occurring between the other tones.

The results showed a high degree of temporal acuity when all the tones were of the same frequency (i.e. Fa/Fb = 1). A pronounced loss of acuity was observed for increases in the frequency difference (Fa - Fb) between the tones A, and the tones B, bounding the interval to be adjusted. This trend extended over a large frequency range,with subjects typically miscalculating the magnitude of the gap separating tones of different frequency. The frequency and time structuring of auditory patterns thus seem to be highly interdependent, with changes occurring in one domain typically causing physical or perceptual changes to occur in the other domain as well. The patterning of time is seen to influence judgements of frequency (Jones,1983, Watson et al. 1975, 1976), and conversely, the tonal patterning of a sequence is found to effect temporal acuity.

To account for the loss of temporal acuity, Fitzgibbons et al. propose that groupings in the frequency domain can be thought of as perceptual "channels", the shifting of attention between which takes some finite amount of time. This processing time delay, if greater than the actual physical "delay" (gap), would result in the gap going by undetected, while the switching of channels is taking place.

The difficulty encountered in integrating patterns across streams, and the inability to attend to many streams simultaneously, lends support to this theory of attention shifting. It may well be that such a mechanism is responsible for related perceptual phenomena such as stream segregation, temporal order identification, and time interval estimation.

An example manifesting interdependence of these phenomena is to be found in the "scale illusion" experiment carried out by Diana Deutsch (1975), in which she presented a musical major scale simultaneously in ascending and descending form, with the successive tones of each version alternating from ear to ear. The majority of listeners perceived this stimulus as comprising high tones following a V-shaped contour in one ear, and the low tones segregating into an inverted V-shaped contour emanating from the other earphone.

This complicated stimulus thus seems to have presented a variety of options to the perceptual system. By the law of good continuation, the scale contours should have been perceived veridically, but the conflicting switching of tones from ear to ear made these contours discontinuous. Faced with the odds, the system seems to have decided on grouping by ear channels. However, the frequency of the notes keeps changing at each ear. The confusion seems to have been resolved by deciding to group by ear, but using frequency proximity as the organizing dimension, thus resulting in stream segregation within channels.

Not being able to track these changes in time, it appears that the attentional mechanism sacrifices the more taxing veridical perception, and settles for a relatively steady illusory percept, resulting in confusion of order and contour.

The processing of information via real physical channels such as ears or spatially separated loudspeakers, or via perceptual channels formed on the basis of some similarity criterion as described above, has long been known of (Broadbent, 1958) and infact has been exploited to decipher complicated stimuli such as mixtures of voices, segregating them on the basis of filtering, spatial separation or another separating variable.

The lack of the ability to relate material processed via different channels was also described by Broadbent, on the basis of an interesting experiment requiring subjects to recall a series of digits, some of which were presented aurally, while some were presented visually. Broadbent reported that the digits could only be recalled in a channel by channel fashion, with the visual and auditory series being reported separately. The stimulus was thus segregated on the basis of qualitative similarity.

Indeed, the difference in "quality" as a means of segregating stimuli is stressed by all researchers involved in the study of auditory perception, and particularly those involved in the investigation of the dynamic aspects of pattern perception, yet there has been a paucity of actual research in this area. The reason for this lack of enterprise is not surprising, considering the lack of a precise definition of what it is that contributes to the perceived "quality" of a sound.

The euphemistic term "timbre", officially defined to be the auditory dimension responsible for all aspects of sound that cannot be described as "pitch" or "loudness", continues to function as an obscure, general purpose label, used often, studied seldom.

2.9 "Timbre" as an Attribute Differentiating Sequence Events

The definition of "timbre" as approved by the American Standards Association (1960) (quoted from Plomp, 1969) states it to be : "that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar". This negatively phrased definition, while indicating what timbre is "not", is qualified by an appended note that acknowledges 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".

This extended definition of "timbre" as a multidimensional attribute of a sound simultaneously makes two paradoxical claims: First, the official statement implies that timbre is a feature perceived in isolation from the pitch and loudness of a sound. The second statement however claims that timbre is infact dependent on the frequency spectrum and the sound pressure, and their variation over time; the same variables upon which the pitch and loudness are dependent. It thus appears that these oft quoted "auditory dimensions" are infact related, and may well interact with each other, in trading off their dependence on frequency, amplitude, and time.

The physical correlates of timbre that contribute to its perceived constancy over a wide variety of circumstances still remain to be identified experimentally. The fact that "a saxophone is readily identified as such, regardless of the pitch or dynamic it is playing", or "whether it is heard over a distortion ridden pocket-sized transistor radio or directly in a concert hall", poses the challenge of describing this invariance in terms of the physical structure of sound (Risset and Wessel, in Deutsch, 1982).

The observation that instruments retain their separate timbral identities even when typically played in reverberant rooms, separately or in an orchestral context, further addresses the issue of what exactly imparts the sturdiness to an instrumental timbre. In the "real world" environment of the cluttered living room, or a large concert hall, the phase relations of the tones and their partials are smeared and the intensities drastically changed due to reinforcement or attenuation due to reflections and absorption encountered in transmission.

Due to the ephemeral nature of the dynamic aspects of timbre, such as the attack-decay characteristics of the waveform envelope, and amplitude variation over time, it has long been considered that these transitory features could not be providing very resilient cues for the perception of timbre, and therefore the relative amount of the various harmonics making up the Fourier spectra of different instrument tones, must be the main criterion responsible for timbral differences.

Descriptions of instrumental timbres based solely on this specification of an average spectrum for an instrument playing a note of a particular pitch, have repeatedly been provided by Physicists investigating averaged spectra of different instruments (Culver, 1956, Olson,1967, Askill,1979), yet these investigators fail to mention that even the "averaged" spectrum of an instrumental tone differs as a function of the place and position where it is measured,and the time span over which it is measured. Benade (1983) points out that instrumental spectra are extremely variable indeed, yet the timbre associated with the instrument appears to be perceived as invariant. On the basis of room averaged spectra obtained for different instruments playing different notes in different positions, Benade (1981) attempts to explain the apparent paradox between the variability of the stimulus and the stability of the response as being dependent on the properties of "good" instruments, and the capabilities of the auditory processor.

Benade claims that sounds containing "not too few, and not too many" harmonics, with strong lower partials and rapidly falling uppers, are generally perceived as a single, compact, perceptual "entity", with the partials fused and not heard as separate components. He observes that diverse "well carrying" instruments (such as the oboe, violin, trumpet, piano, saxophone and bassoon) share the common feature of having a rolloff in amplitude of the harmonics of about 18 dB/Octave, beginning with the fifth or sixth harmonic. Such a spectral rolloff of the higher harmonics serves to reduce the inter-component masking in the upper critical bands, while also reducing "roughness" and other perceptually confusing effects, thus facilitating the tracking of pitch and tone color. A spectral rolloff of $1/f^3$ seems to be characteristic of these "good" instruments (where f = number of harmonics). Instruments that have less than or more than six harmonics, and exhibit a steeper or more gradual rolloff than $1/f^3$, are observed as being damped quickly, and showing a weakened trackability of pitch and timbre. Instruments such as the flute, cornet and vibraphone fall into this category, and are seen to have less "carrying" power, and tend to "get lost" in ensembles.

The complex physical behavior of musical instruments and the ensuing spectrum of the sound produced thus seem to be highly correlated with the timbre perceived. The more similar the sound spectra of two complex tones, the more similar they are in timbre (Plomp, 1976). Using multidimensional techniques for studying the relationship between the spectral space and perceptual space for vowels, Plomp arrived at a similar conclusion for vowels, showing that the more similar the sound spectra of two vowels, the more frequently they are confused.

Plomp (1969) asserts that there is a lot to be understood about timbre perception, from studies investigating speech perception. Indeed, the sounds of speech are themselves aggregates of phonemes that differ in "quality" (fricatives, plosives, vowels etc.) based on differences in the mode of production. In particular, the production and perception of vowels seems to have great implications for the perception of timbre. The fact that vowels are perceived as invariant despite being articulated with different fundamental frequencies, seems to have a direct bearing on the observation that by and large, for moderate ranges of frequency, musical instruments preserve their timbral identities while executing notes of different "pitch". This preservation of timbre over changes of frequency, presents the dilemma of attributing the invariance to a spectral envelope that is fixed in terms of its position on the (absolute) frequency scale, or an envelope that retains its "shape" to provide the same amplitude relationships between harmonics of tones of different fundamental frequency produced by the same instrument.

The spectral envelope of a sound is typically characterized by pronounced resonance peaks, showing enhanced amplitudes at different frequencies. These resonance regions were termed "formants" by Hermann (1890) (described in Winckel, 1967), and have since been used as referring to the resonances of the vocal tract, that weigh different frequency components by different amounts, yielding the sounds characterizing vowels. This specification of the spectral peaks as a property of a sound "source" (typified by the existence of a "resonator"), rather than the property of a particular "sound" was not always differentiated, and led Hermann and Helmholtz to debate over the issue of whether "formants are built up harmonically to the fundamental", or whether they are "unmovable and absolute frequencies which can also be inharmonic to the fundamental". In discussing this issue, Winckel (1967) proposes the compromise that both these concepts can be accomodated when a resonance "region" rather than a sharp, discrete spectral peak is considered to be the "formant".

Subsequent experiments however, in particular one by Slawson (1968) on the dependence of vowel quality and musical timbre on spectrum envelope and fundamental frequency, demonstrate that timbre depends mainly on the absolute position of the envelope of the amplitude pattern along the frequency scale, rather than on the absolute position relative to the fundamental frequency. The notion of a spectral envelope that preserves the relative amplitudes of different harmonics with respect to the fundamental, thus has to be abandoned in favor of a "fixed" spectrum model, that establishes the amplitudes of absolute frequencies, rather than their relative levels. Plomp (1976) goes on to claim that "a necessary condition for similarity of timbre for sounds with different fundamental frequencies is that they have similar spectral envelopes in terms of absolute frequency, though this similarity will necessarily be less for greater differences of fundamental frequency".

2.10 Timbre as a "Cue" in the Grouping of Sequence Events

Studied even less often than the concept of "timbre" itself, is the idea that the timbres of sounds juxtaposed in time may interact in a manner similar to the pitch interaction observed in the stream segregation phenomena, and the intensity interaction observed in phenomena encompassing the pulsation threshold studies, the continuity effect, the roll effect (van Noorden, 1976), and various types of masking.

Since timbre is typically dismissed as being a multidimensional attribute that depends on the frequency, amplitude, and temporal relations between the components of complex sounds, it is only natural to assume that the type of phenomena manifested due to interaction of these "primary" dimensions, should be exhibited on manipulation of the higher order timbral features as well. The inherent lack of independence in manipulating timbre by variation of the primary dimensions seems to be the major hurdle in timbre-based sequence research. The few studies that have dared to venture into the problem filled world of timbral investigation in a sequential context, are described below.

McAdams and Bregman (1979), describe an experiment performed by McAdams (1977),in which he studied the effect of timbre on stream segregation, using a "simplistic steady-state timbre derived from summed sinusoids."

Given a sequence of four sinusoidal tones that were close enough in frequency for them to be perceived in a single coherent stream, McAdams and Bregman claim that "it is possible to cause a subset of these tones to segregate into its own stream by inducing a timbre difference." This difference was provided by McAdams by the enrichment of some of the tones by the addition of a higher frequency sinusoid (the third harmonic). The enriched tones were found to perceptually segregate from the pure tones as is shown in figure 2.8, forming a sub-sequence differing from the sub-sequence of pure tones left behind, presumably on the basis of the timbral difference between the two as perceived by a listener.

McAdams however found the "anomalous" result that the conviction with which the tones segregated, varied, depending

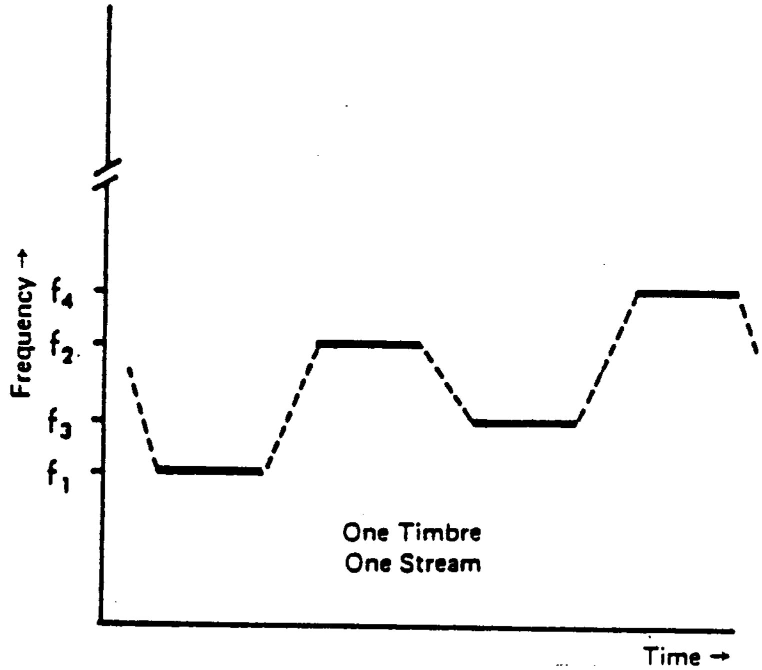
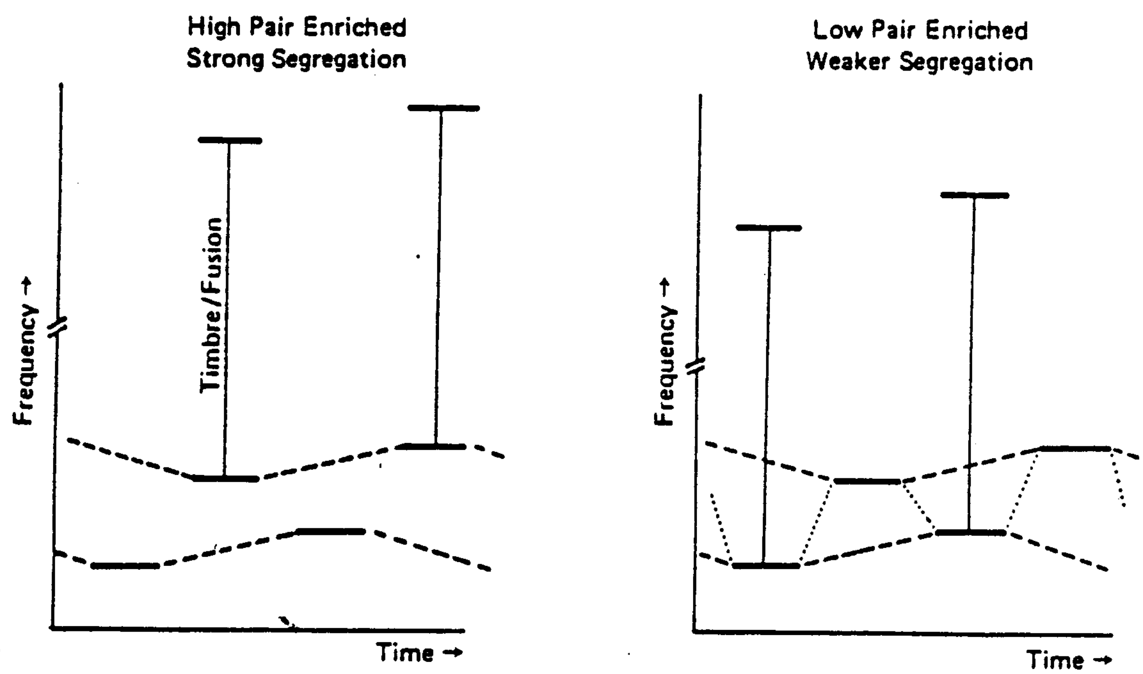


Figure 2.8 Stimulus used by McAdams (1977) (from McAdams & Bregman, 1979) showing the effect of "timbre" on stream segregation. In the top picture, the four iso-timbral tones of the sequence are seen to form a single coherent stream. The lower picture shows the segregation of the sequence into streams, due to harmonic enrichment of two tones of the four tone sequence.



on whether the high tones or the lower tones of the sequence had been enriched. He contends that this happened due to the interaction between pitch and timbre of the tones. In particular, he feels that the perceived pitch "height" of the tones in conjunction with the timbral attribute "brightness", serves to encourage or inhibit stream segregation.

While pitch "height" seems to refer to the relationship between physical frequency and perceived pitch (corresponding to the translational component of the helical structure proposed as a model for pitch, by Shepard (1965) (described in detail in Deutsch 1982, chapter 11), the attribute "brightness" seems to be a function of the location of the center of the energy distribution on the frequency continuum (Lichte (1941), discussed in Plomp (1976, page 104). The timbral attribute "sharpness" extensively studied by von Bismarck (1974) also seems to refer to the same quality, being dependent on "(1) the position of energized spectral regions, and (2) the magnitudes of energy in those regions." The spectral "center of gravity" mentioned by McAdams as being related to the "brightness", refers to the average of this relative amplitude weighting of the spectral components of a sound.

McAdams uses these concepts in speculating that the increased segregation obtained on enrichment of the high pair of tones in the sequence studied by him was probably a result of the inadvertent increase in pitch height, brought about by the addition of the high components to the high tones, pulling them even higher in the pitch domain, away from the lower pair. The enrichment of the lower tones by the addition of the higher harmonics on the other hand, increased their pitch height, thereby bringing them closer in pitch to the higher tones. The proximity in pitch thus achieved seems to have attenuated the effect of segregation based on the timbral difference. The segregation was however still perceptible in contrast to the isotimbral case in which either all the tones were pure, or all were enriched.

The experiment by McAdams reported above, raises the curious question as to why the tone pairs fused with the high harmonic at all, given neighbouring tones that were much nearer in frequency, and so should have resulted in streaming based on frequency proximity. The high components added should similarly have formed their own stream, since they were closer in frequency to each other, than to the four tones of the original sequence. The fact that this did not happen, and the far frequency simultaneous sinusoids fused together instead, leads one to conclude that there must be something about the fact that the fused tones occurred "simultaneously", that segregates them from the rest of the pattern.

This dilemma between "fusion" and "fission" can be explained to some extent on the basis of a study on the building of "timbre" by Bregman and Pinker (1978).

Viewing "stream segregation" to be a parsing mechanism that serves to simplify a complex auditory input typically characterized by a waveform containing many simultaneous frequency components at the same time, into streams that mentally represent the activity of a single source over time, these authors address the important question of why it is that we fuse the sounds of instrumental tones into a single coherent percept, while inhibiting the formation of these linkages between different instrument tones.

Answering such a question entails that the auditory system be able to analyze the messy input waveform to determine which series of frequency components arose over time from the same source and should be integrated into a sequential stream, perceptually distinct from other such streams. This would be necessary in the listening of music, where the "melody" represents such a sequential stream and requires the listener to keep "track" of what a particular instrument is doing over time. In order to follow the time course of such an instrumental timbre, the auditory system would first have to recognize the timbre associated with the sound source. It would therefore have to determine which set of simultaneous components arose from one source and should be fused into the timbral structure of one instrument, and which of the frequency components concurrently present, belong to different instruments.

In investigating this competition between the "sequential", frequency organization and the "simultaneous", timbral organization, Bregman and Pinker asked subjects to judge the stream organization and the timbre of a repeating stimulus comprising a pure tone A, alternating with a two tone "complex" composed of a pair of "more or less" synchronous pure tones B and C (shown in figure 2.9). The frequency of tone A was varied from being quite close to that of B, to being considerably higher. The relative synchrony of onset and offset of B and C was also varied, and these two variables were put in competition with each other, to determine if tone B joined tone A by virtue of frequency proximity, forming a sequential stream, or whether the synchrony in onset between B and C caused them instead to be fused into a compound tone (BC) having a "richer" timbre. To further provide an impetus to the timbral fusion of B and C, the harmonic relationship between them was varied by varying the frequency of C, to provide a "consonant" relationship that ought to encourage spectral fusion.

On the basis of measures such as "perceived richness", and "roughness versus complexity", used as response ratings by their subjects, to describe the perceived grouping between adjacent and overlapping tones, Bregman and Pinker conclude

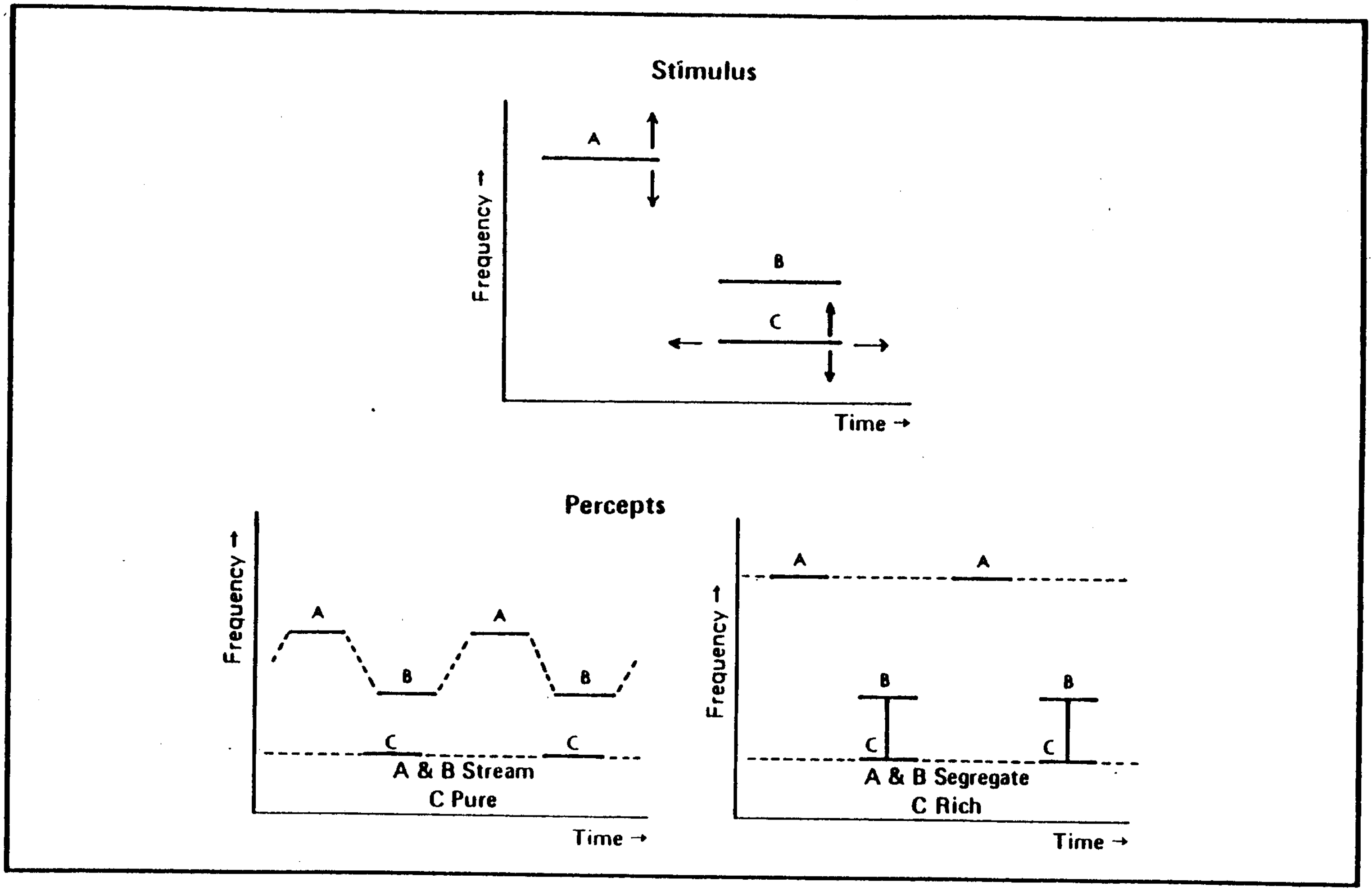


Figure 2.9 Stimulus used by Bregman & Pinker (1978) in studying the competition between the sequential streaming of tones A and B, and the fusion of tones B and C, as a function of the frequency separation between A and B, and relative synchrony of B and C. (From McAdams & Bregman, 1979).

that sequential and simultaneous grouping effects can compete, with frequency separation being a strong determinant of the sequential effect, and onset/offset asynchrony being an important determinant of the simultaneous effect.

In further investigation of the phenomena in which there may be a dilemma in choosing between a sequential or a simultaneous grouping, Dannenbring and Bregman (1978) extended the experiment of Bregman and Pinker to determine the streaming if any that would take place between a complex tone comprising three harmonically related sinusoids, alternating in rapid succession with one of its components. It was found that factors such as intensity relationship of the elements, the frequency of the alternating pure tone relative to those of the complex, and the temporal synchrony of the components of the complex tone, all affected the tendency of an element to be pulled out of the complex tone to form a horizontal stream with its unattached counterpart.

While the general conclusions seem valid and almost evident in the light of everyday experience, the response measures used by these investigators are at best obscure, and present a difficult task to the listener to make qualitative and subjective judgements such as degree of "richness". Also, it is not quite clear why the phenomenon described by these experimentors should be viewed to be based entirely on stream segregation phenomena, when the procedure that they use is infact the typical set up for the "probe tone" studies done routinely in Psychoacoustic research (Plomp, 1976, p3), in which a pure tone is alternated with the "test" stimulus, (typically a complex tone), and the subject asked to make judgements of frequency. The "Pulsation Threshold" studies also use probe tones in doing scans of neural activity extension due to complex tone excitation. The parameter under study is usually the intensity "level" of the probe tone needed to transcend the border between a pulsing effect due to the alternation with the complex tone, and the illusion that the probe tone infact continues through the occurrence of the stimulus tone (equivalent to the "continuity" effect described by van Noorden (1976)

The studies reviewed above reiterate the importance of "context in auditory perception. While the studies on streaming covered earlier in this review showed that properties of the individual auditory "events" themselves determined what would be "streamed" together in a sequential organization, the later studies indicate that the reverse may also be true. Thus, the sequential organizations formed between auditory events have been shown to affect the perceptual properties (such as timbre and pitch) of the events themselves.

The observed interaction between sounds and their components leading to the formation of linkages, and the interaction between the type of link established and other perceptual features of the sounds, thus led McAdams and Bregman (1979) to make the daring claim that " timbre is a perceived property of a stream organization rather than the direct result of a particular waveform, and is thus context dependent ". From the examples discussed above, it does indeed seem that the percept of an "event" might change with circumstance, given other events with which it may be compared. The "fused" attributes such as "pitch" and "timbre" of a complex sound, taken to represent its most basic features in isolation, may undergo a drastic "reorganization" to yield a changed percept, given the context of neighboring sounds.

The studies by Bregman et al. described above, offer some evidence for this adaptive mechanism, showing that pure tones whose synchronous and harmonic relationship is such that their "fusion" is favored under isolated conditions giving a timbral sensation such as "richness", may be perceived as separate, independent tones, if some other organization presents stronger evidence that they belong to separate sequential streams.

The question addressed to McAdams (1977) experiment about why the components of the enriched tones "stuck together", and did not stream sequentially with the alternate pure tones that were closer in frequency, now seems to be answered. The stimulus used by him basically deciphers into being a four note melody, played by two alternating "timbres", with alternate notes being "high" and "low". Here the timbral organization seems to be preserved intact , after tradeoffs between the synchronicity of the compound tone components, and the frequency proximity of the adjacent tones have taken place. Once this timbral identity is established, stream segregation takes place, favoring a sequential organization based on the timbral difference. The fact that McAdams found a pitch-timbre interaction that resulted in different degrees of segregation, seems to imply that the contribution of a tone in "building up timbre", is not that sacred, and that given adjacent tones that are nearer in frequency to a tone of the complex, than its own harmonics are to it, may lead the tone to show an affinity for the nearer frequency, non-simultaneous tone, than for its own far-frequency, synchronous harmonic.

A similar reasoning, based on intelligent observation, rather than rigorous experimentation, was offered by Helmholtz in his famous treatise (1885/1954) in explanation of the fact that we hear musical sounds such as tones, as perceptually identifiable entities, and not as a bunch of partials (spectral components), yet we separate similar sets of superposed sounds that arise from different sources, into sets of sounds coming from the same source. Some comments made by Helmholtz regarding this inconsistency in analytic behavior , are quoted below :

> "... now there are many circumstances which assist us first in separating the musical tones arising from different sources, and secondly, in keeping together the partial tones of each separate source. Thus when one musical tone is heard for some time before being joined by the second, and then the second continues after the first has ceased, the separation in ~~time~~ sound is facilitated by the succession of time. ..."

> ... "We have already heard the first musical tone by itself, and hence know immediately what we have to deduct from the combined effect for the effect of this first tone. Even when several parts proceed in the same rhythm in polyphonic music, the mode in which the tones of different instruments and voices commence, the nature of their increase in force, the certainty with which they are held, and the manner in which they die off, are generally slightly different for each." (p59).

> "... all these helps fail in the resolution of musical tones into their constituent partials. When a compound tone commences to sound, all its partial tones commence with the same comparative strength ; when it swells, all of them generally swell uniformly ; when it ceases, all cease simultaneously. Hence no opportunity is generally given for hearing them separately and independently." (p 60)

The value of receiving a progression of tone superpositions in providing the information that is critical for the discrimination of the pitch and timbre of complex tones is further emphasized by Roederer (1979) Underscoring the statements made by Helmholtz, Roederer asserts that the sequential context of a group of complex tones, provides the crucial ""primary cues" given by the coherent fluctuations in timing and frequency of each tone" and also provides ""higher order" secondary information extracted from the melodic lines" that show the progression of the event attributes varying over time.

Too great a variation between the successive events of these real-world sequences leads to phenomena such as stream segregation, with similar events spread over the sequence bunching up to form subsequences. The conventional music usage of a one-on-one mapping between melody and timbre, has led to the large scale conditioning of music listening to be restricted to iso-timbral melodic sequences that may however be followed or accompanied by other timbral sequences. The use of many changes of timbre within a sequence has been developed to some extent in the music of Webern, Schoenberg, Ligeti, and a few other "modern" composers. This exploration of musical space, often under the title "klangfarbenmelodie", has been quite successful, but by and large, the role of "timbre" as a viable dimension in the contrasting of notes within a melody remains unexploited.

A few studies in which the timbral attributes of sounds have been studied in a sequential context, or conversely, where the sequencing of sounds has been studied in terms of the timbral character of the sounds, have been done by Grey (1978), and Wessel (1979) respectively.

The study by Grey (1978) compared the timbre discrimination of isolated tones with discrimination in various musical contexts, both single voiced and multivoiced. Twelve different contexts were studied, four different patterns for each of the "voicings". The listeners were required to judge if the timbre changed or remained the same during a trial. Two different synthesized timbres were used for each of three instruments (the clarinet, trumpet, and bassoon), differing in the details of the synthesis scheme used in their construction.

The motivation behind undertaking such a study of timbral melodic patterns, as explained by Grey, highlights some important issues. Some of these are quoted below :

> "There are clearly many facets of timbre presented in typical musical contexts that do not exist for isolated tones. The listener is allowed to compare the different spectra of notes from the same instrument taken at different pitches, and the "composite spectral map" of the instrument may be a factor in normal timbre recognition and evaluation ... A variety of articulatory patterns are available that do not exist in isolated notes. The various and idiosyncratic ways that notes are joined together in musical phrasing may be important in the perception of instrumental timbre.
>
> Also, perception of a musical context may be different from the perception of isolated phenomena because of various temporal performance constraints. ... The ability to discriminate two timbres may be affected by the complexity of their context. Likewise, the set of criteria used to form a similarity judgement between two isolated tones may be quite different from those used to rate the similarity in musical contexts".

In keeping with expectations, Grey did indeed find that discrimination of the timbral versions of a single instrument, varied with the general context provided by the absence or presence of other temporally juxtaposed tones. Although the "scheme" differentiating the synthesis of different versions of the same timbre was the same for all instruments, the discrimination performance was seen to vary differently for the different instruments. While the clarinet and trumpet type tones were found to be best discriminated in an isolated context, the bassoon was found to be best discriminated in a sequential context, for both single, and multivoiced patterns.

On the basis of the synthesis scheme used by him to specify the physical structure of the sounds, and the responses made in discrimination, Grey speculates that since spectral factors were dominant in the perception of bassoon tones, and temporal factors in perception of the other two instruments, it may be that single voiced musical patterns enhance the spectral differences that exist between versions of a timbre, while isolated contexts allow the listeners to compare the temporal details of the tones more clearly.

The study by Wessel (1979) goes a step further in manipulating the timbres of the tones within a sequence. Exploiting the fact that spectral differences between versions of a timbre can cause them to be discriminated, Wessel constructed a three tone repeating sequence, over which the spectral energy distribution changed alternately from note to note. Such manipulation resulted in the creation of a sequence in which the notes changed timbre one after another, as depicted in figure 2.10.

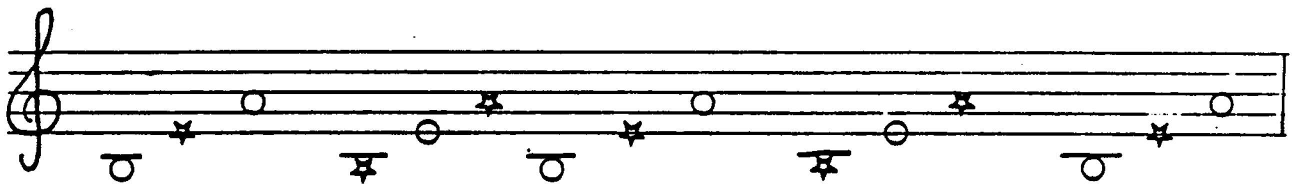


Figure 2.10 Melodic sequence used by Wessel (1979), in which alternate notes of a three note melodic pattern are played with different timbres. Timbral segregation of the sequence ensues as a function of the spectral differences between the tones.

In studying the perception of this timbre alternating three note repeated pattern, Wessel noted that "the nature of the relationships among the elements of the pattern is of primary importance in their perception. Thus, if the timbral distance between the adjacent tones of the pattern is small, the repeating ascending pitch line is seen to dominate attention. However, when the timbral difference is enlarged along the "spectral energy distribution" axis, the perceptual organization of the pattern is radically altered. The melodic line is seen to split at the wide timbral intervals, and two "interwoven descending lines" begin to emerge from the pattern. Wessel attributes this intriguing observation, to a segregation phenomenon based on the large spectral energy distribution between the alternating timbres. similar to the streaming phenomena studied by Bregman et al.

The fascinating description of the timbre based segregation observed by Wessel is however not supplemented by any information on the type of measure the listeners response was based on, the task set before a listener, and specific details of "how" exactly the spectral energy of the different timbres was distributed.

The present experiment aims to provide some measure of this variable used to differentiate between timbres. Furthermore, on the basis of the timbre based segregation of the three note sequence and the four note sequence described by McAdams, the present study also aims to provide evidence for the segregation of sequences in which the "pitch" factor is accounted for using the concept of the "missing fundamental", as was used by van Noorden (1971/1975).

Van Noorden however used only two tones, in an ABAB...type alternating tone sequence, and found that the missing fundamentals did not stream together. He reports however (1971 a,page 11), that bringing the frequency components of the residue pitches close together did lead to some degree of fusion of the components of the two notes. This fusion, based on "frequency" rather than on "pitch", led van Noorden to speculate that "the fusion must originate before the final pitch extractors come into action".

To determine if this was so due to the lack of a common "spectral region", or simply due to the lack of a viable "choice" because of the lack of context of any other tones in the sequence, more tones have been included in the sequences under study. The results of some preliminary experimentation done with these aims in mind, are described in the next chapter.

3. KEY QUESTIONS

3.1 Pilot Experimentation

All the studies on timbre reviewed in the preceding chapter, and particularly the few dealing with temporally juxtaposed sounds of different timbre, indicate that the perceptual attributes of sounds often undergo a dramatic alteration when presented in the dynamic context of a repeating sequence. Sequencing permits comparisons between sounds to be made, and therefore allows "common factors" between sounds to be discovered. The length of time afforded by the process of sequencing, and further augmented by the process of repetition of a sequence allows the similarities and differences between sounds to be used as "criteria" in forming categories and other efficient organizations. Changes along the dimensions being used as such "cues" thus lead to changes in the perception of these organizations.

In order to observe and verify the effects of rate, timbral assignment, and pitch layout of a sequence of complex tones in bringing about the streaming phenomenon, a pilot experiment was designed, in which these parameters were varied, and ensuing changes in perception noted.

Thus, in the "timbre-melody" experiment, two different instrumental "timbres" were constructed, using the synthesis scheme depicted in figure 3.1. The relative amplitudes of the harmonics chosen to be included in the spectra of the two timbres were decided upon on the basis of prior experience with synthesis of pseudo-real timbres. Thus, to obtain a reed-like timbre (Tr), the spectrum defined for the violin-like timbre (Tv) was simply "stretched", to give the same relative amplitude weightings to the six harmonics chosen. However, the harmonics chosen were not the same as those chosen for the "violin". Rather, alternate harmonics were dropped out, so that only the odd harmonics were present in the spectrum of the reed, giving a spectral pattern resembling that typically observed for reed sounds (Benade, 1976).

With the aim of studying "real world sequences of real world sounds", a melody of thirteen notes was composed within the bounds of a fairly small pitch range (less than an octave, from B3 = 247 Hz., to G4 = 392 Hz). This narrow range melody was first played within the context of a single timbre, at rates varying from a note and a half a second (corresponding to an inter-event-interval of 675 msec, to a rate of about thirteen and a half notes a second, (corresponding to an IEI of 75 msec). An in-between rate of about four notes a second (IEI of 225 msec), was also used.

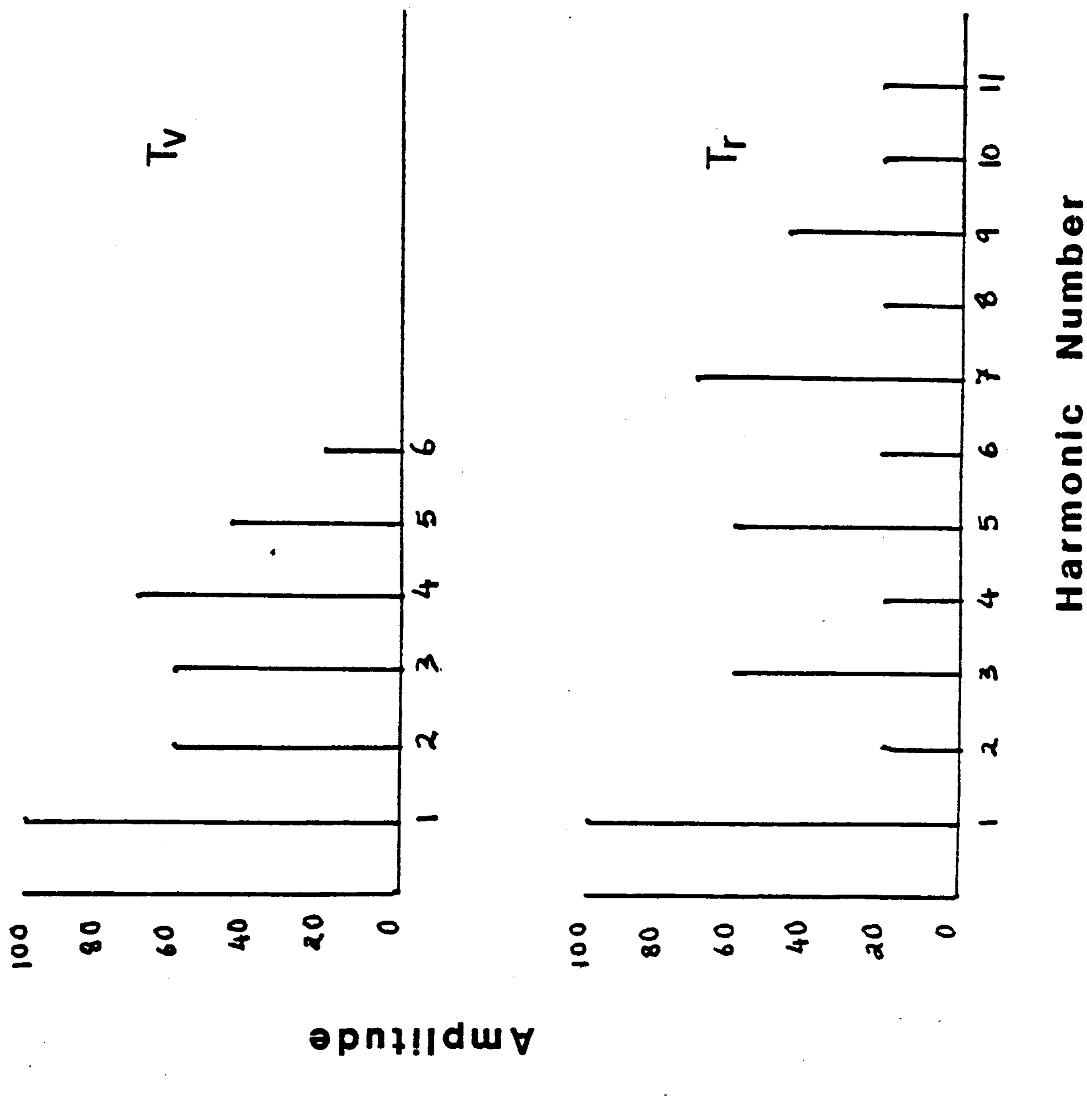


Figure 3.1 The spectral specification of timbres (Tv and Tr) synthesized for the "timbre-melody" experiment. The "relative" amplitudes of the different harmonics used are plotted against the corresponding harmonic numbers.

The melody was found to be perceived "intact" when played at both the slow rates by either of the timbres. At the fastest rate however, the percept was confusing and some degree of segregation between the comparatively larger pitch intervals of the melody began to be observed. Since the melody used in the experiment was fairly "tight", in that it was confined to a fairly narrow pitch range, the segregation was not very distinct in this iso-timbral case.

The same melody was then played at the different rates again, but with the timbres alternating between "violin" and "reed", from note to note, as shown in figure 3.2. This time, the melody was found to segregate into distinct subgroups at the quick rate. Some degree of segregation was perceived at the in-between rate of an IEI of 225 msec , was also observed. These observations were informally confirmed by three independent listeners. When it came to the task of reporting "what" was heard however, the responses were vague and non-committal. It appeared that even though two separate timbral sub-melodies could be heard, something "more" seemed to be going on in the background, in the form of a phrase of low bass notes that seemed to occupy a "stream" of their own.

This strange interaction of the timbres was further investigated by exchanging the assignment of the timbres to the notes of the melody, to see if the streamed percept remained the same. This was not found to be the case.

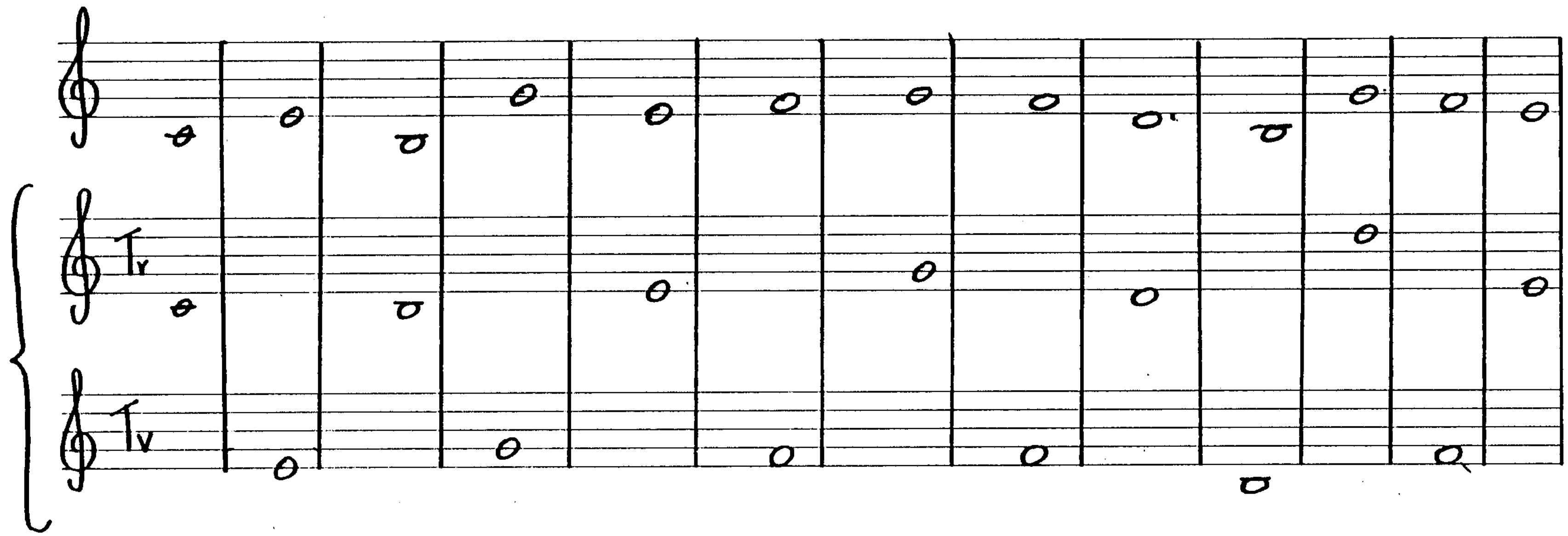


Figure 3.2 Timbral assignment of the melody used in the "timbre-melody" experiment. Different notes of the melody are articulated by different timbres.

The timbres were then assigned to the notes of the melody in a random, arbitrary fashion, instead of in the alternate assignment. The streamed percept was felt to change in keeping with the changed context of temporal proximity, but the unaccounted-for low tones still seemed to lurk in the background. It thus appeared that the structure of a "timbre", defined in terms of its harmonic spectrum, was interacting with that defined for the other "timbre", both of which seemed to depend upon the "pitch" that they were assigned in the melody.

These interesting, but "confounded" observations, led to an invesigation of the timbral structures as they had been defined. It was seen that the first three components (the first, third, and fifth harmonics) of the reed timbre were of the same frequency and amplitude as the first, third, and fifth harmonics of the violin tone. The reed tones however also had higher harmonics, to compensate for their lack of lower harmonics. The partially overlapping harmonic structure thus seems to have been responsible for the strange segregation effects observed.

The ~~"Timbre-Split"~~ experiment undertook the task of observing the interaction between complex tones that had components with the same frequency and amplitude in common in order to be able to explain the interaction of the timbral structures described above.

This was done by making a six component complex tone having fundamental frequency 500 Hz, and a three component complex tone having the fundamental frequency 1000 Hz. Since 1000 Hz tone is the octave of 500 Hz, the harmonics of the former are inherently present amongst those of the latter. These "corresponding" harmonics of the two tones were made to be of equal amplitude. The two tones were then played one after another, to see if the percept resembled any of the strange observations of the previous experiment. The time intervals between the tones were varied as before. It was observed that for the slower intervals (675 msec, and 225 msec), the percept was simply that of a tone being followed by another tone an octave above. As the time interval was further reduced to 75 msec however, it was found that this "trivial" percept suddenly underwent a change, and three tones were heard, a deep low tone, seemingly overlapped by two high tones an octave above. The low tone sounded quite "hollow", and infact very similar to the low "ghost" tones heard in the tibral streaming experiment.

This experiment seemed to imply that the octave tone was getting "streamed" with its corresponding counterparts from the 500 Hz complex, resulting in the dropping out of the 500 Hz fundamental and its odd harmonics, that were perceived as the "low" tone. The stimulus used is shown in figure 3.3.

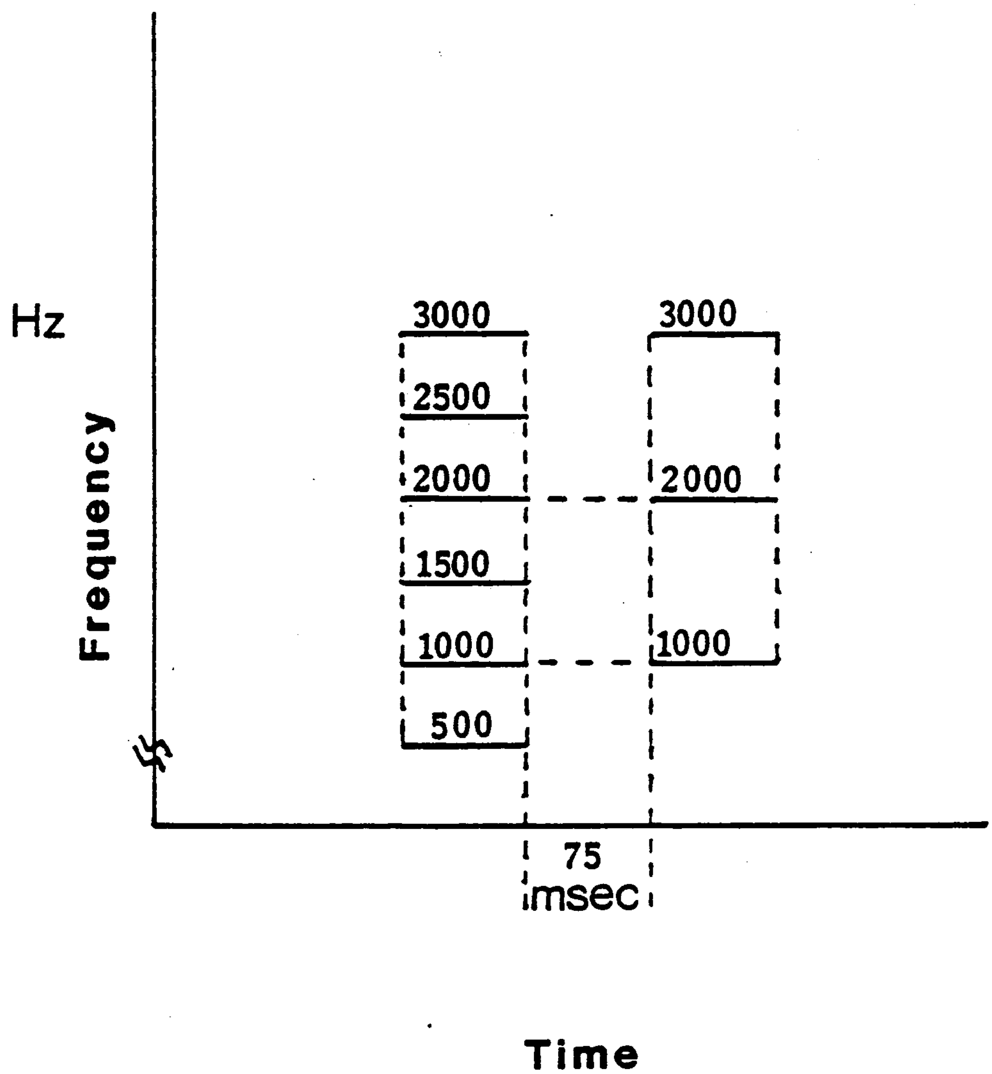


Figure 3.3 Sequence of two complex tones an octave apart, and separated by 75 msec used to elicit component streaming in the "timbre-split" experiment.

3.2 ~~Major Observations~~

These pilot exploratory experiments thus brought to light a number of interesting phenomena that deserve further attention. The main observations and consequent issues that need to be further investigated are listed below:

1). Timbre defined on the basis of a spectral envelope giving the relative amplitudes of harmonics relative to a fundamental gives a sound quality that seems to remain intact over the range of fundamental frequencies used (247 to 392 Hz). This preservation of timbre observed for sounds having different fundamentals but the same relative weighting of harmonics is at odds with the observations of Plomp (1976), and Slawson (1968), according to which an "absolute" spectrum seems to be responsible for timbral invariance.

2). All "timbres" are not equal in strength in functioning as a cue for the perceptual segregation of sound sequences. Some are more salient than others in their powers of persuasion in cohering sounds to form groups. This difference in "salience" is probably related to the spectra of the sounds, in particular the different strengths of different harmonics. Since the timbral differences in this experiment were provided on the basis of different relative amplitudes of the frequency components, this observation seems to be in keeping with Benade's observations on the "carrying power" of instruments (1981). McAdams and Bregman (1979) also reported a similar observation on the differing degree of segregation of two component complex tones. Their sounds however differed in the frequency of the component tones used, and may therefore be attributed to differences in "spectral gravity" of the sounds.

3). The streamed melodies that were obtained on segregation by timbre, often seemed to be accompanied by low ostinato-like tones in the background. Adding up the number of tones heard in the separate streams, often yielded a number greater than the number of tones that were actually present (13). This seems to suggest that some complex interaction must take place between the harmonics of adjacent tones, as they do not simply segregate into notes characterized by a timbral identity.

4). An interactive process trading off pitch and timbre against each other seems to be indicated, since interchanging the timbres that articulate different pitches results in a changed stream percept. If the complex tones were resolute entities with sturdy timbres, the streamed percept would have remained the same, except that the timbres would have been exchanged between the streams.

5). Streaming on the basis of absolute frequency at the component level seems to be implied, rather than on the basis of "pitch" at the "note" level. This is indicated by the "timbre-split" experiment, in which some components of a complex tone were perceived as a part of another tone that followed in rapid succession.

Other related observations revealed that "repetition" of the sequences being studied was often necessary, to elicit "streaming", in keeping with Bregman's (1978) report on the cumulative nature of stream segregation. Another factor that influences the response to such phenomena, is the "attention" of the listener, as claimed by van Noorden (1971). In experiments that assume segregation by default, and then ask the listener to judge between different aspects of the streams that they are "a priori" assumed to be perceiving, can be considered to be the case where the listener is operating above the "temporal coherence boundary" in the "fission region" described by van Noorden.

The tones used in the pilot experiments discussed here were all synthesized digitally, using the "Speech Microscope", an analysis/synthesis software package (Vemula, 1979). The experimentation was done on a Rap III type interactive computer system. All the sounds were presented monaurally, and eventually only to the right ear.

The design of the experiment that was finally done to investigate some of the major issues discussed here, is described in detail in the next chapter.

4. DEVELOPMENT AND DESIGN OF PRESENT EXPERIMENT

4.1 Rationale behind Choice of Sequence Used in Study.

The pilot experiments described in the preceding chapter had yielded the interesting but perplexing observation that "timbre" could indeed provide a strong cue in the segregation of melodic sequences played at rapid rates (between 4 and 14 tones/sec). However, the streamed percept was often "more" than what would be predicted by the timbral assignment of different tones of the melody. Instead of breaking up into "clean" sub-melodies played by the instrumental timbre that was specified for the notes, the melody seemed to segregate on the basis of a combined interaction between the specific "pitch" (note) being played, and the "timbre" with which it was orchestrated. A different orchestration, assigning a different timbre to the same pitch, resulted in a different streamed percept. The pitch and timbral attributes of the sequence tones thus seemed to be confounded with each other.

Since the timbres used in pilot experimentation, were defined on the basis of different relative amplitudes of the same set of six harmonics of the fundamental frequency of the "note" being synthesized, it was difficult to come to any conclusions about the possible factors causing the tones to segregate. Differences in amplitude, and the interaction of the frequencies of the adjacent harmonics, could both be providing the segregation "cue" needed to make separate groups of tones at the rapid rates of presentation employed.

Since the same set of harmonics was used to make sounds of different timbre but the same pitch (determined by the fundamental frequency), there was a complete spectral overlap in the frequency region where the harmonics of the different timbres resided. The "shape" of the spectral envelopes for the different timbres however varied, being determined by the different amplitudes of the different harmonics. The dependence of both "pitch" and "timbre" on the same frequency region thus made interpretation of the individual effects of these dimensions on group formation very difficult to ascertain.

As a means of separating the physical variables responsible for timbre and pitch differences, Hirsh (1984) suggested using widely spaced frequency regions for making tones. Such tones could thus differ in the spectral pitches of the harmonics, but could be made to have the same "virtual" pitch, which was fixed by using different harmonics of the same "missing" fundamental.

This perceptual phenomenon of being able to equate the "pitch" of a complex tone to its fundamental frequency, even if the sound is actually devoid of energy at the fundamental, has been accounted for by many modern theories of "pitch perception" (Wightman, 1973, Goldstein, 1973, Terhardt, 1974).

The Wightman model does a power-spectrum analysis of the input waveform, followed by a Fourier transformation to the frequency domain. A pitch extractor is then assumed by him to operate upon positions of maximal activity in the resultant pattern, and the frequency corresponding to the highest amplitude component is assigned to be the "pitch" of the sound.

Goldstein´s "optimum processor" model is assumed to operate on the frequencies of all the resolved harmonics, in a form akin to a harmonic "comb". Assuming periodic stimuli with adjacent harmonics, the model computes the fundamental frequency on the basis of the harmonic numbers present (all multiples of the fundamental frequency), and assigns the gap between the "teeth" of the best fitting comb to be the fundamental.

Terhardt´s model proposes that the "virtual" pitch of a tone complex can be deduced from the component spectral pitches, by evoking a "trace" of the current spectral pitches on a "matrix" of spectral pitch cues that is "learned" due to the repeated exposure to complex stimuli during the development period. The "traces" left by the groups of spectral pitches then get activated every time that the matrix is activated by a complex signal, and the strongest of the evoked spectral pitch cues is inferred as the "virtual" pitch. (The procedure is basically one of sub-harmonic matching. The sub-harmonic which is the highest common factor of the components of the complex, is taken to be the "virtual" pitch). An algorithm specifying the calculation procedure for this "virtual pitch" of complex tones has been discussed in detail by Terhardt (1979), keeping into account factors such as the influence of sound pressure level (SPL), and partial masking between the components.

To avoid the potential influence of such differences in amplitude, the spectra of the complex tones synthesized for the present experiment, were specified to be of "equal" amplitude.

The pitch attribute of these equal amplitude tones was accounted for by the value of the fundamental frequency required, four selected harmonics of which were used in the synthesis. The frequency "spacing" between the harmonics thus determined the "nominal virtual pitch" of the harmonic complex constructed.

Since the complex tones synthesized for the experiment were all harmonic multiples of a (missing) fundamental frequency, and the harmonics used were all adjacent to each other, and of equal amplitude, it seems to be quite safe to conclude as has been assumed above, in keeping with the Goldstein model, that the pitch is equated to the frequency of separation between the harmonics.

Tones differing in "timbre" or sound quality were then constructed, by merely shifting the spectrum defined, along the frequency axis. This was done by specifying four equal amplitude harmonics for each timbre. The harmonic "number" (m = fm/fo, where fm is the frequency or "spectral" pitch of the harmonic itself, and fo is the frequency of the fundamental), was varied from timbre to timbre, so that each "timbre" contained a different set of four harmonics. Thus, the timbre T1 thus defined, had the first, second, third, and fourth harmonics of the fundamental frequency fo.

Since the first harmonic (following the Physics convention) is the fundamental itself, this timbre contained the "fundamental" in its spectrum. Timbre T2 on the other hand, was built on the second harmonic, and contained the second, third, fourth, and fifth harmonics of the fundamental frequency fo, which was therefore "missing" in the spectrum. Similarly, the spectrum of timbre T3 contained the harmonics 3, 4, 5, and 6, of the (missing) fundamental, timbre T4 contained the harmonics 4, 5, 6, and 7, timbre T5 was built on the fifth harmonic, and contained the harmonics 5, 6, 7, and 8, of the (missing) fundamental, and the spectra of timbres T6 and T7 contained the harmonics 6 through 9, and 7 through 10 respectively.

The harmonic spectra for sounds having these "timbres" are illustrated in figure 4.1.

The spacing between these harmonics, was equal to the fundamental frequency equivalent to the "pitch" that was desired. A set of seven musical "pitches" spanning over a range of two octaves, was constructed, to give a large contrast between pitch "intervals" to be used in the sequences. The set Pn of the seven "pitches" (n = 1, 2, 3, 4, 5, 6,or 7), are displayed in figure 4.2, along with the corresponding fundamental frequency, and the name of the musical note with which they are associated.

Thus, a set of forty-nine ~~different~~ sounds was constructed, each "pitch" being articulated by each timbre, to give a total of 7x7 different (timbre, pitch) pairs. The set therefore comprised seven sets of seven tones that had the same (virtual) pitch, but differed in their "quality", defined by the location of the complex tone spectrum along the frequency axis. Tones of the same pitch, but differing in timbre, or conversely, tones of the same timbre, but differing in pitch, could thus be juxtaposed in a sequential context as desired, and the tradeoffs between these

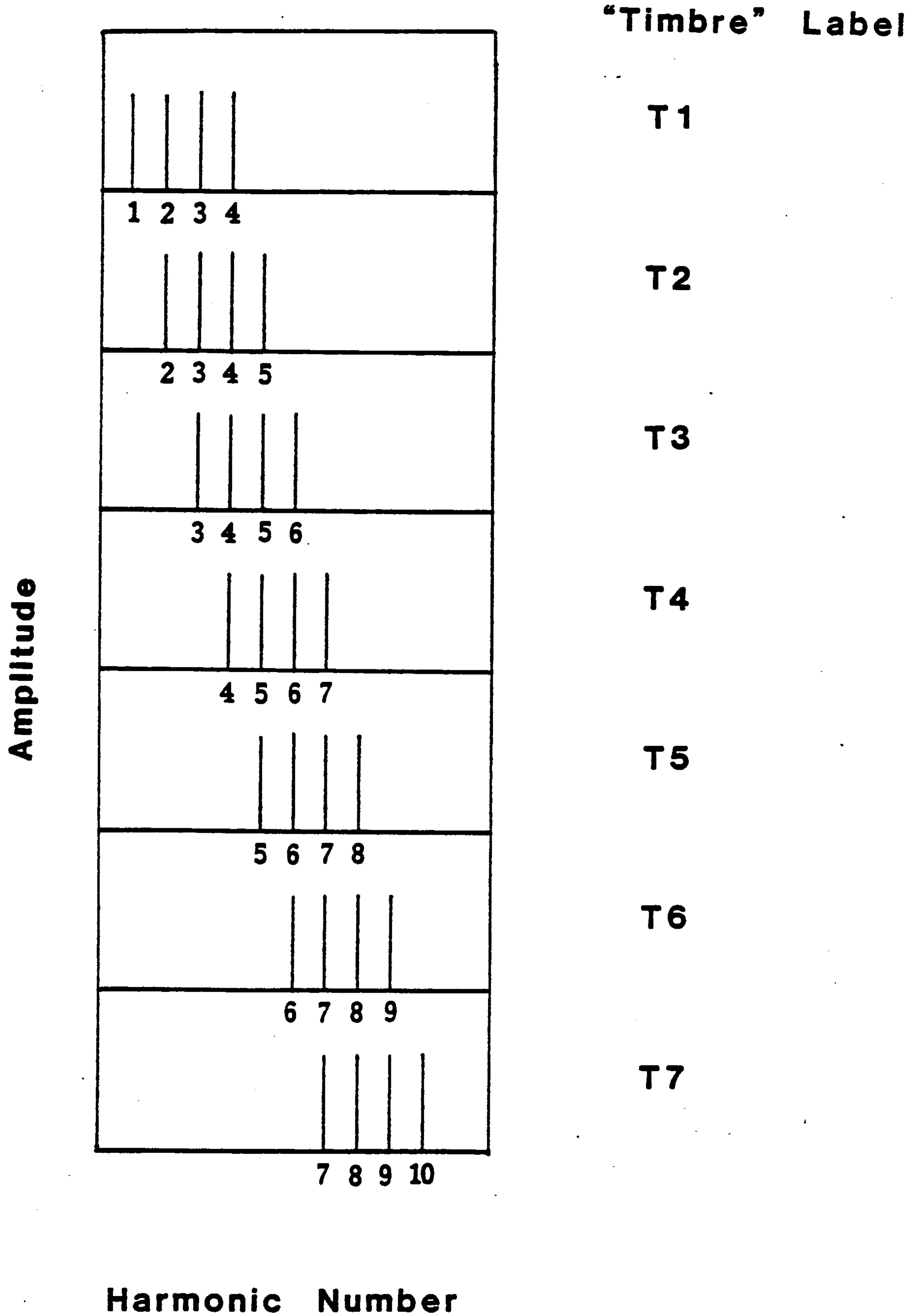


Figure 4.1 Spectra of the "timbres" created by shifting the four equal-amplitude harmonics along the frequency axis.

dimensions studied.

Given a set of forty-nine different sounds, a myriad sequences may be constructed. By deciding on sequences of only four tones, this unfathomably large number gets reduced, but still remains large, enabling a total of 5764801 (49x49x49x49) different sequences to be made. Choosing a set of sequences for close investigation thus becomes a difficult task, since such large numbers of stimuli are quite unmanageable to deal with in an experimental situation. This oft-encountered problem in research on auditory perception, was here dealt with by fixing some parameters of the type of sequence to be studied, thereby reducing the number of "choices" available.

With this aim in mind, forty-nine different four tone sequences were constructed, following the fixed format :

T2P1 TmP1 T2Pn TmPn

Viewing this "specially designed" sequence carefully, reveals that two tones having the same pitch P1, are followed by another pair of tones, that have the common pitch Pn. The timbres of the tones within each "pitch pair" are however different. Alternate tones on the other hand, are seen to have the same timbre. This sequence thus contrasts the iso-timbral tones in terms of their pitches, while simultaneously contrasting equal-pitch tones in terms of their timbres. While the temporal contiguity and the equality in pitch would seem to favor a "pitch-based" grouping, the similarity of the alternate timbres would compete in trying to offer a "cue" that would cause the pitch groups to break up, and timbral groups to be formed.

In four tone sequences of this type, the first tone was fixed to be the lowest pitch (corresponding to P1= middle C = 262 Hz). The timbre of the first tone was also fixed to be the timbre T2, comprising the harmonics of the missing fundamental corresponding to P1. The second tone of the sequence had the same "pitch" as the fixed first tone, but the timbre could be changed to give different sequences. The third tone had the same "timbre" as the first tone (i.e. T2), but the pitch could be changed to give different sequences. Having decided on these two variable tones, the last tone of the sequence was uniquely defined to have the timbre of the second tone, and the pitch of the third tone.

Thus, given the "set" of timbres:

$$Tm : m = (1, 2, 3, 4, 5, 6, 7)$$

corresponding to the timbres T1, T2, T3, T4, T5, T6, and T7, containing four harmonics built on the number used to "label" the timbre, and given the "set" of pitches:

Pn : n = (1, 2, 3, 4, 5, 6, 7)

corresponding to the pitches P1, P2, P3, P4, P5, P6, and P7, spanning a two octave range from C4 to C6 (illustrated in figure 4.2), the specification of the values m and n, each being allowed seven values, would yield a total of forty-nine different sequences of this type. These forty-nine sequences could then be presented in different blocks, each sequence being repeatedly presented within a trial to elicit "streaming", and the type of grouping perceived by different listeners could be investigated.

4.2 Synthesis of the Complex Tones Used

All the tones used in the main experiment were digitally synthesized on the Eclipse S/200 computer by Data General. An "additive" synthesis program called "MIX 1" (Morgan, 1983), was used to generate sinusoidal components of a desired frequency, that could be assigned different amplitudes and phase angles. These "pure" tones were then summed to give a complex tone, representative of the set of four harmonics used. The duration of the synthesized tone, and a "cosine" shaped amplitude "envelope" (over time), could also be specified as desired. This was done by specifying an overall duration for the tone, and then specifying the time

PITCH LABEL	Fo (Hz)	NOTE	INTERVAL from P1
P1	262	C4	Unison
P2	330	E4	Third
P3	392	G4	Fifth
P4	524	C5	Octave
P5	660	E5	Tenth
P6	784	G5	Twelfth
P7	1048	C6	Two Octaves

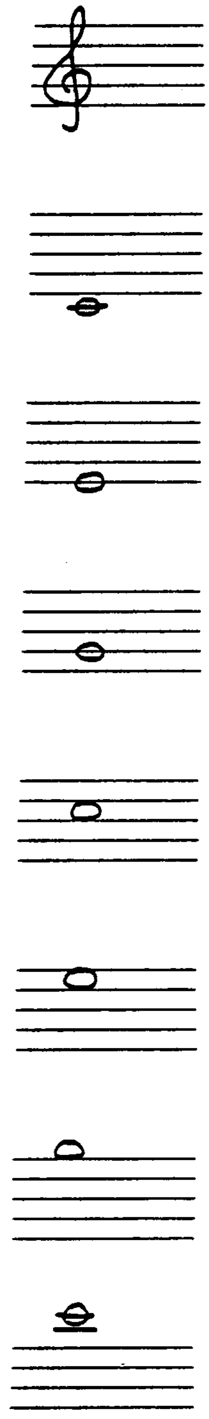


Figure 4.2 The set of "pitch" values (Pn) used in the experiment. The corresponding fundamental frequency, the musical name, and the interval with P1 as reference are also shown.

over which the attack (rise) and the decay (fall) of the tone were to take place, leaving a steady-state "sustain" portion in the middle. A total of eleven harmonics per complex tone could be individually specified, using this program. The frequency range of the sounds was however restricted to a limiting value of 10,000 Hz, beyond which distortion would result due to the "aliasing" phenomenon exhibited on exceeding half the value of the "sampling frequency" typifying such equipment (20 kHz. in this case).

The forty-nine different sounds described earlier, were synthesized using this program , by specifying each harmonic for each tone, in terms of its absolute frequency. As mentioned earlier, the amplitudes and phase angles of all the components were specified to be the same. Similarly, temporal features such as the duration and the rise and fall times, were made the same for all tones. Each tone was made to be 50 msec. long, with 10 msec. rise and fall times. To obtain complex tones of different "pitch", the distance between the harmonics was varied, by varying the absolute frequency of adjacent tones. The synthesized tones were stored on a "RAP" formatted disk, that was used in conjunction with the RAP 111 computer system interfaced with the Eclipse computer.

Photographs showing the waveform of the synthesized sounds corresponding to T1P1, and T2P1 (without fundamental) are shown in figure 4.3. The time scale (abscissa) represents 5 msec/div. The trace therefore shows almost the entire duration of the sound (45 msec of the 50 msec duration). The amplitude rise and fall at the beginning and end of the sounds, and the flat middle period is visible.

The top picture shows the waveform of a complex tone containing the first harmonic or fundamental frequency 262 Hz, as well as the second, third, and fourth harmonics. The bottom picture shows the same fundamental periodicity, but the first harmonic (the fundamental) is infact absent, and the sound is composed of the second, third, fourth, and fifth harmonics. The verification that the fundamental period of this sound corresponds to the fundamental frequency of the harmonics used, is reassuring since the design of this monaural experiment relies on the validity of the phenomenon of the missing fundamental.

Though the traces of other sounds are not shown, they were observed on the oscilloscope, and the fundamental periodicity verified.

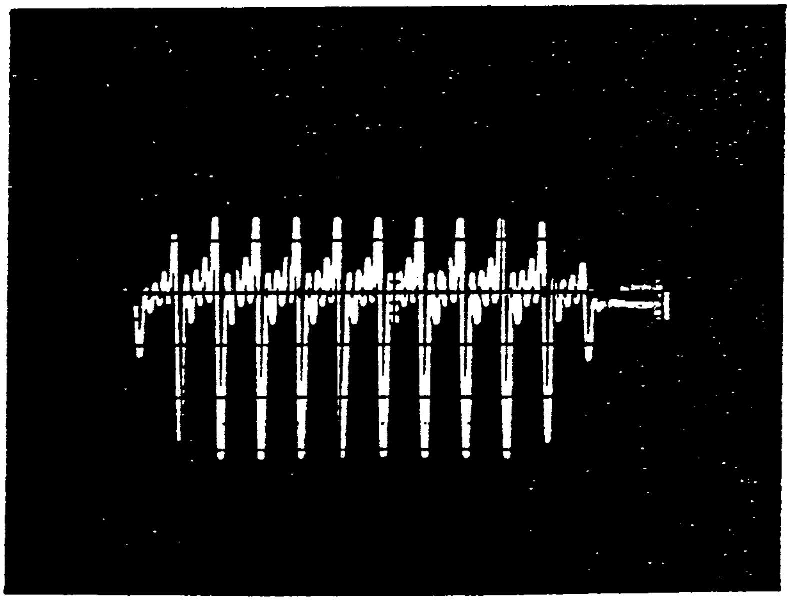
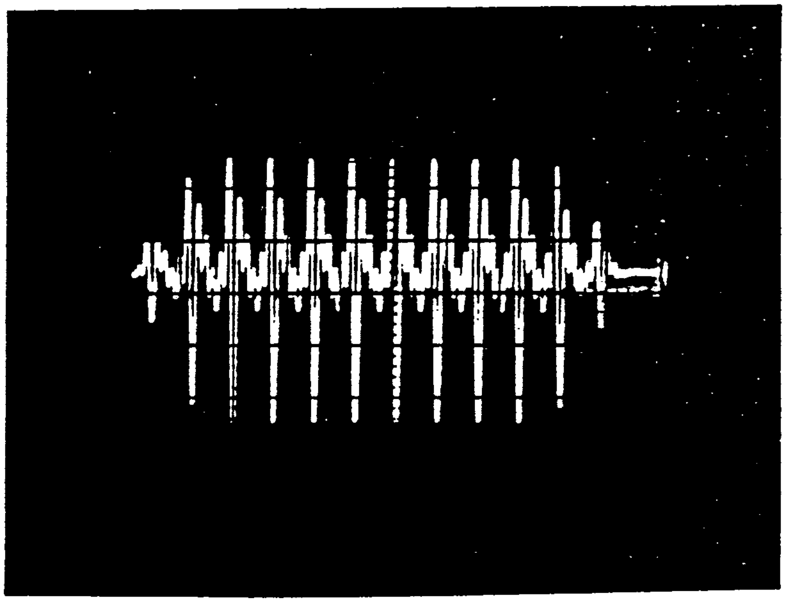


Figure 4.3 Oscillographic pictures of the waveforms of tones T1P1 (top picture) and T2P1 (bottom picture), showing a common fundamental periodicity, despite absence of the fundamental in T2P1.



4.3 Sequence Construction

The synthesized sounds stored on the Rap disk were retrieved and used in the manufacture of the forty-nine sequences described before, using a software package "PARAPET" (Engebretson and Hakkinen, 1979) which was basically a "Programmable Analog Recording And Playback system with Editing and paper Tape" facilities. This was executed on the RAP 111 computer system, comprising a Nova type computer with RAP analog, disk and station driver subsystems.

This sequencing program, basically enabled sounds to be laid out in a particular order, separated by fixed "track" lengths. The sequences made were therefore "frozen" in time, and the "rate of presentation of tones" was kept fixed once it had been specified. On the basis of observations made during pilot experimentation, it had been noted that the grouping of sounds, such as is manifested in the stream segregation phenomena, begins to be perceived when sequences of sounds are presented at rates as rapid as between 4 to 14 tones/sec. These rates correspond roughly, to inter-event-intervals of 225 and 75 msec. respectively. The onset-to-onset interval separating the events in the present sequences, was thus fixed at a "middle-value" of 150 msec. This corresponded to a silent interval of 100 msec. between tones of 50 msec. duration. This rate of presentation was then kept fixed for all the sequences to be used in the experiment. The pitch "layout" of the sequences is illustrated in figure 4.4. The seven pitch intervals possible within the sequences, were orchestrated using the seven timbres to give a total of forty-nine such sequences, each having a different timbral relation between the alternate notes.

4.4 Choice of Task to be Set before the Subjects

The task desired of the subjects, was the report of their perception of the relations if any, between the four tones of the rapidly repeating sequence presented to them. Since there is no inherently "correct" way of grouping sounds, the response of the subject reflects a purely perceptual grouping, based upon some criterion adopted in differentiating between sounds, comparing them, and subsequently categorizing them into groups where possible. From preliminary sessions with different subjects, it had been observed that certain salient groupings do indeed begin to emerge from the sequences, the type and clarity of grouping varying across the set of sequences used.

In order to observe this variability in the response dependent on "type" of sequence, and to see if the type of grouping perceived was consistent across subjects, a

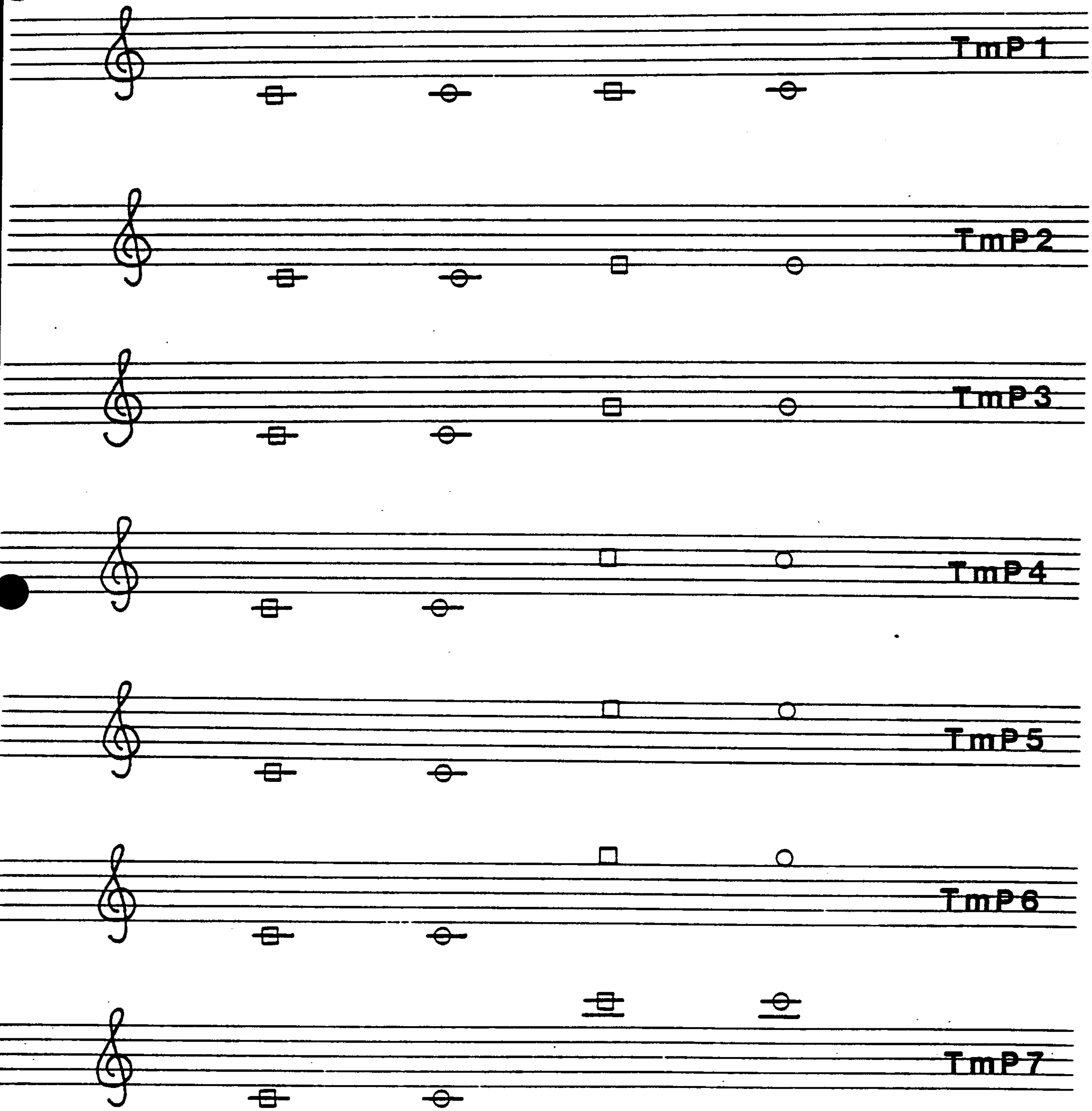


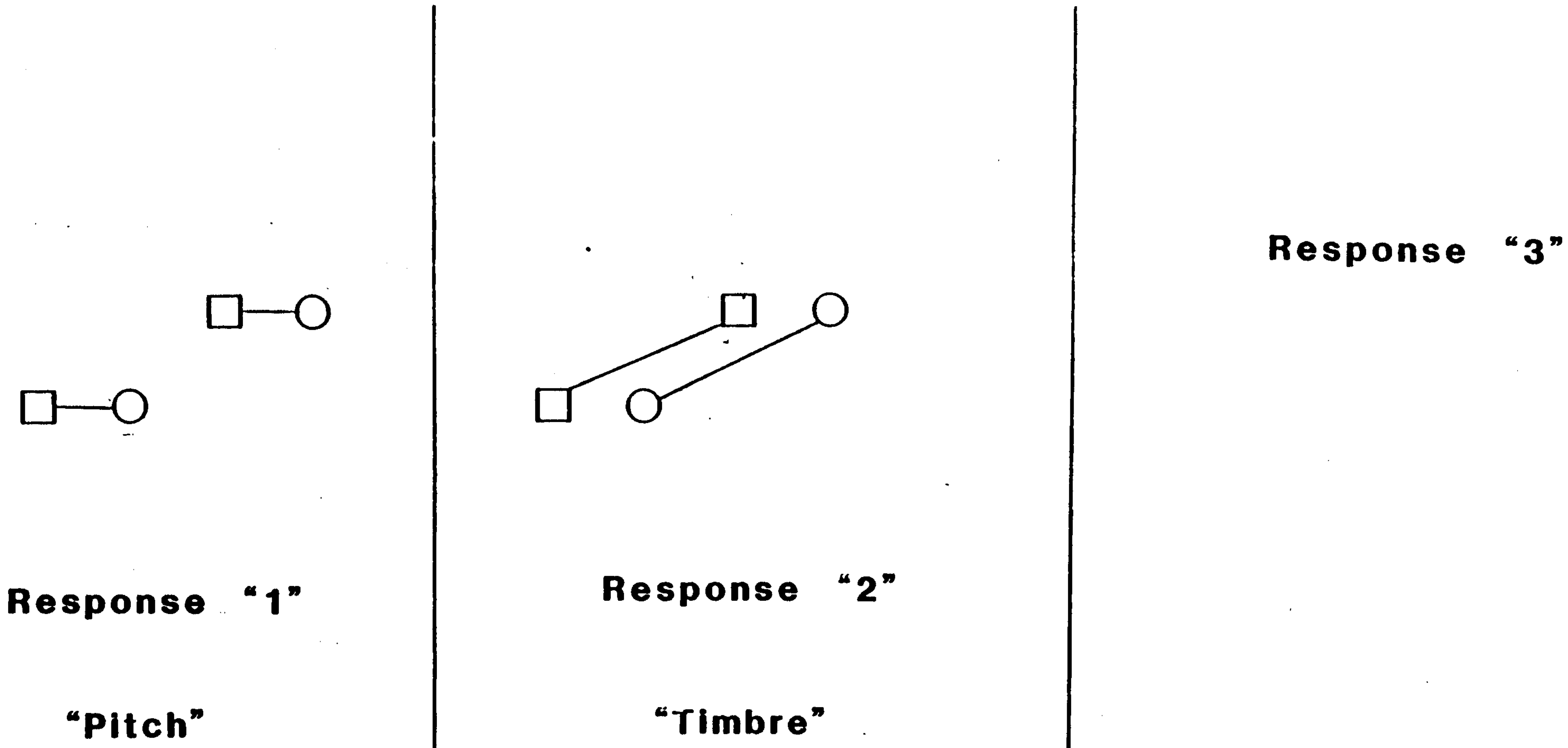
Figure 4.4 Pitch layout of sequences used in the experiment. The seven pitch intervals shown were orchestrated by seven timbres, giving forty-nine different sequences.

pseudo-identification task was decided upon. The subjects were given three alternatives, one denoting a grouping based on pitch, one denoting a grouping based on timbral similarity, and a third alternative that basically reflected a "dustbin" category in which to place stimuli that did not seem to be grouped in the other two ways.

Since the pitch layout of the sequence basically comprised a pair of "low" tones followed by a pair of high tones, a grouping by pitch would cause veridical perception of this layout, with the tones being perceived in the sequence "low-low-high-high", with a jump in pitch occurring between the first pair of sounds, and the second pair, for all sequences other than those having unison (equal pitch) as the pitch interval. Such a response was assigned the key `1´, which when pressed would imply that the subject had used such a pitch based measure in grouping the sounds.

A timbre based grouping on the other hand, would cause the pitch layout of the sequence to be perceptually confused, because of the overriding timbre cue, thus resulting in the perception of two melodic intervals being repeated, in the fashion "low-high-low-high", with the true "order" of the sequence tones being "captured" within the timbre pairs. Such a grouping was assigned the key `2´, on the subject response terminal. The assignment of the response labels to these type of groupings is shown in figure 4.5.

Figure 4.5 Response labels used to denote different perceptual groupings : Response `1´ denotes a pitch based grouping, Response `2´ a timbre based grouping, and Response `3´ denotes uncertainty.



The "unsure" category was provided because, one of the motives of the experiment would be to relate a grouping response to the physical structure of the sequence. If the subject had no choice but to group the sounds into categories `1´ or `2´, the criterion on which such a choice would be based would not be very stringent, and very different stimuli could be grouped together, simply due to "a lack of choice". Since alternative responses could not be provided for all types of "partial" and vague groupings that might be perceived, the aim has been to "ferret out" those sequences that evoked definite pitch-based, or timbre-based responses, so that an analysis of the sequence might reveal a common thread between features of the sequences, and the grouping observed. Since the ambiguous category (represented by key `3´) was always available for getting rid of confusing sequences, it was hoped that the criteria adopted by the subjects in using the more "definitive" categories would be quite conclusive on their part.

5. EXPOSITION : THE EXPERIMENT

5.1 Description of Stimuli used for Presentation.

The stimulus presented to a subject during a single trial of an experimental block basically consisted of a repeating four-tone sequence presented over headphones, to the right ear. The sequences presented over a block were chosen from the set of forty nine sequences described earlier (chapter 4). The actual set of sequences used in a block, and the number of times that a particular sequence appeared in a block being varied, depended upon the experimental condition being studied. While the inter-event interval between the sequential tones was fixed at 150 msec as described, the (silent) time interval between consecutive repetitions of the sequence was fixed at 250 msec. The experimental "length" of one sequence was thus 750 msec. The time course of sequence presentation is illustrated in figure 5.1 . This 750 msec. sequence was then repeated twelve times, yielding a 9 sec. long stimulus, presented during a single trial. The intensity of the stimulus was adjusted to a level deemed to be comfortable by the subjects during initial "warm- up" sessions in which they were presented a short sample block of different sequences. The level was then kept fixed for the remainder of the session.

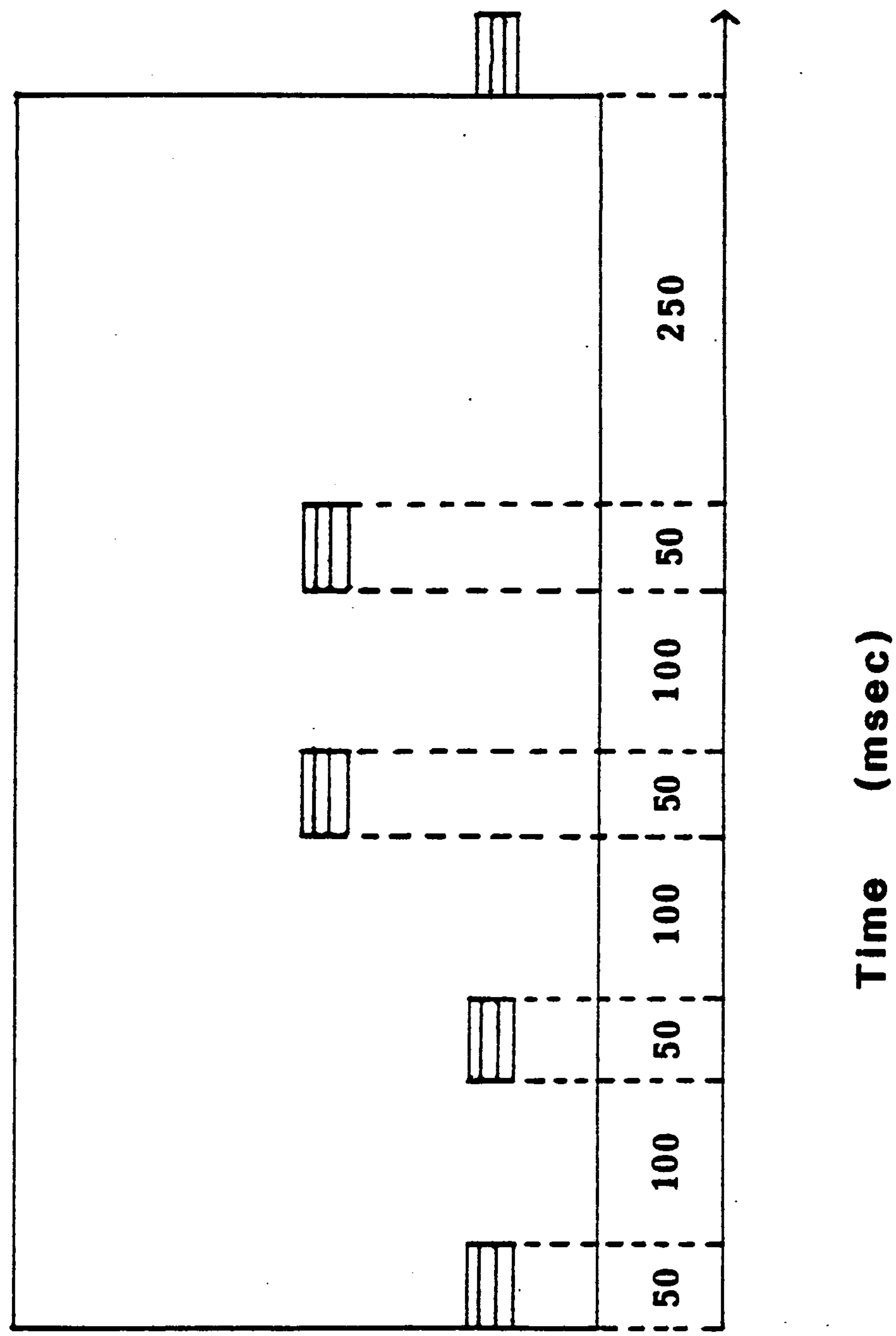


Figure 5.1 Time layout of sequence presented to a subject in a trial. Tones of 50 msec duration are separated by silent intervals of 100 msec. A 250 msec silent interval precedes the repetition of the sequence.

5.2 ~~Apparatus used in the Experiment.~~

An experimental control program "PSPADE" (Hakkinen 1981, modified by Goldwasser, 1984), permitted the experimental sequences that had been constructed and stored on the disk, to be retrieved later from the disk and played back over headphones, after passage through a variety of audio signal processing equipment. This auxilliary equipment basically comprised analog amplification equipment interacting with the computer via the digital-to-analog (and vice versa) circuitry of the RAP interface. This included a RAP system programmable, digital attenuator to facilitate manipulation of intensity level during the course of an experiment. The signal putout by the computer was finally received at five different locations, corresponding to five computer terminals, three of which were "dummy" terminals enslaved by a "master" terminal, these four terminals (Hazeltine 1420) being located in the testing room, which was acoustically treated to attenuate sound. An optional master terminal in the experiment control laboratory could be used in lieu of the other if desired. The signal underwent a final transduction at the output of Koss Pro/4AA headphones connected to the terminals, and arrived at the ear of the listener. The response of a listener, received as an ASCII character encoding the key pressed at the dummy terminal traversed the reverse path to the computer, and was recorded on the fixed disk in the lower drive of the computer. A printer (Centronics 704) connected to the computer was employed to obtain an instant "trial-by-trial" record of the responses made by a subject.

5.3 ~~Procedure and Conditions.~~

The PSPADE experiment control program was set up with parameters for the relevant session, and "cued" in readiness, to present the desired block to the subject when he/she was ready to receive it. This was indicated by the subject's pressing of any key on the terminal. After being presented the 9 second stimulus described above, the subjects were given 5 seconds in which to indicate their choice of the three alternatives, that they had been provided. This was done by assigning the keys `1´, `2´ and `3´ on the response terminal, labels that corresponded to the three alternatives described in section 4.4 (page 118). Thus, key `1´ was representative of a grouping based on pitch, key `2´ designated a streamed grouping based on timbre, and key `3´ represented the obscure "unsure" category.

The subjects were explained the designation of the keys and were given examples of sequences most likely to fall into the three categories in a warm-up session. Rather than being told which sequence corresponded to which type of grouping, the subjects were told what to "listen" for in terms of the pitch and rhythmic layout of the sequence, and to make their own decision about what category a sequence seemed to fit into. Thus, a "pitch-based" grouping (key `1´) was described as one in which two high notes seemed to follow two low notes in rapid succession in the form of a "stuttering" melodic interval "low-low-high-high", with the pitch interval occurring between the second and third tones. The "timbre-based" grouping (key `2´) was similarly described as one in which a low note seemed to be "linked" with a higher note in the form of a melodic interval. This same interval would seem to be repeated twice, but by different (instrumental) voices. The four tone sequence would thus seem to be of the form "low-high-low-high", with each interval ("low-high") occurring at half the tempo of the pitch-based sequences. This rhythmic confounding of the pitch intervals in a "timbre-based" grouping turned out to be a good way of determining the timbre streaming, and the subjects could well understand and relate to the phenomenon on the basis of this cue. These descriptions, while sufficing as an explanation of most of the sequences, could not appropriately be used to describe the sequences containing the unison pitch interval (all notes of pitch P1), since they did not have ordinal relations of the type "low" versus "high". These sequences were thus described separately as a "special case" of the more general groupings. A pitch-based grouping of the unison interval sequences would thus simply appear to be a quick sequence of the same four notes, while a timbre-based grouping would seem to be a slower repetition of a note, played by two instruments interrupting each other. The "unsure" category (key `3´) was explained as being available for labelling sequences that were confusing and did not seem to fall into either of the other categories. Thus sequences in which three notes seemed to be linked together, while one remained isolated, or where no particular grouping appeared to dominate, could be disposed off in this category.

Once the Subjects felt comfortable in making their decisions based on the alternatives available to them, they were launched on the "actual" block of the experimental condition being studied. To aid the remembering of the designated labels, a visual cue sheet was provided (illustrated in figure 5.2), that was placed above the keyboard of the terminal. The Subjects were instructed to use these keys to indicate their choice by pressing the appropriate key, and were told that they "must" respond, and not let the sequences slip by unlabelled. On elapse of the response interval, another trial was presented and the procedure repeated till the number of trials assigned to a block had all been presented. The order in which the set of sequences under study was presented, was randomized within a block. The type of sequences used within a block, and the number of occurrences of each was varied from session to session, depending on the condition being studied:

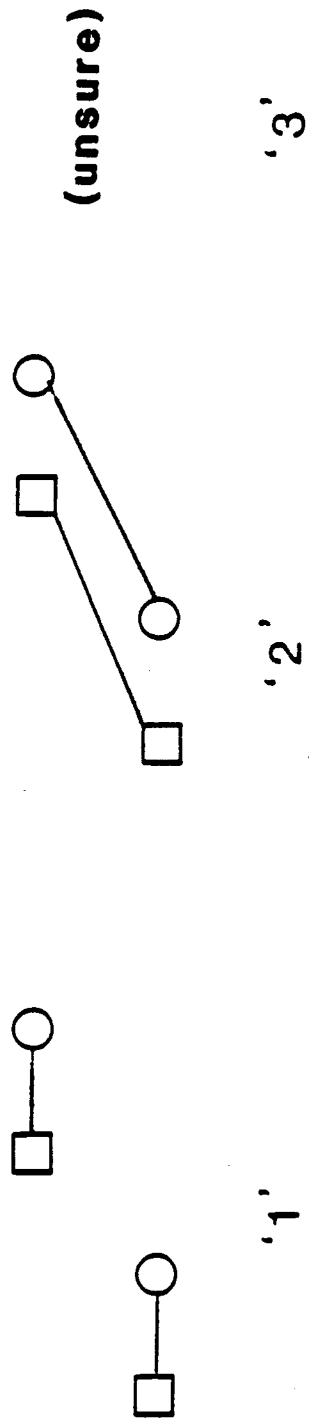


Figure 5.2 Cue sheet provided above response terminal to aid subjects in remembering the assignment of response keys to the type of grouping perceived.

In condition 1, the pitch interval between the first pair of tones and the second pair of tones was kept constant from trial to trial, while the timbres assigned to the second and fourth tones changed from trial to trial, covering the entire range of timbres from T1 through T7 within a single block. The blocksize for this condition was 49 trials, enabling each timbre to be represented seven times.

In condition 2, the timbral contrast between the tones of the first pair of tones and the second pair of tones (T2-Tm), was kept fixed, while the pitch interval between the first pair and the second pair varied from trial to trial. The blocksize for this condition was 56 trials, enabling the representation of the seven pitch steps P1 through P7 articulated in two timbral contexts : T2-Tm, and the isotimbral T2-T2 used as a reference. The fourteen different sequences were thus presented four times each.

In condition 3, the timbral differences and the pitch distance between the pairs of tones were interplayed, providing a mixed context within a block. In one sub-condition, the timbre and pitch difference was proportional, in that as the pitch interval between the first pair of tones and the second pair of tones increased, the difference in timbre (as manifested by difference in frequency region of the harmonics) correspondingly increased. In another sub-condition, the timbre and pitch differences varied inversely with respect to each other. In this case, the smallest pitch interval P1-P1 (unison) was articulated with the greatest difference in timbre (T2-T7) between the notes, and the largest pitch interval (P1-P7), corresponded to the timbral difference (T2-T1). The smallest timbre difference (i.e zero), corresponding to (T2-T2) was however correlated with the pitch interval (P1-P6). In these various conditions, the seven reference isotimbral sequences over the entire pitch range, were again provided. The blocksize was therefore also 56 (= 14 x 4).

In condition 4, the entire sequence set containing 49 different sequences, was presented in a 49-trial block. The subjects thus had to make their decisions on the basis of a single presentation of each sequence. This was repeated for another block, containing the same sequence set presented in a different (randomized) order.

This entire procedure of administering the trials, randomization, the length of the response interval, and the collection of responses, was monitored by the computer, programmed to do so by the PSPADE program mentioned above.

In addition to the "three alternative foced choice" task, the subjects were each given a "comment sheet" on which to describe their reaction to the different stimuli, in particular those that fell into the third "unsure", ambiguous category. While this option was entirely voluntary and not emphatically demanded of the subjects, it was provided in order to gain some insight on the other types of groupings or relationships perceived between sounds, that could not be included in the explicitly labeled categories (`1´ and `2´). Most of the subjects complied with this request, conscientiously providing trial-by-trial reports. These were later analysed in order to see the correspondence between the comments made by different subjects about a particular sequence, the physical structure of the synthesized sounds and the sequential context in which they were presented. These observations are reported in the "results" section of the next chapter.

5.4 Subjects

Six subjects, three male and three female, were used in the experiment. They were all between the age of 20 and 35 years, and had normal hearing. Two of the six (MJ and WK) had undergone extensive musical training, and were in fact "professional" musicians, MJ a composer, and WK a double bass player. In addition, MJ claimed to have "perfect pitch", and did indeed identify all the musical intervals used in the experiment correctly. The other four subjects were not rigorously trained musicians, but had all had varying degrees of musical training (between two and eight years). WK and PS had prior experience as subjects in psychoacoustic experiments, and DG in other psychological experiments. The subjects were tested using the same order of conditions on consecutive sessions, sometimes individually, and sometimes in conjunction with another subject. A typical session lasted two hours, and data were collected for sessions spread over a period of four weeks. The subjects were paid for their participation.

6. RESULTS

6.1 Data Collection and Analysis.

A trial-by-trial record of the responses made by the subjects was obtained by having the printer connected to the computer provide a "hard copy" of the experimental block that was concurrently being administered and monitored. The printout thus obtained provided information about both the stimulus used in a particular trial, and the type of response made by the subject (`1´, `2´, or `3´). The stimulus being presented was identified by the number of the beginning "track" of the disk on which the sequences were stored. Given the track numbers of the stimuli, as recorded on the printout, one could easily deduce the type of sequence that had been presented on a given trial, and note the type of label that had been attached to it by the listener.

Such trial-by-trial records were obtained for each subject, for every block of the different conditions. The number of times that a subject assigned a particular label to a particular sequence, was determined. Given the total number of times that a particular sequence was presented in a block, the proportion of times the various response keys were used to label it was determined, for each subject.

Since the aim of this study was to observe the relationship between the "type" of grouping perceived by an "average" or "typical" listener, and the physical structure of the sequence, the proportions of the various responses made in labeling the various sequences, were averaged across listeners. These data are given in tables 1.1 through 4.7, which also show in detail the responses of each listener, as well as the average proportions of the times each label was used as a response to a given sequence. The relative proportions in terms of "percent of the number of trials", that a particular type of response was made, are also given in the tables, for every block of each condition.

In order to observe the relationship between the grouping strategy used by a typical listener, and the structure of the sequences, the number of times that each response label was used (expressed as % of the number of total trials that each sequence was presented), was plotted against the "type" of sequence used. Figures 6.1.1 through 6.4.7 show this functional mapping between stimulus and response, for the different conditions studied.

6.2 Data Interpretation

The graphs plotted for the different experimental conditions imposed, reveal a number of interesting relationships between the pitch and timbral layout of the sounds of the sequence, and the grouping perceived by the average listener.

For ~~condition 1~~, in which the pitch relationship between the first pair of tones and the second pair of tones was kept fixed, while the timbre of the second and the last tone changed over trials, the graphs for the different "fixed" pitch intervals indicated that by and large, listeners tend to use the pitch relationship between sounds to form groups, if the qualitative difference between the sounds is not too great. However, as the timbre of the second and fourth sounds is made markedly different from that of the first and third sounds, (corresponding here, to a timbral "distance" greater than T4), the grouping strategy suddenly seems to switch from being one based on pitch to being one based on timbral similarity. This is manifested by a sudden "swap" of response label for timbral distances greater than that between T4 and the "anchored" timbre T2 of the first and third tones. Figures 6.1.1 through 6.1.3 corresponding to conditions 1.1 through 1.3 exhibit this change of strategy occurring at around a timbral distance of T4, rather distinctly.

Figures 6.1.4 through 6.1.7 corresponding to conditions 1.4 through 1.7, where the fixed pitch intervals were greater than or equal to an octave (P4 through P7), show a decrease in the tendency to group by timbre.

Table 6.1.1   Condition 1.1

TIMBRES VARIED OVER BLOCK, GIVEN P1

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
T1P1	1	-	7	7	7	-	-	21	0.500	50.0
	2	4	-	-	-	5	5	14	0.333	33.3
	3	3	-	-	-	2	2	7	0.167	16.7
T2P1	1	7	7	7	7	7	7	42	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T3P1	1	5	4	7	7	1	4	28	0.667	66.7
	2	-	-	-	-	6	-	6	0.143	14.3
	3	2	3	-	-	-	3	8	0.190	19.0
T4P1	1	-	-	4	3	-	-	7	0.167	16.7
	2	7	6	1	4	7	7	32	0.762	76.2
	3	-	1	2	-	-	-	3	0.071	7.1
T5P1	1	-	1	5	3	-	-	9	0.214	21.4
	2	7	5	1	4	7	7	31	0.738	73.8
	3	-	1	1	-	-	-	2	0.048	4.8
T6P1	1	-	-	4	3	-	-	7	0.167	16.7
	2	7	7	2	4	5	7	32	0.762	76.2
	3	-	-	1	-	2	-	3	0.071	7.1
T7P1	1	-	-	1	1	-	-	2	0.048	4.8
	2	7	7	4	6	7	7	38	0.905	90.5
	3	-	-	2	-	-	-	2	0.048	4.8

RESPONSE \`1´ denotes "pitch" based grouping
RESPONSE \`2´ denotes "timbre" based grouping
RESPONSE \`3´ indicates uncertainty.

Figure 6.1.1 Graph showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 1.1, in which the pitch interval between the tones of the sequence was a unison. Response 1 indicates a pitch based grouping, Response 2 a timbre based grouping, and Response 3 denotes uncertainty.

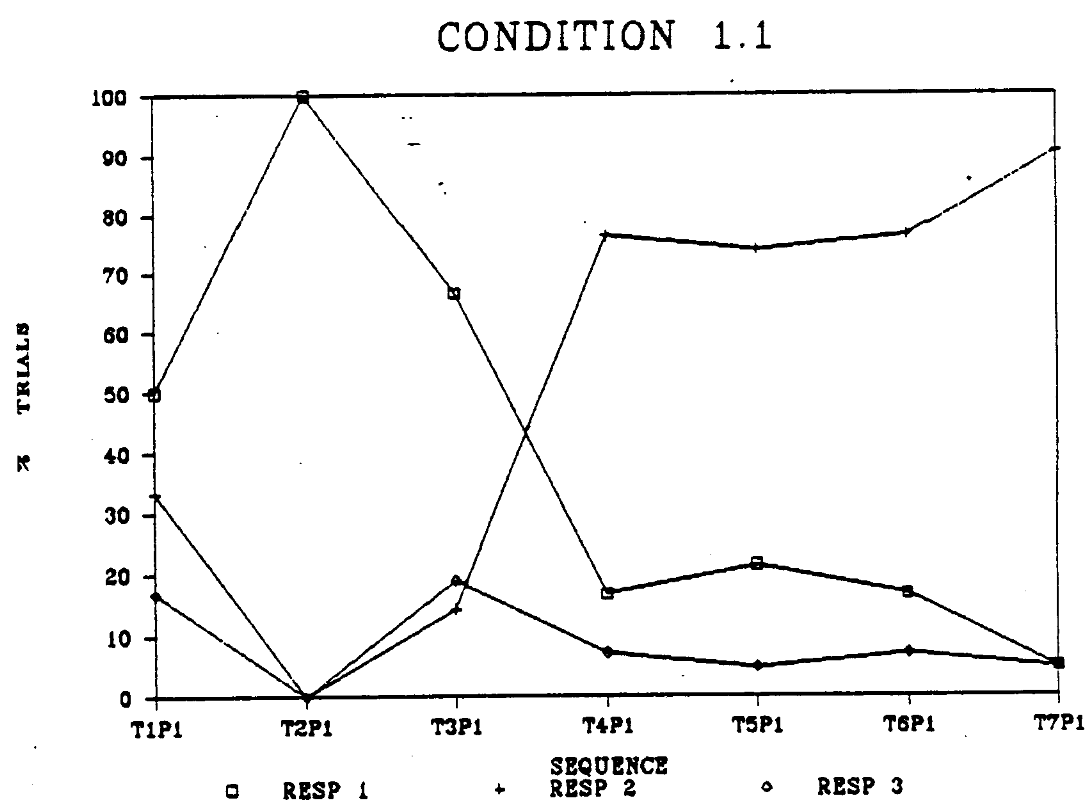


Table 6.1.2 Condition 1.2

TIMBRES VARIED OVER BLOCK, GIVEN P2

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
T1P2	1	-	7	1	1	4	-	13	0.310	31.0
	2	7	-	6	4	1	4	22	0.524	52.4
	3	-	-	-	2	2	3	7	0.167	16.7
T2P2	1	7	7	7	7	7	7	42	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T3P2	1	7	7	4	7	6	1	32	0.762	76.2
	2	-	-	1	-	1	-	2	0.048	4.8
	3	-	-	2	-	-	6	8	0.190	19.0
T4P2	1	-	-	5	2	5	-	12	0.286	28.6
	2	5	7	2	1	1	6	22	0.524	52.4
	3	2	-	-	4	1	1	8	0.190	19.0
T5P2	1	-	-	4	1	5	-	10	0.238	23.8
	2	7	7	3	6	-	7	30	0.714	71.4
	3	-	-	-	-	2	-	2	0.048	4.8
T6P2	1	-	-	2	-	2	-	4	0.095	9.5
	2	7	7	4	7	2	7	34	0.810	81.0
	3	-	-	1	-	3	-	4	0.095	9.5
T7P2	1	-	-	2	-	2	-	4	0.095	9.5
	2	7	7	5	7	3	7	36	0.857	85.7
	3	-	-	-	-	2	-	2	0.048	4.8

RESPONSE `1´ denotes "pitch" based grouping
RESPONSE `2´ denotes "timbre" based grouping
RESPONSE `3´ indicates uncertainty.

Figure 6.1.2 Graph showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 1.2, in which the pitch interval between the tones of the sequence was a third. Response 1 indicates a pitch based grouping, Response 2 a timbre based grouping, and Response 3 denotes uncertainty.

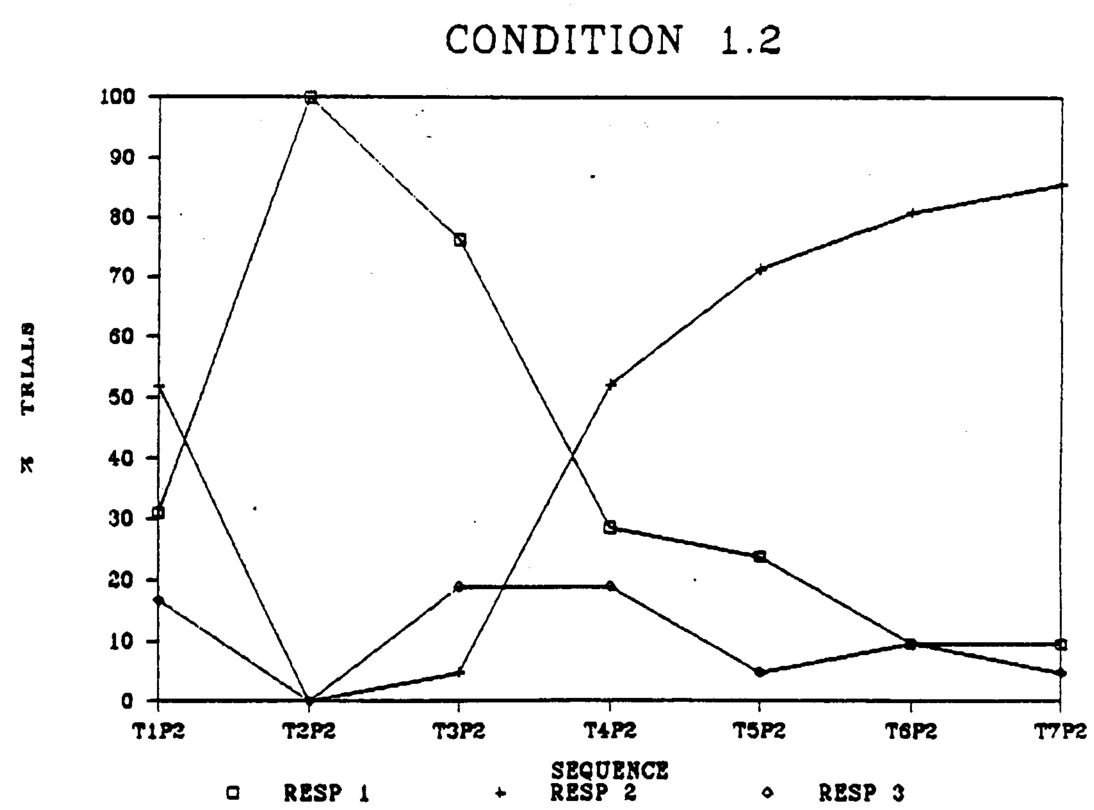


Table 6.1.3 Condition 1.3

TIMBRES VARIED OVER BLOCK, GIVEN P3

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
T1P3	1	-	7	4	-	2	-	13	0.310	31.0
	2	7	-	3	4	4	4	22	0.524	52.4
	3	-	-	-	3	1	3	7	0.167	16.7
T2P3	1	7	7	7	7	7	7	42	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T3P3	1	7	5	2	7	7	3	31	0.738	73.8
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	2	5	-	-	4	11	0.262	26.2
T4P3	1	3	-	1	5	5	2	16	0.381	38.1
	2	-	3	1	-	-	-	4	0.095	9.5
	3	4	4	5	2	2	5	22	0.524	52.4
T5P3	1	1	-	1	1	1	-	4	0.095	9.5
	2	4	6	5	-	4	7	26	0.619	61.9
	3	2	1	1	6	2	-	12	0.286	28.6
T6P3	1	-	-	-	-	-	-	0	0.000	0.0
	2	5	6	7	6	4	7	35	0.833	83.3
	3	2	1	-	1	3	-	7	0.167	16.7
T7P3	1	-	-	-	-	2	-	2	0.048	4.8
	2	7	7	7	6	5	7	39	0.929	92.9
	3	-	-	-	1	-	-	1	0.024	2.4

RESPONSE `1´ denotes "pitch" based grouping
RESPONSE `2´ denotes "timbre" based grouping
RESPONSE `3´ indicates uncertainty.

Figure 6.1.3 Graph showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 1.3, in which the pitch interval between the tones of the sequence was a fifth. Response 1 indicates a pitch based grouping, Response 2 a timbre based grouping, and Response 3 denotes uncertainty.

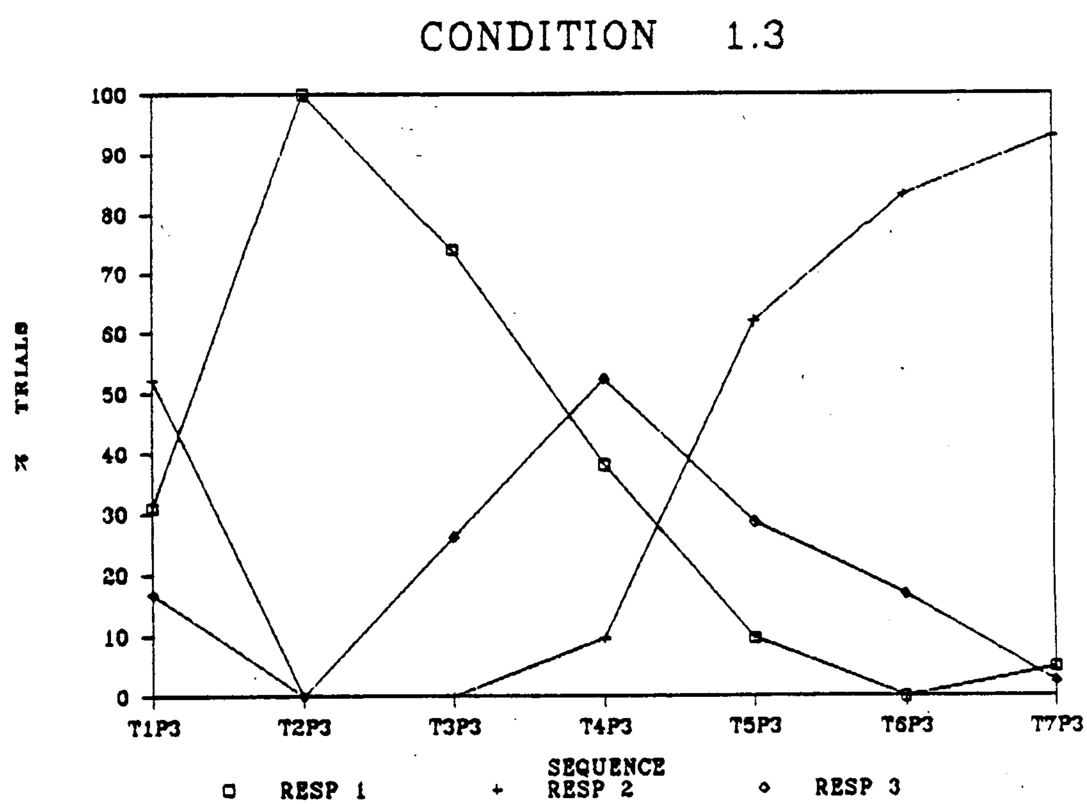


Table 6.1.4 Condition 1.4

TIMBRES VARIED OVER BLOCK, GIVEN P4

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
T1P4	1	1	7	6	1	-	-	15	0.357	35.7
	2	-	-	-	-	3	4	7	0.167	16.7
	3	6	-	1	6	4	3	20	0.476	47.6
T2P4	1	7	7	7	7	7	7	42	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T3P4	1	7	6	6	7	7	7	40	0.952	95.2
	2	-	-	1	-	-	-	1	0.024	2.4
	3	-	1	-	-	-	-	1	0.024	2.4
T4P4	1	4	2	4	7	7	7	31	0.738	73.8
	2	-	-	1	-	-	-	1	0.024	2.4
	3	3	5	2	-	-	-	10	0.238	23.8
T5P4	1	2	-	-	6	7	7	22	0.524	52.4
	2	-	6	4	-	-	-	10	0.238	23.8
	3	5	1	3	1	-	-	10	0.238	23.8
T6P4	1	2	-	1	7	6	3	19	0.452	45.2
	2	3	7	5	-	1	2	18	0.429	42.9
	3	2	-	1	-	-	2	5	0.119	11.9
T7P4	1	-	1	-	4	6	2	13	0.310	31.0
	2	4	6	7	1	-	3	21	0.500	50.0
	3	3	-	-	2	1	2	8	0.190	19.0

RESPONSE `1´ denotes "pitch" based grouping
RESPONSE `2´ denotes "timbre" based grouping
RESPONSE `3´ indicates uncertainty.

Figure 6.1.4 Graph showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 1.4, where the pitch interval between the tones of the sequence was an octave. Response 1 indicates a pitch based grouping, Response 2 a timbre based grouping, and Response 3 denotes uncertainty.

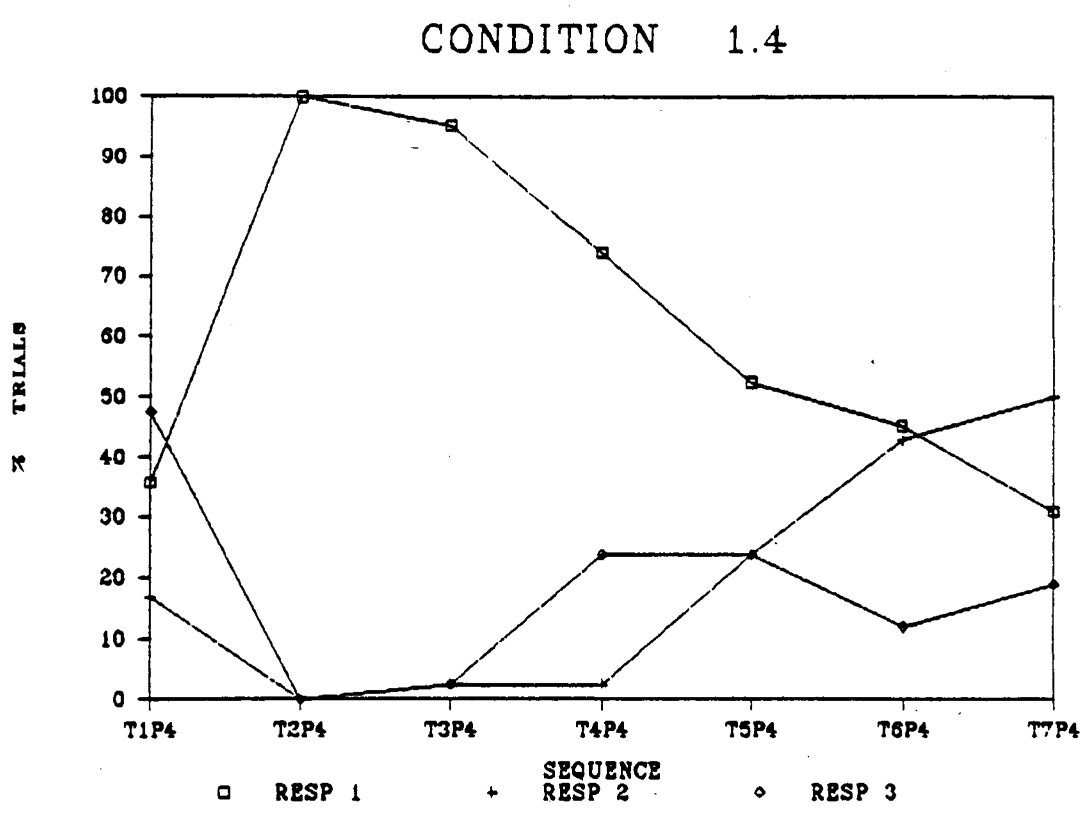


Table 6.1.5 Condition 1.5

TIMBRES VARIED OVER BLOCK, GIVEN P5

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
T1P5	1	7	7	7	-	5	1	27	0.643	64.3
	2	-	-	-	-	-	6	6	0.143	14.3
	3	-	-	-	7	2	-	9	0.214	21.4
T2P5	1	7	7	7	7	7	7	42	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T3P5	1	6	6	7	7	7	7	40	0.952	95.2
	2	1	-	-	-	-	-	1	0.024	2.4
	3	-	1	-	-	-	-	1	0.024	2.4
T4P5	1	7	-	-	7	7	7	28	0.667	66.7
	2	-	3	5	-	-	-	8	0.190	19.0
	3	-	4	2	-	-	-	6	0.143	14.3
T5P5	1	4	-	1	7	6	7	25	0.595	59.5
	2	-	7	5	-	-	-	12	0.286	28.6
	3	3	-	1	-	1	-	5	0.119	11.9
T6P5	1	2	2	-	6	4	5	19	0.452	45.2
	2	5	4	6	-	2	1	18	0.429	42.9
	3	-	1	1	1	1	1	5	0.119	11.9
T7P5	1	-	-	-	5	5	2	12	0.286	28.6
	2	7	7	7	-	2	3	26	0.619	61.9
	3	-	-	-	2	-	2	4	0.095	9.5

RESPONSE \`1´ denotes "pitch" based grouping
RESPONSE \`2´ denotes "timbre" based grouping
RESPONSE \`3´ indicates uncertainty.

Figure 6.1.5 Graph showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 1.5, where the pitch interval between the tones of the sequence was a tenth. Response 1 indicates a pitch based grouping, Response 2 a timbre based grouping, and Response 3 denotes uncertainty.

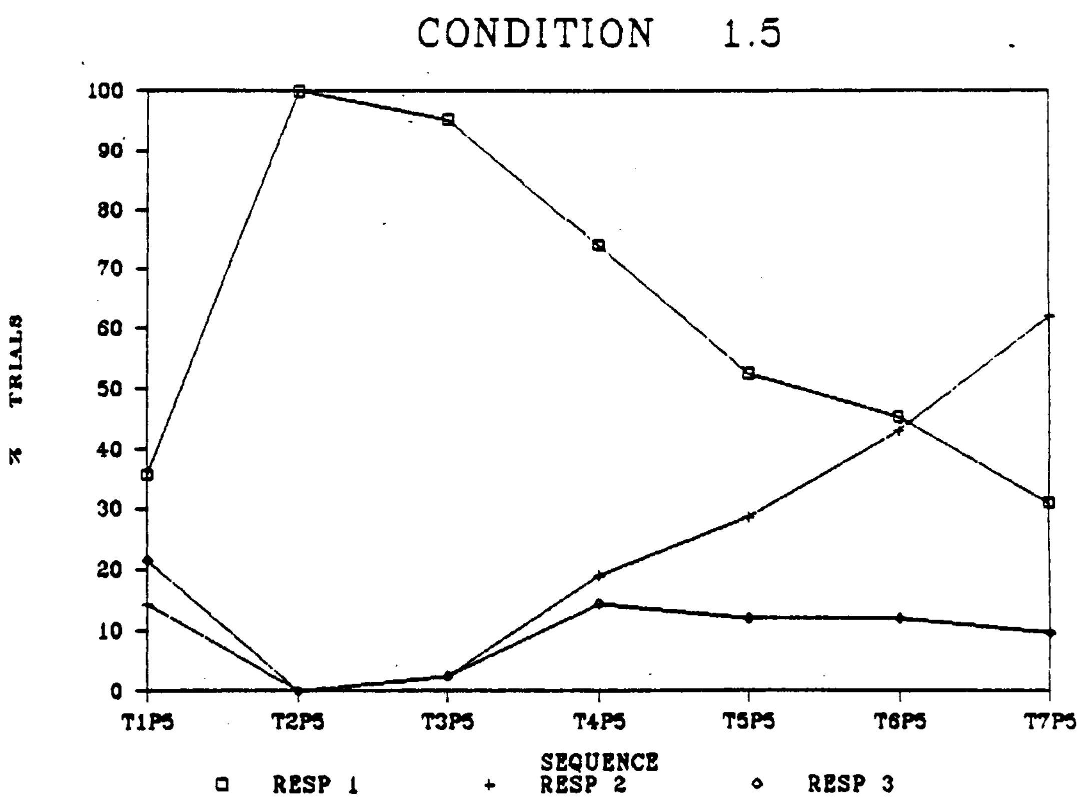


Table 6.1.6 Condition 1.6

TIMBRES VARIED OVER BLOCK, GIVEN P6

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
T1P6	1	7	6	7	-	-	6	26	0.619	61.9
	2	-	-	-	-	2	-	2	0.048	4.8
	3	-	1	-	7	5	1	14	0.333	33.3
T2P6	1	7	7	7	7	7	7	42	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T3P6	1	7	7	6	7	7	7	41	0.976	97.6
	2	-	-	1	-	-	-	1	0.024	2.4
	3	-	-	-	-	-	-	0	0.000	0.0
T4P6	1	7	-	6	7	7	7	34	0.810	81.0
	2	-	4	-	-	-	-	4	0.095	9.5
	3	-	3	1	-	-	-	4	0.095	9.5
T5P6	1	3	-	2	7	6	5	23	0.548	54.8
	2	1	7	4	-	-	-	12	0.286	28.6
	3	3	-	1	-	1	2	7	0.167	16.7
T6P6	1	7	-	-	7	7	6	27	0.643	64.3
	2	-	7	6	-	-	-	13	0.310	31.0
	3	-	-	1	-	-	1	2	0.048	4.8
T7P6	1	7	-	-	7	7	7	28	0.667	66.7
	2	-	7	7	-	-	-	14	0.333	33.3
	3	-	-	-	-	-	-	0	0.000	0.0

RESPONSE `1´ denotes "pitch" based grouping
RESPONSE `2´ denotes "timbre" based grouping
RESPONSE `3´ indicates uncertainty.

Figure 6.1.6 Graph showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 1.6, where the pitch interval between the tones of the sequence was a twelfth. Response 1 indicates a pitch based grouping, Response 2 a timbre based grouping, and Response 3 denotes uncertainty.

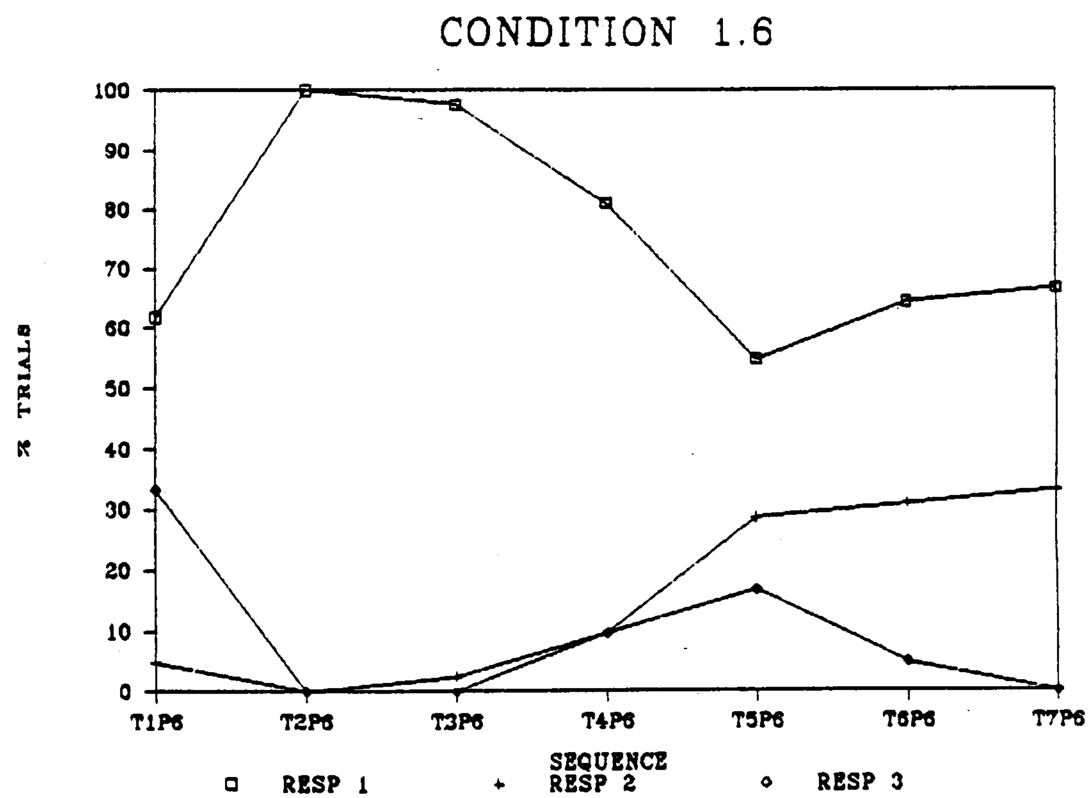


Table 6.1.7 Condition 1.7

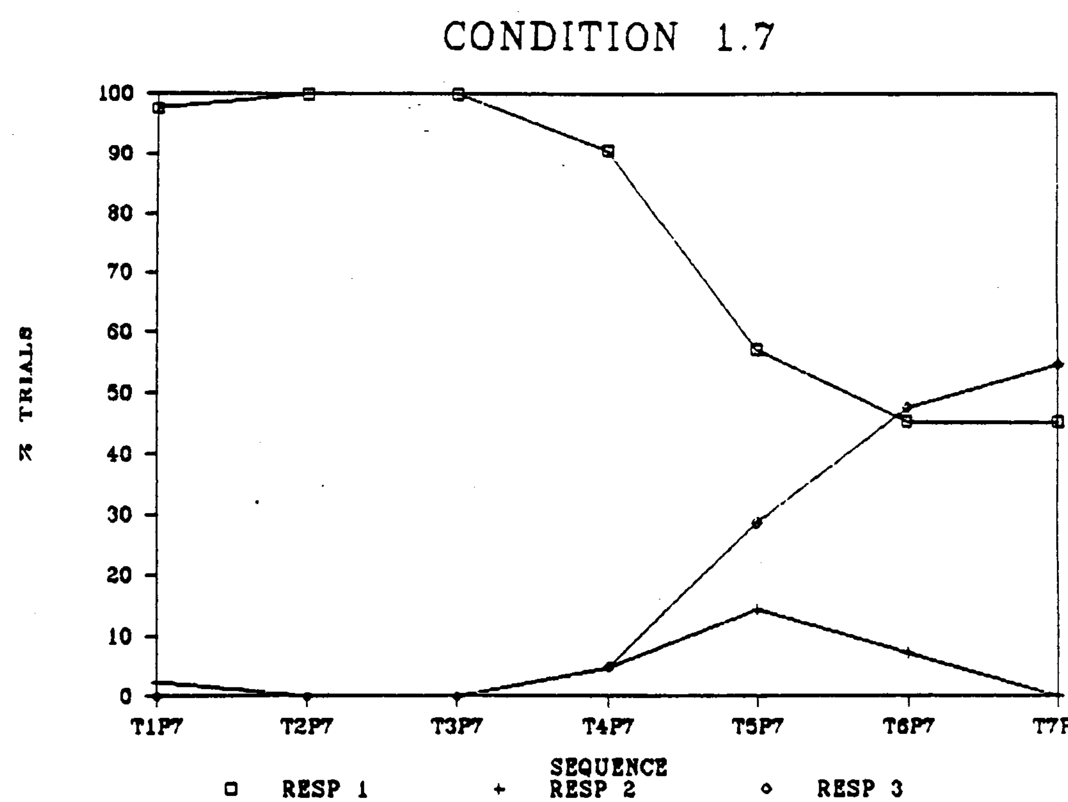
TIMBRES VARIED OVER BLOCK, GIVEN P7

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
T1P7	1	6	7	7	7	7	7	41	0.976	97.6
	2	1	-	-	-	-	-	1	0.024	2.4
	3	-	-	-	-	-	-	0	0.000	0.0
T2P7	1	7	7	7	7	7	7	42	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T3P7	1	7	7	7	7	7	7	42	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T4P7	1	7	7	3	7	7	7	38	0.905	90.5
	2	-	-	2	-	-	-	2	0.048	4.8
	3	-	-	2	-	-	-	2	0.048	4.8
T5P7	1	6	1	-	7	7	3	24	0.571	57.1
	2	1	-	5	-	-	-	6	0.143	14.3
	3	-	6	2	-	-	4	12	0.286	28.6
T6P7	1	5	1	2	-	7	4	19	0.452	45.2
	2	-	1	2	-	-	-	3	0.071	7.1
	3	2	5	3	7	-	3	20	0.476	47.6
T7P7	1	5	7	-	-	7	-	19	0.452	45.2
	2	-	-	-	-	-	-	0	0.000	0.0
	3	2	-	7	7	-	7	23	0.548	54.8

RESPONSE \`1´ denotes "pitch" based grouping
RESPONSE \`2´ denotes "timbre" based grouping
RESPONSE \`3´ indicates uncertainty.

Figure 6.1.7 Graph showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 1.7, where the pitch interval between the tones of the sequence was two octaves. Response 1 indicates a pitch based grouping, Response 2 a timbre based grouping, and Response 3 denotes uncertainty.



For the fixed pitch intervals P4 (octave) and P5 (tenth), the timbre grouping still persists, but at a lower level of certainty than was exhibited for the P1 (unison), P2 (third), and P3 (fifth) pitch intervals. For the interval corresponding to P6 (twelfth), the response `1´ reflecting a pitch based grouping, seems to pervade over the whole range of timbres. For the interval corresponding to P7 (two octaves), the greatest degree of uncertainty is exhibited, with the response label `3´ ("unsure") being the dominant response beyond the timbral distance T4.

For condition 2, in which the timbral "distance" between the sounds was kept fixed, while the pitch interval between the first and second pair was varied within a block, the data plotted reveal similar facets of the tradeoffs between pitch and timbre distances, but from another viewpoint.

For condition 2.1, in which the timbral distance (T2-T1) was kept fixed while the pitch interval between pairs varied from trial to trial in a random fashion, the graph plotted in figure 6.2.1, for the data given in table 6.2.1 shows that grouping by timbre (represented by the response `2´), was dominant upto a pitch interval of an octave (P4). At this pitch interval, the grouping strategy used by the subjects seems to have undergone a sudden change to being one based on pitch (response label `1´). The uncertainty (response label `3´) in grouping is also seen to increase beyond this pitch interval.

The graph plotted for condition 2.2, (figure 6.2.2,table 6.2.2), shows the trivial case (T2-T2), in which there was no difference in the timbre of the sounds. Consequently, the subjects used the only cue available to them in grouping the sounds, viz. the pitch distance between the first and second pairs. A 100 % grouping by pitch (response `1´), was therefore obtained for this iso-timbral case, as was to be expected.

The data plotted for condition 2.3 (figure 6.2.3 ,table 6.2.3) show that for the fixed timbral distance (T2-T3), the timbral cue was not too distinct in initiating group formation. While a timbre grouping was chosen in about 67 % of the trials for a zero pitch distance P1 (unison), uncertainty prevailed for pitch intervals corresponding to P2 (a third), and P3 (a fifth). For intervals greater than and equal to an octave (P4), grouping by pitch was dominant, reaching almost 96 % for P5 (a tenth).

The data obtained for conditions 2.4 and 2.5 (figures 6.2.4 and 6.2.5,tables 6.2.4 and 6.2.5) show that grouping by timbre was dominant for intervals less than an octave (P4), at which point there was a sudden swap in grouping scheme, with the pitch cue taking over.

Table 6.2.1 Condition 2.1

PITCH CHANGE OVER BLOCK, GIVEN TIMBRE T1

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
P1T1	1	-	-	1	-	1	-	2	0.083	8.3
	2	-	4	3	4	2	4	17	0.708	70.8
	3	4	-	-	-	1	-	5	0.208	20.8
P2T1	1	-	1	-	-	3	-	4	0.167	16.7
	2	3	3	4	4	-	4	18	0.750	75.0
	3	1	-	-	-	1	-	2	0.083	8.3
P3T1	1	-	-	1	-	4	-	5	0.208	20.8
	2	4	4	3	4	-	4	19	0.792	79.2
	3	-	-	-	-	-	-	0	0.000	0.0
P4T1	1	3	3	-	1	4	2	13	0.542	54.2
	2	-	1	3	-	-	2	6	0.250	25.0
	3	1	-	1	3	-	-	5	0.208	20.8
P5T1	1	4	1	2	2	4	2	15	0.625	62.5
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	3	2	2	-	2	9	0.375	37.5
P6T1	1	4	2	4	2	4	1	17	0.708	70.8
	2	-	1	-	-	-	-	1	0.042	4.2
	3	-	1	-	2	-	3	6	0.250	25.0
P7T1	1	3	3	4	2	4	4	20	0.833	83.3
	2	-	1	-	-	-	-	1	0.042	4.2
	3	1	-	-	2	-	-	3	0.125	12.5

RESPONSE \`1´ denotes "pitch" based grouping
RESPONSE \`2´ denotes "timbre" based grouping
RESPONSE \`3´ indicates uncertainty.

Figure 6.2.1 Graph showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 2.1, in which the "timbral-distance" between the sequence tones was kept fixed to be (T2-T1), while the pitch interval varied over the trials in a block. Response 1 indicates grouping by pitch, Response 2 by timbre, and Response 3 denotes uncertainty.

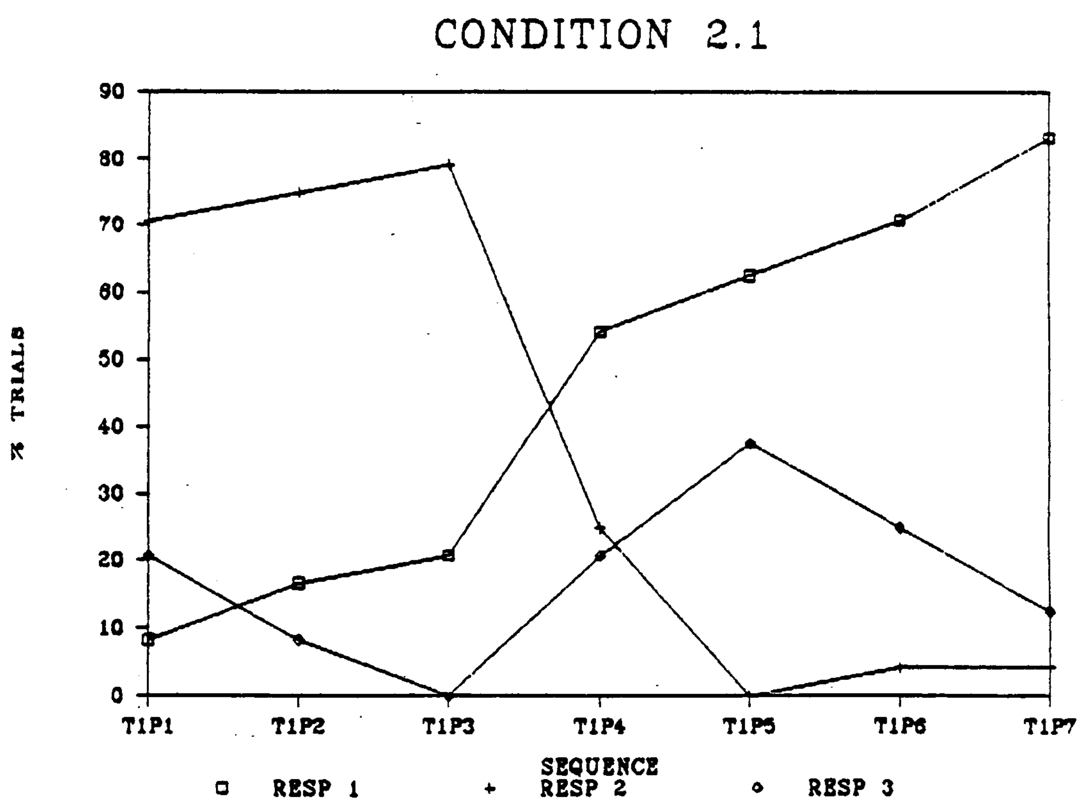


Table 6.2.2 Condition 2.2

PITCH CHANGE OVER BLOCK, GIVEN TIMBRE T2

		SUBJECTS								
SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
P1T2	1	4	4	4	4	4	4	24	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
P2T2	1	4	4	4	4	4	4	24	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
P3T2	1	4	4	4	4	4	4	24	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
P4T2	1	4	4	4	4	4	4	24	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
P5T2	1	4	4	4	4	4	4	24	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
P6T2	1	4	4	4	4	4	4	24	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
P7T2	1	4	4	4	4	4	4	24	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0

RESPONSE \`1´ denotes "pitch" based grouping
RESPONSE \`2´ denotes "timbre" based grouping
RESPONSE \`3´ indicates uncertainty.

Figure 6.2.2 Graph showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for the iso-timbral condition 2.2. A 100 % grouping by pitch (Response `1') is thus obtained, since such sequences lacked the "timbral-distance" cue.

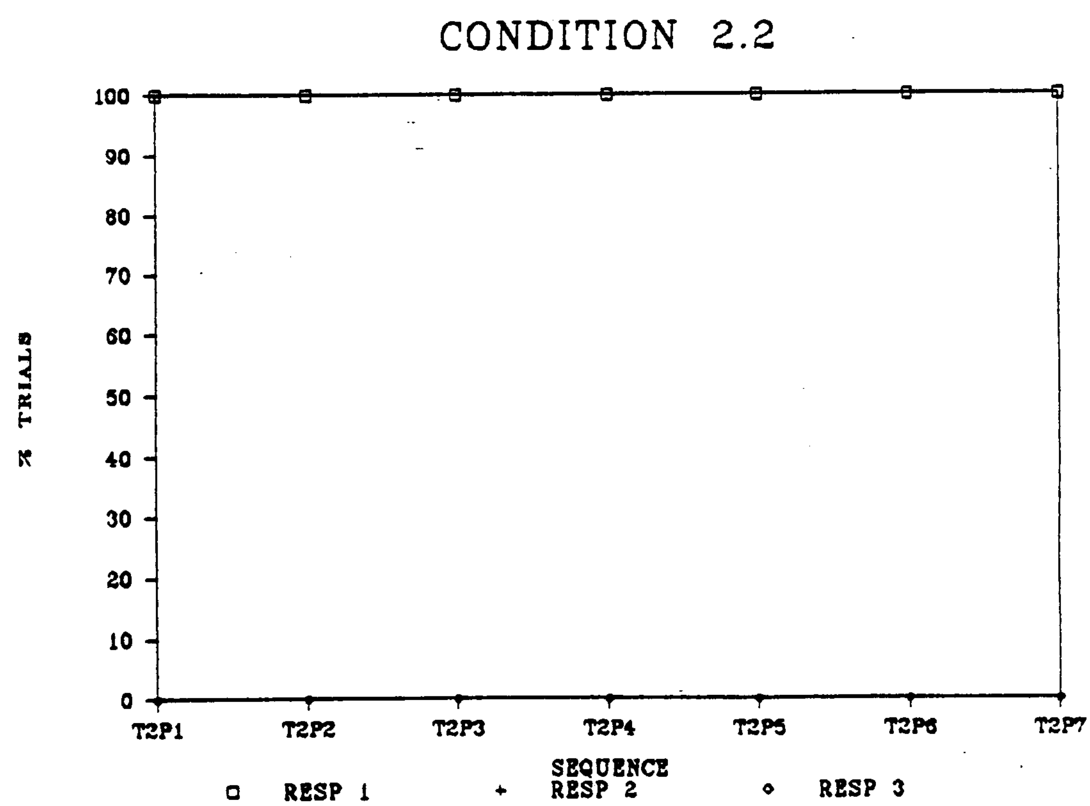


Table 6.2.3 Condition 2.3

PITCH CHANGE OVER BLOCK, GIVEN TIMBRE T3

		SUBJECTS								
SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
P1T3	1	2	-	-	-	-	-	2	0.083	8.3
	2	1	4	4	1	2	4	16	0.667	66.7
	3	1	-	-	3	2	-	6	0.250	25.0
P2T3	1	3	1	1	4	-	-	9	0.375	37.5
	2	1	1	3	-	1	3	9	0.375	37.5
	3	-	2	-	-	3	1	6	0.250	25.0
P3T3	1	3	-	2	3	2	-	10	0.417	41.7
	2	-	-	2	-	-	1	3	0.125	12.5
	3	1	4	-	1	2	3	11	0.458	45.8
P4T3	1	4	1	1	4	4	3	17	0.708	70.8
	2	-	1	3	-	-	-	4	0.167	16.7
	3	-	2	-	-	-	1	3	0.125	12.5
P5T3	1	4	4	3	4	4	4	23	0.958	95.8
	2	-	-	1	-	-	-	1	0.042	4.2
	3	-	-	-	-	-	-	0	0.000	0.0
P6T3	1	3	3	2	4	4	4	20	0.833	83.3
	2	-	-	1	-	-	-	1	0.042	4.2
	3	1	1	1	-	-	-	3	0.125	12.5
P7T3	1	4	2	1	4	3	4	18	0.750	75.0
	2	-	-	3	-	-	-	3	0.125	12.5
	3	-	2	-	-	1	-	3	0.125	12.5

RESPONSE \`1´ denotes "pitch" based grouping
RESPONSE \`2´ denotes "timbre" based grouping
RESPONSE \`3´ indicates uncertainty.

Figure 6.2.3 Graph showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 2.3, in which the "timbral-distance" between the sequence tones was kept fixed to be (T2-T3), while the pitch interval varied over the trials in a block. Response 1 indicates grouping by pitch, Response 2 by timbre, and Response 3 denotes uncertainty.

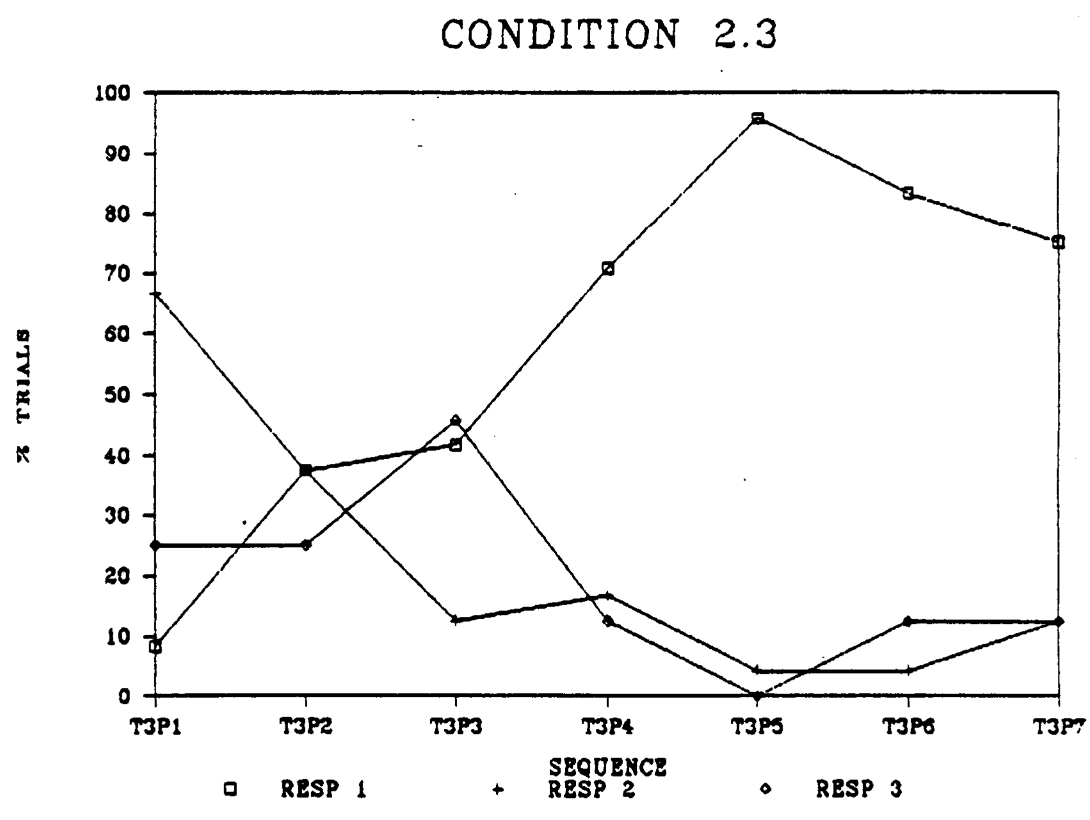


Table 6.2.4 Condition 2.4

PITCH CHANGE OVER BLOCK, GIVEN TIMBRE T4

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
		SUBJECTS								
P1T4	1	-	-	-	-	-	-	0	0.000	0.0
	2	4	4	4	3	4	4	23	0.958	95.8
	3	-	-	-	1	-	-	1	0.042	4.2
P2T4	1	-	-	-	2	2	-	4	0.167	16.7
	2	4	3	4	1	-	4	16	0.667	66.7
	3	-	1	-	1	2	-	4	0.167	16.7
P3T4	1	-	1	1	3	3	-	8	0.333	33.3
	2	3	1	3	-	-	4	11	0.458	45.8
	3	1	2	-	1	1	-	5	0.208	20.8
P4T4	1	4	4	2	2	4	4	20	0.833	83.3
	2	-	-	2	-	-	-	2	0.083	8.3
	3	-	-	-	2	-	-	2	0.083	8.3
P5T4	1	4	3	-	3	3	4	17	0.708	70.8
	2	-	-	2	-	-	-	2	0.083	8.3
	3	-	1	2	1	1	-	5	0.208	20.8
P6T4	1	4	3	1	4	4	4	20	0.833	83.3
	2	-	-	2	-	-	-	2	0.083	8.3
	3	-	1	1	-	-	-	2	0.083	8.3
P7T4	1	4	4	-	4	2	3	17	0.708	70.8
	2	-	-	3	-	-	-	3	0.125	12.5
	3	-	-	1	-	2	1	4	0.167	16.7

RESPONSE \`1´ denotes "pitch" based grouping
RESPONSE \`2´ denotes "timbre" based grouping
RESPONSE \`3´ indicates uncertainty.

Figure 6.2.4 Graph showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 2.4, in which the "timbral-distance" between the sequence tones was kept fixed to be (T2-T4), while the pitch interval varied over the trials in a block. Response 1 indicates grouping by pitch, Response 2 by timbre, and Response 3 denotes uncertainty.

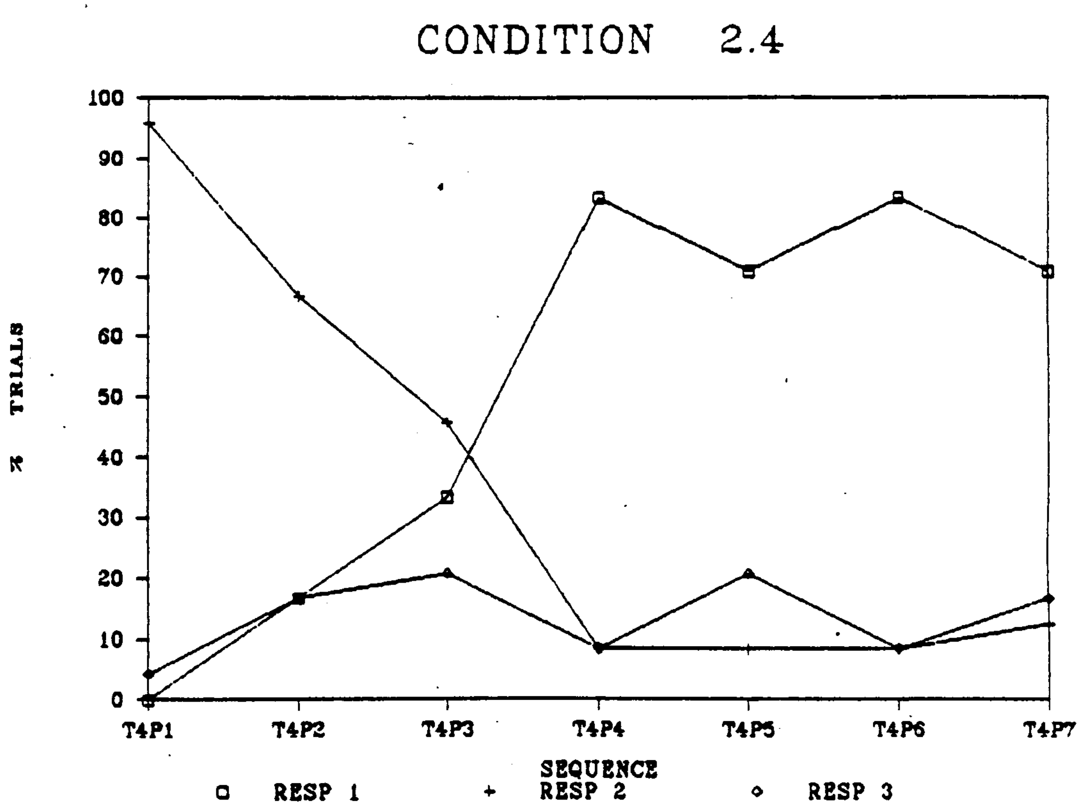


Table 6.2.5 Condition 2.5

PITCH CHANGE OVER BLOCK, GIVEN TIMBRE T5

		SUBJECTS									
SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
P1T5	1	-	-	-	-	-	-	0	0.000	0.0
	2	4	4	4	4	4	4	24	1.000	100.0
	3	-	-	-	-	-	-	0	0.000	0.0
P2T5	1	-	-	-	-	2	-	2	0.083	8.3
	2	4	4	4	4	-	4	20	0.833	83.3
	3	-	-	-	-	2	-	2	0.083	8.3
P3T5	1	1	-	-	-	4	-	5	0.208	20.8
	2	3	3	4	3	-	4	17	0.708	70.8
	3	-	1	-	1	-	-	2	0.083	8.3
P4T5	1	2	-	1	4	4	2	13	0.542	54.2
	2	-	1	3	-	-	-	4	0.167	16.7
	3	2	3	-	-	-	2	7	0.292	29.2
P5T5	1	2	1	2	3	1	2	11	0.458	45.8
	2	-	1	2	-	1	-	4	0.167	16.7
	3	2	2	-	1	2	2	9	0.375	37.5
P6T5	1	2	3	1	4	3	3	16	0.667	66.7
	2	1	-	3	-	-	-	4	0.167	16.7
	3	1	1	-	-	1	1	4	0.167	16.7
P7T5	1	4	4	-	4	4	3	19	0.792	79.2
	2	-	-	4	-	-	-	4	0.167	16.7
	3	-	-	-	-	-	1	1	0.042	4.2

RESPONSE \`1´ denotes "pitch" based grouping
RESPONSE \`2´ denotes "timbre" based grouping
RESPONSE \`3´ indicates uncertainty.

Figure 6.2.5 Graph showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 2.5, in which the "timbral-distance" between the sequence tones was kept fixed to be (T2-T5), while the pitch interval varied over the trials in a block. Response 1 indicates grouping by pitch, Response 2 by timbre, and Response 3 denotes uncertainty.

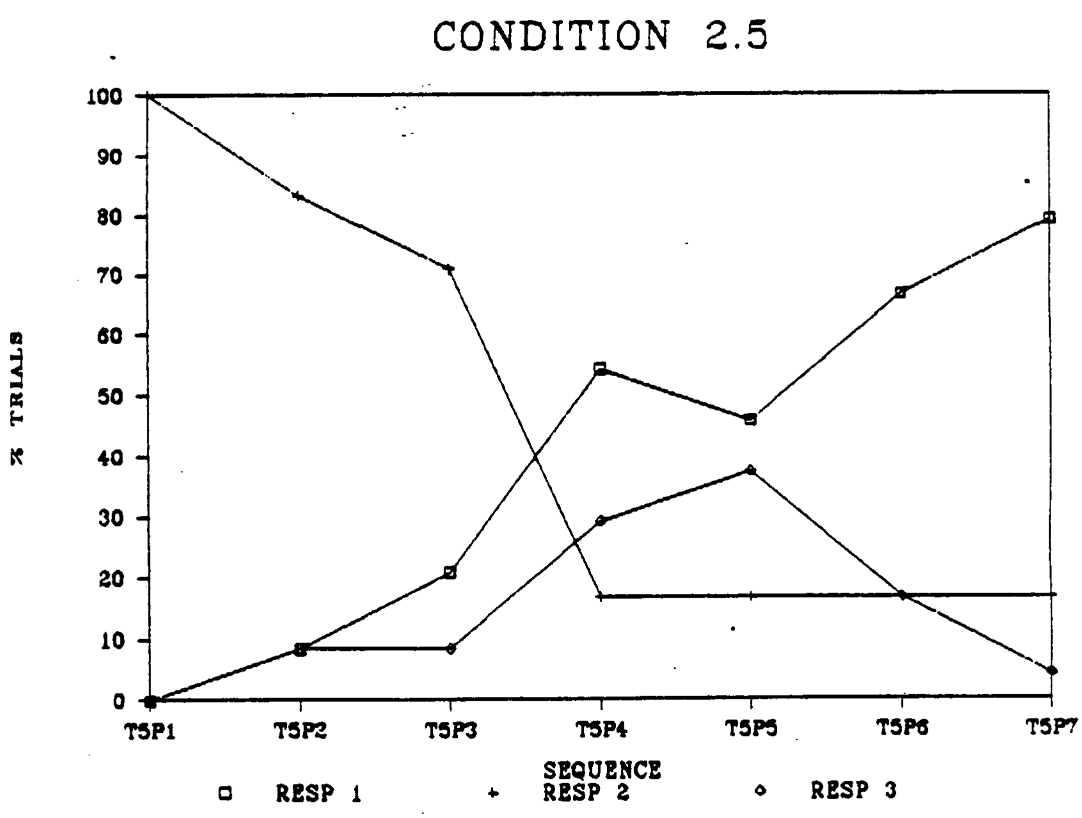


Table 6.2.6 Condition 2.6

PITCH CHANGE OVER BLOCK, GIVEN TIMBRE T6

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
P1T6	1	-	-	-	-	-	-	0	0.000	0.0
	2	4	4	4	4	3	4	23	0.958	95.8
	3	-	-	-	-	1	-	1	0.042	4.2
P2T6	1	-	-	-	-	-	-	0	0.000	0.0
	2	4	4	4	4	2	4	22	0.917	91.7
	3	-	-	-	-	2	-	2	0.083	8.3
P3T6	1	-	-	1	-	1	-	2	0.083	8.3
	2	4	4	3	4	1	4	20	0.833	83.3
	3	-	-	-	-	2	-	2	0.083	8.3
P4T6	1	2	-	-	3	3	1	9	0.375	37.5
	2	2	4	4	-	-	-	10	0.417	41.7
	3	-	-	-	1	1	3	5	0.208	20.8
P5T6	1	4	1	-	1	3	2	11	0.458	45.8
	2	-	2	3	-	-	-	5	0.208	20.8
	3	-	1	1	3	1	2	8	0.333	33.3
P6T6	1	4	1	-	-	3	3	11	0.458	45.8
	2	-	1	3	-	-	-	4	0.167	16.7
	3	-	2	1	4	1	1	9	0.375	37.5
P7T6	1	4	3	-	1	4	4	16	0.667	66.7
	2	-	1	4	-	-	-	5	0.208	20.8
	3	-	-	-	3	-	-	3	0.125	12.5

RESPONSE `1´ denotes "pitch" based grouping
RESPONSE `2´ denotes "timbre" based grouping
RESPONSE `3´ indicates uncertainty.

Figure 6.2.6 Graph showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 2.6, in which the "timbral-distance" between the sequence tones was kept fixed to be (T2-T6), while the pitch interval varied over the trials in a block. Response 1 indicates grouping by pitch, Response 2 by timbre, and Response 3 denotes uncertainty.

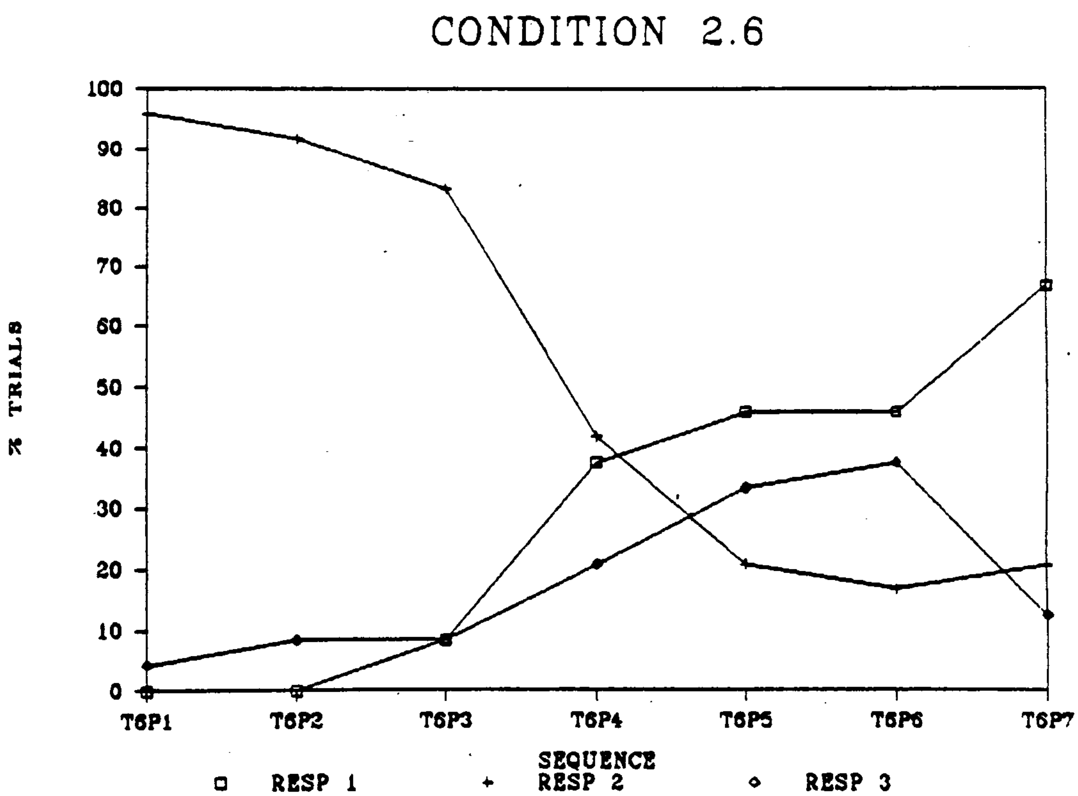


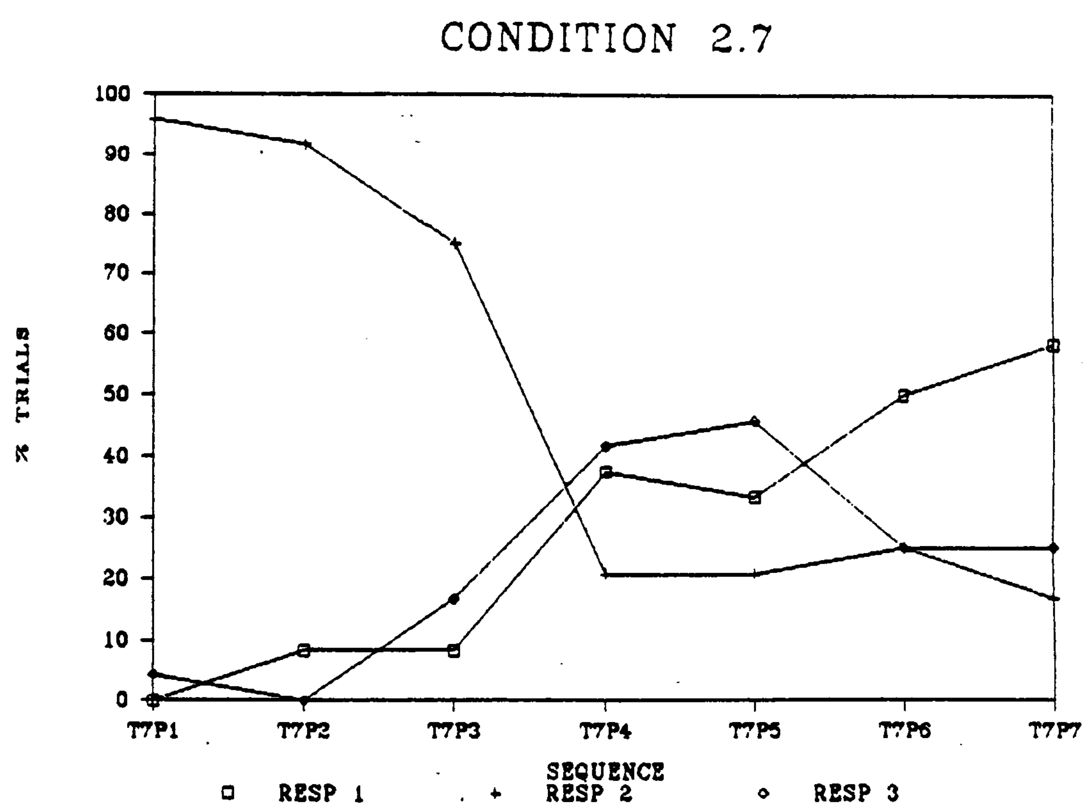
Table 6.2.7 Condition 2.7

PITCH CHANGE OVER BLOCK, GIVEN TIMBRE T7

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
		SUBJECTS								
P1T7	1	-	-	-	-	-	-	0	0.000	0.0
	2	4	4	4	4	3	4	23	0.958	95.8
	3	-	-	-	-	1	-	1	0.042	4.2
P2T7	1	-	-	-	-	2	-	2	0.083	8.3
	2	4	4	4	4	2	4	22	0.917	91.7
	3	-	-	-	-	-	-	0	0.000	0.0
P3T7	1	-	1	-	-	1	-	2	0.083	8.3
	2	4	2	4	4	-	4	18	0.750	75.0
	3	-	1	-	-	3	-	4	0.167	16.7
P4T7	1	3	1	-	2	1	2	9	0.375	37.5
	2	-	1	4	-	-	-	5	0.208	20.8
	3	1	2	-	2	3	2	10	0.417	41.7
P5T7	1	1	-	-	1	4	2	8	0.333	33.3
	2	-	1	4	-	-	-	5	0.208	20.8
	3	3	3	-	3	-	2	11	0.458	45.8
P6T7	1	2	-	-	3	4	3	12	0.500	50.0
	2	-	2	4	-	-	-	6	0.250	25.0
	3	2	2	-	1	-	1	6	0.250	25.0
P7T7	1	2	3	-	4	4	1	14	0.583	58.3
	2	-	-	4	-	-	-	4	0.167	16.7
	3	2	1	-	-	-	3	6	0.250	25.0

RESPONSE \`1´ denotes "pitch" based grouping
RESPONSE \`2´ denotes "timbre" based grouping
RESPONSE \`3´ indicates uncertainty.

Figure 6.2.7 Graph showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 2.7, in which the "timbral-distance" between the sequence tones was kept fixed to be (T2-T7), while the pitch interval varied over the trials in a block. Response 1 indicates grouping by pitch, Response 2 by timbre, and Response 3 denotes uncertainty.



Conditions 2.6 and 2.7 (figures 6.2.6, 6.2.7 ,tables 6.2.6, 6.2.7), also showed a similar trend, with the timbre cue being traded off for the pitch cue at the octave interval (P4). However, in these two cases, the pitch grouping was not as "salient" as for the previous conditions, this fact being manifested by a lower number of `1´ responses, and an increased amount of `3´ ("uncertain") responses.

For condition 3, in which changes in both pitch and timbre could occur from trial to trial, the response varies greatly, depending upon the way the timbre and pitch changes are mixed within a block.

Thus, for the condition 3.1 (figure 6.3.1 table 6.3.1), in which the pitch intervals and timbral distances are proportional, in that large pitch intervals are correlated with larger timbral distances, the plotted data show an almost unilateral tendency to group the sounds by pitch (between 60 % and 100 % of the trials) for pitch changes of greater than and equal to a third (P2), accompanied by timbral distances from T2 (i.e zero distance) through T7 (maximum). The sequence T1P1, having a zero pitch interval P1 (unison), accompanied by the timbre T1 contrasted with the anchor timbre T2, however seemed to be grouped by timbre (response `2´), for 71% of the trials on which it had been presented, as indeed it was grouped on previous occasions under condition 2.1.

Table 6.3.1 Condition 3.1

BOTH TIMBRE AND PITCH VARIED OVER BLOCK (PROPORTIONAL)

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
T1P1	1	-	-	1	-	4	-	5	0.208	20.8
	2	3	4	2	4	-	4	17	0.708	70.8
	3	1	-	1	-	-	-	2	0.083	8.3
T2P2	1	4	4	4	4	4	4	24	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T3P3	1	4	1	-	4	4	2	15	0.625	62.5
	2	-	-	2	-	-	-	2	0.083	8.3
	3	-	3	2	-	-	2	7	0.292	29.2
T4P4	1	4	3	4	4	3	4	22	0.917	91.7
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	1	-	-	1	-	2	0.083	8.3
T5P5	1	4	2	4	3	3	4	20	0.833	83.3
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	2	-	1	1	-	4	0.167	16.7
T6P6	1	3	1	3	4	2	3	16	0.667	66.7
	2	-	1	-	-	-	-	1	0.042	4.2
	3	1	2	1	-	2	1	7	0.292	29.2
T7P7	1	4	3	2	4	4	2	19	0.792	79.2
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	1	2	-	-	2	5	0.208	20.8

RESPONSE `1´ denotes "pitch" based grouping
RESPONSE `2´ denotes "timbre" based grouping
RESPONSE `3´ indicates uncertainty.

Figure 6.3.1 Graph showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 3.1 , in which the pitch interval between the tones and the timbral distance were both varied in conjunction with each other, with large pitch intervals being articulated by larger timbral distances. Response 1 indicates grouping by pitch, Response 2 by timbre, and Response 3 denotes uncertainty

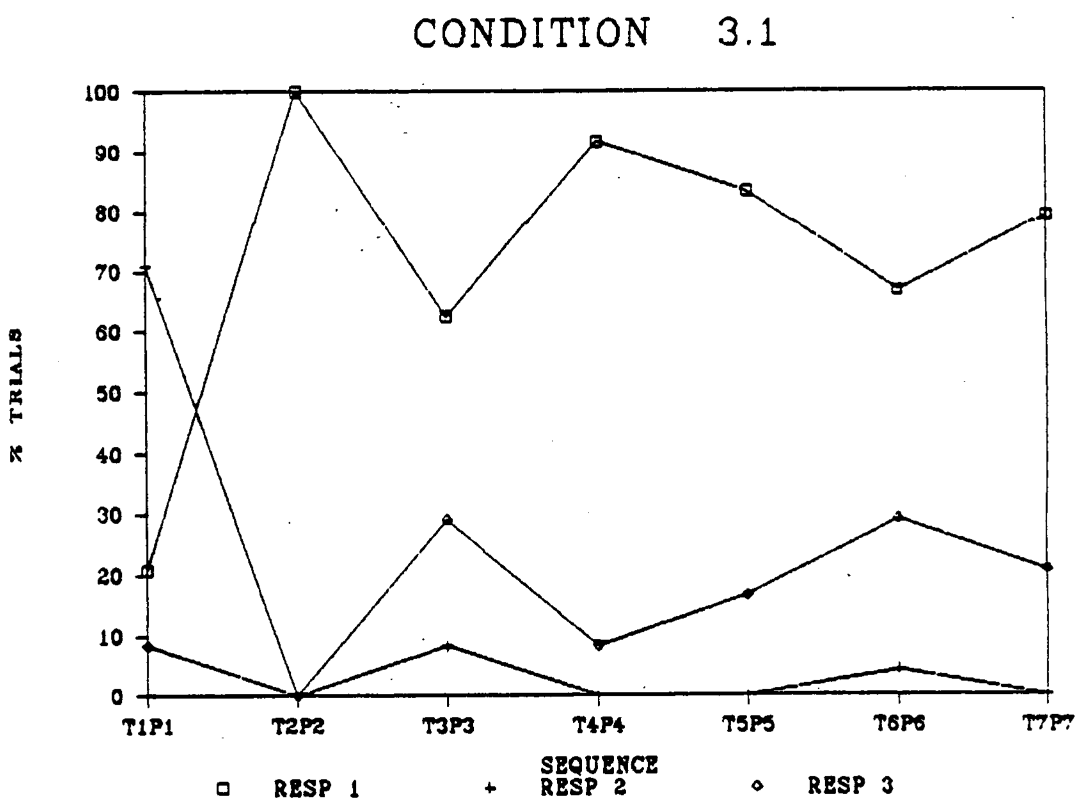


Table 6.3.2 Condition 3.2

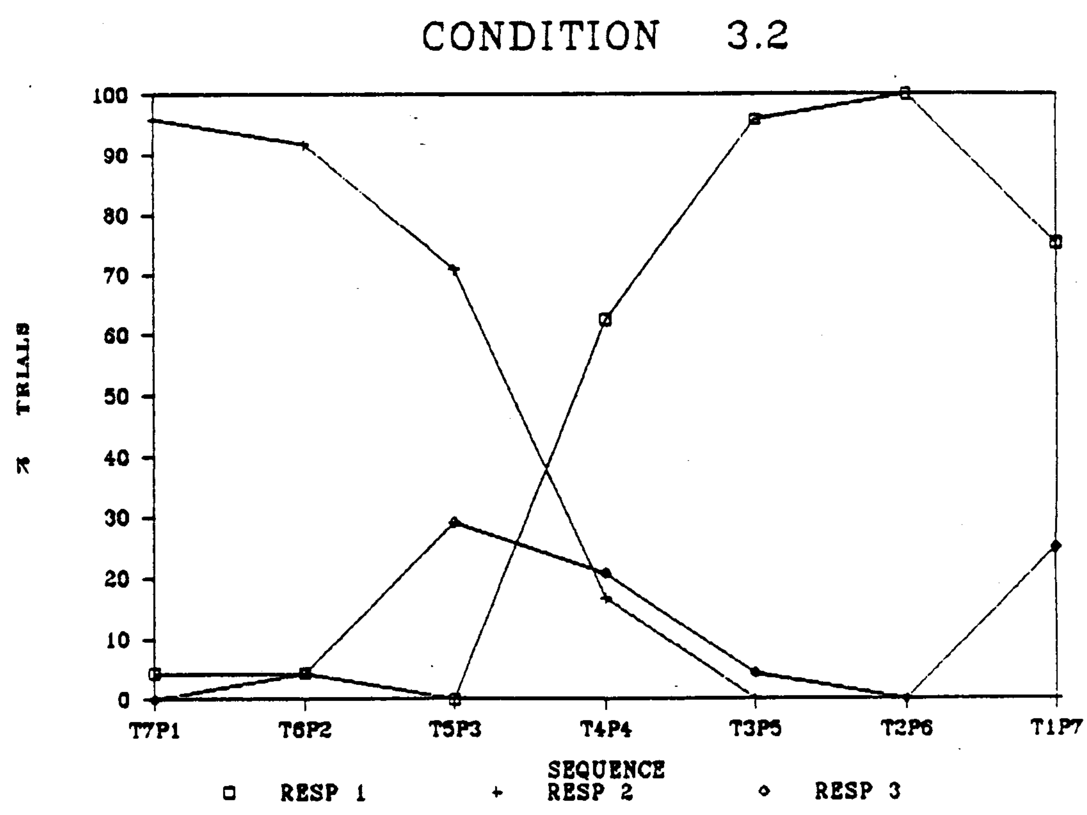
BOTH TIMBRE AND PITCH VARIED OVER BLOCK (INVERSE SET)

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
T7P1	1	-	-	1	-	-	-	1	0.042	4.2
	2	4	4	3	4	4	4	23	0.958	95.8
	3	-	-	-	-	-	-	0	0.000	0.0
T6P2	1	-	-	1	-	-	-	1	0.042	4.2
	2	4	4	3	4	3	4	22	0.917	91.7
	3	-	-	-	-	1	-	1	0.042	4.2
T5P3	1	-	-	-	-	-	-	0	0.000	0.0
	2	1	4	2	4	2	4	17	0.708	70.8
	3	3	-	2	-	2	-	7	0.292	29.2
T4P4	1	4	-	-	4	3	4	15	0.625	62.5
	2	-	2	2	-	-	-	4	0.167	16.7
	3	-	2	2	-	1	-	5	0.208	20.8
T3P5	1	4	3	4	4	4	4	23	0.958	95.8
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	1	-	-	-	-	1	0.042	4.2
T2P6	1	4	4	4	4	4	4	24	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T1P7	1	4	4	3	-	4	3	18	0.750	75.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	1	4	-	1	6	0.250	25.0

RESPONSE `1´ denotes "pitch" based grouping
RESPONSE `2´ denotes "timbre" based grouping
RESPONSE `3´ indicates uncertainty.

Figure 6.3.2 Graph showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 3.2, in which the pitch interval between the tones was varied along with the timbral distances, but in an inverse relationship, so that large pitch intervals were articulated by small timbral distances, and vice versa.



For condition 3.2 (figure 6.3.2,table 6.3.2), in which the pitch and timbre intervals within a sequence varied inversely with respect to each other, so that small pitch intervals were articulated by tones separated by large timbral distances, the data show the interesting combination of the previous two conditions.

From condition 1, the timbral distance T4 had been observed to be the transition point where the grouping scheme underwent an abrupt change from being pitch based to being timbre based. In condition 2, the pitch interval P4 (octave) had similarly been seen to be a transition pitch, beyond which the grouping strategy changed from being timbre based to being pitch based. Since condition 3.2 represents a superposition of these two conditions, the graph reveals the interesting interaction of the two conditions, with timbre being the dominant cue for timbral distances less than T4, and pitch being the dominant cue for pitch intervals greater than an octave (P4). The sequence represented by T4P4 thus seems to be an "in-between" transition sequence, the parameters of which lie between values that begin to encourage grouping of the sequence into subgroups ; on the basis of timbre, if given a slight increase in timbral distance, and on the basis of pitch, given a larger pitch interval.

In ~~Condition 4~~, every sequence of the stimulus set was presented only once within a block, with the order of presentation of the sequences randomized. This was basically done to determine if the response made by a subject to a particular sequence was "true", in the sense that the subject would always react to that sequence in the same way, regardless of the context of other sequences within a block, or if infact the number of times that a sequence appeared in a block caused learning of a type of pattern, given the context of other such patterns within a block, and therefore biased the response in some way. Two such blocks were presented to the subjects, and the responses averaged over blocks, and over subjects, to obtain the most general label that was used as a response to a particular sequence. The percent of trials on which a particular type of response was used, have been plotted against the "type" of sequence, to give a clue to the relationship between sequence structure, and imposed grouping, under this difficult condition.

In general, the graphs 6.4.1 and 6.4.2 plotted for the pitch intervals equivalent to the unison and a third respectively, show almost the same trend as the corresponding graphs drawn for conditions 1.1 and 1.2, though the degree of uncertainty is increased, not surprisingly, for the full set condition.

Table 6.4.1 Condition 4

FULLSET (ALL SEQUENCES PRESENTED)

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
T1P1	1	-	-	-	-	2	-	2	0.167	16.7
	2	-	2	-	2	-	1	5	0.417	41.7
	3	2	-	2	-	-	1	5	0.417	41.7
T2P1	1	2	2	2	2	2	2	12	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T3P1	1	-	-	1	-	-	1	2	0.167	16.7
	2	-	2	-	1	2	-	5	0.417	41.7
	3	2	-	1	1	-	1	5	0.417	41.7
T4P1	1	-	-	-	-	-	-	0	0.000	0.0
	2	2	2	1	2	1	2	10	0.833	83.3
	3	-	-	1	-	1	-	2	0.167	16.7
T5P1	1	-	-	1	-	-	-	1	0.083	8.3
	2	2	2	1	2	2	2	11	0.917	91.7
	3	-	-	-	-	-	-	0	0.000	0.0
T6P1	1	-	-	-	-	-	-	0	0.000	0.0
	2	2	2	-	2	1	2	9	0.750	75.0
	3	-	-	2	-	1	-	3	0.250	25.0
T7P1	1	-	-	1	-	-	-	1	0.083	8.3
	2	2	2	1	2	-	2	9	0.750	75.0
	3	-	-	-	-	2	-	2	0.167	16.7

RESPONSE \`1´ denotes "pitch" based grouping
RESPONSE \`2´ denotes "timbre" based grouping
RESPONSE \`3´ indicates uncertainty.

Figure 6.4.1 Graphs showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 4, in which the fullset of sequences was presented in a single block. The subset of responses for a pitch interval of a unison are shown.

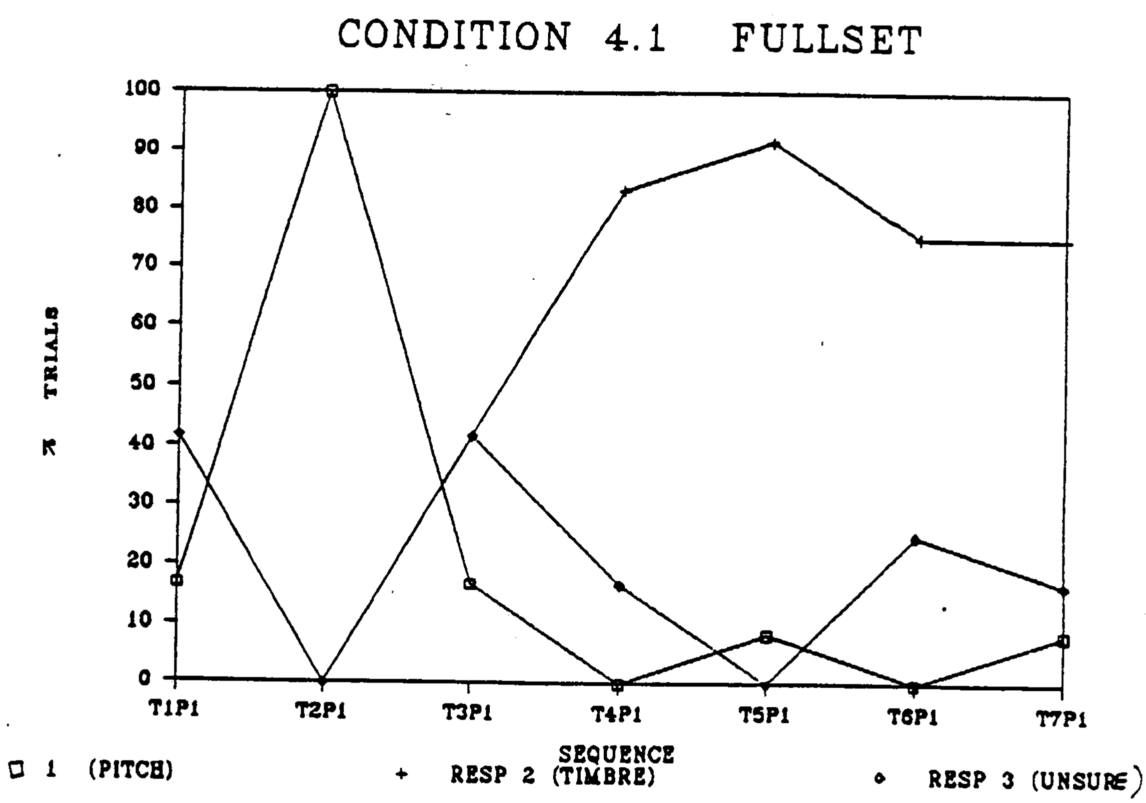


Table 6.4.2 Condition 4 (continued)

FULLSET

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
T1P2	1	-	-	2	-	1	-	3	0.250	25.0
	2	1	2	-	2	-	2	7	0.583	58.3
	3	1	-	-	-	1	-	2	0.167	16.7
T2P2	1	2	2	2	2	2	2	12	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T3P2	1	-	-	2	2	2	1	7	0.583	58.3
	2	-	1	-	-	-	-	1	0.083	8.3
	3	2	1	-	-	-	1	4	0.333	33.3
T4P2	1	-	-	2	-	-	-	2	0.167	16.7
	2	2	1	-	1	2	2	8	0.667	66.7
	3	-	1	-	1	-	-	2	0.167	16.7
T5P2	1	-	-	1	-	-	-	1	0.083	8.3
	2	2	2	-	2	2	2	10	0.833	83.3
	3	-	-	1	-	-	-	1	0.083	8.3
T6P2	1	-	-	1	-	-	-	1	0.083	8.3
	2	2	2	1	2	1	2	10	0.833	83.3
	3	-	-	-	-	1	-	1	0.083	8.3
T7P2	1	-	-	1	-	-	-	1	0.083	8.3
	2	2	2	-	2	1	2	9	0.750	75.0
	3	-	-	1	-	1	-	2	0.167	16.7

RESPONSE `1´ denotes "pitch" based grouping
RESPONSE `2´ denotes "timbre" based grouping
RESPONSE `3´ indicates uncertainty.

Figure 6.4.2 Graphs showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 4, in which the fullset of sequences was presented in a single block. The subset of responses for a pitch interval of a third are shown.

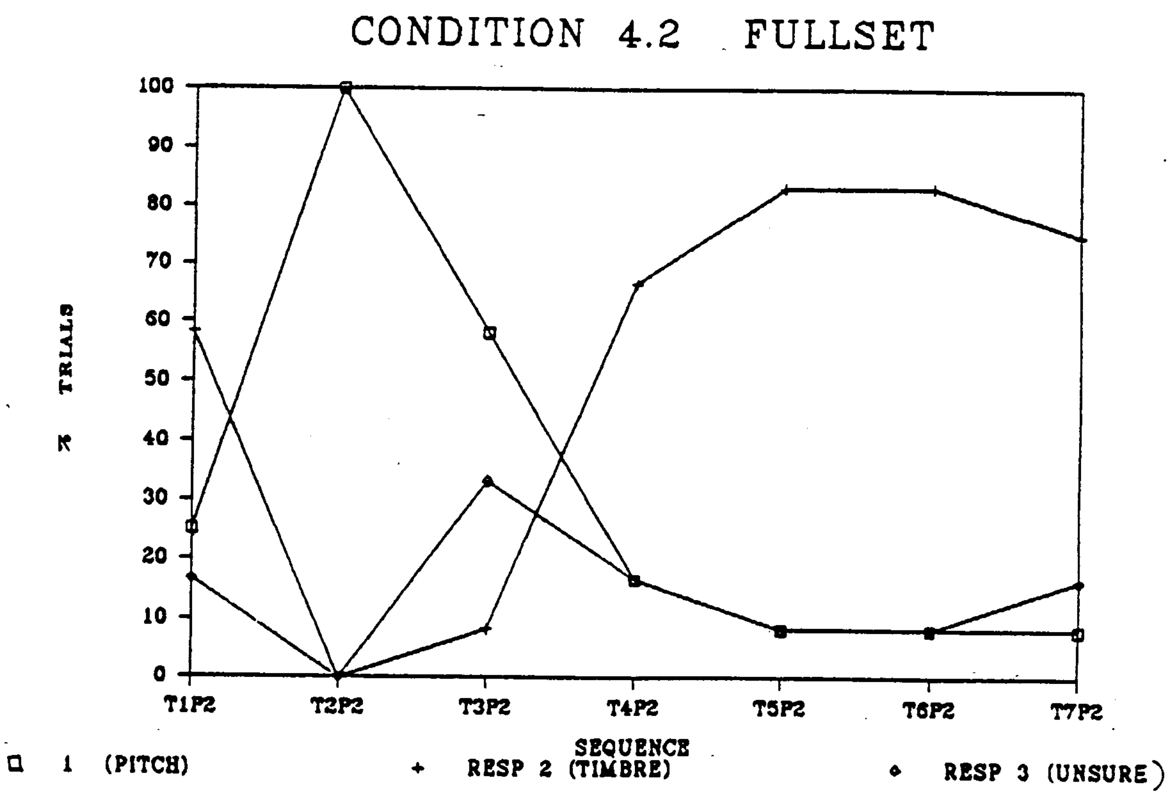


Table 6.4.3 Condition 4 (continued)

FULLSET

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
T1P3	1	-	-	2	-	2	-	4	0.333	33.3
	2	1	2	-	1	-	2	6	0.500	50.0
	3	1	-	-	1	-	-	2	0.167	16.7
T2P3	1	2	2	2	2	2	2	12	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T3P3	1	2	2	2	2	2	2	12	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T4P3	1	2	1	2	1	2	-	8	0.667	66.7
	2	-	1	-	-	-	2	3	0.250	25.0
	3	-	-	-	1	-	-	1	0.083	8.3
T5P3	1	-	1	1	-	1	-	3	0.250	25.0
	2	2	1	-	2	1	2	8	0.667	66.7
	3	-	-	1	-	-	-	1	0.083	8.3
T6P3	1	-	-	2	-	-	-	2	0.167	16.7
	2	2	2	-	2	1	2	9	0.750	75.0
	3	-	-	-	-	1	-	1	0.083	8.3
T7P3	1	-	-	1	-	1	-	2	0.167	16.7
	2	2	2	-	2	-	2	8	0.667	66.7
	3	-	-	1	-	1	-	2	0.167	16.7

RESPONSE \`1´ denotes "pitch" based grouping
RESPONSE \`2´ denotes "timbre" based grouping
RESPONSE \`3´ indicates uncertainty.

Figure 6.4.3 Graphs showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 4, in which the fullset of sequences was presented in a single block. The subset of responses for a pitch interval of a fifth are shown.

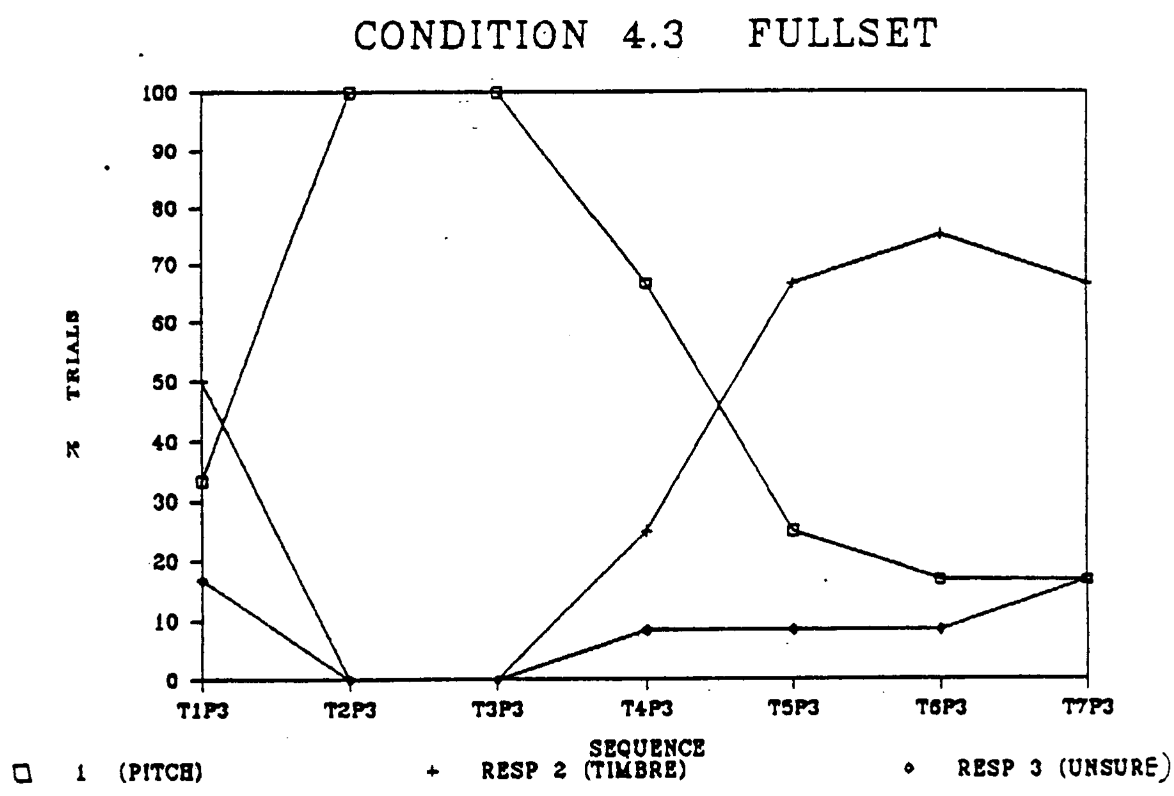


Table 6.4.4 Condition 4 (continued)

FULLSET

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
T1P4	1	2	2	2	-	2	1	9	0.750	75.0
	2	-	-	-	1	-	-	1	0.083	8.3
	3	-	-	-	1	-	1	2	0.167	16.7
T2P4	1	2	2	2	2	2	2	12	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T3P4	1	2	2	2	2	2	2	12	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T4P4	1	2	2	2	2	2	2	12	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T5P4	1	1	1	2	2	2	1	9	0.750	75.0
	2	-	1	-	-	-	-	1	0.083	8.3
	3	1	-	-	-	-	1	2	0.167	16.7
T6P4	1	1	1	2	2	-	1	7	0.583	58.3
	2	-	-	-	-	-	-	0	0.000	0.0
	3	1	1	-	-	2	1	5	0.417	41.7
T7P4	1	2	-	2	2	-	1	7	0.583	58.3
	2	-	2	-	-	-	1	3	0.250	25.0
	3	-	-	-	-	2	-	2	0.167	16.7

RESPONSE \`1´ denotes "pitch" based grouping
RESPONSE \`2´ denotes "timbre" based grouping
RESPONSE \`3´ indicates uncertainty.

Figure 6.4.4 Graphs showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 4, in which the fullset of sequences was presented in a single block. The subset of responses for a pitch interval of an octave are shown.

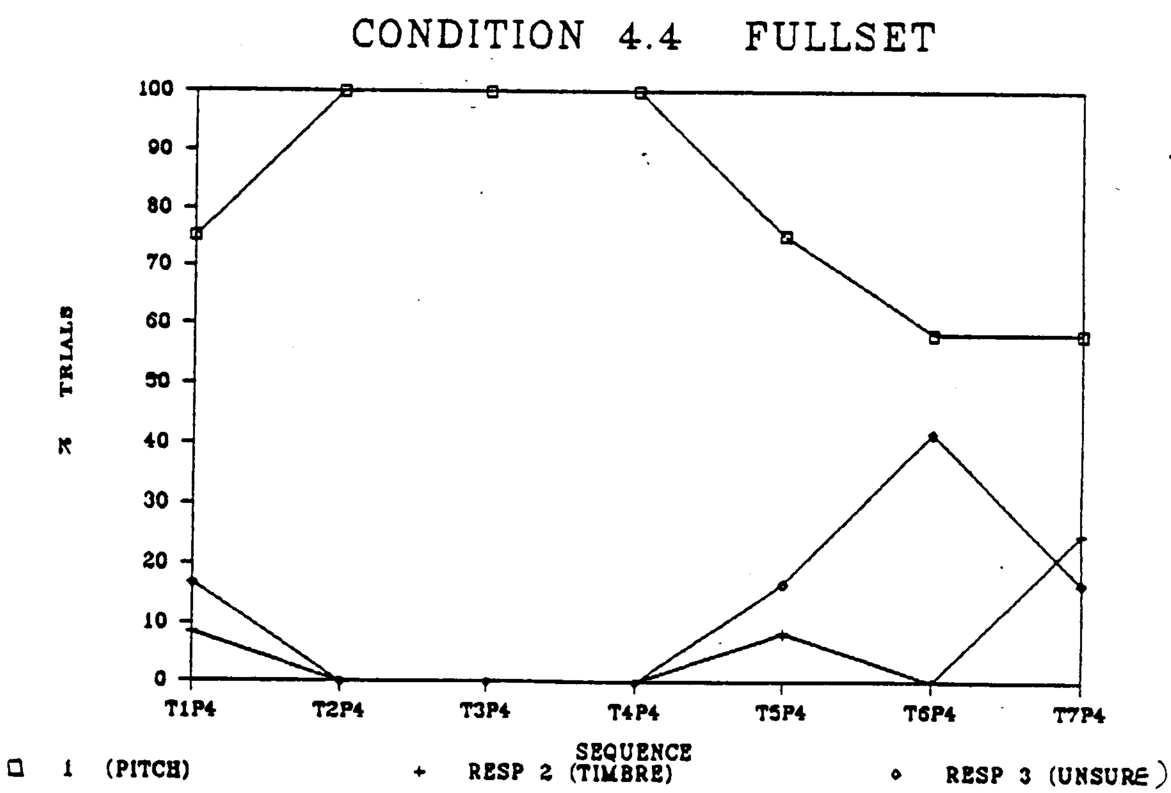


Table 6.4.5 Condition 4 (continued)

FULLSET

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
T1P5	1	2	2	1	-	2	1	8	0.667	66.7
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	1	-	-	1	2	0.167	16.7
T2P5	1	2	2	2	2	2	2	12	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T3P5	1	2	2	2	2	2	2	12	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T4P5	1	2	1	2	2	2	2	11	0.917	91.7
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	1	-	-	-	-	1	0.083	8.3
T5P5	1	2	1	2	2	2	2	11	0.917	91.7
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	1	-	-	-	-	1	0.083	8.3
T6P5	1	1	2	2	2	2	2	11	0.917	91.7
	2	-	-	-	-	-	-	0	0.000	0.0
	3	1	-	-	-	-	-	1	0.083	8.3
T7P5	1	2	1	2	1	2	1	9	0.750	75.0
	2	-	1	-	-	-	-	1	0.083	8.3
	3	-	-	-	1	-	1	2	0.167	16.7

RESPONSE `1´ denotes "pitch" based grouping
RESPONSE `2´ denotes "timbre" based grouping
RESPONSE `3´ indicates uncertainty.

Figure 6.4.5 Graphs showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 4, in which the fullset of sequences was presented in a single block. The subset of responses for a pitch interval of a tenth are shown.

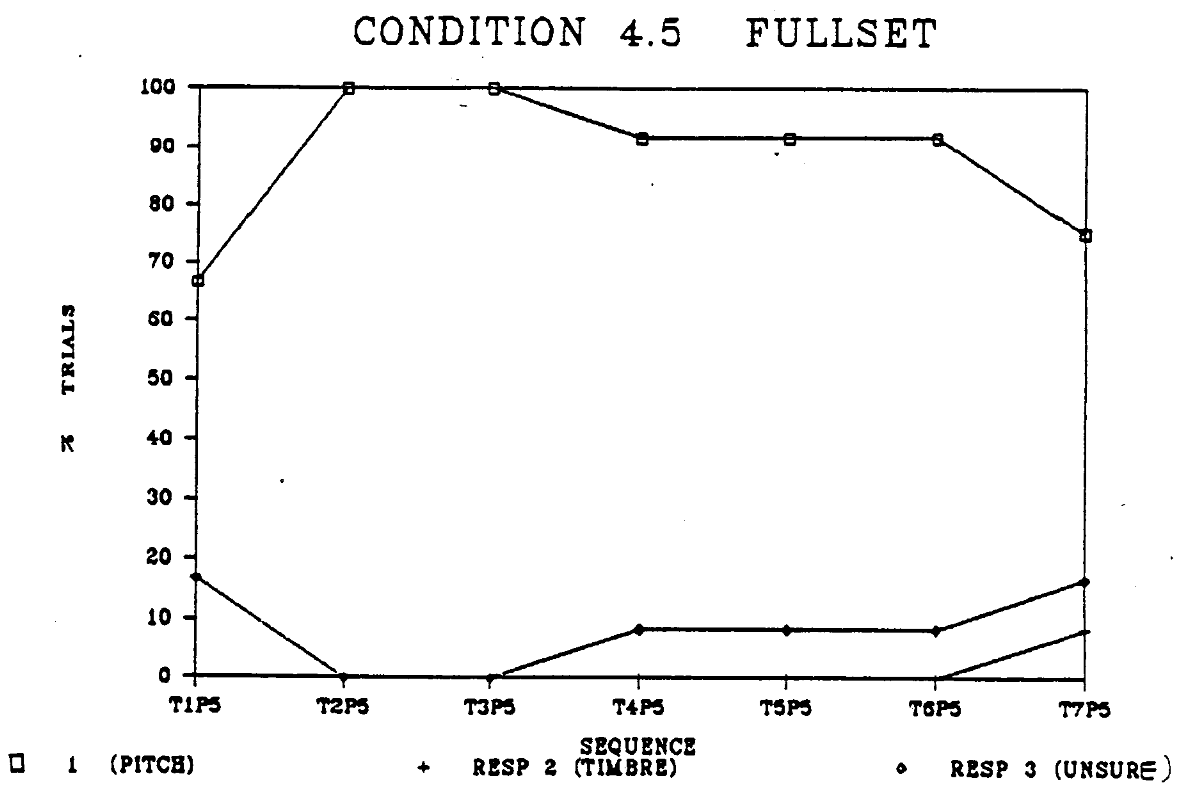


Table 6.4.6 Condition 4 (continued)

FULLSET

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
T1P6	1	2	2	2	-	2	1	9	0.750	75.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	2	-	1	3	0.250	25.0
T2P6	1	2	2	2	2	2	2	12	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T3P6	1	2	2	2	2	2	2	12	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T4P6	1	2	1	2	2	2	2	11	0.917	91.7
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	1	-	-	-	-	1	0.083	8.3
T5P6	1	2	2	2	2	2	2	12	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T6P6	1	2	-	2	2	2	2	10	0.833	83.3
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	2	-	-	-	-	2	0.167	16.7
T7P6	1	1	2	2	1	2	2	10	0.833	83.3
	2	1	-	-		-	-	1	0.083	8.3
	3	-	-	-	1	-	-	1	0.083	8.3

RESPONSE \`1´ denotes "pitch" based grouping
RESPONSE \`2´ denotes "timbre" based grouping
RESPONSE \`3´ indicates uncertainty.

Figure 6.4.6 Graphs showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 4, in which the fullset of sequences was presented in a single block. The subset of responses for a pitch interval of a twelfth are shown.

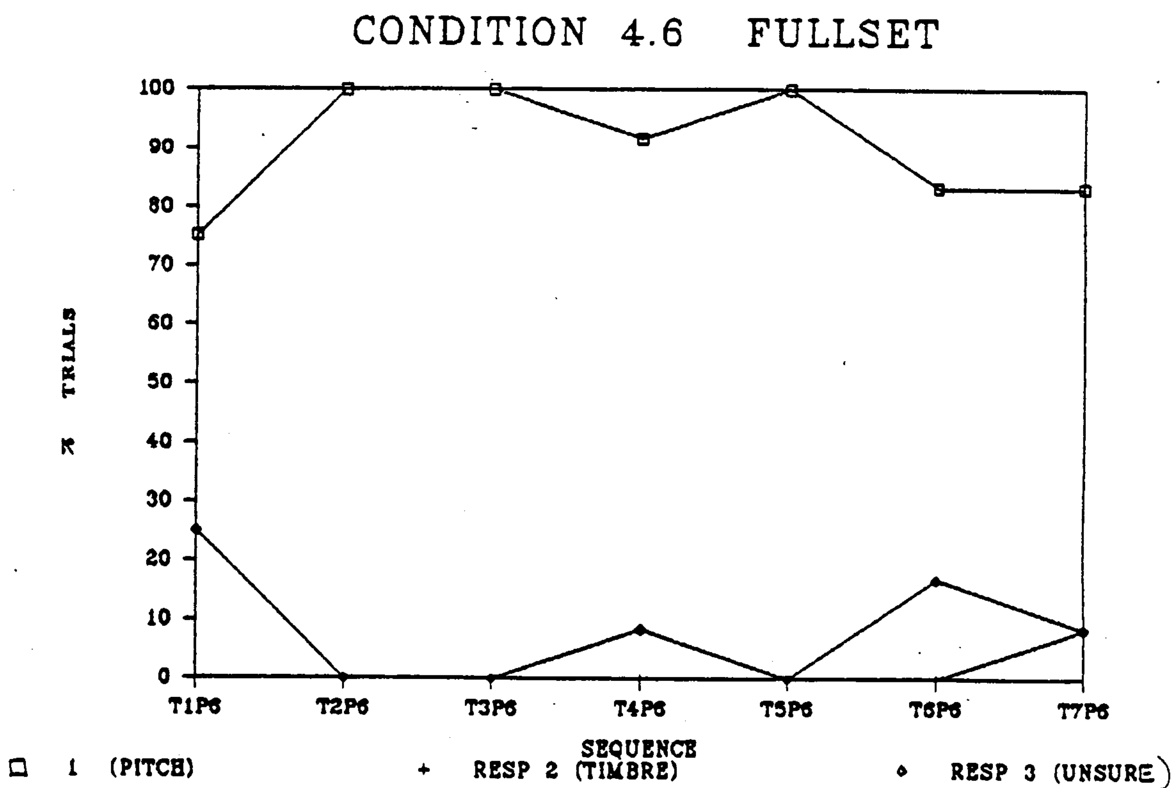


Table 6.4.7 Condition 4 (continued)

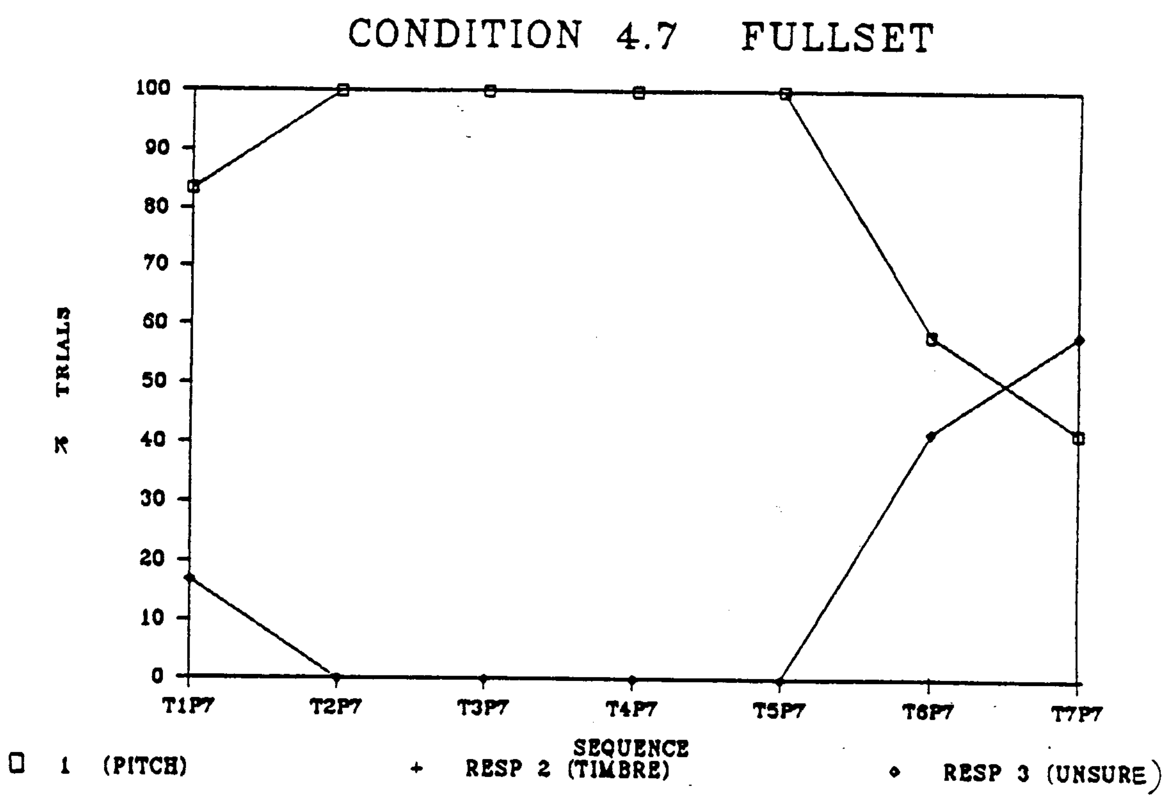
FULLSET

SUBJECTS

SEQUENCE	RESP	DG	MJ	JA	PG	WK	PS	TOTAL	AVG	%
T1P7	1	2	2	2	-	2	2	10	0.833	83.3
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	2	-	-	2	0.167	16.7
T2P7	1	2	2	2	2	2	2	12	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T3P7	1	2	2	2	2	2	2	12	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T4P7	1	2	2	2	2	2	2	12	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T5P7	1	2	2	2	2	2	2	12	1.000	100.0
	2	-	-	-	-	-	-	0	0.000	0.0
	3	-	-	-	-	-	-	0	0.000	0.0
T6P7	1	1	1	2	2	1	-	7	0.583	58.3
	2	-	-	-	-	-	-	0	0.000	0.0
	3	1	1	-	-	1	2	5	0.417	41.7
T7P7	1	1	1	2	-	1	-	5	0.417	41.7
	2	-	-	-	-	-	-	0	0.000	0.0
	3	1	1	-	2	1	2	7	0.583	58.3

RESPONSE \`1´ denotes "pitch" based grouping
RESPONSE \`2´ denotes "timbre" based grouping
RESPONSE \`3´ indicates uncertainty.

Figure 6.4.7 Graphs showing the proportion of responses (expressed as % of total trials) used to label the sequences shown along the abscissa, for condition 4, in which the fullset of sequences was presented in a single block. The subset of responses for a pitch interval of two octaves are shown.



Graph 4.3 however, while following the main features of the corresponding graph 1.3, shows a reduced degree of uncertainty at the previously designated "in-between" timbre T4. It thus seems that this timbre was perceived as different enough from T2 given only a single trial for comparison, but was confused when given repeated occurrences. Graphs 6.4.4, 6.4.5, and 6.4.6 however, exhibit greater uncertainty in response for these high pitched sequences, and therefore an increased tendency to group by pitch. Graph 6.4.7 follows this same trend, but now shows the absolute lack of the use of timbre (response `2´), as a cue for grouping.

To summarize the results observed in the graphs, and to provide a quick visual reference of the type of grouping picked in general by an average listener, a matrix displaying the entire sequence set is displayed in table 6.5. The abbreviated labels used for sequences, and the majority response for each sequence, are given in each cell of the matrix. The letter `T´ denotes the fact that the majority grouping chosen for that particular sequence was one based on "timbre" as it has been defined here. Similarly, the letter `P´ has been used to designate a majority grouping based on the pitch cue. Cases in which there was no majority of response label, or if the ambiguous label `3´ (unsure) dominated, have been assigned a dashed label to indicate the lack of any dominant segregation cue for that sequence.

Table 6.5 SEQUENCE MATRIX

TIMBRAL DISTANCE vs. PITCH INTERVAL WITHIN A SEQUENCE

	T2-T1	T2-T2	T2-T3	T2-T4	T2-T5	T2-T6	T2-T7
P1-P1	(T1P1) T	(T2P1) P	(T3P1) -	(T4P1) T	(T5P1) T	(T6P1) T	(T7P1) T
P1-P2	(T1P2) T	(T2P2) P	(T3P2) -	(T4P2) T	(T5P2) T	(T6P2) T	(T7P2) T
P1-P3	(T1P3) T	(T2P3) P	(T3P3) -	(T4P3) -	(T5P3) T	(T6P3) T	(T7P3) T
P1-P4	(T1P4) -	(T2P4) P	(T3P4) P	(T4P4) P	(T5P4) P	(T6P4) -	(T7P4) -
P1-P5	(T1P5) P	(T2P5) P	(T3P5) P	(T4P5) P	(T5P5) -	(T6P5) -	(T7P5) -
P1-P6	(T1P6) P	(T2P6) P	(T3P6) P	(T4P6) P	(T5P6) P	(T6P6) P	(T7P6) P
P1-P7	(T1P7) P	(T2P7) P	(T3P7) P	(T4P7) P	(T5P7) P	(T6P7) -	(T7P7) -

To decide if pitch or timbre was the dominant mode of grouping, a very simple measure has been adopted. Each sequence has been viewed under all the conditions in which it was presented. If a particular type of response label was consistently used on more than 50 % of the trials in every case that the sequence occurred, that response label has been inferred as having the majority, and the corresponding mode of grouping (pitch or timbre), has been declared as dominant.

While this overall response measure is indeed "rough", and a lot of the finer details of competition amongst the cues are sacrificed, it does provide a convenient way of viewing the entire sequence set as a mapping between stimulus and response.

A survey of the sequence matrix reveals the same basic features of the stimulus set as have been discussed above. The second column of the matrix , displays the iso-timbral (T2-T2) sequences that were grouped by pitch a 100 % of the time. The sequences lying below the row demarcated by the pitch interval P1-P4 show either uncertainty in response, or grouping based on pitch. The dominant timbral grouping is seen to occupy the top right region of the matrix, demarcated by the timbre T4, and the pitch interval P4, as was noted before. The seemingly "odd" observation that the timbre T1 which is separated by a very small timbral "distance" from T2, always resulted in a timbre based grouping for pitch intervals less than an octave (P4) is corroborated here, as shown in the top left column of the matrix. On the other side, the timbre T3, which is also separated by a very small "distance" from T2, is seen to make sequence grouping very non-commital for pitch intervals less than the octave (P4).

A detailed analysis of these "first order" observations from the sequence matrix, is undertaken in the next chapter, and the observations related to the physical properties of the stimulus set.

7. CONCLUSIONS

7.1 Results Viewed in Terms of Stimulus Structure

Given the data and the general trend exhibited by the average listener in grouping different sequences on the basis of relations perceived between the sequence events, it seems to be in order to now drop the use of the obscure terms "timbral distance", and "pitch interval", and view the sequences in terms of the finer, more definitive attributes of the individual events.

Since the so called "timbres" were actually four component complex tones with equal amplitude spectra, each "timbre" may be considered to be a band of spectral energy localized in a "harmonic region" along the frequency axis. Shifting the center of this "band" of spectral energy, results in the transposition of the spectral "center of gravity" of the tone complex, resulting in a difference in the quality or the timbre perceived, akin to the "brightness" mentioned by Lichte (1941), and the "sharpness" studied by von Bismarck (1974), as described in section 2.10. However, rather than going into the semantic differentiation of the "timbres" thus created, the explanation of the results is attempted in terms of the interaction between the harmonics of adjacent tones, due to proximity of the frequency region inhabited by them.

The pitch attribute of the sequence tones is interwoven by "default" with the timbral attribute as it is here defined, since the amplitudes and temporal features such as the duration, and temporal envelope of the tones were equalized in the synthesis of the tones, leaving only "frequency" as the variable dimension. Shifting the "band" of harmonics upward in frequency region in the creation of the "different" timbres, could thus lead to an increase in the perceived pitch "height" of the tones, as was referred to by McAdams (1979). The pitch "value" assigned to the tones is however assumed to depend only on the frequency difference separating the consecutive harmonics, regardless of their absolute position along the frequency scale. This "virtual" pitch thus corresponds to the frequency of the "missing" fundamental, in the case of timbres T2, T3, T4, T5, T6, and T7, and to the frequency of the fundamental (or the "spectral pitch" of the first harmonic),in the case of timbre T1.

Viewed in these terms, the individual sound "events" making up the various sequences, differ from each other (as far as their synthesis scheme goes), only in the width of the spacing between their spectral components, and in the overall placement of the equal amplitude spectrum along the frequency axis. While the relative width of the spacing between the spectral components determines the pitch relationship between sounds, the absolute frequencies of the particular harmonics used in the spectra of the different sounds, determines their "timbre". The grouping of the sequences may therefore be attributed to tradeoffs between the relative and absolute frequency parameters of the complex tones.

Keeping in mind that the "anchor" tone T2P1 which was the first tone in all the sequences, is constructed by the addition of the second, third, fourth, and fifth harmonics of the fundamental frequency corresponding to the pitch P1 (middle C = 262 Hz), it appears that a grouping based on timbre is the dominant percept for sequences in which the contrasted timbre of the second and fourth tones was one that had two or fewer harmonics in common with the timbre T2, and for which the contrasted pitch interval between the first pair of tones and the second pair was less than an octave. The timbre dominated region in which such sequences occurred, appears to be the upper right half of the sequence matrix, demarcated by the "timbral distance" T2-T4, and the "pitch interval" P1-P4.

While the timbres T4 and T5, belonging to this region, have respectively two and one harmonic overlapping with the harmonics of the anchored timbre T2, the "far" timbres T6 and T7 have no spectral overlap with the harmonics making up T2, given the same frequency separation between the harmonics.

For the pitch intervals corresponding to P1 (unison, C4 =262 Hz), P2 (third, E4 =330 Hz), and P3 (fifth, G4 =392 Hz), the harmonic spectra of these timbres (T4 through T7) seem to be far enough apart along the frequency axis, to cause them to segregate when placed adjacent to the reference timbre T2. For these "small" pitch intervals, the spacing between the harmonics of the first pair of tones, and the second pair of tones, is not too great. Since the pitch interval is less than an octave (P4), the spacing in the second pair is less than twice as wide as the spacing for the first pair. Given these conditions, where the adjacent timbres are far apart in frequency region, and alternate timbres occupy the same region with not too wide a difference in harmonic spacing, seems to cause the alternate timbres to be perceived as occupying a single "stream".

The perception of a timbre dominated grouping between the alternate tones of the sequences inhabiting this region of the sequence matrix, results in the loss of veridical perception of the serial order of the tones of the sequence. Instead of hearing two pairs of tones following each other in quick succession, with a change in pitch (if any), occurring between the second and third tones, two "canoning" melodic intervals superposed on each other and articulated by different timbres are perceived. This timbral segregation is confounded with the rhythm of the sequence, resulting in the halving of the "tempo", when the alternate notes (separated by twice the inter-event interval) are grouped together.

In contrast to the region just discussed, the lower portion of the sequence matrix, demarcated by the pitch interval P1-P4, (an octave), shows pitch to be the dominant cue in grouping, across the entire range of timbres employed. For these "far" pitch sequences, the bandwidth of the spectra defining the tones differs greatly between the first and second tone pairs. The much wider spacing of harmonics for the higher pitched second pair, pushes them higher and higher along the frequency axis, smearing the perception of the absolute frequency position of the spectra. The difference in timbre that is dependent on this positioning of the spectral envelope is thus obscured, and grouping by pitch dominates.

In the bottom-right section of this region, where both the timbral "distance", and the "pitch interval" are large, uncertainty of response often dominates. Considering that the sequences inhabiting this region are composed of sounds differing greatly in spectral bandwidth, and separated by large distances between the location of their spectra, it seems that the sounds become so far removed in the overall perceived pitch, that the last two sounds seem to belong together simply due to their "high" character, while the lower sounds of differing timbre, making up the first pair, are perceived as being different enough to not belong together. This confusion thus leads to the subjects' exploting the "unsure" category provided, in response to these "far pitch/far timbre" sequences.

The sequences T6P7 and T7P7, which occupy the highest frequency region in the whole set, were indeed expected to elicit the "unsure" response, since they contained physically distorted high frequency components, resulting from the limitations of the digital synthesis equipment. Perceptually, these sounds appeared to be very shrill and squeak-like and were therefore typically grouped as "high" (response `1´), or ambiguous (response `3´). The domination of uncertainty for these sequences thus mollifies their having been used despite acknowledgement of their distorted character.

The second column of the sequence matrix, representing the iso-timbral sequences T2-T2 displays a 100 % grouping based on pitch, for the entire range of pitch intervals used. This trivial result is actually quite informative, since it confirms the fact that the fixing of the spectral envelope along the frequency axis, results in the invariance of timbre (Slawson,1968, Plomp,1976). Since there is complete overlap of the harmonics for the adjacent tones making up the two pairs in these sequences, the sounds differ only in their "bandwidth". A difference in fundamental frequency is thus implied, and consequently "pitch" is the only cue available in initiating grouping.

On either side of this iso-timbral column, lie the "near" timbres T3 and T1, each having three components in common with T2. The sequences made of sounds separated by the near-timbre T3, show a consistently ambiguous response for pitch intervals less than an octave (P4). Since the "dominant" third, fourth, and fifth harmonics (Ritsma,1967) are common between the timbre T3, and the anchored timbre T2, it seems that being adjacent to each other in the sequence, they probably get "streamed" together. The timbral difference due to the different fourth component, thus becomes very difficult to perceive. Also, since the spectra of the adjacent sounds within a pair occupy nearly the same location along the frequency axis, they seem quite similar in timbre, and therefore appear to be difficult to categorize.

The timbre T1 on the other hand, also shares three components with the anchor timbre T2 (the second, third, and fourth harmonic of the fundamental of whichever pitch is being studied), yet it appears to elicit a unanimous grouping by "timbre" for pitch intervals less than an octave (P4) .

Close investigation of this apparently paradoxical result, reveals that timbre T1 is quite special, in that it actually contains the "fundamental" in its spectrum. None of the other timbres were designed to have energy at the fundamental, so that their timbral effects could be studied independently from the pitch aspect, by making timbre differences a function of "absolute frequency". Since changes in fundamental frequency would lead to changes in the pitch perceived, the fundamental frequency was preserved in timbres having the same pitch, in an "invisible" format. This was done by using harmonic multiples of the fundamental, the frequency spacing between which is the same as the frequency of the fudamental (as discussed in the "design" section of chapter 4). By varying the absolute frequency of the harmonics, while preserving their relative spacing, the timbral aspect of the sounds could therefore be studied in isolation of the pitch aspect, which was "accounted for" by equating the "virtual pitch" (Terhardt, 1979).

The percept arising due to the inclusion of the fundamental, was however very different from that of any other "timbre". Sequences containing T1 as the timbre of the second and fourth sounds were invariably perceived as "very different" by the listeners, and a strong connection perceived between the second and fourth notes. This connection was however described by the subjects to be "a low ostinato-like (melodic) interval in the background , accompanying a higher sequence of four notes".

The "comment" sheet provided to the listeners proved to be a great way of studying any concurrence in their perception of the sequence. To qualify what they thought was strange about sequences containing T1 for small pitch intervals, five of the listeners drew "pictures" of the sequence as they felt they had perceived it. The pictures drawn by different subjects at different times, interestingly enough, turned out to be almost exact replicas of each other. A typical picture drawn for these sequences is illustrated in figure7.1. The sixth listener refrained from drawing pictures, but commented that "the pitch seemed to drop an octave or something..." and attributed her response to such sequences as being probably due to her "bias for the lows", since she was a double bass player !

The perception of sequences containing sounds built on the first harmonic (the fundamental), juxtaposed next to sounds that were built on the second harmonic (T2), thus appeared to be perceived as two overlapping sequences (or "streams"), a higher four tone sequence occurring at double the tempo of a lower two tone sequence, that was basically a two note melodic interval. Since the low note accompanying the first pair seems to "reach out" to connect with its counterpart in the second pair, the veridical perception of temporal order is disrupted, and the slower, low tone pair appears to be occurring simultaneously, syncopated with the higher set of four tones.

The fact that six tones were perceived in total, given a four tone sequence, seemed to be a consequence of the special nature of such sequences, analogous to the perception

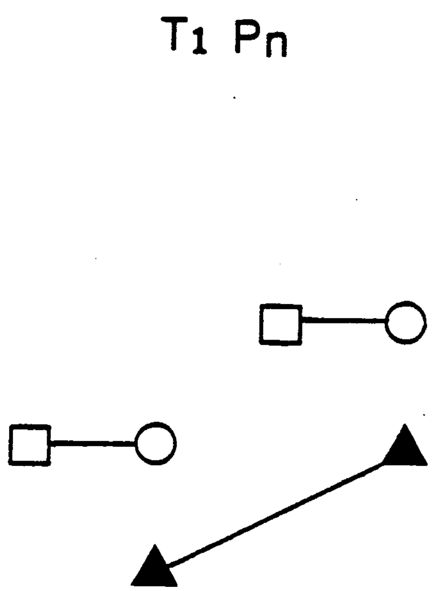


Figure 7.1 Illustration of the type of "picture" most often drawn by subjects in response to sequences that contained the timbre T1 for alternate tones of a sequence in which the pitch interval was not too great.

of three tones, two high tones, occurring simultaneously with one tone an octave below, when given a pair of complex tones an octave apart in pitch, and 75 msec. apart in time, as was observed in the "timbre-split" pilot experiment described earlier.

The timbre based grouping observed for sequences employing T1 in a near pitch context, thus seems to have been based on the streaming together of the harmonics 2, 3, and 4 of timbre T1 and timbre T2, and the streaming together of the fundamentals (first harmonics) of the two occurrences of timbre T1.

The fact that such streaming was possible but not perceived between the repeated occurrences of the fifth harmonic in T2, seems to indicate, that the "better resolved" low frequencies of a complex tone stimulus (Schouten, Ritsma, and Cardozo, 1962) dominate in providing cues to form groups based on frequency difference.

Some evidence to support this speculation was obtained on the analysis of some of the timbrally grouped sequences of the stimulus set on a more "molecular" level, or as is proposed here, on a finer "microsonic" level. Rather than going into a detailed textual account of this evidence, it has been provided in pictorial form (figures 7.2, 7.3, 7.4)

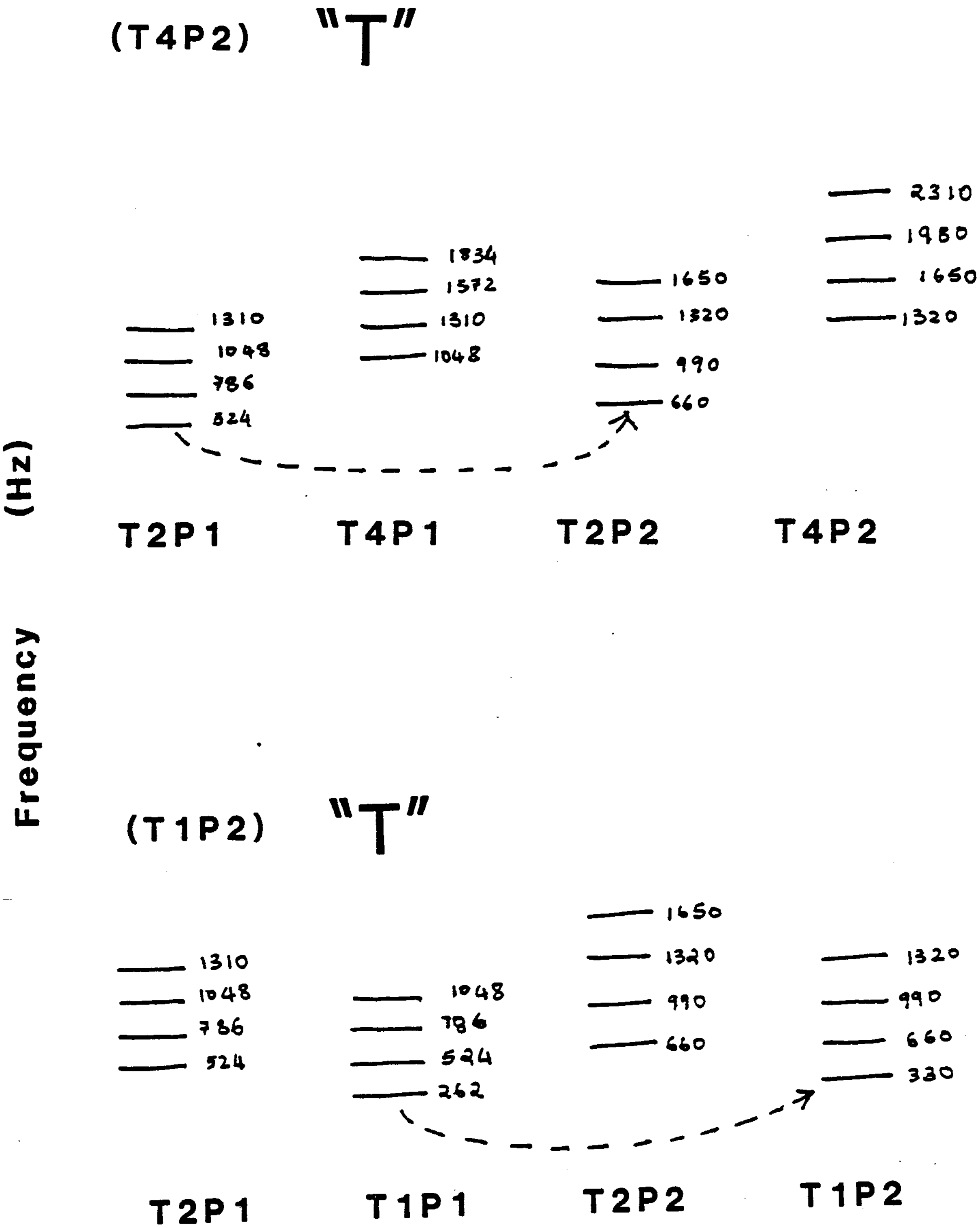


Figure 7.2 Magnified analysis of the sequences (T4P2) and (T1P2), showing the interaction between components of the sequence tones.

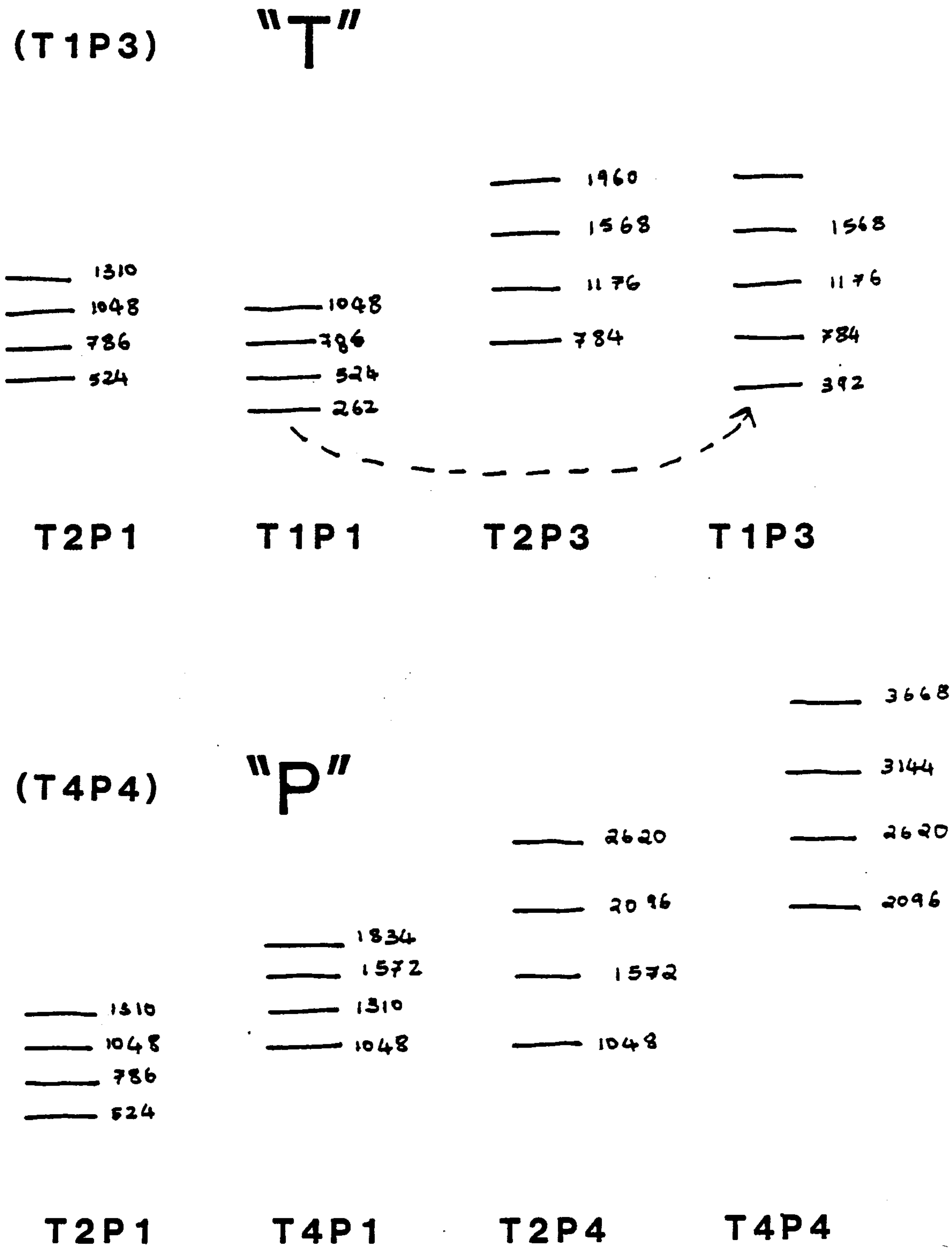


Figure 7.3 Magnified analysis of the sequences (T1P3) and (T4P4), showing the interaction between the components of sequence tones.

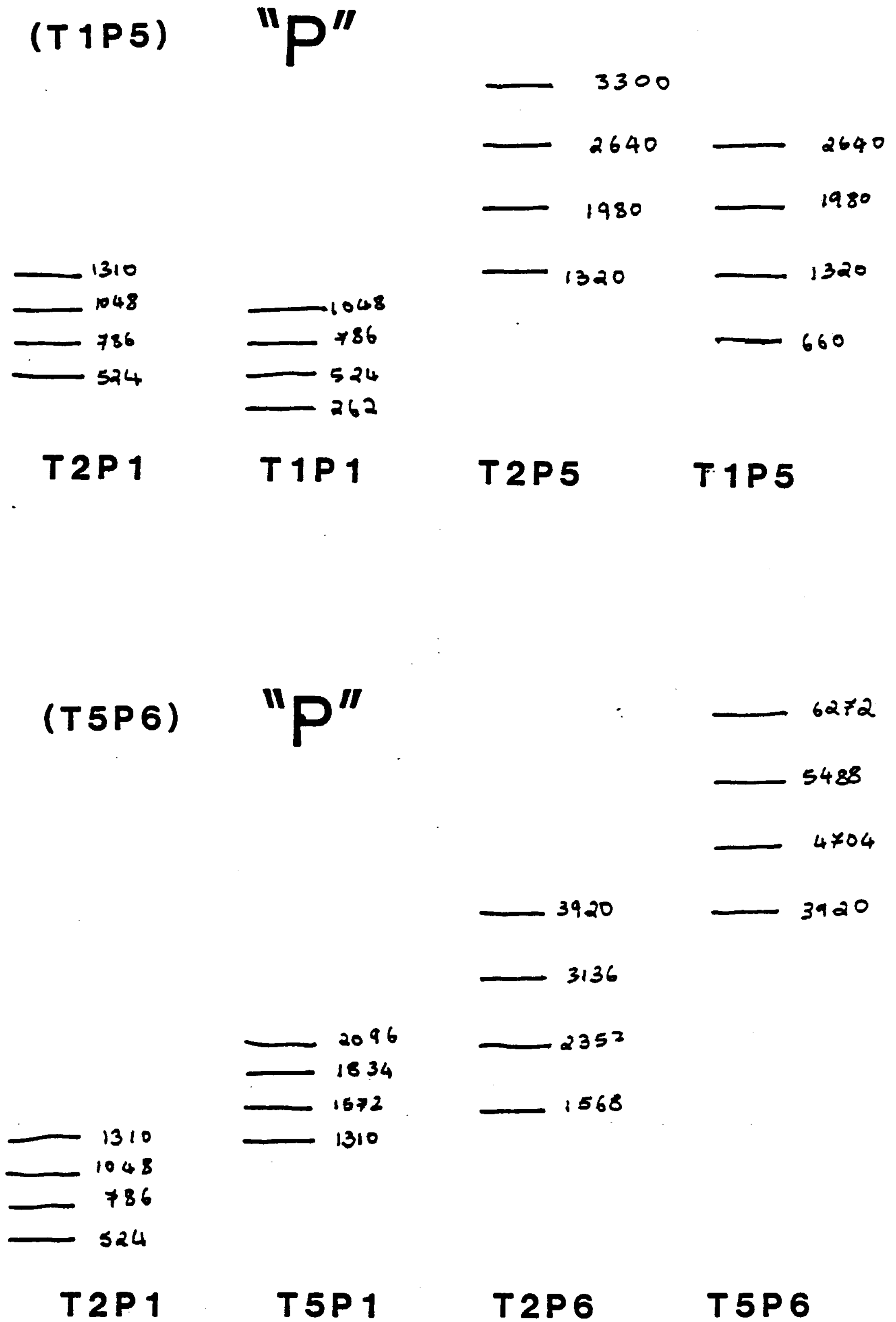


Figure 7.4 Magnified analysis of the sequences (T1P5) and (T5P6), showing the interaction between the components of sequence tones.

The sequences being looked at with this magnified analysis, are depicted in spectrographic format, though to facilitate representation, the ordinate (frequency axis) has not been assigned a specific scale. Rather, the components of the complex tones are drawn stacked upon each other, with their "absolute" frequencies specified alongside, to enable exact comparisons in frequency to be made. The numbers of the various harmonics in each tone can easily be deduced from the label defining the timbre, shown along the abscissa.

In all the examples considered, a common pattern begins to emerge within the sequences that were grouped on the basis of timbre, those that were grouped on the basis of pitch, and the ambiguous sequences.

On looking at the timbrally grouped sequences (T4P2), (T1P2), and (T1P3), it is observed that a "connection" is formed between two tones, whenever the lowest component in either of the tones can find another component close enough in frequency. Not having enough data to specify "how close is close", it is speculated, merely on the basis of these observations, that in each case where such connections were made, a low component was involved, that was closer in frequency to a component in the other note, rather than to the other components within the note. The "note" barrier thus seems to have been trespassed in such cases, and linkages formed between the "components" of the tones, in keeping with the stream segregation phenomenon based on frequency proximity, rather than on the basis of pitch proximity.

In accordance with this claim, it is observed that a grouping based on the so called "pitch interval" existed when all tones were "satisfied" within their own complexes or had the appropriate "nearest frequency" components juxtaposed within the adjacent (same pitch) tones, so that they did not need to wander across tone boundaries to find companionship, that resulted in "streaming" that disrupted the perception of temporal order.

It should be noted that the "pitch" based grouping referred to here, occurs between the tones that are temporally juxtaposed within a pair. Since the grouped tones lie next to each other, the perception of temporal order is preserved intact. It is only the "timbre" based streaming in these sequences that disrupts temporal order perception, since the timbrally similar tones occur alternately and are not juxtaposed in time. The connections formed thus overlap over the other tones, making serial order perception a confusing task.

Another interesting feature observed from the results, is the repeated encountering of T4 as the timbre that demarcates the "timbre stream boundary" , a timbral distance greater than which causes timbre-based streaming for pitch intervals less than an octave. The pitch interval P4, corresponding to this octave interval, demarcates the "pitch stream boundary" for these sequences, with grouping based on "pitch" occurring for larger pitch intervals (that are not too far apart in timbral distance).

Translating these results in terms of the primary attributes of the sounds, it appears that what has been studied is basically the tradeoff between two different facets of the frequency dependent characteristics of the sounds.

Since the tones were synthesized to have equal amplitude harmonic spectra, and were of equal duration, and had the same temporal amplitude envelope (the cosine pedestal), it seems to be safe to rule out temporal and intensity factors as being responsible for the phenomena observed. This leaves only the frequency dimension to be considered, in the specification of both, the "pitch", and the "timbral" attributes of the sounds.

Recalling the discussion on fixed, and relative spectrum of a sound , in determining the invariance of timbre (Slawson,1968, Plomp,1976, section 2.9), it was concluded that a spectral envelope "fixed" at a particular location of the frequency continuum, was responsible for determining the timbre of a sound, rather than the specification of a spectral envelope that preserved the relative amplitudes of the harmonics.

The sounds that were synthesized for this study satisfy both these criteria simultaneously, since a flat (equal amplitude) harmonic spectrum has been chosen, that is assigned a particular "timbre" , by the specification of the frequency region of the harmonics. "Different" timbres thus only differ in the location of their harmonics along the frequency axis, the number of harmonics, their amplitudes, and the spacing between them all being the same, for a sound that is specified to be equally intense, and that is equated in "pitch". The harmonic occupation region is the only variable that changes across timbres, and can thus be thought of as being solely responsible for the timbral attributes of the sounds.

The pitch attribute on the other hand, seems to primarily be a function of the spacing between the harmonics of the spectrum defined, corresponding to the frequency of the "missing" , or present, fundamental. However, the overall "pitch height" associated with a sound also seems to be a function of the location of the "spectral center of gravity", the same attribute that has been used to define "timbre" here. Thus timbres having harmonics that occupy higher regions of the frequency domain, may often be perceived (as indeed they sometimes were), as being sharper, brighter, etc (Lichte, 1941 ,von Bismarck, 1974).

The type of sounds used in this study can therefore be described by the specification of two parameters : (1), the location of the spectral "band" of energy along the frequency axis, and (2), the bandwidth of this spectral band. While the former parameter is responsible for the "timbre" of the sounds comprising the four chosen harmonics, the latter parameter determines the "pitch" of the sound by the specification of spacing between the harmonics. While the former parameter denotes an "absolute" measure of frequency region, the latter parameter reflects the more "relative" aspect of frequency , manifested in the perception of pitch.

7.2 Tradeoffs in Frequency.

Both "where", and "how much" mechanisms thus seem to be needed to specify sounds such as these, that are supposedly equal in other ways. The segregation phenomena observed in this study seem to result due to the interplay between these two aspects of the frequency domain ; the absolute, and the relative.

Some evidence for the relationship between these two facets of the frequency spectrum, and the tradeoffs that occur between them is provided in the interesting observation that the values of these parameters, that are responsible for bringing about segregation of the sequence into subsequences, are roughly proportional to each other.

Thus T4, the timbre built on harmonics 4, 5, 6, and 7, of some (missing) fundamental frequency, is seen to have two harmonics in common with the anchored reference timbre T2, built on the harmonics 2, 3, 4, and 5, of the specified fundamental. The ratio of the distances along the frequency axis corresponding to the absolute frequency of the lowest components in each of these timbres, is thus (4:2), or `2´. Similarly, the pitch interval between P1, the reference "pitch", and P4, the pitch demarcating the streaming boundary, is seen to be the octave (C5:C4), corresponding also, to the ratio of `2´, for the relative frequency parameter, describing the pitch relationship between the first pair of tones, and the second pair.

Whether this "octavial" ratio of the frequency of the base harmonics of T4 and the reference timbre T2 can be taken to be the "spectral region" counterpart of the "spectral width" ratio that determines the more conventional "pitch" difference between P4 and P1, needs to be corroborated with further experimental evidence. The results of this study however, imply the plausibility of this speculation.

If additional evidence was obtained, it may well become possible to find analogous "intervallic" relationships between the spectral "distance" and spectral "width", for other frequency ratios representative of a "fifth", "fourth", "third" etc. The interval corresponding to "unison" in terms of "spectral distance", corresponded in the study to the timbral distance T2-T2, which did not cause timbral segregation for any pitch interval. The similarity with the within-timbre, non-streaming "unison" in the "spectral width" (pitch) domain is again implied. The "pitch" unison however was seen to segregate given a large timbral distance between the tones, while the "timbre" unison was found to remain coherent over the entire pitch interval range studied.

This indicates that the overall "absolute" frequency parameter might be dominant over the "relative" frequency parameter that defines pitch. Alternatively, on the basis of the repeated observation of T4P4 being the "breakpoint" of sequence segregation, a tradeoff between the salience of the spectral region, and the spectral width, might take place, with one being dominant up until some value, and the other taking over as the dominant cue beyond that point.

Quantification of these two facets of the frequency relations between sounds and their components, presents an exciting "glimpse" of a day when the ambiguous term "timbre" might be quantified in terms of its dependence on the "finer-grained" relations in the amplitude, time, and frequency domains. While the exact details of the complicated interactions of these primary dimensions may never be fully quantifiable, the acquisition of knowledge about the type of tradeoffs that do take place between subsets of them no longer seems to be a never-to-be-realized "impossible dream".

8. CODA

8.1 Revelations

In the light of the results obtained, and the conclusions drawn from the present study, it is interesting to "take a look" at work done in the past that has investigated similar phenomena, and see if the present results account for some previously unexplained observations, or if old work acquires new significance under new light. Some related work germane to this issue is briefly discussed below.

Since the sounds used in this study were designed along the lines of the pitch model proposed by Goldstein (1973), it is interesting to see how the present results relate to an earlier experiment of his (Houtsma and Goldstein, 1971) in which similarly constructed tones were used in a melodic context.

In studying the human auditory system's ability to recognize simple melodies comprising sequences of periodic sounds devoid of fundamental energy, Houtsma and Goldstein used a musical interval identification task , that required their subjects to identify a two note (melodic) interval presented monotically or dichotically. Each "note" comprised two consecutive harmonics of the "missing fundamental" of the pitch value desired. In changing from one note to the next, the fundamental frequency of the pair of harmonics for the first note was changed to that desired for the second. In addition, the harmonic "number" (n), of the lower harmonic, of the simultaneous pair of adjacent harmonics was changed randomly from note to note, and trial to trial, to prevent the simple "tracking of fundamental frequency" that would have been equivalent to the veridical tracking of the frequency of one of the harmonic frequencies actually presented. The extent to which "n" could change over a trial was however restricted to the range (n-1, n, n+1), for different "actual" values of the "average" harmonic number "n". Each note lasted 500 msec, and the two were separated by a silent interval of 250 msec.

The similarity of stimuli between the present experiment and that of Houtsma and Goldstein is immediately apparent. The stimuli used in the present experiment differed in that four tones were used instead of two, and in each tone, four consecutive harmonics were used instead of the two used here. However, the present experiment used the same type of pitch relation as used by these investigators, with the first two tones being of the same pitch, and the second two being the same pitch, in the format of a "stuttering" melodic interval (low-low-high-high). The number of the lower (base) harmonic similarly changed from note-to-note in the present experiment, but in a systematic fashion, unlike the random change over n-1, n, and n+1 as used by Houtsma and Goldstein.

The aim of Houtsma and Goldstein was simply to see if listeners could perceive and consequently identify melodic intervals constructed using two tones made of two harmonics, presented either to the same ear, or to different ears as a function of the harmonic number `n´, and the (missing) fundamental frequency of the consecutive notes. The aim here has however been not to "identify" per se the intervals themselves. Rather, the subject is required to report a "percept" that reveals how the four notes making up the single interval appeared to be grouped ; whether as a stuttering interval with a single change in pitch between the first and second pair of equal-pitch sounds, or whether as two occurrences of the same interval, played seemingly by two different instruments, overlapping in time.

The results of both the experiments are informative to each other. Houtsma and Goldstein found that both, the dichotic, and monotically presented stimuli were equally well recognized by subjects, leading them to propose a higher order "central" processing mechanism for the perception of pitch. They also found the interesting result that the best performance was achieved with the lowest harmonic numbers. In conjunction with the observations of Plomp (1964), these experimentors imply that the lower partials of complex tones are "better resolved", and therefore hypothesize that "both of the harmonics of the sounds employed in the experiment are processed through separate channels of the cochlear output, to obtain successful identification".

These observations of Houtsma and Goldstein appear to be corroborated by the results of the present experiment. The "lower conponent" dominated streaming discussed in the last chapter gets an impetus from these results. It seems that the better resolved low components might have "first rights" on picking the proximal frequency components that they stream with, given the entire set of sounds in terms of constituent components. A hierarchy of this type of streaming was suggested on the basis of the stimuli analyzed in the last chapter, which may thus be accounted for in terms of the limitations of the frequency analyzing capability of the ear. The perceptual groups are then probably formed at a later, more central stage, based on the hierarchical output of the peripheral analysis.

The fact that Houtsma and Goldstein do not report any changes in "quality", or timbre of their sounds despite note-to-note changes in spectrum "location", as ought to be the case on the basis of the results obtained in the present experiment, seems to be a consequence of two major differences in their paradigm, and that used here :

(1) Only consecutive harmonic numbers (n-1, n, n+1) have been used by them. The spectra of the two sounds are thus never too far from each other, so that the subtle change in timbre may go by unnoticed, particularly given (2),the sequence of two tones comprising the melodic interval was very brief, and not presented in the repeating format that enables comparisons of features to be made, and consequently elicits "streaming". Had this been done, it is plausible that the timbral changes due to differences in harmonic numbers used, would have emerged.

A recent experiment reported by Houtsma and Canning (1983) is an interesting extension of the first author's earlier work, in that it uses stimuli silmilar to those described above, but the complex tones made (sans the fundamental), are "superposed" in time, and the listeners required to identify the "simultaneous" (harmonic) intervals thus formed. Since the real world is full of sounds with complicated spectra overlapping in time, these investigators hoped to determine the extent to which the auditory system is capable of perceptually separating such simultaneous complex sounds in the absence of other cues such as the variation in temporal envelopes, the inherent vibrato (FM) and tremolo (AM) rates, and other dynamic factors.

Each complex tone was synthesized by the addition of three consecutive harmonics, but the triplet of harmonic "numbers" was chosen randomly from the six numbers 2 through 7, for either tone. The two complex sounds that were thus created, varied in their degree of spectral overlap with each other, depending on the triplet of harmonics chosen to make each tone.

On averaging their results for correct identification of intervals over different combinations of harmonic numbers chosen for the two tones, Houtsma and Canning found that unisons yielded the highest percent correct scores, followed by fifths, thirds, and fourths in that order. While this is an interesting observation about the pitch relations of the two tones, it does not bring to light the actual frequency relations between the components of the tones. Fortunately, Houtsma and Canning describe some of their observations on the relation between the spectral structure and the response of their listeners to some extent.

Since they had observed identification performance to be higher when the tone complexes were presented dichotically (one set to each ear), than when they were presented to the same ear, these authors assumed the deficit to be of a peripheral origin, and therefore looked for clues to explain the difference in performance, in terms of the spectra of the sounds. In keeping with their speculation, the worst performance in the diotic case was found for tones that showed spectral overlap leading to interference between the component tones of the two complexes. In the dichotic case this deterioration in identification was not observed, leading the authors to conclude that the recognition task was done in two stages ; one in which the complex of all available stimulus harmonics was divided into subgroups corresponding to each harmonic complex tone, and a second stage in which the fundamental frequencies were estimated from the subgroups. While the latter stage is assumed to be of a "central" character, the first stage is considered by them to be peripheral. Harmonics that are spectrally close, are thought to interfere with each other, rendering them unavailable for further processing. This constraint on spectral resolution within each ear, could account for the unhampered performance in the dichotic case, since spectrally close harmonics sent to different ears would still be available for processing.

While this experiment deals with the perception of simultaneous complex tones unlike the sequential stimulus used in the present study, it has important implications for all complex tone research, since a complex tone is by definition already a pattern of "simultaneous" tones. The same reasoning of "better spectral resolution" obtained here can thus be used to account for the fact that tones that had non-overlapping spectra, were kept perceptually apart, while those that had some degree of spectral overlap, showed a decided tendency to fuse together into a "stream".

An extension of the present experiment, placing the advantage of separate dichotic channeling of a spectrally overlapping sequence in competition with the lack of spectral overlap within an aural channel, might provide the decisive evidence needed to make these claims with conviction.

Since the present experiment deals with complex tones in sequence, it differs from most of the work on streaming reported earlier, where either pure tones were used, or simple alternating sequences comprising a repeated pair of sounds was used. It is therefore difficult to make direct comparisons with the earlier work. However, it is evident that all the earlier results are valid, and often manifested in the complicated interaction of the complex tones and their components. Thus, the observation that stream segregation results between similar sequence events, given the context of contrasting events temporally juxtaposed in time, is still upheld. This grouping of events according to some organizing principle operating along some auditory dimension, is found to result in the deterioration of the ability to perceive patterns that trespass stream boundaries, and serial order perception is seen to get disrupted. The organizing dimension in this experiment is the frequency domain, with the difference in spectral region providing the "discrete jump in frequency" that often demarcates the across-stream events. The observation that "contour" of a pattern affects the occurrence or inhibition of streaming, is also exhibited in this experiment to some extent. If one looks at the "magnified" pictorial analysis of the sequences T4P4 and T5P6 in figures 7.3,7.4, it can be seen that here the different sequence events proceed upward in the frequency domain, in a continuous fashion. The fact that these sequences did not undergo timbral streaming despite large timbral "distances" may be a consequence of this "good continuation".

8.2 A Final Note

This study has been devoted to the investigation of the relations between the physically specifiable structure of sounds in sequence, and the type of perceptual structure that is imposed on such an array by a typical listener. The motivation in studying these psychoacoustic relations has basically been to unravel the mysteries of the type of processes that go on in the perception of music.

It is interesting to see how the auditory dimensions and their interdependence studied in this thesis relate to the world of "real" sequences of "real" sounds that are the building blocks of an auditory art like music.

Although the sounds of music are produced via a variety of instruments having their own particular "quality" and "register", it is surprising that traditionally, the forms of music have observed a silent separation between "pitch" and "timbre". While pitch has been thought of as the primary vehicle of the melody, and therefore has been regarded as the quantitative aspect of music, "timbre" has been relegated to the secondary position of a "quality" that is not crucial to the preservation of the musical piece, but merely serves to augment it aesthetically.

As Erickson (1975), observes; "in music theory timbre and pitch have been kept sharply separated. Pitch has been structure, timbre has been ornament". He goes on to assert his belief that the treatment of pitch and timbre as independent, separable dimensions in music, is "fiction", and on the basis of his musical experience, and corroboration by results of psychoacoustic experiments, implies that these dimensions infact appear to be highly related.

The result of this "psychoacoustic experiment", while providing one more piece of evidence to this should-be-apparent claim, also provides an example of the success with which a "melody of timbres", reminiscent of Schoenberg´s vision of "Klangfarbenmelodie", can actually be realized.

The so called "timbral streaming" observed for some of the sequences used in this experiment, demonstrates that while the "discontinuous" nature of changing timbres may seem disruptive of the "fluidity" that is traditionally desired of music, it is possible for appropriate magnitudes of "rate of execution", and "extent of timbral difference" to be used in creating a melody that is connected across the interfering timbres. Conversely, the differences that initiate this "streamed" grouping, can also be varied to provide more similarity and consequently more coherence between sounds that have contrasting timbres.

Recently, the Serial composers, (e.g Cage, Glass, Reich, Riley, Adams) have had more luck in getting audiences that are willing to partake of the relations between all the dimensions of sound, than their predecessors (Schoenberg, Webern, Stravinsky) had. These composers envision a "total sound space" that is "limited only by the ear" (Cage, 1962), within which any and all of the parameters of sound can be "closely controlled in the compositional process" (McDermott, 1966).

While admiring the ambition of these visionary composers, it is the contention here, that perhaps this "vying for supremacy" between the dimensions, is a fruitless struggle. It seems from the study of sounds presented here, that "other" dimensions such as timbre, are infact highly correlated with the frequency dimension, which in turn is dependent on the time dimension (rate of vibration, rate of repetition), and is also influenced to some extent, by the intensity dimension.

Given these relationships, it may be possible to use them in "supplementary" roles with each other, in making melodies or other musical creations, that can depend on either pitch or timbre or loudness or duration for structure, and the other dimensions as embellishments.

A profound statement made by Schoenberg (quoted from Erickson, 1975), regarding this competition between musical materials, has much to offer in the way of advice, to zealots of musical style, both the traditional and the modern ;

> ".. I cannot readily admit that there is such a difference, as is usually expressed between timbre and pitch. It is my opinion that the sound becomes noticeable through its timbre, and one of its dimensions is pitch. In other words: the larger realm is the timbre, whereas pitch is one of the smaller provinces. The pitch is nothing but timbre measured in one direction. If it is possible to make compositional structures from timbres which differ according to height (pitch), structures which we call melodies, sequences producing an effect similar to thought, then it must also be possible to create such sequences from the timbres of the other dimension from what we normally and simply call timbre. Such sequences would work with an inherent logic, equivalent to the kind of logic which is effective in the melodies based on pitch. All this seems a fantasy of the future, which it probably is. Yet I am firmly convinced that it can be realized ...".

The audience the ear, the "logic" is stream,

the evidence is here, to fulfil the dream...

BIBLIOGRAPHY

Askill, J (1979) Physics of Musical Sounds. D. Van Nostrand Company, New York, Cincinnati.

Benade, A.H (1976) Fundamentals of Musical Acoustics. Oxford University Press, New York.

Benade, A.H (1981) "Spectral Similarities of Tones". (S37) Journal of the Acoustical Society of America, 69.

Benade, A.H (1983) "From Instrument to Ear in a Room: Direct or via Recording". Preprint by the Audio Engineering Society, October, 1983.

Bingham, H.V.D (1910) "Studies in Melody". Psychological Monographs, 12.

von Bismarck, G (1974) "Sharpness as an Attribute of the Timbre of Steady Sounds." Acustica, 30, 159-172.

Broadbent, D.E, (1958) Perception and Communication, Pergamon, New York.

Broadbent, D.E, & Ladefoged, P, (1959)"Auditory Perception of Temporal Order", Journal of the Acoustical Society of America, 31,(11), 1539.

Bregman, A.S (1978 a) "Auditory Streaming: Competition among Alternative Organizations". Perception and Psychophysics, 23, 391-398.

Bregman, A.S (1978 b) "Auditory Streaming is Cumulative". Journal of Experimental Psychology (Human Perception and Performance), 4, 380-387.

Bregman, A.S, & Campbell, J (1971) "Primary Auditory Stream Segregation, and the Perception of Order in Rapid Sequences of Tones", Journal of Experimental Psychology, 89, 244-249.

Bregman, A.S, & Dannenbring, G.L (1973) "The Effect of Continuity on Auditory Stream Segregation". Perception and Psychophysics, 13, 308-312.

Bregman, A.S, & Pinker, S (1978) "Auditory Streaming and the Building of Timbre". Canadian Journal of Psychology, 32, 19-31.

Bregman, A.S, & Rudnicky, S.I (1975) "Auditory Segregation: Stream or Streams?" Journal of Experimental Psychology (Human Perception and Performance), 1, 263-267.

Cage, J (1962) "Silence". (p.9) Middletown, Connecticut.

Cole, R.A, & Scott, B (1974) "Toward a Theory of Speech Perception". Psychological Review, 81, 348-374.

Cuddy, L.L, & Cohen, A.J (1976) "Recognition of Transposed Melodic Sequences". Quarterly Journal of Experimental Psychology, 28, 255-270.

Cuddy, L.L, Cohen, A.J, & Mewhort, D.J.K (1981) "The Perception of Structure in Short Melodic Sequences". Journal of Experimental Psychology, 7, 869-883.

Culver, C.A (1956) Musical Acoustics. McGraw Hill, New York.

Dannenbring, G.L, & Bregman, A.S (1978) "Streaming versus Fusion of Sinusoidal Components of Complex Tones". Perception and Psychophysics, 24, 369-376.

Davies, J.B (1978) The Psychology of Music. Stanford University Press, Stanford, California.

Deutsch, D (1972) "Octave Generalization and Tune Recognition". Perception and Psychophysics, 11, 411-412.

Deutsch, D (1975) "Two Channel Listening to Musical Scales". Journal of the Acoustical Society of America, 57, 1156-1160.

Deutsch, D., (ed.) (1982) The Psychology of Music. Academic Press Series in Cognition and Perception, New York.

Divenyi, P.L, & Hirsh, I.J, (1974) "Identification of Temporal Order in Three Tone Sequences."Journal of the Acoustical Society of America,56, 144-151.

Divenyi, P.L, & Hirsh, I.J, (1975) The Effect of Blanking on the Identification of Temporal Order in Three-Tone Sequences, Perception and Psychophysics, 17,(3), 246-252.

Divenyi, P.L, & Hirsh, I.J, (1978) "Some Figural Properties of Auditory Patterns, Journal of the Acoustical Society of America,64,(5).

Dorman, M.F, Cutting, J.E, & Raphael, L.J (1975) "Perception of Temporal Order in Vowel Sequences with and without Formant Transitions". Journal of Experimental Psychology (Human Perception and Performance), 104, 121-129.

Dowling, W.J (1968) "Rhythmic Fission and Perceptual Organization". Journal of the Acoustical Society of

America, 44, 369.

Dowling, W.J (1972) "Recognition of Melodic Transformations: Inversion, Retrogade, and Retrogade Inversion". Perception and Psychophysics, 12, 417-421.

Dowling, W.J (1973) "The Perception of Interleaved Melodies". Cognitive Psychology, 5, 322-337.

Dowling, W.J (1982) "Melodic Information Processing and Its Development". Chapter 13 in D.Deutsch (ed.) The Psychology of Music, Academic Press, New York.

Dowling, W.J & Fujitani, D.S (1971) "Contour, Interval, and Pitch Recognition Memory for Melodies". Journal of the Acoustical Society of America, 49, 524-531.

von Ehrenfels, C (1890/1942) in E.G Boring: Sensation and Perception in the History of Experimental Psychology". Appleton-Century-Crofts, New York.

Erickson, R (1975) Sound Structure in Music. University of California Press, Berkeley, Los Angeles, London.

Espinoza-Varas, B (1983) "Integration of Spectral and Temporal Cues in Discrimination of Nonspeech Sounds: A Psychoacoustic Analysis". Journal of the Acoustical Society of America, 74, (6), 1687-1694.

Estes, W.K (1972) "An Associative Basis for Coding and Organization in Memory. In A.W Melton & E.Martin (eds.), Coding Processes in Human Memory, Winston, New York.

Fitzgibbons, P.J, Pollatsek, A, & Thomas, I.B (1974) "Detection of Temporal Gaps Within and Between Tonal Groups". Perception and Psychophysics, 16, (3), 522-528.

Goldstein, J.L (1973) "An Optimum Processor Theory for the Central Formation of the Pitch of Complex Tones". Journal of the Acoustical Society of America, 54, 1496-1516.

George, G (1970) Tonality and Musical Structure. Praeger, New York, Washington.

Grey, J.M (1978) "Timbre Discrimination in Musical Patterns". Journal of the Acoustical Society of America, 64, 467-472.

Green, D.M (1969) "Masking with Continuous and Pulsed Sinusoids." Journal of the Acoustical Society of America, 46, 939-946.

Handel, S (1974) "Perceiving Melodic and Rhythmic Auditory Patterns". Journal of Experimental Psychology, 103, 922-933.

Heise, G.A, & Miller, G.A (1951) "An Experimental Study of Auditory Patterns". American Journal of Psychology, 64, 68-77.

von Helmholtz, H.L.F, (1885/1954) On the Sensations of Tone. (2nd Edition), Dover, New York, (Translator A.J Ellis).

Hirsh, Ira J, (1959) Auditory Perception of Temporal Order, Journal of the Acoustical Society of America, 31, (6),759-767.

Hirsh, I.J,& Sherrick, C.E (1961) Perceived Order in Different Sense Modalities, Journal of Experimental Psychology, 62, 423-432

Hirsh, I.J, et Fraisse, P (1964) Simultaneite et Succession de Stimuli Heterogenes, L'Annee Psychologique, Presses Universitaires de France.

Hirsh, Ira J, (1974) Temporal Order and Auditory Perception, in H.R Moskowitz et al. (eds), Sensation and Measurement, D.Reidel (pubs). 251-258.

Hirsh, Ira J, (1975) Temporal Aspects of Hearing, in D.B Tower, (ed) Human Communication and its Disorders, Raven, New York.

Hirsh, Ira J, (1976) Order of Events in Three Sensory Modalities, in S.K Hirsh, D.H Eldredge, I.J Hirsh, S.R Silverman (eds.), Hearing and Davis, Washington University, St.Louis.

Hirsh, Ira J, (1984) Personal Communication.

Houtsma, A.J.M, & Goldstein, J.L (1972) "The Central Origin of the Pitch of Complex Tones : Evidence from Musical Interval Recognition." Journal of the Acoustical Society of America, 51, 520-529.

Houtsma, A.J.M, & Canning, J.M (1983) "Pitch Perception of Simultaneous Complex Tones". Annual Progress Report of the IPO (Institute for Perception Research), Eindhoven, Holland.

Jones, M.R (1976) "Time, Our Lost Dimension". Psychological Review, 83, (5), 323-355.

Jones, M.R (1978) "Auditory Patterns: Studies in the Perception of Structure". Chapter 7 in the Handbook of Perception, Vol.VIII (E.C Carterette & M.P Friedman eds.), Academic Press, New York.

Jones, M.R (1983) "The Patterning of Time and Its Effects on Perceiving". Paper presented at the Conference on Timing and Time Perception, New York City, May, 1983.

Jones, M.R, Kidd, G, & Wetzel, R (1981) "Evidence for Rhythmic Attention". Journal of Experimental Psychology (Human Perception and Performance), 7, (5), 1059-1073.

Koffka, K (1935) The Principles of Gestalt Psychology. Harcourt-Brace, New York.

Kolers, P.A (1962) "Intensity and Contour Effects in Visual Masking". Vision Research, 2, 277-294.

Liberman, A.M, Cooper, F.S, Shankweiler, D.P, & Studdert-Kennedy, M (1967) "Perception of the Speech Code". Psychological Review, 74, 431-461.

Lichte, W.H (194I) "Attributes of Complex Tones". Journal of Experimental Psychology, 28, 455-480.

Massaro, D.W (1972) "Preperceptual Images, Processing Time, and Perceptual Units in Auditory Perception". Psychological Review, 79, 124-145.

McAdams, S.E, & Bregman, A.S (1979) "Hearing Musical Streams". Computer Music Journal, 3,(4), 26-43.

McDermott, J.V (1966) "The Articulation of Musical Space in the Twentieth Century". Ph.D Dissertation (University of Pennysylvania), University Microfilms (pubs.), Ann Arbor, Michigan.

Miller, G.A (1962) "Decision units in the Perception of Speech". IRE Transactions on Information Theory, IT-8,81-83, (Cited in Warren (1982) Auditory Perception, Pg. 182).

Miller, G.A, & Heise, G.A (1950) "The Trill Threshold". Journal of the Acoustical Society of America, 22, 637-638.

Nabelek, I.V, Nabelek, A.K & Hirsh, I.J (1970) "Pitch of Tone Bursts Changing Frequency". Journal of the Acoustical Society of America, 48, 536-553.

Nabelek, I.V, Nabelek, A.K & Hirsh, I.J (1973) Pitch of Sound Bursts with Continuous or Discontinuous Change of Frequency. Journal of the Acoustical Society of America, 53, 1305-1312.

Neisser, U (1967) Cognitive Psychology. Appleton-Century Crofts, New York.

Neisser, U & Hirst, W (1973) "Effect of Practice on the Identification of Auditory Sequences." Perception and Psychophysics, 15, 391-398.

van Noorden, L.P.A.S (1971 a) "Rhythmic Fission as a Function of Tone Rate". Annual Progress Report of the IPO (Institute for Perception Research), Eindhoven, Holland.

van Noorden, L.P.A.S (1971 b) "Discrimination of Time Intervals Bounded by Tones of Different Frequencies." Annual Progress Report of the IPO (Institute for Perception Research), Eindhoven, Holland.

van Noorden, L.P.A.S (1974) "Temporal Coherence in Random Tone Sequences". Annual Progress Report of the IPO

(Institute for Perception Research), Eindhoven, Holland.

van Noorden, L.P.A.S (1975) "Temporal Coherence in the Perception of Tone Sequences". Ph.D Dissertation, Institute for Perception Research, Eindhoven, Holland (pubs). (Discussed by McAdams & Bregman, 1979).

van Noorden, L.P.A.S (1976) Minimum Differences of Level and Frequency for Perceptual Fission of Tone Sequences ABAB". Journal of the Acoustical Society of America, 61, (4), 1041-1045.

Norman, D.A (1967) "Temporal Confusions and Limited Capacity Processors". Acta Psychologica, 27, 293-297, North Holland Publishing Company, Amsterdam.

Olson, H.F (1967) Music, Physics, and Engineering. (2nd Edition), Dover, New York.

Orr, D.B, Friedman, H.L, & Williams, J.C (1965) "Trainability of Listening Comprehension of Speeded Discourse". Journal of Educational Psychology, 56, 148-156.

Ortmann, O (1926) "On the Melodic Relativity of Tones." Psychological Monographs, 35, (62).

Plomp, R (1964) "The Ear as a Frequency Analyzer." Journal of the Acoustical Society of America, 36, 1628-1636.

Plomp, R (1970) "Timbre as a Multidimensional Attribute of Complex Tones." In : Frequency Analysis and Periodicity Detection in Hearing. (Plomp and Smoorenburg, eds.) 397-414, Sijthoff, Leiden.

Plomp, R (1976) Aspects of Tone Sensation. Academic Press, New York, London.

Roederer, J.G (1979) Introduction to the Physics and Psychophysics of Music. Springer-Verlag (2nd Edition), New York, Berlin, Heidelberg.

Risset, J.C, & Wessel, D.L (1982) "Exploration of Timbre by Analysis and Synthesis". Chapter 2 in D.Deutsch (ed.) The Psychology of Music, Academic Press, New York.

Ritsma, R.J (1967) "Frequencies Dominant in the Perception of the Pitch of Complex Sounds". Journal of the Acoustical Society of America, 42, 191-198.

Schouten, J.F, Ritsma, R.J, & Cardozo, B.L (1962) "Pitch of the Residue". Journal of the Acoustical Society of America, 34, 1418-1424.

Shepard, R.N (1982) "Structural Representations of Musical Pitch". Chapter 11 in D.Deutsch (ed.) The Psychology of Music, Academic Press, New York.

Slawson, A.W (1968) "Vowel Quality and Musical Timbre as Functions of Spectrum Envelope and Fundamental Frequency". Journal of the Acoustical Society of America, 43, 87-101.

Singh, P (1983) "Transition Loss: Practice Room to Concert Hall, Music = f(space)".Unpublished Report.

Stroud, J.M (1955) "The Fine Structure of Psychological Time" In H. Quastler (ed.), Information Theory in Psychology, Free Press: Glencoe, Illinois

Terhardt, E (1974) "Pitch, Consonance, and Harmony". Journal of the Acoustical Society of America, 55, 1061-1069.

Terhardt, E (1979) "Calculating Virtual Pitch". Hearing Research, 1, 155-182.

Thomas, I.B, Cetti, R.P, & Chase, P.W (1971) "Effect of Silent Intervals on the Perception of Temporal Order for Vowels". Journal of the Acoustical Society of America, 49, 584.

Thomas, I.B, Hill, P.B, Carroll, F.S, & Garcia, B (1970) "Temporal Order in the Perception of Vowels" Journal of the Acoustical Society of America, 48, 1010-1013.

Vernon, M.D (1937) Visual Perception. Macmillan, New York.

Warren, R. P (1982) Auditory Perception: A New Synthesis. Pergamon Press, New York.

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.

Watson, C.S, Wroton, H.W, Kelly, W.J, & Benbassat, C.A (1975) "Factors in the Discrimination of Tonal Patterns, I. Component Frequency, Temporal Position, and Silent Intervals". Journal of the Acoustical Society of America, 57, 1175-1185.

Watson, C.S, Kelly, W.J, & Wroton, H.W (1976) "Factors in the Discrimination of Tonal Patterns, II. Selective Attention and Learning under Various Levels of Stimulus Uncertainty". Journal of the Acoustical Society of America, 60, 1176-1186.

Wertheimer, M (1923/1982) "Untersuchung zur Lehre von der Gestalt II". Cited in D.Deutsch (ed.) Psychology of Music, (Pg.100, Chapter 4: Grouping Mechanisms in Music).

Wessel, D.L (1979) "Timbre Space as a Musical Control Structure". Computer Music Journal, 3,(2), 45-52.

Wightman, F.L (1973) "The Pattern Transformation Model of Pitch". Journal of the Acoustical Society of America, 54, 407-416.

Winckel, F (1967) Music, Sound and Sensation. Dover Publications, New York.

Zwicker, E (1961) "Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen)". Journal of the Acoustical Society of America, 33, (2), 248.