National Library of Canada
Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services Branch
Direction des acquisitions et des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

The influence of log-frequency parallel gliding upon perceptual fusion.

by

Martine Turgeon

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Master of Science.

Department of Psychology
McGill University
Montreal, Quebec, Canada

1994, March, 21

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file   Votre référence

Our file   Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-94535-4

Canada

Preface

The completion of that master thesis was done under the supervision of Albert S. Bregman, at the department of psychology of McGill University. The experimental work related to that thesis would not have been possible without the technical help of Pierre A. Ahad. I am also grateful for the statistical advices provided by Rhonda Amsel and Yoshio Takane. I would also like to acknowledge the assistance of Todd Mondor, Robert Zatorre and John Macnamara for comments on a draft of this thesis.

Abstract

It is generally recognized that simple harmonic ratios among partials promote their perceptual fusion. However, the influence of parallel gliding in log frequency upon fusion is not understood. The present experiment investigated fusion in relation to different types of parallel and non-parallel log-frequency motion of three concurrent tonal glides. The main hypothesis was that parallel motion on log-frequency-by-time coordinates favors fusion. It was reasoned that a higher degree of fusion of the glides evokes fewer auditory images. Fusion was thus measured by asking eighteen subjects to rate the number of distinct sounds perceived in various gliding stimuli. On test trials, subjects received a pair of stimuli in succession and had to judge which one contained more sounds and to rate the size of the difference on a 7-point scale. Each stimulus was a complex of three sinusoidal tones, gliding in frequency. Each 1400-ms three-glide complex was either increasing or decreasing in frequency, and the spacing among its components was either small, medium or large. The stimuli were aligned in one of five ways: 1) harmonically related, parallel, and therefore unequally spaced in log frequency, 2) inharmonic, parallel, and equally spaced in log frequency, 3) inharmonic, parallel, and unequally spaced in log frequency, 4) non-parallel and diverging in log frequency, and 5) non-parallel and converging in log frequency. Results showed that more sources were perceived under the three parallel conditions than under the two non-parallel ones ($\underline{p} < 0.00001$). Moreover, as the spacing between gliding partials increased, more distinct sounds were heard ($\underline{p} < 0.00001$). These results suggest that both spectral spacing and non-parallel log-frequency motion segregate concurrent glides. A multiple regression analysis showed that parallel log-frequency gliding promotes fusion over and above the contribution of average spectral spacing and harmonicity ($p < .0001$). The observed data thus support the hypothesis that a common direction of log-frequency gliding among partials promotes their fusion into a coherent percept.

Résumé

Il est généralement reconnu que des rapports harmoniques simples favorisent la fusion perceptuelle de sons tonals purs(sinusoïdes). Cependant, l'influence d'un mouvement parallèle de fréquence logarithmique sur leur fusion n'est pas connu. Cette étude avait pour but d'explorer la fusion perceptuelle de trois sons tonals purs simultanés d'après leur trajectoire de changement sur une échelle de fréquence logarithmique. L'hypothèse principale était qu'une direction commune de changement de fréquence logarithmique fusionne les sons tonals purs. Plus des sons fusionnent, moins ils devraient évoquer d'entités perceptuelles distinctes. Selon cette assumption, le dégré de fusion a été mesuré en demandant à 18 sujets d'évaluer le nombre de composantes distinctes percues dans divers stimuli. Lors d'un essai expérimental, une paire de stimuli était présentée à chaque auditeur en succession. La tâche de ce dernier consistait à determiner lequel des deux stimuli contenait le plus de composantes distinctes, ainsi qu'à évaluer la grandeur de cette différence sur une échelle de 1 à 7. Chaque stimulus était composé de trois sons tonals purs qui descendaient ou qui montaient en fréquence, et dont l'espace spectral étaient petit, moyen ou grand. Les trois sons tonals de chaque stimulus suivaient un parmis cinq types de trajectoire de changement: 1) des trajectoires parallèles maintenant des rapports harmoniques simples, mais pas l'espace entre les sinusoïdes sur une échelle de fréquence logarithmique, 2) des trajectoires parallèles maintenant l'espace de fréquence logarithmique constant entre les sinusoïdes, mais pas de rapports harmoniques simples, 2) des trajectoires parallèles sur une échelle de fréquence logarithmique ne maintenant ni l'espace, ni les rapports harmoniques entre les sinusoïdes, 4) des trajectoires convergentes, donc non parallèles sur une échelle de fréquence logarithmique, et 5) des trajectoires divergentes, donc non parallèles sur une échelle de fréquence logarithmique. Les résultats ont démontré que plus de composantes distinctes sont percues lorsque les trois sons tonals purs suivent des trajectoires de fréquence logarithmique parallèles que losqu'ils suivent des trajectoires non parallèles ($\underline{p} < 0.00001$). De plus, le nombre de composantes distinctement percues augmente en fonction de l'espace spectral entre les sinusoïdes ($\underline{p} < 0.00001$). Ces

résultats suggèrent que l'espace spectral et les trajectoires non parallèles sur une échelle de fréquence logarithmique contribuent à séparer perceptuellement les ondes sinusoïdales simultanées. Une analyse de regressions multiples a démontré qu'un mouvement parallèle sur une échelle de fréquence logarithmique favorise la fusion perceptuelle au-dela de la contribution de l'espace de fréquence moyen entre les sinusoïdes et de leur relations harmoniques ($p < .0001$). Ces observations soutiennent donc l'hypothèse qu'une direction commune de changement de fréquence logarithmique parmis des sons tonals purs favorise leur fusion en une unité perceptuelle cohésive.

Introduction

In a normal listening environment, many events take place at a given moment. Consequently, the pressure wave received by the ear is itself the product of the activity of multiple sound emitters. Moreover, a complex waveform emanates from each acoustic event. This waveform consists of many frequency components changing over time. Some processes must thus determine which parts of the sensory data provide information about individual events.

The ability to separate the acoustic array into its many sources of sound allows a cognitive system to pay attention to single environmental events. This has adaptive value since it is a prerequisite for extracting meaningful information about a given source, such as whether it is dangerous or innocuous. In order to describe the acoustic environment, the auditory processes must construct separate mental descriptions of the distinct sources that have contributed to the signal. This appears to be realized by the operation of some mechanisms that systematically decomposes the mixture of sounds that reaches the ear. As a result some auditory streams are formed (1981, 1984, 1990). Their role is to psychologically describe the pattern of temporal activity of discrete sources of sound. The parsing process uses different pieces of sensory evidence and analyses them according to some basic principles of perceptual organization. An important goal of the field of auditory perception is to determine the nature of these rules.

Many experiments have been conducted to determine the rules of auditory organization (Bregman, 1990). These heuristics can be divided into two groups: those that promote sequential grouping and those that encourage simultaneous grouping. In sequential grouping, sounds that have arisen at different times from a single source are connected into a common stream. Simultaneous grouping processes connect spectral components that are emanating from a given source at a particular time. Various heuristics compete with and reinforce one another in order to reach the best possible perceptual description of the acoustic input (Bregman, 1978a, 1978b, 1984, 1990).

Bregman proposed a theoretical framework known as "auditory scene analysis" (Bregman, 1981, 1984, 1990). This theory presupposes that the gestalt principles regulating visual-pattern formation also operate along acoustic dimensions. Three major heuristics implicated in visual processing were originally proposed by Gestalt psychologists: the proximity, the similarity and the common fate principles. The proximity rule states that the closer elements are to each other along some dimension(s), such as color, the more likely they are to be grouped together into a common perceptual unit. The similarity rule asserts that sharing features along some dimension(s), such as shape, favors perceptual grouping. Finally the common fate rule states that common change among subsets of elements of a scene, such as following a parallel spatial trajectory, promotes their grouping into a common figure.

Many empirical data support Bregman's theory of auditory organization. It seems that one heuristic governing the processes of sequential stream formation is the proximity of acoustic elements along the temporal and spectral dimensions of sound (Bregman, 1984, 1990; Bregman & Campbell, 1973; Bregman & Pinker, 1978; Deutsch, 1975; Dowling, 1973; Norden, 1975, 1977). Another sequential-grouping rule seems to be the similarity in spectral composition (Singh, 1985, 1987, 1990; Wessel, 1979), spatial location (Judd, 1977) and fundamental frequency (Bregman, 1990; Singh, 1987, 1990). In simultaneous grouping, common amplitude-modulation (AM) among acoustic components appears to be a critical grouping cue (Bregman, Abramson, Doehring & Darwin, 1985; Hall & Grose, 1990; McAdams & Wessel, 1981). A common attack pattern or onset also appear to be a spectral-organization heuristic (Bregman & Pinker, 1978; Rasch, 1978; Singh, 1985).

Some heuristics in audition are associated with sequential grouping and others with simultaneous grouping. For example, while common AM is likely to group the concurrent partials of a given voice into one syllable (Bregman, Abramson, Doehring & Darwin, 1985), similarity in the locus and spacing of harmonics among successive sung syllables probably links them to form a coherent melody (Singh, 1985, 1987). However, a given parsing rule actually affects both sequential and synchronous grouping. For instance, in an orchestral performance, similarity in the locus of spectral peaks of intensity is probably critical for both holding together a melody (sequential grouping) from a single violin, and a chord (simultaneous grouping) from all the violins in the string section (Singh, 1985, 1987; Wessel, 1979). Nevertheless, the classification of parsing rules into either sequential or simultaneous grouping processes is a useful operational distinction.

As a general rule, any vibrating body with a clear pitch, whether it be the vocal chords or the wooden reed of a saxophone, produces a set of sinusoidal frequencies or harmonics that are all integer multiples of a common frequency. Such a sound is called a complex tone and its first harmonic, that is, the lowest frequency component is the fundamental (F0). If the fundamental is called f, the harmonics have the frequencies f, 2f, 3f, 4f, 5f,...etc. In simultaneous grouping, partials that are harmonically related to one another fuse together into a single sound (Bregman, 1990; De Witt & Crowder, 1987; McAdams, 1984; Moore, 1989). The resulting experience is a single pitch corresponding to the F0 and an overall timbral quality richer than each isolated partial. To the extent that simultaneously played complex tones have more harmonics in common, they blend together more strongly in a chord percept (Bregman, 1990). This phenomenon appears to be due to an auditory mechanism whose function is to group partials into families of harmonics that are each based on a common F0. By grouping together partials that are harmonically related, the auditory system takes advantage of the fact that, in nature, components that show this type of relation often have originated from a common source, such as the crying of a baby or the howling of a dog.

At this point, it is important to differentiate two phenomena, namely, perceptual grouping and perceptual fusion. Fusion occurs when many acoustic elements perceptually merge together into one sound. However, sounds can be grouped together into a coherent stream without fusing into a single percept. Global properties emerge when components fuse together to form a higher-order sound organization, such as a new pitch and richer timbre. Thus powerful fusion occurs when individual frequency components "melt" together into a single pitch, corresponding to their common F0. The opposite of fusion is therefore segregation or the ability to perceptually isolate a component in a mixture of sounds. However, many sounds can group together without blending into one. For example, each chord has a characteristic pitch and timbral quality, and yet it is still possible for an attentive ear to perceive the discrete pitch and timbre of its three or four individual tones. A chord percept thus illustrates partial fusion in which the individual sounds can still be heard if an attempt is made, but where the grouped sounds also acquire a unique quality.

If one considers that it is very unlikely by chance for the constituent parts of distinct world events to simultaneously undergo a proportional change along some dimension(s), the following heuristic would be adaptive in the parsing of acoustic data: "if different parts of a spectrum change in the same way at the same time, then fuse them into a common auditory event". This heuristic constitutes an auditory version of the common-fate principle of visual perception. Various sources of empirical evidence support the hypothesis that such a rule operates in auditory processing. For instance, it appears that common temporal patterns of intensity and frequency fluctuations help bring about the grouping of acoustic components and their fusion into a cohesive percept. First, synchrony of onset and offset has been shown to promote the simultaneous grouping of partials (Bregman, 1984; Bregman & Pinker, 1978; Rasch, 1978). Moreover, some experimental evidence suggests that fusion is encouraged by a synchronous and parallel rise in intensity between pure tones (Turgeon, 1992). In addition, there is strong evidence for spectral integration based on common AM (Bregman, Abramson, Doehring & Darwin, 1985; Hall & Grose, 1990; McAdams & Wessel, 1981). Finally, it appears that a parallel motion on log-frequency-by-time coordinates between concurrent tonal glides favor their grouping (Halpern, 1977; Lipscombe, 1992a, 1992b) and their fusion into a common perceptual unit (Bregman & Doehring, 1984).

The question of whether any common pattern of temporal change whatever promotes grouping and fusion is still an unresolved issue. While there is consistent evidence that acoustic components undergoing a parallel change in intensity tend to be grouped and fused together, it is debatable whether this holds in the frequency domain. The issue of whether coherent frequency change is in itself a simultaneous-grouping cue will be considered for both, fast frequency modulation (FM) and slow frequency gliding. Indeed, different types of coherent frequency change among components might effect their grouping differentially. "FM" is customarily used in reference to rapid and periodic changes over small frequency ranges, but this is not inherent to its meaning. "FM" refers to any rate of frequency change and thereby encompasses aperiodic frequency motion, as well as slow frequency gliding over large frequency ranges. Typically, FM involves short-duration signals, usually below 500 ms. Moreover, modulation commonly involves rates from 5 to 15 Hz and frequency depth below 10% of carrier frequencies (Carlyon, 1990; McAdams, 1989; Schooneveldt & Moore, 1988). On the other hand, gliding components, as used in this laboratory, typically cover a large range of frequencies, namely, over 500 Hz. Moreover, gliding takes place over relatively long periods, usually longer than 500 ms (Bregman & Doehring, 1984; Bregman, Halpern, Halikia & Doehring, 1986; Halpern, 1977; Lipscombe, 1992a, 1992b; Tougas & Bregman, 1990).

First, let us consider coherent versus incoherent FM as it is typically studied. There is evidence that supports the role of coherent FM in promoting auditory grouping and fusion (Chalikia & Bregman, 1990; Grose & Hall, 1990; McAdams, 1984) and other data that do not support it (Carlyon, 1990; Gardner & Darwin, 1986; McAdams, 1989; Schooneveldt & Moore, 1988). Coherently modulated components of a complex sound change frequency in the same direction at the same time. A coherent FM that is particularly interesting is the one that maintains the frequency ratios of the partials. Past research has shown that this form of modulation induces a natural voice-like quality to synthesized sung vowels (Chowning, 1980; McNabb, 1981). However, whether other forms of coherent FM also produce the same effect still remains to be thoroughly determined (McAdams, 1989).

Carlyon proposed that there is no across-frequency mechanism for FM coherence over and above the one detecting harmonicity. This was based on results showing that listeners cannot discriminate among coherent and incoherent FM of complex tones unless that coherence proceeded from harmonicity among partials (Carlyon, 1990). In other words, it seems that detection of FM incoherence is contingent upon detection of the resulting mistuning of the target component. This can be explained by the fact that any sustained natural perturbation of the F0 is imparted proportionally to all of the harmonics. Consequently, when the F0 of a harmonic sound is modulated, all the components change frequency in the same direction at the same time. However, a set of inharmonic partials can undergo parallel frequency change as a result of the maintenance of their inharmonic ratios throughout modulation. Hence, although harmonicity implies FM coherence, harmonicity does not follow from FM coherence. Upon observation that FM coherence encourages grouping only when the modulated partials are related by a simple ratio (harmonic) or are close to being harmonic, i.e., are quasi harmonic, it appears that FM coherence is not in itself an acoustic cue for simultaneous grouping (Carlyon, 1990).

Natural FM often involves very rapid parallel fluctuations in log frequency. Instances of such FM are the musical vibrato and the normal vibrato produced in speech. Whether slower parallel log-frequency gliding obey the principles of auditory organization that Carlyon proposed remains to be investigated. Musical portamento, as well as voice inflections involve slow pitch gliding. Some experimental evidence suggest that a common pattern of frequency gliding fosters grouping. Indeed, a gliding partial more strongly resists being captured into a sequential stream with a preceding glide when it is related by simple harmonic ratios to two synchronous glides (Bregman & Doehring, 1984). Moreover, parallel log-frequency gliding among partials promotes their grouping into one uniform gliding percept (Bregman, Halpern, Halikia & Doehring, 1986; Halpern, 1977; Lipscombe, 1992a, 1992b).

Some experiments have investigated the contribution of a parallel log-frequency trajectory among glides upon their perceptual grouping (Bregman, Halpern, Halikia & Doehring, 1986; Halpern, 1977; Lipscombe, 1992a, 1992b; McPherson, Ciocca & Bregman, in press; Tougas & Bregman, 1990). Fusion implies grouping, but does not follow from it; hence conclusions from these experiments do not directly apply to the role of common log-frequency change upon fusion. The following question remains to be thoroughly investigated: "Is fusion promoted by a parallel motion among acoustic elements along log-frequency-by-time coordinates?". If so, does parallel log-frequency gliding have to maintain simple harmonic relations among partials in order to promote their fusion into a cohesive unit? The main purpose of the current experiment is to investigate these issues.

Perceptual fusion is not a new phenomenon. Many composers have already exploited its virtues in orchestration through the colorful spectra of various instrumental timbres. Actually, the characteristic timbre of each note emanating from an instrument arises from the powerful fusion of many harmonics, each of a specific intensity. Composers undoubtedly take advantage of that phenomenon. Indeed, unless one is a fancier of computer music, one rarely has the occasion to listen to a musical performance made up of pure tones. While a complex tone is composed of a set of completely fused harmonics, a chord is composed of a set of partly fused complex tones. However, complex tones and chords share a property: the raw acoustic materials for fusion are steady-state components, at least nominally. Are there some empirical data relevant to the fusion of dynamically changing acoustic components? Some experimental findings do bear on that issue

(Bregman & Doehring, 1984; Halpern, 1977; Lipscombe, 1992a, 1992b; McAdams, 1984, 1989). I will briefly go over each of these experiments.

McAdams (1984) found that FM which maintained the frequency ratios among a set of harmonics fuses them into a cohesive sound. He further investigated the question of whether maintaining the frequency ratios among a set of coherently modulated partials was necessary to fuse them or whether moving them in the same direction at the same time would be sufficient (McAdams, 1984). Subjects were asked to make comparisons among complex tones with different modulation schemes. While one tone had a modulation that maintained constant frequency ratios among 16 partials, the other had a modulation that maintained a constant frequency difference among them. A constant-ratio modulation keeps the distance on a logarithmic scale among partials constant, whereas a constant-difference modulation does not. It also maintains the harmonic relations among the partials. For frequencies above 500 Hz, the basilar membrane of the inner ear resolves frequency components roughly on a log-frequency continuum (Bregman, 1990; Moore, 1989). Parallel log-frequency motion thus maintains the relative distances between places of maximum stimulation on the membrane due to various harmonics. This implies that linear changes in log frequency correspond to linear changes in perceived pitch. Thus, a constant-ratio modulation should produce a uniform change in perceived pitch. When the rms modulation depth was at least 12 cents (0.7% of the centre frequency), listeners more often chose the constant-difference tone as having more sources or distinguishable entities in it. On the other hand, a constant-ratio modulation produced a unified sound even at rather large modulation depths, that is, up to 56 cents (3.3 % of the centre frequency). These findings demonstrated that the maintenance of a constant frequency ratio is an important aspect of FM coherence. However, this experiment did not provide conclusive evidence as to whether the critical property for fusion is the maintenance of log-frequency spacing between modulated components or the maintenance of their harmonic relations.

McAdams conducted a further experiment on the effect of FM coherence upon the segregation of concurrent synthesized vowels (McAdams, 1989). In that experiment, listeners had to judge the perceptual salience of a vowel presented simultaneously with two other vowels. Frequency-modulated vowels were rated as more prominent than their unmodulated counterparts. However, unexpectedly, listener's prominence ratings were not related to whether the F0 of the target vowel was modulated by the same or by a different waveform as those of the other two concurrent vowels. According to these data, while modulation among components seems to promote their fusion, the role of modulation coherence per se is dubious.

While some of McAdams' results suggest that very rapid parallel fluctuations in log frequency among partials fuse them together in tones (McAdams, 1984), others, using vowels, are inconclusive (McAdams, 1989). Other findings suggest that FM coherence is effective at promoting the perceptual grouping of components only when they are, or are close to being, related by simple harmonic ratios (Carlyon, 1990). In other words, what seems to be critical is harmonicity among frequency-modulated components as opposed to their following a common trajectory of change in log frequency. The present experiment explored whether or not this interpretation is valid for slower coherent log-frequency change.

An experiment by Bregman and Doehring (1984) studied the influence of a parallel log-frequency trajectory upon fusion. The rationale of the study was based on the assumption that if a gliding partial is not fused with other synchronous glides, it should be freer to be grouped into a sequential stream with earlier sounds that resemble it. They used a sequential capturing paradigm in which a three-component glide was preceded by a captor glide (see Figure 1, p. 15). The glides were all sinusoids gliding in a straight line on log-frequency-by-time coordinates. A pair composed of a captor glide followed by a complex of three simultaneous glides alternated rapidly. The captor had the same frequency and slope as the target glide. The target was the central glide of the complex. On some conditions, the glides were parallel on a log-frequency scale. In that case, a frequency ratio of an octave (4:2:1) was maintained throughout gliding. In other conditions the middle glide was mistuned from the octave relation by a quarter octave (4:1.68:1). The results showed that it was much easier to capture the target glide into a sequential stream under the mistuned condition. The fact that it is harder to capture the target glide under the parallel condition suggests that the following of a parallel log-frequency trajectory fuses glides. However, it is unclear whether parallel log-frequency gliding is contingent upon maintaining simple harmonic ratios among partials throughout gliding.

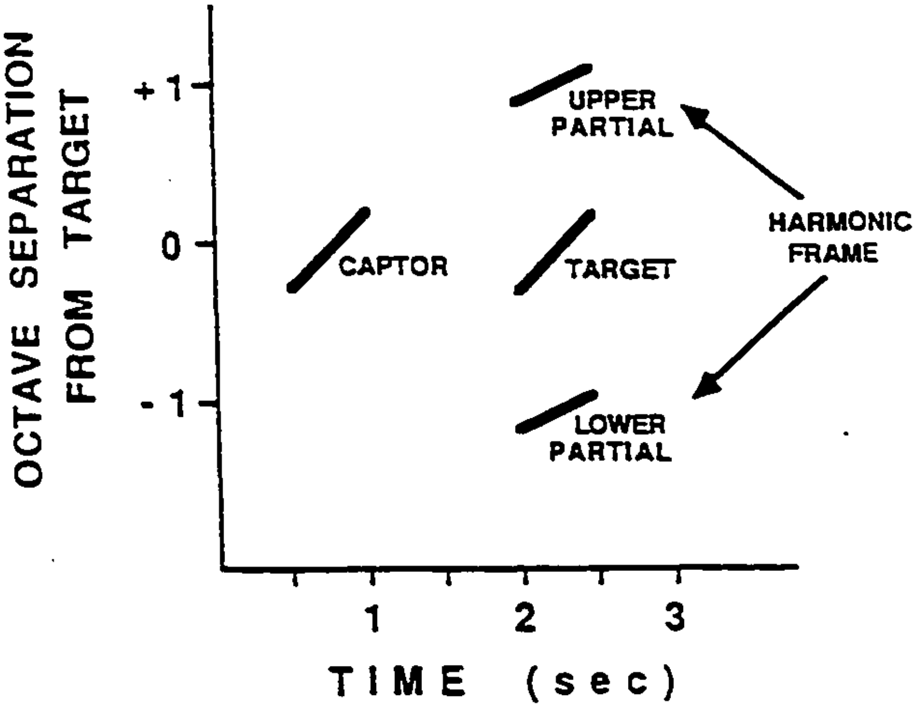

Figure 1: Capturing the central glide in a three-component glide (Bregman & Doehring, 1984).

Other experiments looked at whether parallel log-frequency gliding promotes perceptual grouping (Halpern, 1977; Lipscombe, 1992a, 1992b). The stimuli that were used by Halpern and those used by Lipscombe shared a similar structure. Each stimulus was formed of a crossing set of pure-tone glides, some rising and others falling linearly on log-frequency-by-time coordinates. First, let us describe the experiment by Halpern. Notice in Figure 2, appearing to the left or the right of each glide, the harmonic number of each partial relative to a given F0. Observe that the partials within each subset maintain a fixed harmonic ratio to one another throughout the entire duration of the glide. Consider the sound pattern shown in B.

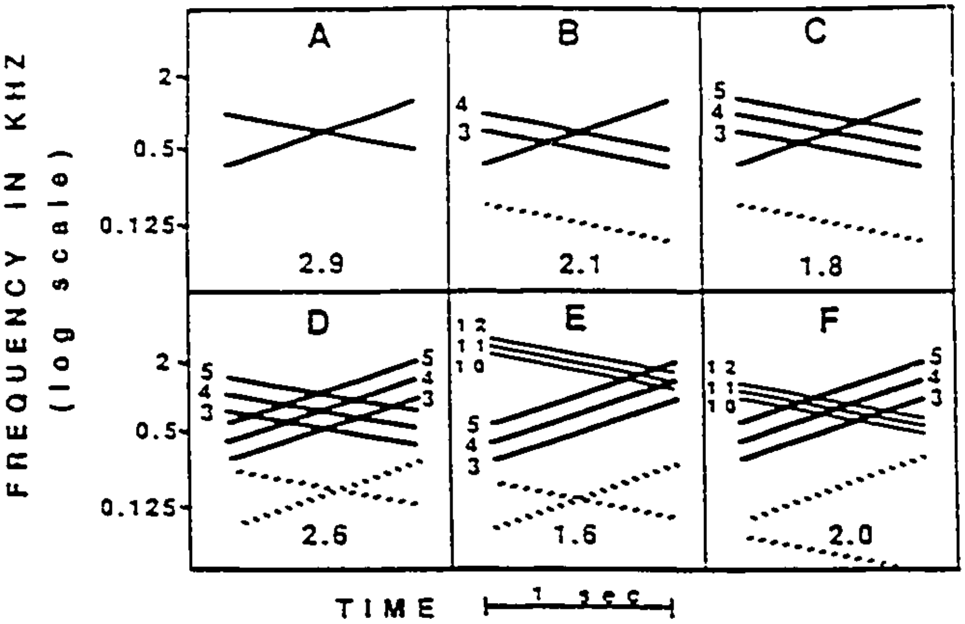

Figure 2: Schema of the six patterns of stimuli used by Halpern (Halpern, 1977).

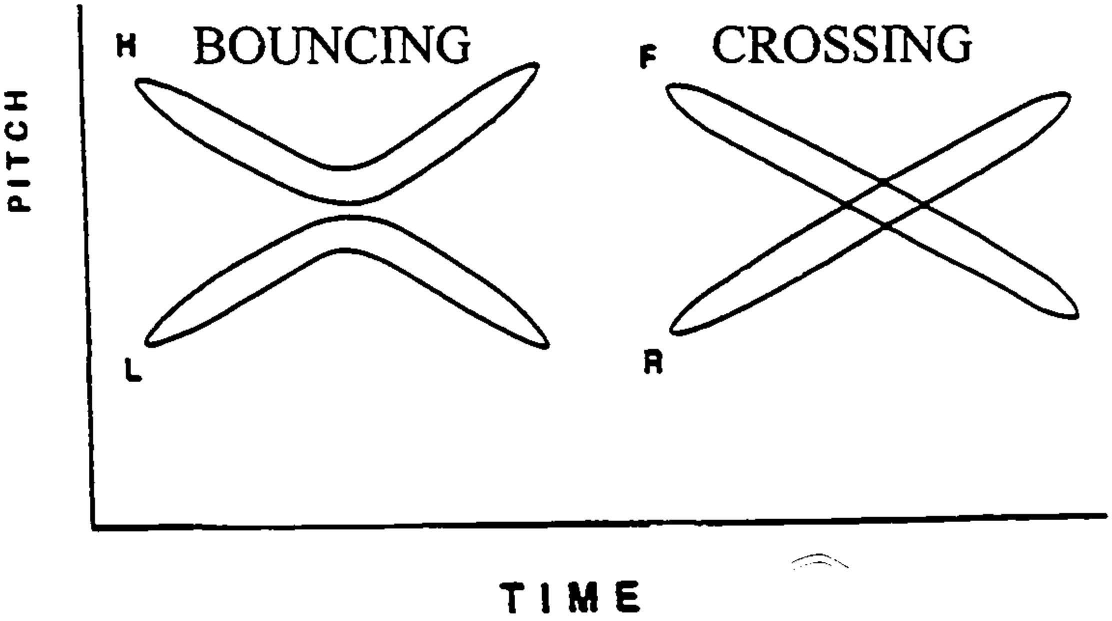

Figure 3: Schematic representation of sounds perceived as bouncing and crossing. 'H' refers to high frequencies and 'L' refers to low ones. 'F' and 'R' stand for the perception of falling and rising in pitch respectively.

How do listeners perceive these gliding patterns of sound ? Three possible percepts could arise: 1) a confused mass of sound constantly shifting in quality, 2) one sound first decreasing and then rising in pitch, accompanied by another which first rises and then decreases in pitch (this is referred to as the percept of "bouncing"), 3) two sounds which cross one another, one rising in pitch and the other falling (this is referred to as the percept of "crossing") (see Figure 3, p.16). In this experiment, sounds were either perceived as bouncing or as crossing. Thus, the two subsets of glides must have been segregated from one another; otherwise, an indiscriminably fuzzy mass of sounds would have been heard. The percept of bouncing took place when the rising and falling subsets shared the same structure, for example two complex tones with identical spectral spacing, as in D in Figure 2. Otherwise, the percept of crossing occurred.

To explain those results, it is useful to look at what happens at the cross-over point. Consider the case of a pure-tone crossing stimulus like that of A in Figure 2. When the falling component reaches the center, the auditory processes have to determine whether the partial is continuing downward or starting to glide upward. In terms of structural similarity there is no information available in that first half of the stimulus as to how each component continues in the second half. Therefore, whenever the rising and falling components are the same, grouping occurs by similarity and proximity in frequency. Thus the percept of bouncing occurs; that is, the descending first half of the stimulus groups with the rising second half and the rising first half groups with the falling second half. However, when there is information available at the cross-over point about structural similarity between the two halves of the stimulus, it overcomes frequency proximity. For example, if the falling first half is more similar to the falling second half than to the rising second half, as in C in Figure 2, then a crossing percept occurs (Halpern, 1977).

The general finding of the perception of crossing when the rising and falling components of the "X" pattern of Figure 2 are the same in harmonic composition and the perception of bouncing when they are different has been replicated many times using slightly different methods (Bregman et al., 1986; McAdams & Bregman, 1979; Tougas & Bregman, 1990). Is this experimental evidence sufficient to conclude that slow parallel gliding motion among partials promotes their perceptual grouping? Unfortunately, it is not. In the Halpern and other replicating experiments, it is unclear whether the grouping within subsets results from the parallel motion of the glides, or the maintenance of harmonic relations between them. To demonstrate that slow parallel motion promotes grouping, the contribution of these two factors must be separated.

Lipscombe attempted to show that, independently of harmonicity, parallel gliding fosters grouping (Lipscombe, 1992a, 1992b). Four alignments were compared to study the degree to which they encouraged crossing or bouncing using the same "X" paradigm (see Figure 3, p. 16). The four alignments were: i) harmonic: parallel motion on log-frequency-by-time coordinates keeping the harmonic ratios constant but not the spacing between glides; ii) equal-log: parallel motion on log-frequency-by-time coordinates, keeping the spacing and the inharmonic ratios between glides constant; iii) shifted: parallel motion on log-frequency-by-time coordinates, keeping the inharmonic ratios constant, but not the spacing between glides; iv) non-parallel: the gliding partials are neither harmonically related nor parallel on log-frequency-by-time coordinates. Spacing among partials was either small, medium or large (see Figure 4, p.19).

Lipscombe's results showed that perceived bouncing was most common among non-parallel glides, followed by the shifted ones. On the other hand, perceived crossing occurred more under the harmonic and equal-log conditions. The percept of crossing emerges when each of two sets of partials, namely, the rising ones and the falling ones, is grouped together into a coherent unit. Thus, the above findings suggest that parallel log-frequency gliding fosters perceptual grouping under certain conditions. However, it seems that the type of parallel gliding and the spectral spacing among components also matters.

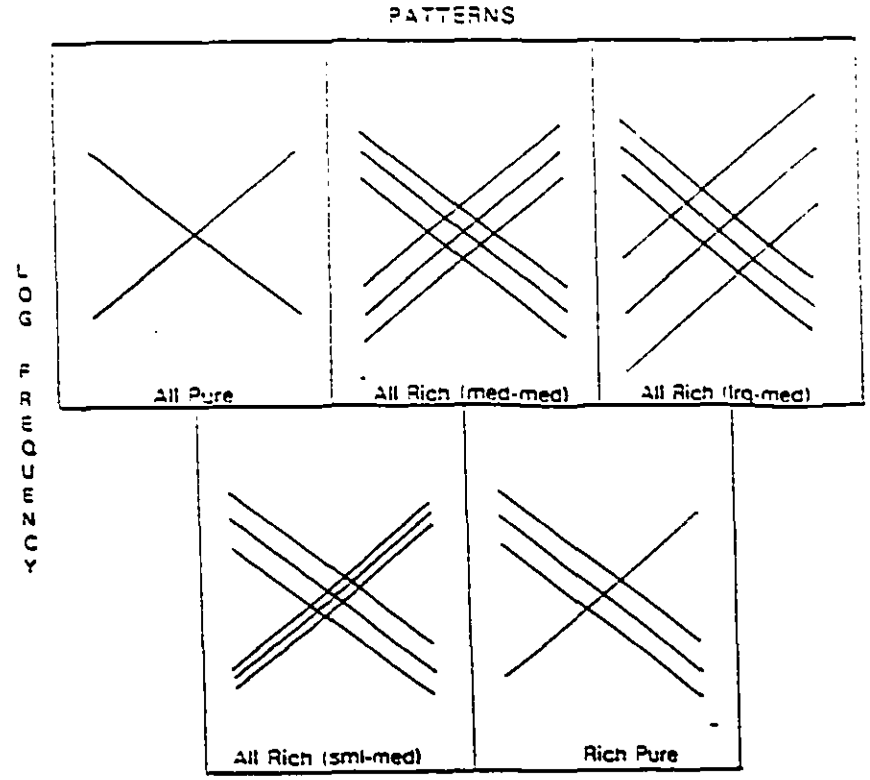

Figure 4: Illustration of the five patterns of stimuli used by Lipscombe (Lipscombe, 1992).

My own experiment is about whether or not parallel log-frequency gliding among partials promotes their perceived fusion. A further question is: "Do some types of parallel motion on log-frequency-by-time coordinates encourage fusion more than others?". In order to investigate the latter issue, the experiment compares the perceived fusion of three parallel conditions, namely, harmonic, equal-log and shifted. These types of parallel log-frequency motion are the same as those used by Lipscombe (1992a, 1992b).

The main hypothesis was that any type of log-frequency parallel motion favors fusion. This was explored by comparing the three above alignments with two non-parallel ones for their ability to fuse components. The types of non-parallel alignments were: i) diverging: non-parallel motion on log-frequency-by-time coordinates in which the log-frequency difference increases during the glide; ii) converging: non-parallel motion on log-frequency-by-time coordinates in which the log-frequency difference decreases during the glide.

How can we measure fusion ? Segregation is the absence of fusion. Thus, as the synchronous components of a three-glide complex fuse more strongly together, fewer distinct sounds should be heard. For example, if the tonal glides do not fuse at all, three sounds should be perceived distinctly. On the other hand, if they completely fuse together, a single sound with a richer timbre should be heard. The experimental task consisted in judging which of two patterns of glides contained more distinct sounds than the other. Moreover, each listener had to rate, on a 7-point scale, how big the difference was between the number of sounds of the two patterns of glides. For each pattern of glides, an overall segregation score was computed from the ratings obtained in all those trials in which it appeared. The assumption underlying this way of scoring is that the segregation-score value is indicative of the degree of fusion of the components in a particular condition. Thus, while a low value on that score indicates strong perceptual fusion, a high one indicates weak fusion.

A particular stimulus comprised two different patterns of glide. On each trial, the stimulus to be judged was presented twice. Each pattern had three glides, either increasing or decreasing in frequency (see Figure 5).

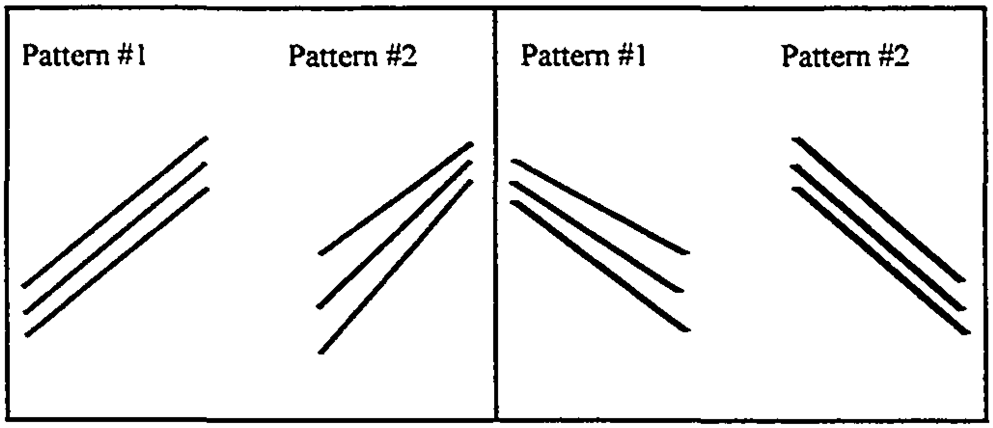

Figure 5 : In any trial, glides were either ascending (left) or descending (right) in log frequency.

Each stimulus pattern followed one of the five possible alignments: harmonic, equal-log, shifted, diverging or converging. To examine the possible influence of absolute frequency difference on grouping, three spacing conditions were examined at each alignment: small, medium and large. Each level of spacing was kept as constant as possible in terms of absolute frequency difference, taking into account the constraints inherent in each type of alignment.

In short, each three-glide pattern corresponded to an experimental condition. Any condition could be described along three dimensions: alignment, spacing and direction. All possible pairwise comparisons between conditions were made within a given direction. Thus upward glides were not compared with downward glides. Therefore, there were no comparisons that could determine whether there was or was not a main effect of direction (rising or falling). This omission allowed the number of comparisons in the experiment to be greatly reduced. There were fifteen upward and fifteen downward conditions (seeFigures 7a and 7b, p. 34 and p. 35). Thus, a given condition was compared with all the other fourteen conditions within the same direction. There were only two replications for every comparison, namely, one for each order of presentation of the two patterns. Thus, to estimate the perceived segregation of a given condition, ratings were averaged across all the twenty-eight comparisons. While the method of comparison allowed a more accurate estimation of segregation for each condition, it induced a slight amount of non-independence among scores. That is, on any given trial, the rating for one pattern of the pair of stimuli constrained the rating for the other.

For each stimulus, a segregation score was computed on the basis of the ratings obtained in all the pairwise comparisons in which it appeared. This score was expressed as a proportion of the maximum score that could possibly be obtained for a condition across all of its comparisons. A zero on this score meant that every time a given condition was paired with another one, the other was rated as having more components and the raw score indicated the largest possible difference between the two conditions. On the other hand, the value, one, indicated that in all of its pairwise comparisons, a particular condition was judged as containing more distinct sounds, and was superior to the other condition by the largest possible value on the rating scale.

In this study, it was expected that the effect of harmonicity in promoting fusion would be replicated. This was verified by comparing harmonic and non-harmonic stimuli in terms of the overall segregation score obtained. Furthermore, increasing the spectral spacing among concurrent partials should yield larger segregation scores. Another prediction was that independently of harmonicity and spacing, listeners would more easily perceive each of the three glides in a non-parallel complex than in a parallel one. To demonstrate that, it had to be shown that over and above harmonicity and spectral proximity, a parallel log-frequency trajectory among partials lowers segregation scores. Finally, the effect of harmonicity, spectral proximity and parallel log-frequency motion upon perceived segregation should hold for both ascending and descending glides.

Let us now examine the research that pertains to the hypothesis that the greater the frequency spacing among glides, the more they segregate. To evaluate the contribution of spectral spacing to the segregation of concurrent glides, the role of spectral spacing in two domains has to be considered: streaming and beating. Streaming occurs when many acoustic elements are grouped together to form one perceptual unit. Such a unit describes the temporal pattern of activity of a single external source (Bregman, 1990). For example, the different successive notes produced by a saxophone are going to be experienced as a single coherent melody. Beating refers to periodic fluctuations in peak amplitude. It occurs when two sinusoids with slightly different frequencies are added together and when the filter system that they pass through does not separate them into different channels (Moore, 1989).

What is the role of spectral spacing in streaming? Spectral proximity among acoustic components has been repeatedly shown to promote their sequential grouping (Bregman &

Campbell, 1973; Deutsch, 1975; Dowling, 1973; Norden, 1975, 1977). Some results also suggest that spectral spacing decreases the fusion of a high AM tone and a low AM tone (Bregman A. S., Abramson J. & Doehring, 1985). It was thus predicted that larger spectral spacing among simultaneous glides would promote their segregation from each other. Moreover, wide spacing and non-parallel gliding were expected to have additive effects in promoting segregation.

What is the role of spectral spacing in beating? When two synchronous components, near in frequency, are superimposed, they resemble a single sinusoid, with frequency equal to the mean frequency of the components, but whose amplitude fluctuates at a rate equal to the frequency difference between the two tones. For example, if a 1000 Hz and 1002 Hz partials are added together, two beats will occur each second. Since in some experimental conditions there were tones relatively close in frequency, this was an important concern. Consider that beating could in itself be a cue for multiple-source detection, despite the fact that the patterns of activity of single sources might not be well resolved, due to interference at the periphery of the auditory system. To avoid beating, the spacing among adjacent components was made large enough so that it remained outside the critical bandwidth specific to the frequency range, throughout the complex glide. This applied to all three-glide stimuli. This method was justified by the notion of critical bands, which assumes that the peripheral auditory system behaves as if it contained a bank of bandpass filters, with continuously overlapping center frequencies, as proposed by Fletcher (1940). According to the power spectrum model proposed by Patterson and Moore (1986), only those frequency components which are in the same critical band can mask each other or interact with each other to produce beats. The critical band frequency range increases as the center frequency of peripheral auditory filters increases (Moore, 1989). The curve of critical bandwidth as a function of center frequency obtained by Zwicker et al. (1957), was used to determine the optimal frequency range of all the small-spacing conditions of the present study.

One of the goals of the study was to verify that the maintenance of frequency ratios among glides fuses them to produce a cohesive sound rich in quality and with a uniform temporal pattern of pitch motion. Another goal was to explore whether there would be any difference in the degree of fusion between the three alignments maintaining the frequency ratios among glides, namely, harmonic, equal-log and shifted. For instance, it might be that the equal-log alignment yields more fusion than the shifted one. Indeed, in each equal-log pattern, the inharmonic ratios and thereby the log-frequency spacing between partials were maintained over the course of the glide. Consider that the lowest frequency involved in equal-log stimuli was 473 Hz. Above 500 Hz, the size of the critical band is roughly equal to a certain proportion of the frequency (Moore, 1989). This is equal to a fixed distance in log frequency. Therefore parallel glides on log frequency are separated by a certain number of critical bands throughout. Moreover, the straightness on log-frequency coordinates should yield a regular change in pitch. By preserving critical-band separation and generating clear pitches, the equal-log alignment should thus induce a regularity of peripheral change. A low segregation score for the equal-log stimuli would thus suggest that the auditory system exploits such a regularity. Finally, non-parallel log-frequency gliding, whether converging or diverging, should segregate the glides. Consequently, each of the three frequency-gliding course should be perceived more distinctively. Moreover, there should be timbres within each sound resembling those of the individual pure tones.

To sum up, the hypotheses of the present study were as follows: 1) the segregation scores should increase from parallel to non-parallel gliding conditions; 2) the segregation scores should increase from harmonic to inharmonic alignments; 3) the segregation scores should increase from closely-spaced to medium-spaced to widely-spaced glides. All of these effects should be observed with both direction of gliding.

Method

i. Subjects

The subjects were 18 adults (9 females, 9 males) ranging in age from 21 to 40 and paid for their participation. All participants were naive to the purpose of the experiment. Each participant performed within 10 dB range of the normal auditory threshold on an audiometric test for the 250-8000 Hz range. This verified their normal hearing for the range of frequencies used in this experiment.

ii. Procedure

The experiment consisted of four sessions; a five-minute practice session, and three twenty-minute experimental sessions. Rest was provided between the sessions. For each subject, the entire experiment took about 90 minutes.

The listeners were told that, on each experimental trial, they would have to tell which of two patterns of sounds had more distinctive parts in it. Moreover, they would have to rate how big the difference was between the number of sounds in the two patterns. The sequence of two gliding stimuli would be repeated twice on each trial to facilitate their judgment. They were going to use a 7-point scale in their ratings. They were told to choose the central point of the scale, that is, 4, when the two gliding patterns contained an equal number of sounds. If the first pattern had more parts than the second, they should use lower ratings. On the other hand, they should use higher ratings if more sounds were distinguished in the second pattern. Thus, while 3 meant that the first stimulus had marginally more sounds than the second, 2 meant that it had somewhat more, and 1 that it had a lot more. Similarly, 5 meant that the second stimulus had marginally more sounds than the first, 6 meant that it had somewhat more, and 7 that it had a lot more. The practice session was to help them to determine the criterion they were going to rely on in the various pairwise comparisons of stimuli. Finally, in their evaluations, they should stick to the same criterion and use the entire range of the 7-point scale.

The latter part of the set of instructions encouraged listeners to establish their criterion on the basis of the variability heard in the gliding patterns of the practice session which contained a representative sample of the various experimental conditions. To make the above set of instructions more concrete, the experimenter presented four audio demonstrations, sampled from the actual pool of experimental stimuli. These examples were chosen by the experimenter so as to include stimuli that differed in alignment, such as converging versus harmonic. If the task was not clearly understood after the demonstrations, listeners could ask questions. As a reminder of the task, the following drawing was displayed above the corresponding digits of the computer keyboard.

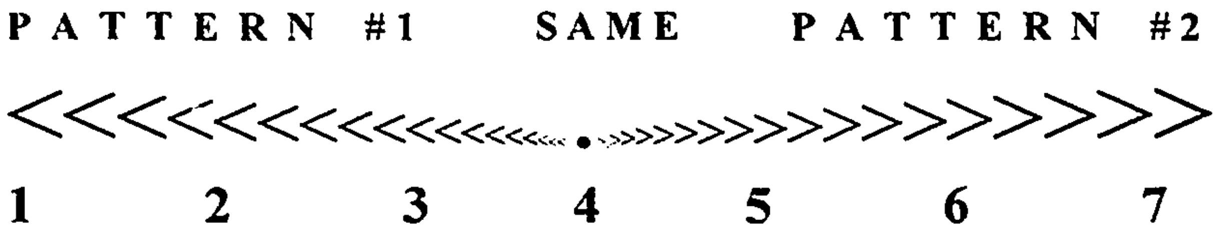

Figure 6: Schema of the 7-point rating scale. The centre of the scale, that is, four, was chosen when the two patterns were perceived as containing the same number of sounds. Lower or higher ratings were used if the first or second pattern was perceived as containing more sounds respectively. The size of the arrow was indicative of the magnitude of the difference in the number of distinct sounds perceived in the two stimuli.

iii. Design

There were three independent variables: i) alignment, varying across 5 levels: harmonic, equal-log, shifted, diverging, converging; ii) spacing, varying across 3 levels: small, medium and large; iii) direction ,varying across 2 levels: up and down. Each three-glide stimulus corresponded to a unique combination of levels of alignment, spacing and direction. There were therefore 15 conditions gliding upward (see Figure 7a, p.34) and 15 gliding downward (see Figure, 7b, p.35).

To verify that the effect of alignment was independent of direction, the frequency range of each glide of the fifteen alignment-by-spacing combinations was the same in both directions of gliding. In other words, a given alignment-by-spacing stimulus going up in pitch was the mirror image of the corresponding one going down in pitch (see Figures 7a and 7b, on p.34 and p.35). There were two orders of presentation for a given comparison among two conditions. This was to counterbalance for a possible order effect.

Within each of the two directions, all possible pairwise comparisons were made between the conditions. Each condition was compared with each of the other fourteen ones sharing the same direction twice, that is, in both orders. Within each direction, combining each of the 15 conditions in pairs gave 105 conditions. Since there were two orders, this gave 210 conditions. These combinations were replicated separately for the falling and rising glides, yielding 420 trials (rising and falling glides were never compared with one another on the same trial). Trials were presented randomly. The total of 420 trials was split in three blocks of 140 trials. The duration of each block was approximately 20 minutes. Prior to the experimental sessions, there was a practice session. This session included a representative sample of 28 stimulus patterns and lasted about 5 minutes.

The variable of main interest was alignment. If indeed parallel motion hinders segregation, that is, promotes fusion, this effect should hold for every direction, and every degree of absolute spacing. To examine the possible influence of absolute frequency difference on grouping, three spacings were combined with the five alignments, namely, small, medium and large. However, the magnitude of absolute frequency spacing for each spacing class (small, medium and large) was not uniform across alignments. An attempt was made to maintain the size of a separation while meeting the criteria that the separation should not be less than a critical band and should satisfy the requirements specific to each alignment. In Table 1, notice that for a given spacing class, the average frequency spacing among a pair of glides in a stimulus differs substantially across the five alignments (see Table 1, p.29)[1]. For instance, the average spacing among widely-spaced glides is 1266.67 Hz for harmonic, 1393.33 Hz for equal-log, 1320.67 Hz for shifted, 2186.67 Hz for diverging and 1512.00 Hz for converging. On the whole, when collapsed across spacing classifications, the average spacing differs across alignments. Indeed, while the average spacing between a pair of glides is 717.78 Hz for harmonic, it is 737.11 Hz for equal-log, 771.56 Hz for shifted, 1649.45 Hz for diverging and 966.22 Hz for converging. Therefore, upon dividing the conditions in a parallel and a non-parallel subsets, one notices a considerable discrepancy between their average spacing, namely, 742.15 Hz and 1307.81 Hz respectively. Since the spacing was not completely orthogonal to the alignment, a method of multiple regression was used to tease apart these effects.

---

1. Note that the average spectral spacing can also be assessed through the geometric mean ratio, that is the antilogarithm of the average of the logarithms of the frequency ratios between the glides. A table of the values of the geometric mean ratios for each condition can be consulted in the Appendix A (see Table 11, p.66).

A. Parallel Conditions	Average Spacing (Hz)	B. Non-Parallel Conditions	Average Spacing (Hz)
01) Harm / Sl / Up	0253.33	19) Div / Sl / Up	1194.67
02) Harm / Sl / Dn	0253.33	20) Div / Sl / Dn	1194.67
03) Harm / Md / Up	0633.33	21) Div / Md / Up	1567.00
04) Harm / Md / Dn	0633.33	22) Div / Md / Dn	1567.00
05) Harm / Lg / Up	1266.67	23) Div / Lg / Up	2186.67
06) Harm / Lg / Dn	1266.67	24) Div / Lg / Dn	2186.67
07) Eq-log / S l / Up	0234.33	25) Conv / Sl / Up	0504.33
08) Eq-log / Sl / Dn	0234.33	26) Conv / Sl / Dn	0504.33
09) Eq-log / Md / Up	0583.67	27) Conv / Md / Up	0882.33
10) Eq-log / Md / Dn	0583.67	28) Conv / Md / Dn	0882.33
11) Eq-log / Lg / Up	1393.33	29) Conv / Lg / Up	1512.00
12) Eq-log / Lg / Dn	1393.33	30) Conv / Lg / Dn	1512.00
13) Shifted / Sl / Up	0306.67		
14) Shifted / Sl / Dn	0306.67		
15) Shifted / Md / Up	0687.33		
16) Shifted / Md /Dn	0687.33		
17) Shifted / Lg / Up	1320.67		
18) Shifted / Lg / Dn	1320.67		
	u = 742.15 Hz		**u = 1307.81 Hz**

Table 1 : Average spectral spacing for each three-glide condition. Given a top(T), a middle(M) and a bottom(B) glide for each complex, a maximum (max) and minimum (min) frequency for each glide, the average spacing (AS) is computed by:

$AS = [(Tmax - Mmax) + (Tmax - Bmax) + (Mmax - Bmax) + (Tmin - Mmin) + (Tmin - Bmin) + (Mmin - Bmin)] / 6$

iv. Stimuli

The temporal structure of each trial is shown in Table 2.

Pattern#1	**Silent**	**Pattern#2**	**Silent**	**Pattern#1**	**Silent**	**Pattern#2**	**Response**
1400 ms	**interval** **100 ms**	**1400 ms**	**interval** **1000 ms**	**1400 ms**	**interval** **100 ms**	**1400 ms**	**(approx.** **2000 ms)**

Table 2 : Schema of the structure and temporal pattern of a trial.

Each stimulus pattern was composed of three simultaneously gliding sinusoidals. These sinusoidals glided linearly on log-frequency-by-time coordinates. The total duration of each stimulus pattern was 1400 ms. On any given trial, the two stimulus patterns were separated by a 100-ms interval. There was also a 1000-ms silence between the two repetitions of the pair. Subjects could take as long as necessary to respond.

Each stimulus began with a 300-ms quarter sine wave amplitude onset, had a steady-state intensity of 60 dB SPL for 1000 ms and ended with a 100-ms reverse quarter sine wave amplitude decay. In each stimulus, the middle component rose to a maximum intensity of 56.4 dB SPL in 300 ms and remained at that intensity until its final 100 ms offset, 1000 ms later. Both, the lower and upper glides reached a maximum intensity level of 44.4 dB SPL, with the same onset, offset and duration as the central glide. A 60-dB SPL steady-state amplitude 1000-Hz tone was used to calibrate intensity. The reason for the choice of such a relative intensity was that it produced the greatest differences, in the degree of perceived fusion of the components, across experimental conditions. This was established in some pilot studies by the experimenter with the help of other researchers from the Speech-and-Hearing Laboratory of McGill University.

The temporal structure of each stimulus is shown in Table 3.

	Onset pattern	**Steady-state pattern**	**Offset pattern**	**Total Duration**
Three-glide complex	**300-ms quarter sine wave**	**1000 ms at 60 dB SPL**	**100-ms reverse quarter sine wave**	**1400 ms**
Central pure-tone glide	**300-ms quarter sine wave**	**1000 ms at 56.4 dB SPL**	**100-ms reverse quarter sine wave**	**1400 ms**
Each of the two flanking pure-tone glides	**300-ms quarter sine wave**	**1000 ms at 44.4 dB SPL**	**100-ms reverse quarter sine wave**	**1400 ms**

Table 3 : Description of the temporal pattern of intensity for the three-glide stimuli.

Each stimulus was comprised of a unique combination of alignment, spacing and direction. There were thirty different stimuli (seeFigures 7a and 7b, on p.34 and p.35). A table of the log-frequency range swept by each glide of these stimuli is presented in Table 4.

Each parallel alignment can be described in terms of the frequency ratios that were maintained throughout gliding (see Table 3,Figures 7a and 7b, on p.31, p.34 and p.35). For the harmonic stimuli, the frequency ratios that were maintained in each of the three spacings were:

A. Parallel conditions	Spectral range from lowest to highest frequency (Hz)	B. Non-parallel conditions	Spectral range from lowest to highest frequency (Hz)
Harmonic/Small		Diverging/Small	
Bottom glide	720 to 2700	Bottom glide	720 to 1350
Middle glide	800 to 3000	Middle glide	800 to 3000
Top glide	880 to 3300	Top glide	880 to 4774
Harmonic/Medium		Diverging/Medium	
Bottom glide	600 to 2250	Bottom glide	600 to 1125
Middle glide	800 to 3000	Middle glide	800 to 3000
Top glide	1000 to 3750	Top glide	1000 to 5426
Harmonic/Large		Diverging/Large	
Bottom glide	400 to 1500	Bottom glide	400 to 750
Middle glide	800 to 3000	Middle glide	800 to 3000
Top glide	1200 to 4500	Top glide	1200 to 6510
Equal-Log/Small		Converging/Small	
Bottom glide	729 to 2734	Bottom glide	360 to 2700
Middle glide	800 to 3000	Middle glide	800 to 3000
Top glide	877 to 3289	Top glide	1273 to 3300
Equal-Log/Medium		Converging/Medium	
Bottom glide	620 to 2325	Bottom glide	300 to 2250
Middle glide	800 to 3000	Middle glide	800 to 3000
Top glide	1032 to 3870	Top glide	1447 to 3750
Equal-Log/Large		Converging/Large	
Bottom glide	473 to 1774	Bottom glide	200 to 1500
Middle glide	800 to 3000	Middle glide	800 to 3000
Top glide	1353 to 5074	Top glide	1736 to 4500
Shifted/Small			
Bottom glide	703 to 2637		
Middle glide	800 to 3000		
Top glide	897 to 3363		
Shifted/Medium			
Bottom glide	583 to 2186		
Middle glide	800 to 3000		
Top glide	1017 to 3814		
Shifted/Large			
Bottom glide	383 to 1436		
Middle glide	800 to 3000		
Top glide	1217 to 4564		

Table 4 : Spectral range swept by each of the three glides of each alignment-by-spacing combination. Note that while upward conditions started from the lower-bound and ended with the upper-bound frequencies, their downward counterparts did the reverse. The central glide is constant across conditions.

large (1:2:3), medium (3:4:5), and small (9:10:11). For each of the three spacings of the equal-log stimuli, the following frequency ratios were constant: large (1:1.69:2.86), medium (1:1.29:1.67), and small (1:1.10:1.20). Finally, the bottom and top glides of the shifted stimuli were obtained by subtracting and adding a constant number of Hz to the bottom end of the 800-3000 Hz glide. This constant was 417 Hz for the closely-spaced stimuli, 217 Hz for the medium-spaced stimuli, and 97 Hz for the widely-spaced stimuli. The following frequency ratios were thus maintained in each of the three spacings of the shifted stimuli: large (1:2.09:3.18), medium (1:1.37:1.74), and small (1:1.14:1.28).

The starting frequencies for the widely-spaced, medium-spaced and closely-spaced diverging stimuli was the same as their harmonic counterparts. As the gliding course proceeded, the log-frequency spacing among partials increased. Actually, there was roughly a six fold, three fold and one-and-a-half fold increment in log-frequency spacing from the beginning to the end of closely-spaced, medium-spaced and widely-spaced diverging glides respectively. Similarly, the terminal frequencies for the widely-spaced, medium-spaced and closely-spaced converging stimuli was the same as their harmonic counterparts. As the gliding course proceeded, the log-frequency spacing among partials decreased. Actually, there was roughly a six fold, three fold and one-and-a-half fold decrement in log-frequency spacing from the beginning to the end of closely-spaced, medium-spaced and widely-spaced converging glides respectively.

To sum up, in the equal-log stimuli, partials maintained an equal log-frequency spacing throughout gliding, thereby keeping constant inharmonic ratios among the components. The harmonic and shifted stimuli maintained the harmonic ratios and the inharmonic ratios among the glides respectively. However, the ratio between the top and the central glide was not equal to the ratio between the central and the bottom ones, as was true for their equal-log counterparts. In the case of non-parallel stimuli, namely diverging and converging partials, the ratios between each

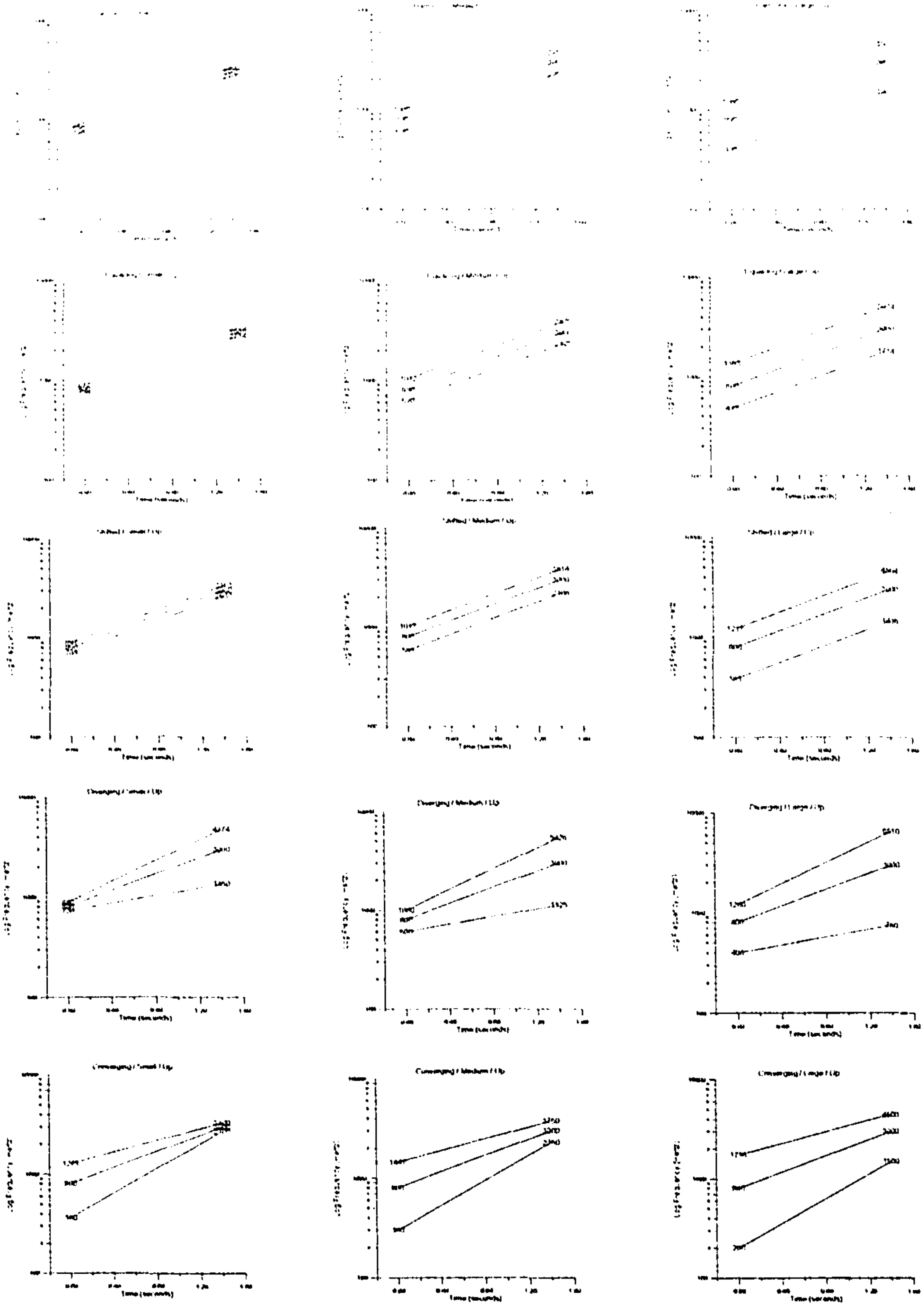

Figure 7a : Schema of all the patterns of stimuli ascending in log frequency. The small, medium and large spacings appear from left to right. The harmonic, equal-log, shifted, diverging and converging alignments are illustrated from top to bottom .

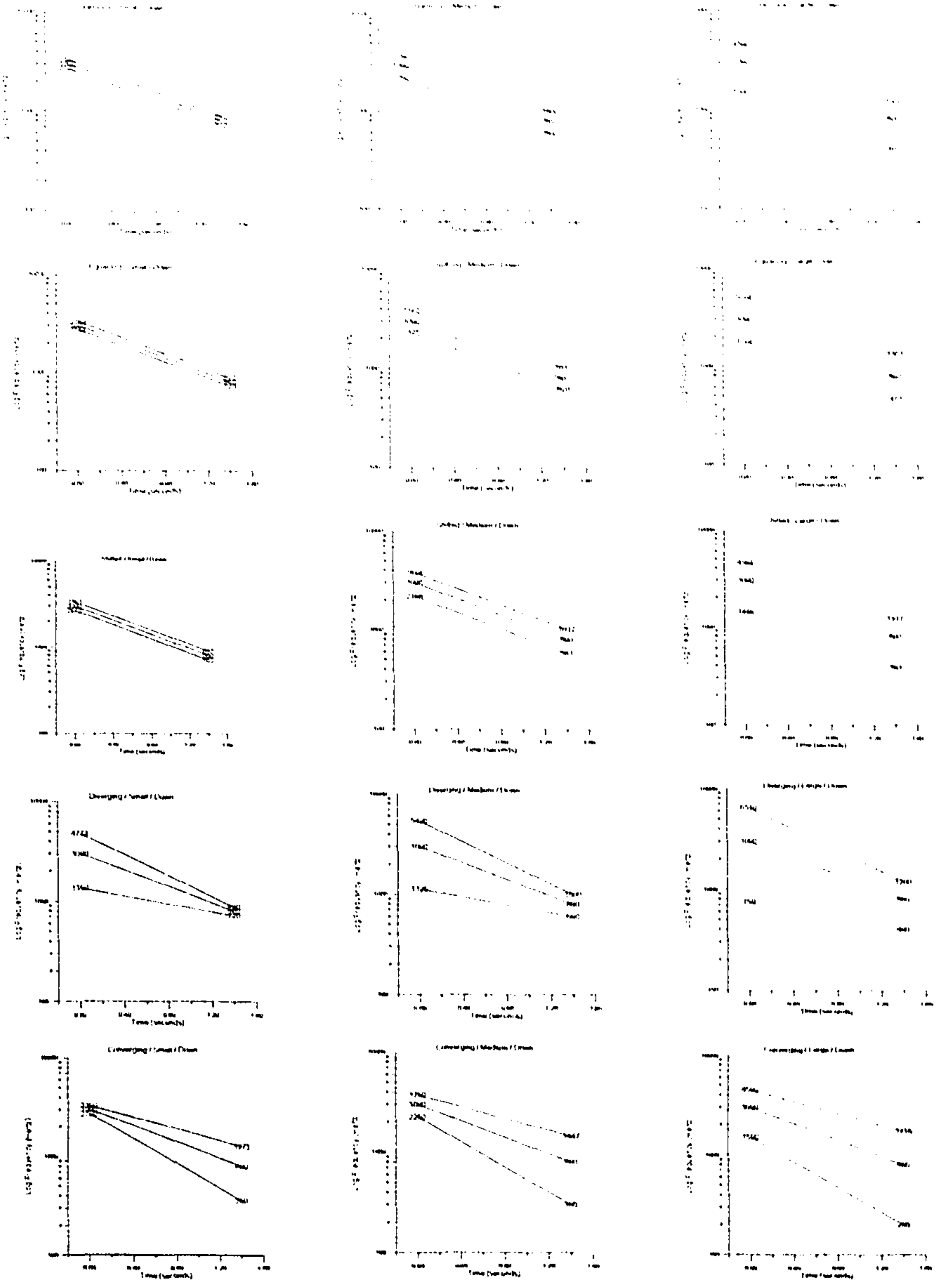

Figure 7b : Schema of all the patterns of stimuli descending in log frequency. The small, medium and large spacings appear from left to right. The harmonic, equal-log, shifted, diverging and converging alignments are illustrated from top to bottom .

pair of adjacent glides were neither maintained nor equivalent for the top-centre and bottom-centre pairs (seeFigures 7a and 7b, on p.34 and p.35).

The different classes of stimuli shared the same intensity and temporal parameters. Moreover, the central glide covered the same frequency range, that is, [800-3000 Hz] in all conditions. While the central glide was constant across conditions, the flanking glides varied (seeFigures 7a and 7b, on p.34 and p.35). The frequency ranges specific to each of the three-glide stimuli are indicated in Table 4. More details about the procedure followed in the synthesis of the various types of stimuli are available in Appendix A.

v. *Apparatus*

All stimuli were synthesized and presented by a PC-compatible 386 computer, using MITSYN Version 8.1 signal processing software (Henke, 1990) This computer controlled a Data Translation DT 2823 16-BIT digital-to-analog converter. The rate of output was 20000 samples per second. Signals were low pass filtered at 8000 Hz through a Rockland Dual Hi/Lo Filter Model 852, with a roll-off of 48 dB/octave, to eliminate aliasing.

Listeners sat in an Industrial Acoustics Company Model 1202 sound-attenuating testing room and listened to stimuli presented binaurally through Sony NR V7 headphones. The experiment was run on-line with the help of a MAPLE Version 1.4 program (Achim, 1991), using ASYST Version 3.1 software. The participants recorded their responses directly into the computer. The intensities of the stimuli were measured using a General Radio Type 1565-B sound-level meter set at "A" weighting and using a flat-plate coupler. The amplitude of each glide was adjusted during synthesis for Fletcher-Munson estimates of equal loudness at 60 dB.

Results

Within the upward and downward set of conditions, each item was compared twice with each of the other fourteen ones, that is, once for each order of the pairs of stimuli. There was thus a total of twenty-eight comparisons for each three-glide complex[2]. On each of the pairwise comparisons, a stimulus obtained a score from one to seven based on the listener's judgments. On the basis of these raw data, a segregation score was computed. This score was a proportion of the maximum score that could be obtained for a given stimulus across all the pairwise comparisons in which it participated. For a given stimulus, a total of 28 trials with a maximum raw score of 7 per trial yields a maximum score of 196. The summary score for a stimulus with 28 comparisons was thus obtained by dividing the sum of its ratings across trials by 196.[3] The mean segregation scores obtained for each condition are presented in the Appendix B (see Table 12, p.67). The summary plots of the distribution of mean segregation scores across the 15 up and 15 down conditions are shown in Figure 8a and 8b respectively.

---

2. Four out of 420 comparisons were inadvertently omitted in the trial table. These four trials were downward comparisons of the harmonic/medium/down with the equal-log/medium/down, and of the equal-log/medium/down with the equal-log/large/down stimuli in the two possible orders of comparison.

3. Two comparisons were absent for both the harmonic/medium/down and the equal-log/large/down conditions. Their summary score was thus obtained by dividing the sum of their respective ratings across 26 trials by 182 (i.e., 26 trials X 7 max. score/trial = 182 max. score). Similarly, for the equal-log/medium/down with four missing comparisons, the summary score was obtained by dividing the sum of its ratings across 24 trials by 168.

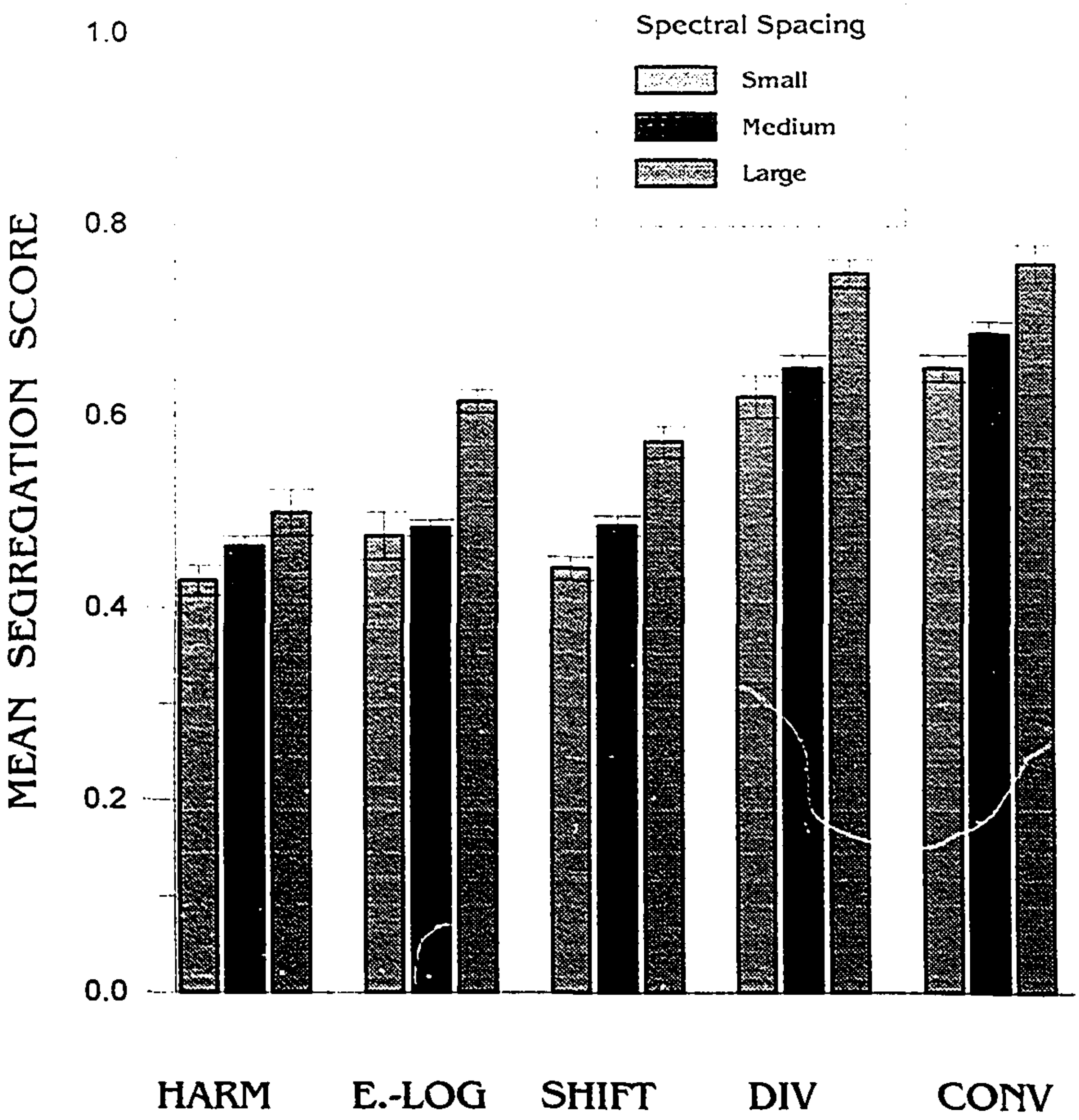

Figure 8a: Distribution of mean segregation scores across upward conditions. 'SMALL', 'MEDIUM' and 'LARGE' are the three levels of spacing. 'HARM', 'E.-LOG', 'SHIFT', 'DIV' and 'CONV' refer to the harmonic, equal-log, shifted, diverging and converging alignments respectively. Scores range from zero to one, where perceived segregation increases with score magnitude. Vertical bars indicate plus or minus one standard error. Each mean score was obtained from 18 subjects.

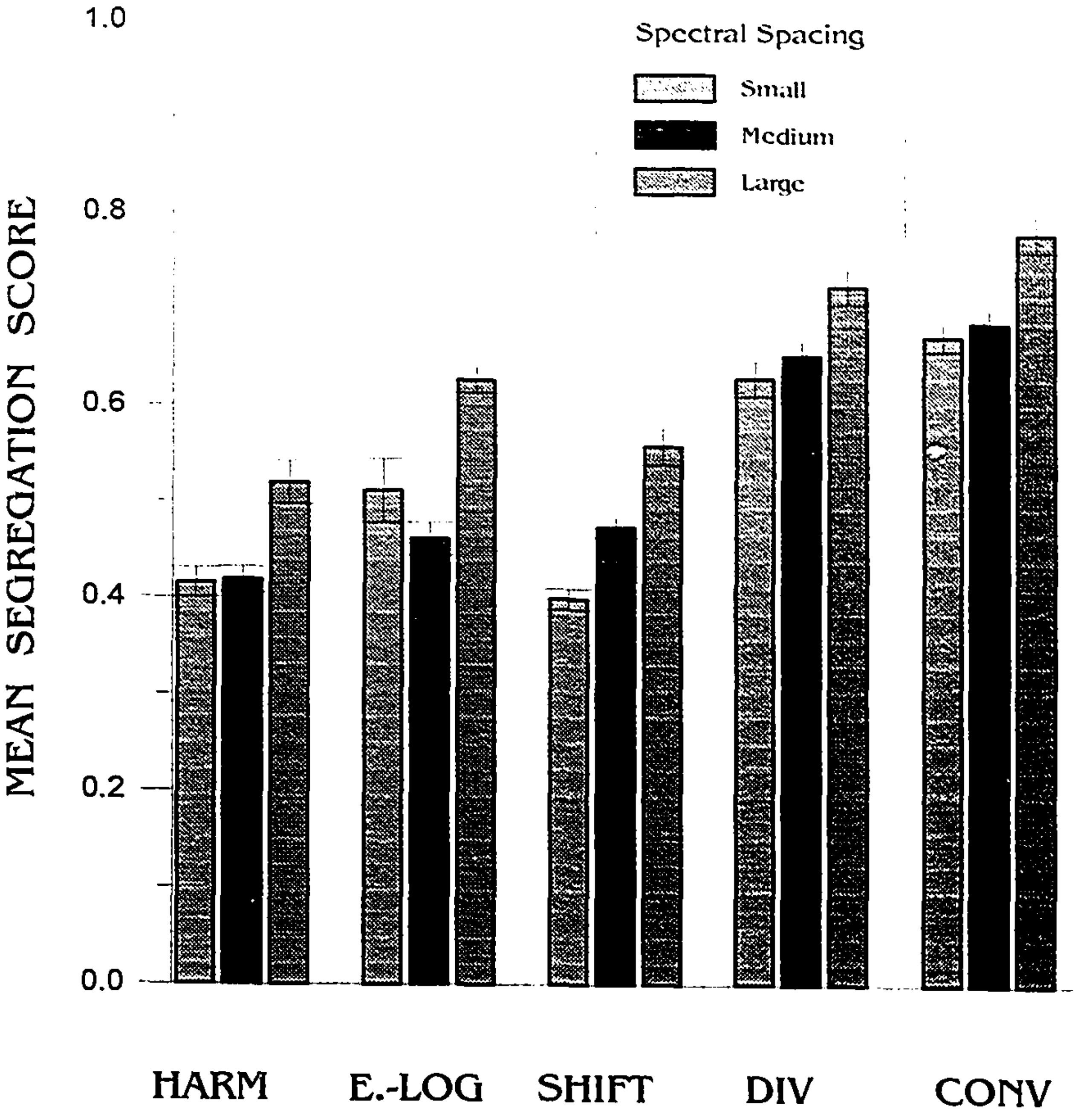

Figure 8b: Distribution of mean segregation scores across downward conditions. 'SMALL', 'MEDIUM' and 'LARGE' are the three levels of spacing. 'HARM', 'E.-LOG', 'SHIFT', 'DIV' and 'CONV' refer to the harmonic, equal-log, shifted, diverging and converging alignments respectively. Scores range from zero to one, where perceived segregation increases with score magnitude. Vertical bars indicate plus or minus one standard error. Each mean score was obtained from 18 subjects.

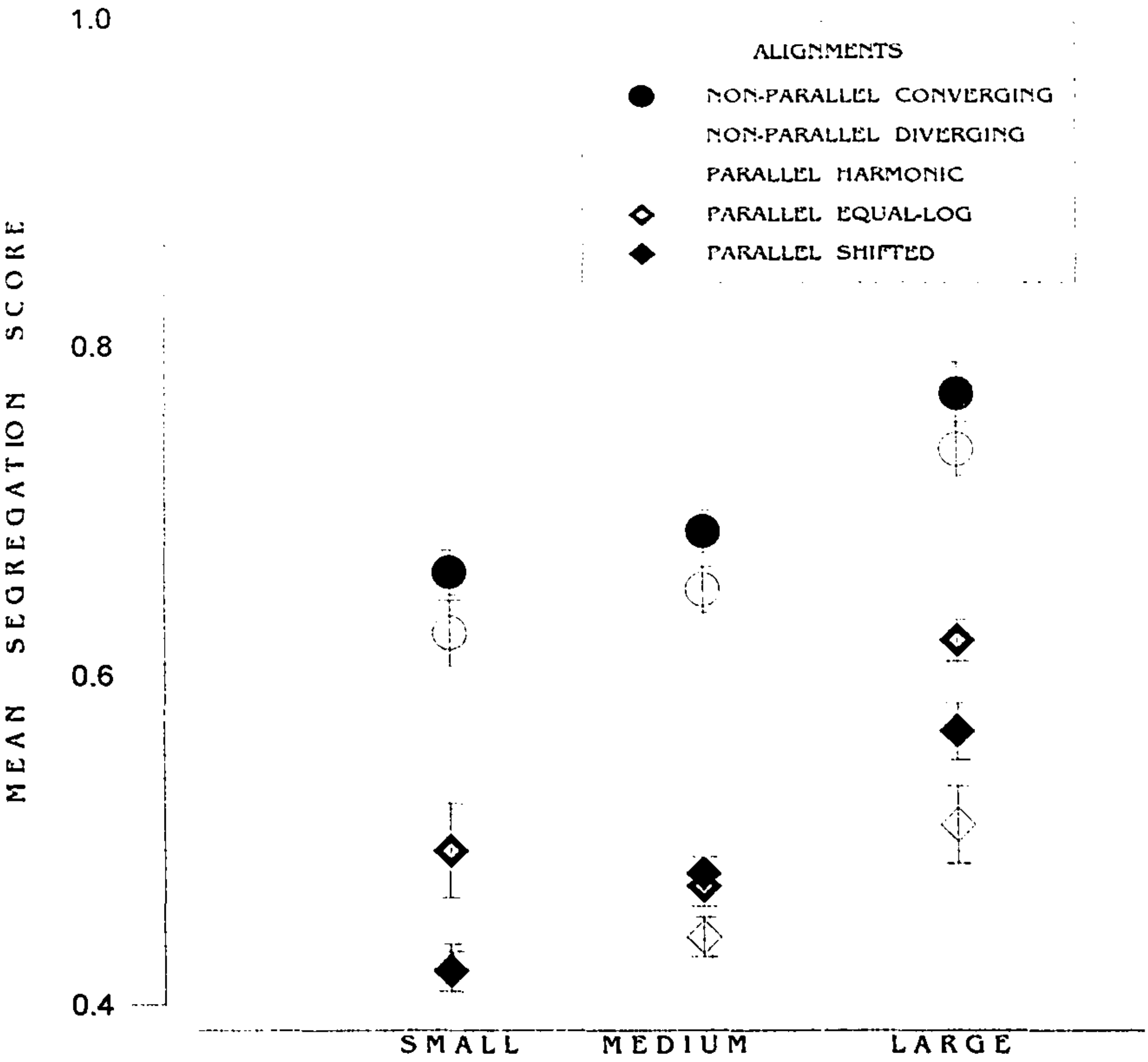

Figure 9:

Mean segregation score of the five alignments as a function of spacing. Scores range from zero to one, where perceived segregation increases with score magnitude. Vertical bars indicate plus or minus one standard error. Each mean score was obtained from 18 subjects.

A plot relating the segregation score for each of the five alignments to spacing is shown in Figure 9. Notice that segregation score increased as a function of both alignment and spacing.

As a first step in exploring the data, each of the sets of upward and downward stimuli was analyzed with a two-way MANOVA using the Greenhouse-Geiser lower-bound correction According to the general MANOVA summary of effects for upward conditions, the alignment main effect [$F = 100.02$, $p < .00001$] and the spacing main effect [$F = 22.93$, $p < .00001$] were very significant (see Table 5). Moreover, there was a two-way interaction between spacing and alignment [$F = 2.96$, $p < .01$].

Effects	df effect	MS effect	df error	MS error	F	Tail p	Greenhouse Geisser p
Alignment	4	0.608761	68	0.00609	100.02	< 1 X 10^{-5}	**< 1 X 10^{-5}**
Spacing	2	0.632621	34	0.01423	22.93	< 1 X 10^{-5}	**< 1 X 10^{-5}**
Al x Spa	8	0.00662	136	0.00223	2.96	0.0046	0.04

Table 5: Two-way MANOVAs for the upward stimuli. The factors were alignment (harmonic, equal-log, shifted, diverging and converging) and spacing (small, medium, large). The significant effects taking into account the lower-bound correction at the 5% level are underlined, while those at the 1% level are bolded.

According to the general MANOVA summary of effects for downward conditions, the alignment main effect [$F = 105.14$, $p < .00001$] and the spacing main effect [$F = 27.44$, $p < .00001$] were very significant (see Table 6). There was also a two-way interaction between spacing and alignment [$F = 3.54$, $p < .01$]. The main effect of alignment and spacing, as well as the alignment-by-spacing interaction effect were thus replicated across the two levels of direction. According to the lower-bound correction, the alignment and spacing main effects were significant at the 1% level and their interaction was significant at the 5% level for both upward and downward stimuli.

Effects	df effect	MS effect	df error	MS error	F	Tail p	Greenhouse Geisser p
Alignment	4	0.74642	68	0.00710	105.14	$< 1 \times 10^{-5}$	**$< 1 \times 10^{-5}$**
Spacing	2	0.36693	34	0.01337	27.44	$< 1 \times 10^{-5}$	**$< 1 \times 10^{-5}$**
Al x Spa	8	0.01179	136	0.00333	3.54	0.0012	0.04

Table 6 : Two-way MANOVAs for the downward stimuli. The factors are alignment (harmonic, equal-log, shifted, diverging and converging) and spacing (small, medium, large). The significant effects taking into account the lower-bound correction at the 5% level are underlined, while those at the 1% level are bolded.

Despite the fact that parallel and non-parallel glides were treated as orthogonal factors in the MANOVAs, a measure of average frequency separation showed that the non-parallel glides were further apart than the parallel ones (see Table 1, p.29). Therefore it was necessary to show that the greater segregation found for non-parallel glides was not due to their greater average separations. This was done in two ways. The first was to perform a set of specific comparisons between selected parallel and non-parallel conditions in which the frequency separation was lower for the non-parallel glides; thus any finding of a greater segregation for these glides could not be due to their frequency separations. The second method was to use a multiple regression analysis in which the average frequency separation for each set of glides was used as a covariate in assessing the effect of parallelness.

Twenty planned comparisons with a Bonferroni adjustment were made (see Table 7, p. 44). Three sets of comparisons were performed. One set compared the levels of the spacing factor against each other. Closely-spaced yielded significantly lower segregation scores than medium-spaced [$F = 63.61$, $p < .00001$] and widely-spaced glides [$F = 230.08$, $p < .00001$]. Moreover, medium-spaced stimuli segregated less than widely-spaced ones [$F = 222.84$, $p < .00001$] (see Table 7c, p.44). Therefore, segregation scores increased as a function of spectral spacing. Another set compared the five alignments in terms of the obtained segregation score (see Table 7a, p.44). The first comparison tested whether non-parallel glides segregated more than their parallel counterparts. The other four compared the segregation scores of the three parallel and the two non-parallel alignments among themselves. The family-wise alpha level was kept constant at 0.05 to yield a 0.01 alpha level per comparison. The analysis indicated that the segregation scores obtained for harmonic, equal-log and shifted conditions combined were lower than those obtained for diverging and converging conditions combined [$F = 231.93$, $p < .00001$]. Thus, the observed trend followed the expected direction of parallel glides fusing more than non-parallel.

Moreover, there were some significant differences in performance among the different levels of parallel alignments. The following pairs of alignment differed significantly from each other: harmonic and equal-log [$F = 31.93$, $p < .0001$], harmonic and shifted [$F = 38.49$, $p < .00001$] and equal-log and shifted [$F = 16.57$, $p < .01$]. In terms of the magnitude of the segregation score obtained, the order from lowest to highest was harmonic, shifted and equal-log. Converging glides tended to yield larger segregation scores than diverging ones [$F = 5.73$, $p < .05$]. However, this difference was not statistically significant with the Bonferroni adjustment [$p > .01$].

Twelve comparisons involved parallel conditions and smaller-spaced non-parallel ones (see Table 7b, p.44). Spacing was defined as the average frequency difference between pairs of glides for each three-glide condition (see Table 1, p.29).

A) Alignment comparisons:	F-value	p	Predicted direction
01) Par. vs Non-Par.	231.93*	0.00001	+
02) Harm. vs E.-Log	31.93*	0.00010	+
03) Harm. vs Shift.	38.49*	0.00001	+
04) E.-Log vs Shift	16.57*	0.00107	0
05) Div. vs Conv.	5.73	0.02705	0
B) Pairwise comparisons in which the average spacing of glides is greater for parallel than for non-parallel stimuli:			
01) Harm./Md. vs Conv./Sl.	210.42*	0.00001	+
02) Harm./Lg. vs Conv./Sl.	25.01*	0.00024	+
03) E.-Log/Md. vs Conv./Sl.	196.50*	0.00001	+
04) E.-Log./Lg. vs Conv./Sl.	5.46	0.03026	+
05) Shift./Md. vs Conv./Sl.	215.78*	0.00001	+
06) Shift./Lg. vs Conv./Sl.	17.13*	0.00095	+
07) Harm./Lg. vs Conv./Md.	55.06*	0.00001	+
08) E.-Log/Lg. vs Conv./Sl.	27.28*	0.00018	+
09) Shift./Lg. vs Conv./Md.	40.41*	0.00001	+
10) Harm./Lg. vs Div./Sl.	9.18	0.00748	+
11) E.-Log/Lg. vs Div./Sl.	0.035	0.83249	+
12) Shift./Lg. vs Div./Sl.	3.49	0.07623	+
			+
C) Spacing comparisons:			
01) Small vs Medium	63.61*	0.00001	+
02) Small vs Large	230.08*	0.00001	+
03) Medium vs Large	222.84*	0.00001	+

Table 7 : Summary of 20 planned comparisons using a Bonferroni adjustment. For each comparison, the number of subjects was 18, the degrees of freedom for the numerator and the denominator were respectively, 1 and 17. For each set of comparisons, A, B and C the family-wise alpha-level was kept constant at 0.05. Thus, the alpha level per comparison was 0.01, 0.004 and 0.017, for the sets A, B and C respectively. Each significant comparison is followed by '*', according to these criteria. The predicted direction is '+' if the observed difference is in the predicted direction, '-', if the observed difference is opposite to the predicted direction, and '0', if there was no predicted difference.

For each spacing-by-alignment condition shown in Table 7b, the results have been combined over the ascending and descending glide directions. The analysis showed that in eight out of twelve such comparisons, the segregation score obtained under non-parallel gliding was significantly higher than under parallel gliding [$p < 0.004$]. That is, for each of those eight significant cases, the probability was less than 0.004, keeping the joint-alpha level at the 5% level for the twelve comparisons. A critical observation is that despite the bias against them, induced by a large average spectral spacing, all parallel glides yielded a smaller segregation score than non-parallel glides closer together on a log-frequency scale. Therefore, all comparisons (significant or not) favoured the hypothesis; two-third of which were statistically significant.

The first comparison in Table 7 showed that parallel and non-parallel conditions were statistically different [$F = 231.93$, $p < 0.00001$]. This difference might be attributed to several characteristics of three parallel glides, not only parallelness, but, harmonicity and smaller average spacing. An ANACOVA is recommended to correct for problems of counfounding (Kleinbaum, Kupper & Muller, 1992). This technique involves a multiple regression model in which the study factors of interest are all treated as nominal variables, whereas the variables being controlled, that is, the covariates, may be measurement on any measurement scale.

An ANACOVA was performed to find out whether there was a significant contribution of parallel log-frequency gliding over and above the contribution of the covariates. The dependent variable was the mean segregation score obtained from 18 subjects. Two predictors were dichotomous variables: harmonic status (harmonic vs. non-harmonic) and parallelness status (parallel vs. non-parallel). Another predictor was the average spacing, that is, the mean frequency difference between glides for each of the thirty conditions (see Table 1, p.29). A model with three predictors, namely average spacing, harmonic status and parallelness status, was tested against a null-hypothesis model with only the average spacing and harmonic status covariates (see Table 8, p.46). Both models included an intercept. The partial correlation coefficient yielded from the model comparison was 0.8. That is, the ratio of the reduction in error sum of squares contributed by the parallelness variable to the error sum of squares obtained by fitting an intercept and the covariates to the data was 0.64. This contribution of parallelness to the proportion of variance accounted for by the multiple regression is strongly significant [$F = 46.33$, $p < 0.001$].

Full Model (H_0):

$$\text{MeanSeg} = B_1(\text{AveSpa}) + B_2(\text{HarmSta}) + B_3(\text{ParSta}) + b_0 ;$$

Restricted Model (H_1):

$$\text{MeanSeg} = B_1(\text{AveSpa}) + B_2(\text{HarmSta}) + b_0 .$$

SS effect	df effect	SS error	df error	R^2	F	p
0.08631	1	0.04844	26	0.64	46.33	0.001

Table 8: Results from a multiple regression analysis for a within-subject design with 30 conditions and 18 subjects. The dependent variable 'MeanSeg' is the mean segregation score across subjects for each condition. 'HarmSta' stands for harmonic status. It classifies conditions into either a harmonic or an inharmonic category. 'ParSta' refers to a dichotomous variable categorizing conditions into either a parallel or a non-parallel class. 'AveSpa' refers to the average frequency spacing between glides for each stimulus. The proportion of variance accounted for by a model with an intercept and three predictor variables (AveSpa, HarmSta, and ParSta) relative to a null-hypothesis model with an intercept and the covariates (AveSpa and HarmSta) was 64 % [$F = 46.33$, $p < 0.001$].

Finally, a stepwise regression procedure was conducted. This procedure involves choosing the variable which decreases the error sum of squares by the largest amount for all possible predictors not already in the equation. This evaluated the ordered contribution of parallelness, average spacing, and harmonicity to the reduction of the observed variance among mean segregation scores as 73%, 12% and 2% respectively (see Table 9, p. 47)[4]. Parallelness, being the first predictor [$R^2 = 0.73$] has the highest correlation [$r = 0.85$] with the mean segregation scores. Parallel log-frequency gliding is thus the strongest predictor for the observed data

Step	Variable entered	Number Removed In	Partial R^2	Model R^2	C(p)	F	p
1	**ParSta**	1	**0.7260**	0.7260	30.8112	**74.2072**	**0.0001**
2	**AveSpa**	2	**0.1245**	0.8505	6.9948	**22.4890**	**0.0001**
3	**HarmSta**	3	**0.0241**	0.8746	4.0000	**4.9948**	**0.0342**

Table 9 : Summary of a stepwise regression analysis for a within-subject design with 30 conditions and 18 subjects. The dependent variable was the segregation score for each condition, averaged over 18 subjects. The first, second and third predictor variables were the conditions' parallelness status (parallel or non-parallel), average spacing and harmonic status (harmonic or inharmonic) respectively.

4. It was verified that there is no significant decrease in the ability to account for variation upon deletion of the interactions from the list of predictors [$F = 2.46$, $p > 0.05$]. The homogeneity-of-regression assumption was thus satisfied (Kleinbaum, Kupper & Muller, 1992).

Discussion

In an attempt to interpret the experimental results, let us examine some critical questions and the data trends relevant to answering them. Firstly, what is the role of spectral spacing in the fusion of concurrent tonal glides? The main effect of spacing upon segregation score turned out to be strongly significant for both upward [$F = 22.93$, $p < 0.00001$] and downward conditions [$F = 27.44$, $p < 0.0001$]. Moreover, the three planned comparisons for the spacing variable showed that segregation scores increase from small to medium [$F = 63.61$, $p < 0.00001$], and from medium to large [$F = 222.84$, $p < 0.00001$]. The stepwise regression analysis with average spacing, parallelness status (parallel or non-parallel) and harmonic status (harmonic or inharmonic) as predictors for mean segregation scores indicated that average spacing accounts for 12% of the variance over and above parallelness status. This is a strongly significant contribution statistically [$F = 22.50$, $p < 0.0001$]. It has been exhaustively shown that spectral proximity is a sequential-grouping heuristic in auditory processing (Bregman & Campbell, 1973; Deutsch, 1975; Dowling, 1973; Norden, 1975, 1977). However, its influence in simultaneous spectral analysis has never been thoroughly demonstrated. The present set of results is thus informative. The relations between the various stimuli's segregation scores obtained from the small, medium and large spacing categories in the MANOVAs as well as from the average-spacing covariate in the regression analysis confirm that in addition to promoting sequential integration, spectral proximity encourages the fusion of simultaneous components.

What is the contribution of average spacing to the segregation scores of the three parallel alignments? First, consider that the parallel alignments have a very similar overall average spacing: 717.78 Hz for harmonic, 737.11 Hz for equal-log and 771.56 Hz for shifted. In terms of their geometric mean ratios, they are almost equivalent, that is: 1.54 for harmonic, 1.51 for equal-log and 1.60 for shifted (see Table 11, p.66). Despite that equivalence of average spacing, the main predicted trend across alignments shows up, namely, harmonic conditions get a significantly lower segregation score than both equal-log [$p < .00001$] and shifted [$p < .00001$]. Thus there does not seems to be any counfounding effect of average spacing within parallel alignments.

The role of spectral spacing in the comparison between converging and diverging glides can also be assessed. Converging glides tended to get a higher segregation score than diverging ones [$F = 5.73$, $p < .05$]. However, the average spacing across converging conditions (966.22 Hz) was lower than across diverging ones (1694.45 Hz). Spectral spacing differences would thus predict the reverse trend, namely, converging partials should segregate more easily than diverging ones since they were in general further apart. It is thus reasonable to conclude that the observed differences among the two non-parallel alignments in perceived segregation can be attributed to their respective alignments, not to some spectral-spacing artifact.

What is the role of parallel log-frequency gliding upon the fusion of partials? Although the comparison of parallel with non-parallel three-glide stimuli was strongly significant [$F = 231.93$, $p < 0.00001$], the alignment and spacing variables were not completely orthogonal. On average, pairs of glides in the non-parallel stimuli were further apart than in the parallel ones: 1307.81 Hz for non-parallel versus 742.15 Hz for parallel. Note however that this difference in average spacing is of a lesser magnitude when it is evaluated using a geometric-mean-ratio criterion, that is: 2.20 for non-parallel versus 1.61 for parallel (see Table 11, p.66). A multiple regression analysis was conducted in an attempt to remove the influence of average frequency spacing upon the mean segregation scores. The analysis showed that there was a significant contribution of log-frequency parallelness over and above the contribution of average frequency spacing and harmonicity to the proportion of variance accounted for by the predictor variables [$F = 46.33$, $p < 0.001$]. Moreover, the stepwise procedure evaluated that the variable classifying conditions into a parallel and a non-parallel classes was the most important predictor for the observed data [$R^2 = 0.73$].

To see whether non-parallel glides segregated more than parallel ones, even when spectral spacing did not favor this result, some selected parallel alignment-by-spacing conditions were compared with non-parallel ones. Each pair in this subset was composed of a parallel condition with a greater average spectral spacing than a non-parallel one. In each of eight out of twelve such comparisons, the parallel condition obtained a significantly lower segregation score than the non-parallel one [$p < .004$]. This suggests that even though in two thirds of these comparisons, the larger spectral spacing among the parallel glides favored their segregation, their alignment encouraged fusion, and the influence of the latter was stronger. Moreover, the multiple regression analysis showed that differences across the conditions' segregation scores were neither entirely due to differences in average spectral spacing nor to harmonicity. Results yielded by the above comparisons as well as those from the multiple regression analysis thus support the main hypothesis that a parallel log-frequency trajectory fuses partials. However, gliding components were never compared with steady-state tones. Consequently, to thoroughly demonstrate that log-frequency parallelness fuses partials, its influence in both steady-state (slope = 0 Hz/sec) and gliding complexes (slope > 0 Hz/sec) will have to be investigated and teased apart from harmonicity.

The interaction of spacing and alignment significantly influenced perceived segregation both within upward-gliding [$F = 2.96$, $p < 0.01$] and downward-gliding stimuli [$F = 3.54$, $p < 0.01$] [5]. However, for both upward and downward stimuli the overall alignment-by-spacing effect was significant at only the 5% level according to the lower-bound correction (see Tables 5 and 6, on p.41 and p.42). The existence of an interaction is not surprising, since the non-parallel glides were further apart than parallel ones on the average. It is thus likely that the influence of

---

[5] Note that the omission of four downward trials might have slightly bias the interaction effect between alignment and spacing (R. Amsel, personal communication, February 15, 1994, McGill University). A conservative interpretation is thus recommended for the alignment-by-spacing interaction of downward stimuli.

the three categories of spacing differed for the non-parallel (diverging, converging) and the overall smaller-spaced parallel alignments (harmonic, equal-log, shifted). It could also be that parallelness and spectral proximity have a synergistic effect upon fusion

Does the maintenance of simple harmonic ratios promote perceived segregation? Firstly, let us look at the role of harmonicity within parallel alignments. Results indicated that harmonic stimuli yield significantly lower segregation scores than both equal-log [$F = 31.93$, $p < 0.00001$] and shifted ones [$F = 38.49$, $p < 0.00001$]. These results are compatible with past empirical evidence for harmonicity as a simultaneous-grouping cue (Bregman & Doehring, 1984; De Witt & Crowder, 1987; McAdams, 1982, 1983a, 1983b, 1984). However, the overall contribution of harmonicity across the five alignments was small. In the stepwise regression analysis, a dichotomous variable classifying alignment conditions in either a harmonic or a non-harmonic category was used in addition to two other predictor variables: average spacing and parallelness status (parallel or non-parallel). The harmonic status contributed to an extra 2% in the reduction of the variance among mean segregation scores over and above average spacing and parallelness status. Although that contribution was significant [$p < 0.05$], it was small compared to the 12% for average spacing and 73% for alignment status.

Harmonicity is known to have strong effects on the fusion of steady-state tones (Bregman, 1990). Why did harmonicity play such a small role in the fusion of gliding tones in the present study? One might speculate that the harmonicity detector, whatever its nature, operates under specific constraints. For instance, it might detect harmonicity among a steady-state series of partials (e.g. 300 Hz, 400 Hz and 500 Hz tones, i.e., 3:4:5 ratio of a 100-Hz F0) more easily than in a dynamically changing version (e.g. 300-1200 Hz, 400-1600 Hz and 500-2000 Hz glides, i.e., 3:4:5 ratio of a 100-400 Hz gliding F0). Moreover, it might function best within certain spectral limits. For instance, while the frequency shift away from a perfect harmonic ratio might be resolved for the widely-spaced shifted glides, it might remain undetected for the closely-spaced ones. Indeed, upon repetitive listening sessions to the stimuli, the experimenter as well as some colleagues from the Speech-and-Hearing Laboratory of McGill failed to hear any difference between the closely-spaced parallel alignments.

Pattern-recognition models have been proposed for the detection of specific harmonic series (Goldstein, 1973; Moore, 1989). These models require the auditory system to recognize the incoming pattern of partials as a set of harmonics that is appropriate for some particular F0. Models of this type might perform a "best fit" calculation on the set of estimated frequencies (Bregman, 1990). The probabilistic nature of such models suggest that imperfect ratios among frequency components can be treated as harmonic. Indeed, a set of steady-state partials shifted from a perfect harmonic ratio by a small percentage of the best-fit F0 evokes a pitch slightly higher than the F0 pitch (Bregman, 1990; Moore, 1989). For instance the set of 107 Hz, 307 Hz and 507 Hz partials (7-Hz shift from a 3:4:5 ratio of a 100-Hz F0) yield a pitch just a bit higher than the 100-Hz pitch. Such a template-matching mechanism might also function optimally with a certain number of partials. For instance, four partials might define better a given harmonic series than three. In turn, a series composed of five harmonics might evoke the F0 pitch more strongly than when it is composed of four. This might go on up to a certain ceiling number of partials. For instance, the harmonicity detector might operate maximally when five or more partials are integer multiples of a particular F0. Therefore, it is possible that in the present experiment more than three gliding partials in the harmonic conditions would have produce a more powerful fusion.

Does the maintenance of an equal log-frequency spacing among inharmonic components matter in perceived segregation? Overall, equal-log conditions obtained larger segregation scores than their shifted counterparts [$F = 16.57$, $p < 0.001$]. The observed order of the types of parallel log-frequency gliding in terms of the magnitude of the segregation score is thus from lowest to highest: harmonic, shifted and equal-log. Thus, contrary to the prediction that shifted stimuli segregate more than equal-log stimuli, a smaller segregation score was obtained by the shifted stimuli on average. This suggests that the maintenance of an equal spacing between adjacent parallel glides on a log-frequency scale is not a regularity exploited by the auditory system to fuse components. It is also possible that unless the shift from perfect harmonic ratios was large enough, the auditory system might have treated the set as approximately harmonic. This would explain why the overall segregation-score difference between parallel alignments (shifted, equal-log and harmonic) is small relative to the difference between parallel and non-parallel alignments (diverging and converging) (see Figure 9, p.40).

Finally, I would like to relate the above results of parallelness gliding to Carlyon's results on FM coherence (Carlyon, 1990). Carlyon's data showed that listeners can only distinguish coherently from incoherently modulated complex tones when FM coherence issued from harmonicity among partials. In other words, highly-trained listeners cannot in general detect FM coherence. I would like to argue that while Carlyon's conclusion is compatible with the view that there is no evolutionary use for a "parallelness detector" given a harmonicity-detection mechanism, the present results suggest that under some particular conditions the auditory system does detect a parallel log-frequency change.

Conclusion

The main hypothesis of the present study was that a parallel trajectory of log-frequency change among partials encourages their perceptual fusion into one cohesive gliding sound. Various sources of empirical evidence support that hypothesis (Bregman & Doehring, 1984; Chalikia & Bregman, 1990; Halpern, 1977; Lipscombe, 1992a, 1992b; McAdams, 1984). However, in the current experiment, two other factors were found to influence perceived fusion: harmonicity and spectral spacing. It was expected that the harmonic alignment would best induce fusion. This hypothesis was based on past experiments indicating that fusion is promoted by simple harmonic ratios among steady-state tones (De Witt & Crowder, 1987; McAdams, 1982, 1983a, 1983b, 1984) and gliding tones (Bregman & Doehring, 1984). Furthermore, the role of spectral proximity in sequential grouping (Bregman, 1990) was expected to play its part in simultaneous grouping. This was supported by some findings suggesting that spectral spacing hinders the fusion of a high AM tone and a low AM tone (Bregman A. S., Abramson J. & Doehring, 1985). All expectations were supported by the data and occurred in both the ascending and the descending three-glide stimuli.

The most plausible interpretation of the results appears to be that parallelness, spacing and harmonicity among partials gliding on a log-frequency continuum, all partially influence perceptual fusion. However, each does not contribute equally to the fusion of glides. Parallelness along log-frequency-by-time coordinates seems to play the most consequential role. The next strongest influence is by the frequency spacing. Finally, the part played by harmonicity is also substantial, though of a lesser magnitude.

Additional research is needed to replicate those results. It would be preferable to use a different methodology insuring that the spacing and alignment truly vary independently of each other. One way to achieve that would be to perform pairwise comparisons of two types: i.

comparisons between stimuli with the same average spacing, but of different alignments; and ii. comparisons between different average spacings of the same alignment. Any other design insuring that within each level of spacing the average spacing is constant would be acceptable. The average spectral spacing could be assessed either on a linear or on a logarithmic scale, but it should be consistent in the experimental design. To rely upon the maximum or minimum spacing criteria would be problematic due to the inherent constraints of parallel and non-parallel alignments. That is, the diverging and converging three-glide stimuli necessarily involve a greater difference between the maximum and minimum values of the frequency range covered than their equal-average-spacing parallel counterparts. This applies whether the spacing is measured on a linear or on a logarithmic frequency scale. Stimuli of one of two possible alignments could be used, namely, parallel or non-parallel. Moreover, two categories of spacings would be sufficient to show whether there is a spectral-spacing effect. By involving fewer stimuli, this simple 2 x 2 design has the advantage of allowing for more comparisons per condition, i.e., more replications per data point.

To cope effectively with constantly changing natural forces, it is crucial for an individual to detect other living creatures, both similar and different. For example, a finely-tuned auditory system should allow one to differentiate the howl of a potentially threatening wolf from both the voice of a rescuing human and the howl of a friendly dog. In nature, series of harmonics sharing the same pattern of temporal change emanate from most individual living sources. Consider that parallel log-frequency change follows from harmonicity; however it does not entail it. The proposition that common spectral change is not a strong grouping cue relative to others, such as frequency proximity and harmonicity thus makes sense. From an evolutionary point of view, what would be the adaptive value of a parallel log-frequency-change detector, if in all probable situations, that information is obtained from another intermediary, such as a template-matching mechanism for harmonic series? Indeed in a natural environment, an harmonicity-detection mechanism would most of the time do the job of grouping "acoustic elements changing in the same way at the same time" into a cohesive perceptual unit. However, it would not be infallible. For instance it could not detect parallel log-frequency change among inharmonic partials. The crucial conclusion of the present experiment is thus that despite the absence of an evolutionary rationale, it seems that the auditory system is engineered in such a way as to use a regularity of log-frequency change among inharmonic components. However, it is important to realize that the present findings were obtained using a specific set of intensities, range of frequencies and rate of frequency change of the three glides (see Tables 3 and 4, on p.31 and p.32) . An important issue thus concerns the extent to which the present conclusions can be generalized beyond the slow rate, frequency ranges, and intensities of the glides used in this experiment.

References

Bregman, A. S. (1978a). Auditory streaming: Competition among alternative organizations. Perception and Psychophysics, 23, 391 - 398.

Bregman, A. S. (1978b). The formation of auditory streams. In J. Requin (Ed)., Attention and Performance VII. Hillsdale, N J : Erlbaum.

Bregman, A. S. (1981). Asking the "what for" question in auditory perception. In M. Kubovy & J. R. Pomerantz, Perceptual Organization. Hillsdale, NJ : Erlbaum.

Bregman, A. S. (1984). Auditory scene analysis. Reprint from Proceedings of the IEEE Seventh International Conference on Pattern Recognition, Montreal, Canada.

Bregman, A. S. (1990). Auditory scene analysis: The perceptual organization of sound. Cambridge, MA: The MIT Press.

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

Bregman, A. S. & Campbell, J. (1973). Primary auditory stream segregation & the perception of order in a rapid sequence of tones. Journal of Experimental Psychology, 89, 19-31.

Bregman, A. S. & Doehring, P. (1984). Fusion of simultaneous tonal glides: The role of parallelness and simple frequency relations. Perception and Psychophysics, 36(3), 14-30.

Bregman, A. S., Halpern, M., Halikia, M., & Doehring, P. (1986). Partitioning of a Set of Simultaneous Tonal Glides. Unpublished manuscript, Psychology Department, McGill University.

Bregman, A. S. & Pinker, S. (1978). Auditory streaming & the building of timbre. Canadian Journal of Psychology, 32, 19-31.

Carlyon, R. P. (1990). Discriminating between coherent and incoherent frequency modulation of complex tones. Journal of the Acoustical Society of America, 89 (1), 329-340.

Chalikia, M. H. & Bregman, A. S. (1990). The perceptual segregation of simultaneous vowels with harmonic, shifted, and random components. Perception and Psychophysics, 46, 487-496.

Chowning, J. M. (1980). Computer synthesis of the singing voice. Sound Generation: Winds, Strings, Computers. Stockholm: Royal Swedish Academy of Music, Publ. #29. Kungl. Musikaliska Akademien.

De Witt, L. A. & Crowder, R. G. (1987). Tonal fusion of consonant musical intervals. Perception and Psychophysics, 41, 73-84.

Deutsch, D. (1975). Two-channel listening to musical scales. Journal of the Acoustical Society of America, 57, 1156-1160.

Dowling, W. J. (1973). The perception of interleaved melodies. Cognitive Psychology, 5, 322-337.

Fletcher, H. (1940). Auditory patterns. Rev. Mod. Phys., 12, 47-65.

Gardner, R. G. & Darwin, C. J. (1986). Grouping of vowel harmonics by frequency modulation: absence of effects on phonemic categorization. Perception and Psychophysics, 40, 183-187.

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.

Grose, H. & Hall, J. W. III. (1990). The effect of modulation coherence on signal threshold in frequency-modulated noise bands. Journal of the Acoustical Society of America, 88, 703-710.

Hall, J. W. III. & Grose, H. (1990). Comodulation masking release and auditory grouping. Journal of the Acoustical Society of America, 88 (1), 119-125.

Halpern, L. (1977). The Effect of Harmonic Ratio Relationships on Auditory Stream Segregation. Unpublished research report, Psychology Department, McGill University, Montreal.

Henke, W. L. (1990). MYTSYN Languages. Belmont, MA: Author.

Judd, T. (1977). An Explanation of Deutsh's Scale Illusion. Unpublished Manuscript, Department of Psychology, Cornell University.

Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1992). Applied regression analysis and other multivariable methods. PWS-KENT Publishing company, Boston.

Lipscombe, J. (1992a). The contribution of common gliding motion to the perceptual grouping of simultaneous sounds. Unpublished undergraduate paper, McGill University, Montreal.

Lipscombe, J. (1992b). A replication of "The contribution of common gliding motion to the perceptual grouping of simultaneous sounds". Unpublished undergraduate paper, McGill University, Montreal.

McNabb, M. (1981). Dreamsong: The composition. Computer Music Journal, 5, 36-53.

McAdams, S. (1982). Spectral fusion and the creation of auditory images. In: M. Clynes (Ed.) Music, Mind and Brain: The Neuropsychology of Music. New York: Plenum.

McAdams, S. (1983a). Acoustic cues contributing to spectral fusion. Proceedings of the International congress of Acoustics, Paris, France, 3, 127-130.

McAdams, S. (1983b). L' image auditive: Une métaphore pour la recherche musicale et psychoacoustique. In: ' Un Portrait ': IRCAM. Paris: IRCAM.

McAdams, S. (1984). Spectral fusion, spectral parsing, and the formation of auditory images. Unpublished doctoral dissertation, Stanford University, California.

McAdams, S. (1989). Segregation of concurrent sounds. I. Effects of frequency modulation coherence. J. Acoust. Soc. Am., 86 (6), 2148-2159.

McAdams, S. & Bregman, A. S. (1979). Hearing Musical Streams. Computer Music Journal, 3 (4), 22-44.

McAdams, S., & Wessel, D. (1981). A general synthesis package based on principles of auditory perception. International Computer Music Conference, Denton, Texas.

McPherson, L. M. P., Ciocca, V. & Bregman, A. S. (in press). Auditory tracking of frequency glides through crossing: Effect of difference in slope. Perception and Psychophysics.

Moore, B. C. J. (1989). An introduction to the psychology of hearing (3 rd ed.). London: Academic Press.

Norden, L. P. A. S. Van. (1975). Temporal Coherence in the Perception of Tone Sequences. Unpublished doctoral dissertation, Eindhoven University of Technology, Eindhoven, The Netherlands.

Norden, L. P. A. S. Van. (1977). Minimum differences of level and frequency for perceptual fission of tone sequences ABAB. Journal of the Acoustical Society of America, 61, 1041-1045.

Patterson, R. D. & Moore, B. C. J. (1986). Auditory filters and excitation patterns as representation of frequency resolution. In Frequency Selectivity in Hearing, (ed. B. C. J. Moore), Academic Press, London and New York.

Rasch, R. A. (1978). The perception of simultaneous notes such as in polyphonic music. Acustica, 40, 21-23.

Schooneveidt, G. P. & Moore, B. C. J. (1988). Failure to obtain comodulation masking release with frequency modulated maskers. Journal of the Acoustical Society of America, 83, 2290-2292.

Singh, P. G. (1985). Spectral spacing and spectral locus as determinants of the perceptual organization of complex tone sequences. Journal of the Acoustical Society of America, Suppl. 1 78, S63.

Singh, P. G. (1987). Perceptual organization of complex-tone sequences: A tradeoff between pitch & timbre? Journal of the Acoustical Society of America, 82 (3), 886-899.

Singh, P. G. (1990). Perceptual correlates of spectral changes in complex tones. Unpublished doctoral dissertation, Washington University, St-Louis, Missouri.

Tougas, Y., & Bregman, A. S. (1990). Auditory streaming and the continuity illusion. Perception and Psychophysics, 47, 121-126.

Turgeon, M. (1992). The implication of synchrony of rise, rate of rise and rate of prior decay in intensity on the perceptual fusion of two pure tones. Unpublished undergraduate paper, McGill University, Montreal.

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

Zwicker, E., Flottorp, G., & Stevens, S. S. (1957). Critical bandwidth in loudness summation. Journal of the Acoustical Society of America, 29, 548-557.

Appendix A

The various types of parallel three-glide stimuli were designed in the following manner. a) The lower-bound frequency of the central glide, that is, 800 Hz, served as the base frequency of an exponential oscillator for the synthesis of each gliding partial. In MITSYN Version 8.1 signal processing software (Henke, 1990), an exponential oscillator is characterized by a frequency input that is internally transformed by exponentiation. The frequency of the oscillator is obtained by the following function: Exponential-Oscillator Frequency = (Base Frequency) x $2^{(\text{Frequency input})}$. Thus, each unit increase in the input raises the frequency one octave. Accordingly, the exponential factor to raise a base frequency of 800 Hz to 3000 Hz (1.875 octave) is equivalent to: $[\log_{10}(\text{Upper-bound freq./Lower-bound freq.})]/\log_{10} 2 = [\log_{10}(3000 / 800)]/\log_{10} 2 = 1.907$. b) Given a 800-3000 Hz range swept by a glide, log-frequency parallel trajectories can be obtained by either adding or subtracting a constant to its exponential factor. Therefore, for the synthesis of the top and bottom glides, a constant was added to the central-glide exponent and subtracted from it, respectively. Different values are required to yield each of the various parallel conditions (see Table 10a, p.64).

The different non-parallel conditions were designed in the following manner. a) The 800-3000 Hz frequency range swept by the central component was the same as in the parallel conditions; hence a constant exponential factor of 1.907 was used across all conditions (see Table 10b, p.65). b) To yield non-parallel log-frequency gliding trajectories, exponential factors specific to each glide were needed. Thus, a different exponent was used for each partial within each non-parallel three-glide complex (see Table 10b, p.65). Note that for diverging and converging glides, the exponential factor was constant across small, medium and large spacings.

Alignment-by-Spacing Condition	Glide	Lower-Bound Frequency(Hz) (Starting freq. for upward glides and terminal freq. for downward glides.)	Exponential Factor
A. Parallel conditions		**Note that the 800 Hz lower-bound frequency of the central glide serves as the base for the exponential oscillator.**	**Exp(central) - k = Exp(bottom) Exp(central) + k = Exp(top)**
Harmonic/Small (9:10:11)	Bottom	720	1.907 - 0.152 = **1.755**
	Central	**800**	**1.907**
	Top	880	1.907 + 0.137 = **2.044**
Harmonic/Medium (3:4:5)	Bottom	600	1.907 - 0.415 = **1.492**
	Central	**800**	**1.907**
	Top	1000	1.907 + 0.322 = **2.229**
Harmonic/Large (1:2:3)	Bottom	400	1.907 - 1.000 = **0.907**
	Central	**800**	**1.907**
	Top	1200	1.907 + 0.585 = **2.492**
Equal-Log/Small	Bottom	729	1.907 - 0.133 = **1.774**
	Central	**800**	**1.907**
	Top	877	1.907 + 0.133 = **2.040**
Equal-Log/Medium	Bottom	620	1.907 - 0.367 = **1.540**
	Central	**800**	**1.907**
	Top	1032	1.907 + 0.367 = **2.274**
Equal-Log/Large	Bottom	473	1.907 - 0.758 = **1.149**
	Central	**800**	**1.907**
	Top	1353	1.907 + 0.758 = **2.665**
Shifted/Small	Bottom	703	1.907 - 0.186 = **1.720**
	Central	**800**	**1.907**
	Top	897	1.907 + 0.165 = **2.072**
Shifted/Medium	Bottom	583	1.907 - 0.456 = **1.451**
	Central	**800**	**1.907**
	Top	1017	1.907 + 0.346 = **2.253**
Shifted/Large	Bottom	383	1.907 - 1.063 = **0.844**
	Central	**800**	**1.907**
	Top	1217	1.907 + 0.605 = **2.512**

Alignment-by-Spacing Condition	**Glide**	**Lower-Bound Frequency(Hz)** (Starting freq for upward glides and terminal freq for downward glides)	**Exponential Factor**
B. Non-parallel conditions		**Note that each glide's lower-bound frequency serves as the base for the exponential oscillator.**	**Note that exponential factors are constant across the 3 spacings for the 3 glides.**
Diverging/Small	Bottom	720	0.907
	Central	800	1.907
	Top	880	2.439
Diverging/Medium	Bottom	600	0.907
	Central	800	1.907
	Top	1000	2.439
Diverging/Large	Bottom	400	0.907
	Central	800	1.907
	Top	1200	2.439
Converging/Small	Bottom	360	2.907
	Central	800	1.907
	Top	1273	1.374
Converging/Medium	Bottom	300	2.907
	Central	800	1.907
	Top	1447	1.374
Converging/Large	Bottom	200	2.907
	Central	800	1.907
	Top	1736	1.374

Table 10 : Parameters used in synthesis of the various alignment-by-spacing conditions.

A. Parallel Conditions	**Geometric Mean Ratio**
01) Harmonic / Small	1.14
02) Harmonic / Medium	1.41
03) Harmonic / Large	2.08
04) Equal-log / Small	1.13
05) Equal-log /Medium	1.40
06) Equal-log / Large	2.01
07) Shifted / Small	1.18
08) Shifted / Medium	1.45
09) Shifted / Large	2.16
	$\underline{u}$ = 1.61
B. Non-Parallel Conditions	**Geometric Mean Ratio**
10) Diverging / Small	1.63
11) Diverging / Medium	2.00
12) Diverging / Large	2.96
13) Converging / Small	1.63
14) Converging / Medium	2.00
15) Converging / Large	2.96
	$\underline{u}$ = 2.20

Table 11: Geometric mean ratio for each three-glide condition. Given a top(T), a middle(M) and a bottom(B) glide for each complex, a maximum(max) and minimum(min) frequency for each glide, the geometric mean ratio(GMR) is computed by:

GMR= antilog[average of logs of ratio] = $\log^{-1}$ [log(Tmax/Mmax) + log(Tmax/Bmax) + log(Mmax/Bmax) + log(Tmin/Mmin) + log(Tmin/ Bmin) + log(Mmin/ Bmin) / 6].

Appendix B

BASIC STATS	MIN	MAX	MEAN	STANDARD ERROR	STANDARD DEVIATION
Harmonic Small Up	0.347	0.612	0.423	0.015	0.066
Harmonic Small Dn	0.316	0.500	0.416	0.013	0.056
Harmonic Medium Up	0.418	0.576	0.464	0.011	0.045
Harmonic Medium Dn	0.335	0.533	0.419	0.013	0.057
Harmonic Large Up	0.321	0.673	0.499	0.024	0.103
Harmonic Large Dn	0.326	0.638	0.520	0.022	0.095
Equal-log Small Up	0.337	0.709	0.475	0.024	0.104
Equal-log Small Dn	0.352	0.801	0.512	0.033	0.140
Equal-log Medium Up	0.423	0.566	0.483	0.008	0.036
Equal-log Medium Dn	0.369	0.607	0.462	0.017	0.071
Equal-log Large Up	0.520	0.724	0.615	0.012	0.050
Equal-log Large Dn	0.511	0.725	0.627	0.013	0.055
Shifted Small Up	0.362	0.556	0.441	0.012	0.051
Shifted Small Dn	0.311	0.459	0.400	0.011	0.046
Shifted Medium Up	0.408	0.556	0.485	0.010	0.044
Shifted Medium Dn	0.413	0.561	0.474	0.010	0.042
Shifted Large Up	0.439	0.709	0.572	0.016	0.069
Shifted Large Dn	0.367	0.668	0.560	0.019	0.079
Diverging Small Up	0.474	0.857	0.620	0.022	0.094
Diverging Small Dn	0.485	0.796	0.631	0.018	0.077
Diverging Medium Up	0.551	0.786	0.649	0.014	0.060
Diverging Medium Dn	0.510	0.770	0.655	0.014	0.060
Diverging/Large Up	0.607	0.852	0.748	0.015	0.063
Diverging Large/Dn	0.582	0.847	0.727	0.018	0.075
Converging Small/Up	0.576	0.781	0.649	0.014	0.059
Converging Small Dn	0.612	0.811	0.675	0.014	0.059
Converging/Medium Up	0.582	0.760	0.685	0.012	0.053
Converging/Medium Dn	0.597	0.832	0.689	0.014	0.058
Converging Large/Up	0.597	0.893	0.759	0.020	0.085
Converging Large/Dn	0.663	0.903	0.783	0.018	0.786

Table 12 : Descriptive statistics for a within-subject design with 30 conditions and 18 subjects.