Music Perception
Winter 1991, Vol. 9, No. 2, 155–184

Stream Segregation and Peripheral Channeling

WILLIAM MORRIS HARTMANN & DOUGLAS JOHNSON
Michigan State University

Two interleaved melodies, with theory tones alternating as ABAB..., can be individually followed and identified if auditory stream segregation takes place. Stream segregation can occur if the tone conditions are favorable, for example, if the tones of the different melodies are in different octaves. Using an interleaved melody identification task, we have measured the extent to which 12 different tone conditions lead to stream segregation. The purpose of the experiment is to discover whether stream segregation is mediated entirely by channeling that is established in the auditory periphery or whether more complicated principles of source grouping are at work. Peripheral channels are defined as either tonotopic (frequency based) or lateral (localized left or right). The data show that peripheral channeling is of paramount importance, suggesting that a set of rather simple rules can predict whether two interleaved melodies will be perceived as segregated or not. The data reveal a secondary effect of tone duration. Otherwise, in the absence of peripheral channeling, the experiments find little or no stream segregation, even in those cases where individual tones should clearly evoke images of different sources. Additional experiments show that interleaved melody identification is made more difficult by a transposition that maximizes the number of melodic crossings, even though the transposition may place the interleaved melodies in different keys. An appendix develops an elementary mathematics of melodic crossings and contacts.

Introduction

Auditory stream segregation was described in a 1950 publication by Miller and Heise: "a pattern that alternates rapidly between high tones and low tones tends to break into two melodies. The high tones group to give one melody, and low tones another, and the listener does not notice that in fact the successive tones from the two melodies alternate." This brief description captures two main aspects of stream segregation: the grouping of tones that fall within a restricted frequency range to form a perceptual stream, and the loss of a sense of temporal order across different streams.

Requests for reprints may be sent to William Morris Hartmann, Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824.

The grouping effect is a very powerful one. If the high and low tones are sufficiently separated in frequency and the tone presentation rate is fast enough, then the grouping into streams is irresistible (van Noorden, 1975). Similarly, the loss of any sense of temporal order across streams can be extreme (Warren, Obusek, Farmer, & Warren, 1969). It may even be difficult to count the number of tones if the tones are segregated into two different streams (Massaro, 1977). To observe stream segregation does not require the controlled conditions of the laboratory; it can be observed under any circumstances. The strength of the effect is part of its appeal to researchers in hearing; it is so strong that one feels that it has to be important.

The actual term *stream segregation* as a name for this effect was introduced by Bregman and Campbell (1971). These authors distinguished between auditory stream formation and peripheral channeling. In their view, a stream is a mental organizational entity; it is not readily definable by the physical properties, such as frequency, that determine peripheral channeling. Subsequent work by Bregman (1978, 1981) identified stream segregation with a process that may be called *source grouping*, a process by which listeners associate tones with imagined sources. Those tones that seem likely to have come from the same sound source are most often grouped into the same stream.

The distinction made by Bregman and Campbell concerning the nature of the stream segregation effect is the basis of the work to be described in this article. We ask, to what extent is stream segregation possible in the absence of peripheral channeling? There is also a converse question: Can listeners form a single perceptual stream from tones that excite different peripheral channels? For definiteness we recognize only two kinds of peripheral channels, those based on frequency (tonotopic) and those based on ear presentation (lateral). These are the two kinds of peripheral channels that have been regarded as dominant by both auditory physiology and psychoacoustics for more than a century.

Results from the literature are rather mixed on these questions concerning the role of peripheral channeling. An early work by Dowling (1968) noted that listeners are able to segregate auditory streams that differ in intensity level. This observation is one piece of experimental evidence that stream segregation is possible in the absence of peripheral channeling. Van Noorden (1975, 1977) corroborated this observation, although he concluded that segregation based on a level difference is weak compared with the segregation obtainable with a frequency difference.

The thesis by van Noorden (1975) particularly emphasized the role that peripheral channeling plays in the formation of auditory streams. Experiments using alternating complex tones, having no low harmonics, with similar or dissimilar pitches and with similar or dissimilar spectra found that streaming is based on frequency contiguity (same peripheral channel); it is not based on contiguity of pitch.

Van Noorden's dichotic experiments showed that tones presented to opposite ears (the other kind of peripheral channeling) are segregated according to ear presentation. Rivalry experiments, however, showed that segregation on the basis of frequency is stronger than segregation on the basis of ear presentation (Deutsch, 1975).

Investigating the second question, the formation of a single stream from tones that occupy different peripheral channels, Bregman and Dannenbring (1973) introduced frequency glides that partially spanned the gap between tones that were separated by several octaves. This experiment can be regarded as an attempt to reduce the stream segregation effect without actually putting the tones in the same peripheral channel. The introduction of the glides as pointers was, however, only partially successful in restoring a single stream. Bregman and Dannenbring concluded that in some cases peripheral channeling may be so strong that it precludes the formation of a single stream (see also Tougas & Bregman, 1985).

A dichotic experiment by Deutsch (1979) showed that peripheral channeling could be defeated. The recognition of a melody, with notes sent randomly to one ear or the other, was much improved by the addition of a drone tone to the contralateral ear. This strongly implicates a higher level process, presumably the identification of a figure and a ground in the mind of the listener.

The literature, therefore, includes evidence that peripheral channeling is of considerable importance in stream segregation, and it includes evidence that peripheral channeling may not be so important after all. Given this rather inconclusive state of affairs, we undertook a series of stream segregation experiments using a set of tones of unprecedented variety. The intent was to use different tone pairings, some exciting the same peripheral channels, others exciting different peripheral channels.

Experiment 1

The goal of the experiment is to provide a quantitative measure of stream segregation. The technique is to measure the ability of listeners to identify interleaved melodies, as in the classic study by Dowling (1973). Of primary interest is the functional relationship between this ability and the tone conditions under which the melodies are heard. If, under a given condition, the listener can easily identify both of the interleaved melodies then we will interpret that datum as evidence that the tone condition is one that promotes stream segregation.

THE TONE CONDITIONS

In contrast to much of the previous work on stream segregation (Bregman & Rudnicky, 1975; Bregman & Pinker, 1978; Dannenbring & Bregman, 1978; Steiger & Bregman, 1982), our experimental focus is less on the context in which tones are heard and more on the tones themselves, on tone pairs that excite different peripheral channels and tone pairs that excite mainly the same peripheral channels.

There are 10 basic conditions in Experiment 1. They are described below with the aid of Figure 1 and Table 1. These show the time sequence and the specific values of the durations and sound levels that apply in the various conditions. In each stimulus sequence, there are always two interleaved melodies, melody A and melody B. Melody A is presented on the odd-numbered tones; melody B is presented on even-numbered tones. Because each melody has 16 tones, there are 32 tones in the sequence. The first tone belongs to melody A (i.e., it is an A tone); the last tone of the sequence belongs to melody B (i.e., it is a B tone). The description of the conditions also includes our opinion as to whether the condition favors stream segregation on the basis of peripheral channeling and some comments relevant to stream segregation on the basis of source grouping principles, as distinct from peripheral channeling.

Condition 0: The Null Standard

In the null-standard condition, tones A and B are identical in every respect. They are both sine tones, and the mean note of each melody is 440 Hz. All other conditions represent departures from the null standard, along some stimulus dimension(s).

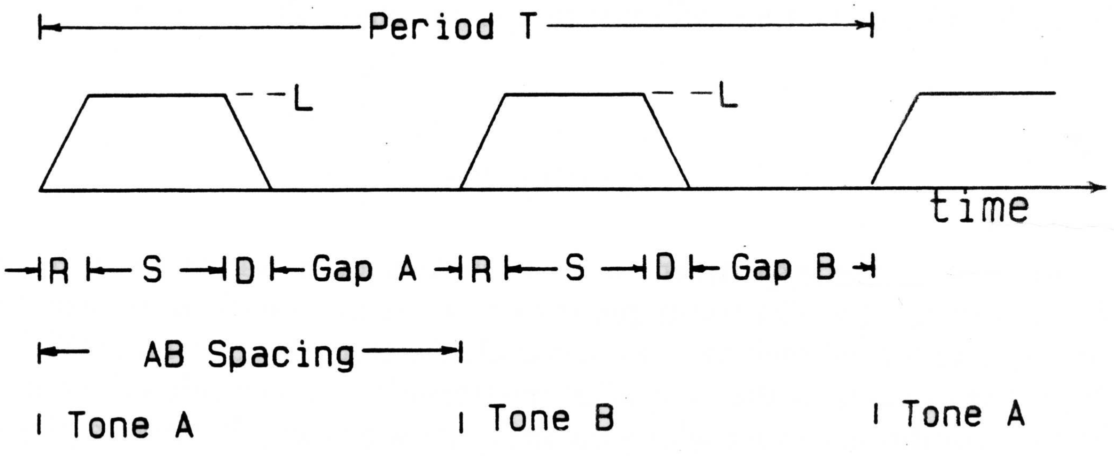

Fig. 1. Tones for the interleaved melody experiment. Time is on the horizontal axis, envelope on the vertical. Table 1 gives numerical values of the rise, sustain, decay, and level parameters for different experimental conditions.

TABLE 1
Timing and Levels for Signal Conditions

Condition	AB	Tone A				Tone B			
		R	S	D	L	R	S	D	L
STD	125	8	67	8	60	8	67	8	60
4A	125	8	67	8	56	8	67	8	64
4B	125	8	67	8	64	8	67	8	56
5A	135	74	1	8	62	8	1	74	62
5B	115	8	1	74	62	74	1	8	62
6A	150	8	108	8	58	8	27	8	66
6B	100	8	27	8	66	8	108	8	58
9A	85	8	67	8	60	8	67	8	60
9B	165	8	67	8	60	8	67	8	60
10A	125	8	67	8	54	8	67	8	54
10B	125	8	67	8	54	8	67	8	54
11A	125	8	67	8[a]	56	8	67	8	60
11B	125	8	67	8	60	8	67	8[a]	56

NOTE. The table gives parameters for the conditions of Experiments 1 and 2. See Figure 1 for parameter definitions. The period T is always 250 msec. The STD parameters apply to conditions 0A 0B, 1A 1B, 2A 2B, 3A 3B, 7A 7B, 8A 8B, and 12A 12B. The units for the AB Spacing (AB), the rise times (R), the sustain times (S), and the decay times (D) are all milliseconds. The levels (L) are in dB SPL.

[a] Add also 1000 msec of reverberation.

Condition 1: Octave Shift

For Condition 1A, the A tones are shifted up by one octave; the B tones remain centered on 440 Hz. The shift of 12 semitones is adequate to eliminate crossing of melody lines for any pair of melodies in the melody table (see Table 2) except one. Based on the experience of others, we expect melody recognition performance to be high (Brokx & Nooteboom 1982; S. McAdams, personal communication, 1988). Strong peripheral channeling is present. For Condition 1B the B tones, namely the even-numbered tones of the sequence, are shifted up by one octave; the A tones remain centered on 440 Hz.

Condition 2: Dichotic Presentation

For Condition 2A, the A tones are sent to the listener's left ear; the B tones are sent to the right. Because left and right ears correspond to different peripheral channels, we expect performance to be high. Because the A and B tones appear to have different locations, they should be grouped as separate sources. Grouping by frequency appears to dominate grouping by ear presentation (Deutsch, 1975). Therefore, one might expect performance to be less good in this condition than in Condition 1.

By including the alternative (2B) condition, with A tones sent to the right ear and B tones to the left, we make the experiment methodologically symmetric.

Condition 3: Timbre Differences

In Condition 3A, the A tones have a complex waveform; the B tones remain sine waves. The complex waveform is a pulse with a 20%duty factor, low-pass filtered at 5000 Hz, −48 dB per octave. Therefore, the first nine harmonics are strong, except for the fifth. The ninth harmonic is 19 dB down compared with the fundamental. The level of the complex tone is set so that it has the same loudness as the sine tone, B.

Experiments by Wessel (1979) show that melodies can be segregated on the basis of timbre. Although the pitches of the two melodies are in the same range, the complex timbres excite different tuned channels in the auditory periphery. Good performance is therefore expected. Because different timbres often correspond to different sources in the environment, a source-grouping process should also be effective. Rivalry experiments by Butler (1979) suggest that timbre differences may not be as effective in promoting stream segregation as frequency-range differences (Condition 1) are.

Condition 4: Level Differences

In the 4A condition, the level of the A sine tones is increased by 4 dB and the level of the B sine tones is decreased by 4 dB, compared with the null standard. The levels then differ by 8 dB. The purpose of this condition is to determine whether a level difference is an adequate basis for the segregation of streams. Clearly the choice of level difference involves a compromise: Some level difference is necessary in order for there to be any effect at all, but a large level difference would obscure the weaker melody. Van Noorden (1977) found that a difference of 3 dB was the smallest level difference for which a segregation effect could be noted in his experiments. Our choice of 8 dB is based on preliminary experiments that found that a difference of 6–8 dB is optimum for hearing both melodies.

To a first approximation, a small level difference does not result in the excitation of different peripheral channels. Some increased spread on a tonotopic coordinate with increasing level can be expected as a second approximation. If peripheral channeling is necessary for stream segregation, then we would expect poor performance. If tones with different loudness are associated with sources at different distances or with instruments of different dynamics, then source-grouping principles would suggest good performance.

Condition 5: Envelope Differences

In Condition 5A, the A tones have a long rise duration and an abrupt decay. The B tones are the mirror images of the A tones with short attacks and long decays. The gap durations are adjusted so that the AB . . . sequence sounds isochronous at the experimental intensity level.

This condition would seem to be an excellent discriminator. The long-term spectra of A and B tones are identical. Apart from neural adaptation effects then, the long-term peripheral excitation should be similar for both tones. Therefore, a peripheral model of stream segregation predicts poor performance. By contrast, other source-grouping principles would seem to predict good performance because temporal envelope is an important factor in the recognition of sources. Tones with long rise durations and relatively shorter decays resemble the tones of an accordion or harmonica. Reversing the envelope durations produces percussive tones, like those of a marimba.

Condition 6: Duration Differences

In Condition 6A, the A tones have a long duration and the B tones have a short duration. The levels are adjusted so that the tones are equally loud, and the gap is adjusted for perceptual isochrony.

Table 1 shows that the duration differences are considerable, a ratio of four to one in the sustained portions. This is intended to promote a strong impression that the A and B tones come from different sources, thereby giving a source-grouping principle a good opportunity to succeed. There was some question whether an overall duration of 45 msec, including rise and decay, is long enough to avoid significant spectral splatter. However, informal listening tests showed that the envelope shape is quite effective in reducing this splatter, so that to a first approximation, poor performance is expected if stream segregation requires peripheral channeling.

Condition 7: Interaural Time Differences (T_d)

In Condition 7A, the A tones are presented with a 500 μsec interaural time difference, with the tone to the left ear leading the tone to the right. The B tones also have a 500 μsec interaural time difference, but with the tone to the right ear leading the tone to the left. In Condition 7B, the headphones are reversed.

An interaural difference of 500 μsec is adequate to lateralize completely tones with frequencies of our stimuli (Durlach & Colburn, 1978). Such lateralization is the likely cause of the segregation effects studied by Kubovy (1982).

Because of different apparent spatial locations, one expects good performance on this task based on source grouping. Performance should be similar, if not identical, to performance on the dichotic task (Condition 2). From the point of view of peripheral channeling, this condition occupies an interesting middle ground. If one supposes that the peripheral channels are delineated entirely by frequency range and left-ear/right-ear primary neurons, then one expects poor performance. If the definition of the periphery is expanded to include the medial superior olive, where interaural time differences are processed (Yin & Kuwada, 1984), then one expects good performance.

Condition 8: Added Noise

In Condition 8A, a high-frequency noise is added to the A tones and a low-frequency noise is added to the B tones. The tones are those of the null-standard condition. The noises are enveloped on and off together with the tones. The high-frequency noise is spectrally flat from 1000 Hz to 2000 Hz. Its low-frequency tail is down by 35 dB at 440 Hz, and its high-frequency tail is down by 30 dB at 5000 Hz. The low-frequency noise is spectrally flat from 100 Hz to 200 Hz, and it is also down by 35 dB at 440 Hz. The levels of the two noises are chosen to produce equal masking at the frequencies of the melody tones (all in the vicinity of 440 Hz); each increases the threshold for a 440-Hz sine tone by 10 dB.

The noise bands occupy different frequency regions, and when heard by themselves, at the presentation rate of the experiment, these bands are segregated into different streams by tonotopic channeling. The purpose of this condition is to discover whether it is possible to mediate the segregation of A and B tones by adding a cue, extraneous to the melodies, that is itself segregated.

The interest in this condition stems from the observation that peripheral channeling can sometimes be so strong as to override all other factors, such as attention (van Noorden, 1975; Bregman & Dannenbring, 1973). Conceivably, adding noise bands will improve performance. On the other hand, the noise bands are irrelevant to the melodies, and other source-grouping principles predict that the noise bands will be streamed separately, leading to a total of three or four perceptual streams. Adding the noise bands then might only cause confusion and diminish performance.

Condition 9: Altered Rhythm

In Condition 9A, the duration of the gap that follows an A tone (and precedes a B tone) is made very short. In order to preserve the overall timing, the gap following the B tone is made correspondingly longer. Preliminary experiments found that partial forward masking is not important here; therefore, no level changes are made for this condition.

There is no additional peripheral channeling in Condition 9, compared with the null standard. There is probably no increased tendency to stream based on the identification of sources either. The purpose of Condition 9 is to discover whether there may be a higher-level grouping process that leads to a streaming advantage, when such grouping can be expected from experience (Dowling, Mei-tak Lung, & Herbold, 1987). The experiments of Handel, Weaver, & Lawson (1983), however, suggest that pairing of the A and B tones, as our altered rhythm does, should lead to diminished performance.

This concludes the list of the 10 basic conditions. There are, in fact, 20 conditions in Experiment 1, because each condition (e.g., 1A) is accompanied by an "alternative condition" (e.g., 1B), where the stimulus characters of A (odd) and B (even) tones are reversed. This methodologic symmetry is introduced because preliminary experiments showed some primacy effect on melody recognition. Only for the null standard, where A and B tones are identical, is the methodologic symmetrization unimportant.

The 10 conditions are the major experimental variable in the experiment. Of interest is whether a listener can segregate melodies better in one condition than in another. However, the different conditions constitute a multidimensional variable, and something must be done to make the comparisons legitimate. Otherwise, a comparison of performance across conditions would only show that some arbitrarily chosen difference in one dimension (for example in frequency range) leads to greater or smaller stream segregation than some other arbitrarily chosen difference in some other dimension (for example in intensity). For this reason, the parameters for the conditions, shown in Table 1, were determined by extensive listening tests with the goal of finding the optimal conditions for segregation. We considered this to be especially important for Conditions 4, 5, 6, 7, and 9, where segregation is not expected on the basis of peripheral channeling. Less attention was given to parameter choices where peripheral channeling was expected. For example, we do not know that an octave separation in frequency range is more effective than an interval of a tenth. The nature of our effort required that we give those conditions where segregation is expected only on the basis of source grouping the best possible chance of succeeding. Table 1 shows that the parameter values chosen for these conditions have extreme differences.

THE MELODIES

The melodies in an interleaved melody recognition experiment must be presented with tones of equal duration and spacing (quarter notes); they must also be recognizable. These two constraints rather seriously limit the choice of melodies in the stimulus set. The melody table (Table 2) lists

TABLE 2
Melody Table

1	Frère Jacques	65	72	7	68.6	2.6	(F)	65 67 69 65 65 67 69 65 69 70 72 72 69 70 72 72
2	Twinkle Twinkle Little Star	65	74	9	69.3	3.2	(F)	65 65 72 72 74 74 72 72 70 70 69 69 67 67 65 65
3	Peter Peter Pumpkin Eater	64	73	9	68.7	2.6	(A)	73 69 71 69 66 69 64 69 73 69 71 69 66 69 64 69
4	Yankee Doodle	64	73	9	69.4	2.6	(A)	69 69 71 73 69 73 71 64 69 69 71 73 69 69 68 64
5	Sur Le Pont d'Avignon	63	75	12	69.3	2.6	(A♭)	68 68 68 68 70 70 70 70 72 73 75 68 67 68 70 63
6	God Rest Ye Merry Gentlemen	64	73	9	69.3	3.1	(F♯ min.)	66 66 73 73 71 69 68 66 64 66 68 69 71 73 73 73
7	Spartan Fight Song	66	75	9	68.9	3.2	(B)	66 67 68 67 66 67 68 68 66 66 68 70 71 75 75 75
8	Surprise Symphony (Haydn)	63	75	12	70.0	3.6	(A♭)	68 68 72 72 75 75 72 72 73 73 70 70 67 67 63 63
9	Hark the Herald Angels Sing	61	73	12	68.7	3.2	(F♯)	61 66 66 65 66 70 70 68 73 73 73 71 70 68 70 70
10	Ode to Joy (Beethoven 9th)	66	73	7	69.4	2.0	(F♯)	70 70 71 73 73 71 70 68 66 66 68 70 70 68 68 68
11	Good King Wenceslas	65	72	7	68.7	2.1	(B♭)	70 70 70 72 70 70 65 65 67 65 67 69 70 70 70 70
12	Man on the Flying Trapeze	64	74	10	68.7	3.6	(A)	64 64 69 71 73 73 73 74 66 66 71 71 64 64 68 69
13	There's a Hole in the Bucket[a]	65	74	9	69.0	2.6	(B♭)	70 72 74 70 65 67 70 65 67 70 65 67 70 72 70 70
14	Aura Lee (Love Me Tender)	66	71	5	68.7	1.4	(A)	68 69 68 69 71 66 71 71 69 68 66 68 69 69 69 69
15	Old McDonald Had a Farm	65	74	9	69.1	3.0	(B♭)	70 70 70 65 67 67 65 65 74 74 72 72 70 70 70 65
16	Bingo Dog	65	74	9	68.8	3.2	(B♭)	65 70 70 65 65 67 67 65 65 70 70 72 72 74 74 70
17	Peer Gynt Morning Song	65	74	9	69.5	2.8	(F)	72 69 67 65 67 69 72 69 72 74 69 74 72 69 67 65
18	Little Brown Jug	63	75	12	68.8	3.4	(B)	63 66 66 66 64 68 68 68 70 70 68 70 71 73 75 75
19	Hail to the Victors	66	71	5	69.3	1.6	(F♯)	70 70 66 68 70 66 68 70 71 71 68 70 71 68 70 71
20	Camptown Races	67	74	7	69.3	2.3	(F)	72 72 69 72 74 72 69 69 69 67 67 67 69 67 67 67
21	Lightly Row	65	72	7	69.3	2.2	(F)	72 69 69 69 70 67 67 67 65 67 69 70 72 72 72 72
22	Doxology (The Old Hundred)	63	73	10	69.4	2.8	(A♭)	68 68 67 65 63 68 70 72 72 72 72 70 68 73 72 70
23	The Blue Tail Fly[a]	63	73	10	69.0	3.2	(A♭)	72 73 72 70 68 65 65 73 65 63 67 70 73 72 68 68

TABLE 2 *continued*

24	Westminster Chimes 64 73 9 69.3 3.3 (A)	73 69 71 64 64 71 73 69 73 71 69 64 64 71 73 69
25	Have You Seen the Muffin Man 65 74 9 69.2 2.7 (B♭)	67 65 70 70 72 74 70 70 69 67 72 72 70 69 65 65
26	This Old Man 65 72 7 68.7 1.8 (E♭)	70 67 70 70 70 67 70 70 72 70 68 67 65 67 68 68
27	Poor Little Buttercup 66 73 7 68.9 2.2 (B)	66 68 71 70 68 66 66 68 71 70 68 66 71 70 71 73
28	Jesu Joy of Man's Desiring[b] 63 75 12 68.9 3.7 (E♭)	63 65 67 70 68 68 72 70 70 75 74 75 70 67 63 65
29	London Bridge Is Falling Down 65 72 7 68.6 1.7 (E♭)	70 72 70 68 67 68 70 70 65 67 68 68 67 68 70 70
30	The Blue Bells of Scotland 63 75 12 69.3 4.2 (E♭)	70 75 75 74 72 70 70 72 75 67 67 68 65 63 63 63
31	All Things Considered 63 75 12 68.7 3.1 (E♭, B)	70 75 72 68 65 70 67 63 66 71 68 64 68 70 71 71
32	Jupiter (Holst Planets) 63 73 10 69.2 2.8 (C♯)	73 68 70 71 70 68 73 68 70 63 65 66 73 68 70 71
33	Twenty Froggies Went to School 63 73 10 69.4 2.8 (A♭)	63 72 72 73 72 70 70 70 63 70 70 72 70 68 68 68
34	Skip to My Lou 65 73 8 68.9 2.5 (F♯)	70 70 66 66 70 70 73 73 68 68 65 65 68 68 71 71

NOTE. A set of 16-note melodies with semitone numbers in MIDI notation. On the second line appear (in order) the maximum and minimum semitone numbers, the difference between the maximum and minimum, the mean semitone number, the standard deviation of the semitone numbers ($N = 16$ weight), and the key. All melodies have a mean of 69 ± 1, corresponding to A 440 Hz.

[a] Ending phrase.

[b] Descant.

34 melodies that some listeners can identify when played individually, not interleaved, with quarter notes.

In fact none of the melodies of the table can be correctly played with only quarter notes. Even in those few cases where the score calls for notes of equal duration, accepted performances give the notes unequal duration or spacing. The most common distortion in the table is to substitute two quarter notes for a half note. Any syncopation (e.g., dotted quarter notes) is, of course, also absent. The best that can be said for these rhythmless melodies is that they can be recognized by some listeners, in many cases by all the listeners, in informal tests.

The melody table includes the following information for each melody: a melody number, a common name, and the notes of the melody on a semitone scale in Musical Instrument Digital Interface (MIDI) notation, where middle C is number 60. On the second line are the minimum and maximum note values and the difference between them, namely the "range." The next entries are the average note value and the standard deviation (N-weight) about that average. We attach no particular importance to the standard deviation at this time. Possibly melodies with a large standard deviation are more readily recognized in an interleaved competition, possibly not. The final entry gives the musical key.

The melodies of the initial stimulus set have the following properties:

1. There are 16 notes in each melody. In most cases this corresponds to a musical phrase, usually the first phrase of the song.
2. The average note for each melody is MIDI 69 plus or minus 1 (i.e., it is concert A 440 Hz) plus or minus a semitone. Equating the mean note values has the effect of causing the greatest number of melodic crossings on the average. A large number of crossings between melodies A and B is expected to make the task more difficult. (See Appendices 1 and 2.)
3. With one exception (Melody 9) the largest of all maximum notes is 75; the smallest of all minimum notes is 63. These extremes differ by 12 semitones, which means that if the notes of melody B are shifted up by an octave, then no melodic crossing can occur between any pair of melodies.
4. The melodies, and the melodies shifted up by one octave, are in ranges where melodic lines occur in instrumental music.

LISTENERS

A total of seven listeners, including the authors, served in Experiment 1. All listeners had normal hearing and all had some musical experience, although most of them were not professional musicians or students of music.

PROCEDURE

The experiments required a listener to identify two interleaved melodies, where both melodies belonged to a set of 12 overlearned melodies. This closed-set procedure was used as a guard against learning effects. The set of 12 melodies was different for each listener in the experiment.

The first step in the procedure was to test listeners individually for melody recognition. A potential listener heard all 34 melodies of Table 2 and was asked to identify the melodies by name. The experimenter noted the speed and confidence of each response. The goal of this step was to establish a list, for the particular listener, of the 12 melodies that were most readily recognized. If a potential listener could not recognize 12 melodies, he or she did not participate in further experiments.

The second step was a training stage. The listener heard interleaved melodies (from the list of 12) separated by one octave (Conditions 1A and 1B), doing the experimental task as described below. The training stage continued until the pair identification score reached 70% on several successive runs, that is, until the listener correctly identified both melodies on 21 out of 30 trials in the run. If after 4 hr of training the listener could not meet the 70% criterion, the listener was dropped from the subject pool.

THE TASK

For a given experimental run, the condition (0A through 9B) was fixed. There were 30 trials (30 interleaved melody pairs) in a run. For each trial, a pair of melodies from the listener's list was selected at random and the listener heard a single presentation of the interleaved melodies, 32 notes, 16 notes per melody. A single presentation makes the task more difficult for two reasons: First, the listener has no opportunity to confirm suspicions with a second hearing; second, there is good evidence that a long stimulus sequence promotes segregation, whether or not the listener intends to segregate (Deutsch, 1982; Anstis & Saida, 1985). By making the sequence as short as possible, we avoided this duration advantage. The listener responded according to which melody or melodies were recognized. Guessing was not encouraged or discouraged. The number of responses for each trial could therefore be zero, one, or two. There was no feedback.

After the listener's response, there was a delay of 1 sec; then the next trial, with another pair of randomly selected melodies, began. There were two constraints on the random selection of the pair, first that melodies A and B of a pair not be the same melody, and second that no melody that appeared in the *n*th trial appeared in the $(n + 1)$th trial. The listener was not informed of the latter constraint. After a total of 30 trials, the run was over. Subsequently a different condition was chosen for the next run.

The different conditions were given in a different order for each listener, apart from Condition 1, which each listener did first. The A and B runs for each basic condition were done sequentially, with half the listeners doing the A run before the B run. Each of the 20 conditions was done once (one run of 30 trials). The set of runs for 20 conditions was called the "first pass." After a listener had completed the first pass, the listener did the "second pass." The conditions of the second pass were in the same order as the first except that the order of A and B runs was reversed. The final data set for each listener then consisted of 40 runs: 60 trials on each of 20 conditions, a total of 1200 trials. It required six or seven experimental sessions, of 2 hr each, to complete the two passes.

EXPERIMENTAL DETAILS

For all conditions, the average presentation rate was 8 tones/sec. This rate is equal to that used by Dowling (1973) and is somewhat higher than the rate of 6.7 tones/sec used by Singh (1987). A rate of 8 tones/sec is high enough that a frequency shift of 12 semitones lies above van Noorden's (1975) temporal-coherence boundary. Therefore, at this rate an octave shift leads to stream segregation, regardless of the listener's attention set. The rate of 8 tones/sec is well below the threshold for the "roll effect" (van Noorden, 1977; McAdams & Bregman, 1979), related to pulsation threshold.

The experiment was controlled by a computer (PDP 11/73). The tones were generated by a digital oscillator using the technique of fractional addressing on a 8 kword buffer that contained a single cycle of a sine tone (Hartmann, 1987). The sample rate was 16 kHz. Tones were converted to analog by a 16-bit digital-to-analog converter and then low-pass filtered at 7 kHz, −115 dB/octave. Tones were presented by Yamaha YH1000 headphones in a soundproof room.

The digital oscillator ran continuously, changing only the frequency of the sine tone as required by the pattern of interleaved notes. The tones were shaped by routing them to two computer-controlled amplifiers, one for the A melody, the other for the B melody. The level and envelope timing were thereby controlled for the two melodies independently. Outputs A and B could be separately processed or routed.

The complex waveform was made by passing the sine tone through a comparator with an adjustable threshold. The 20% duty factor was set by zeroing the fifth harmonic, and its multiples, on a spectrum analyzer. The interaural time delays were created by all-pass filters (see Appendix 3). The listener's choices were collected by a response box, called the "jukebox," with the titles of the melodies next to touch buttons. The response box permitted the direct entry of response data into the computer program that controlled the experiment and performed the initial data analysis.

RESULTS

The raw data for each condition consist of the correct identities of melodies A and B actually presented and the listeners' responses.[1] To determine the relative effectiveness of the 10 conditions as bases for stream segregation, as described earlier, we simply count the number of instances of correct identification of both melodies for a given condition. This is a more sensitive measure than the number of correct responses. No significance is attached to a false alarm, as opposed to no response, in this analysis.

The results averaged over the seven listeners are shown in Figure 2. The horizontal axis shows the conditions, 0 through 9. The vertical axis shows the percentage of correct pair identifications, the first pass by an open circle on the left, the second pass by an open circle on the right. The filled circle in the middle is the average of the two passes. In most cases, performance improved on the second pass.

Figure 3 shows the results averaged over passes for individual subjects, indicated by numbered symbols 1 through 7. Here the average percentage correct, over all 20 conditions, for each subject has been subtracted from the scores for that subject before plotting. Therefore, a point above the zero line indicates that for the given condition a listener is performing better than his average over all the conditions.

Figures 2 and 3 suggest that there is a systematic dependence of performance on tone condition. This was verified with an analysis of variance calculation; the effect of tone condition is highly significant [$F(9, 54) = 15.5, p < .0001$]. Post-hoc comparisons were made among the tone conditions (Newman-Keul's range test). The results are reported in Figure 4. The conditions are arranged from left to right on the page according to their mean scores. The lowest score occurs in the null-standard condition, the next lowest in the added-noise condition, etc. Those conditions not grouped together within one of the boxes are found to be significantly different ($p < .05$).

The range test clearly indicates that peripheral channeling is highly important. Those conditions for which the greatest peripheral channeling

---

[1] This collection of data includes a wealth of information, much of which is not directly pertinent to our study. For example, some experiments in the dichotic condition showed a considerable ear-advantage effect. There is information on the sort of melodic competition that makes a particular melody easy or difficult to recognize, and information on what kinds of confusions among melodies frequently occur.

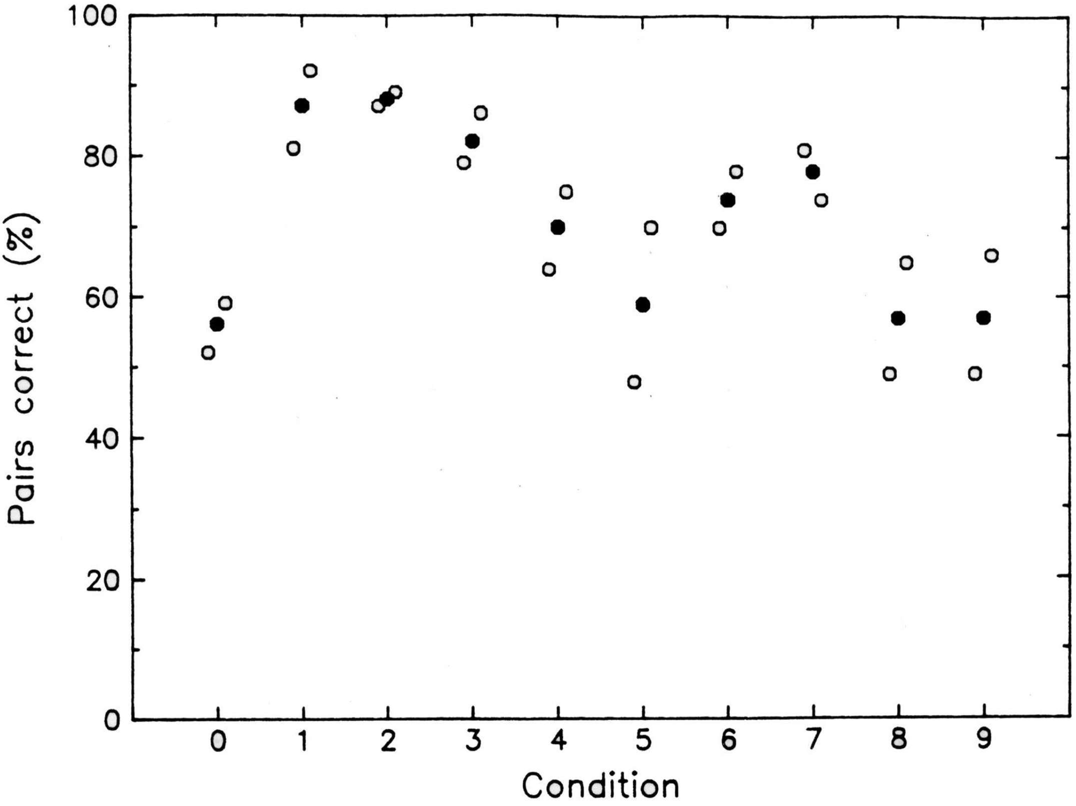

Fig. 2. Percentage of correct pair identifications for 10 conditions of Experiment 1, averaged over seven listeners. Data for the first pass are shown by the open circle on the left, data for the second pass by the open circle on the right; the average is given by the filled circle in the middle.

is expected (1, 2, and 3) are in the same range, the highest. Interaural time difference (ITD) also is in this range, suggesting that expanding the definition of the periphery to include low-level binaural interaction is justified. The ITD condition however, is not distinguishable from two conditions with small peripheral channeling, namely tone duration and intensity differences. Of considerable interest is the fact that streaming performance given different temporal envelopes is not significantly different from performance on the null standard. On the basis of source-grouping principles, one might have expected a different result.

The results of this section indicate that those tone differences that lead to the excitation of different peripheral channels promote stream segregation much more effectively than tone differences that do not excite different channels but which might well evoke the images of different sources, based on other source-grouping grounds. On the other hand, peripheral channeling does not guarantee stream segregation, as shown by the fact that our attempt to induce segregation by adding high-pass and low-pass noise was a total failure.

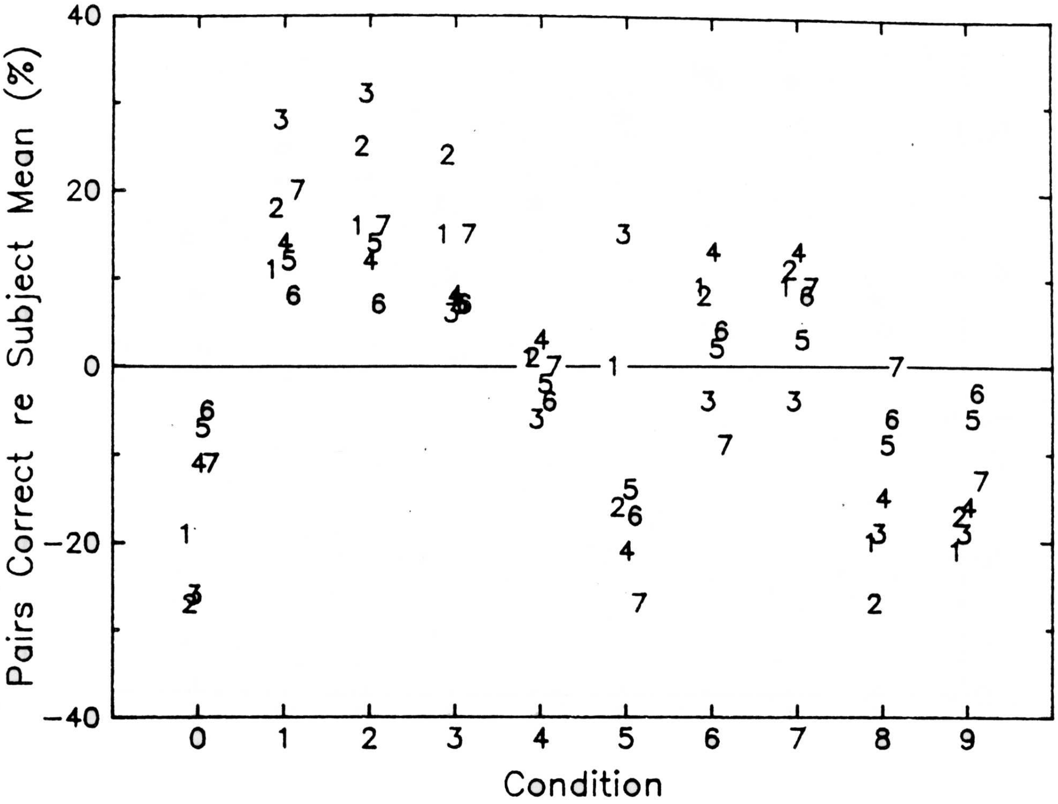

Fig. 3. Pair identification performance for 10 conditions of Experiment 1, averaged over passes, for each of seven listeners labeled 1–7. For each listener, the percentage correct on each condition is plotted with respect to that listener's average over all conditions.

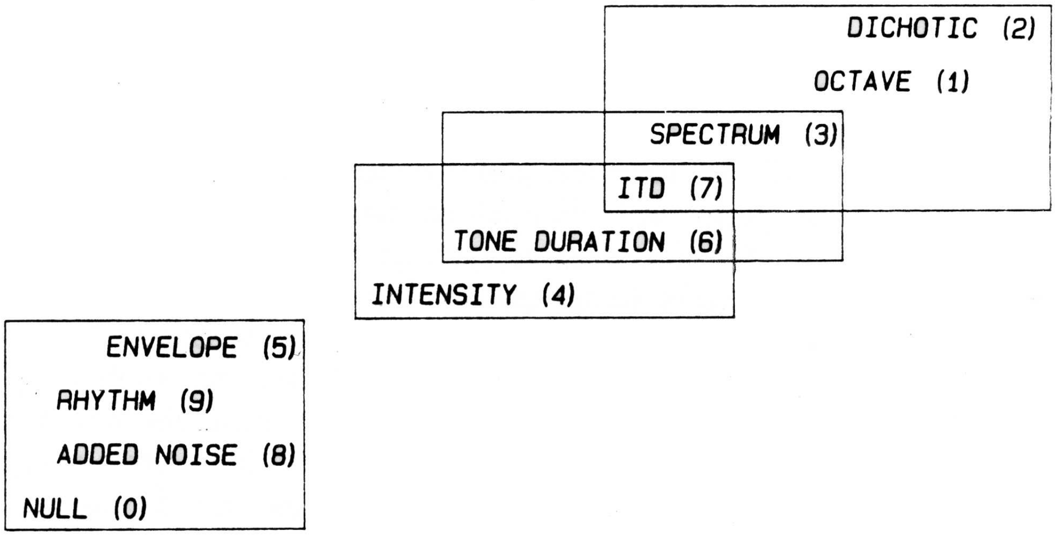

Fig. 4. Newman Keul's range test of the hypothesis that average performance is different for the different conditions of Experiment 1. Two conditions are in the same box if the range test cannot say that performance is different ($p < .05$). Best performance is at the upper right, worst is at the lower left. Numbers in parentheses are condition numbers (e.g., from Figures 2 and 3).

Experiment 2

Experiment 2 is similar in form to Experiment 1. Again, the goal is to discover which tone conditions promote stream segregation and which tone conditions do not, by means of an interleaved-melody identification task. Experiment 2 includes three new conditions, which seemed to be particularly interesting in light of the results of Experiment 1.

THE TONE CONDITIONS

Condition 10: Odd Harmonics versus Even Harmonics

In Experiment 1, Condition 3 tested listeners' ability to segregate a sine tone from a complex tone with a rich spectrum of high harmonics. Performance was high, showing that gross spectral differences are an adequate basis for segregation, probably because different peripheral channels are excited by the two tones. Condition 10 is a first attempt to determine the tuning of this effect. In Condition 10A, the A tones have a fundamental component and odd-numbered upper harmonics whereas the B tones have a fundamental component and even-numbered harmonics. The A tones are square waves; the B tones are made by putting a sine tone into a precision half-wave rectifier, then high-pass filtering the output with a one-pole filter. Therefore, both tones have the same spectral envelope, asymptotically decreasing with frequency at a rate of 6 dB/octave. Both tones are also low-pass filtered at 5 kHz, −48 dB/octave. The difference between the tones is only in the odd and even harmonic numbers. In Condition 10B, the spectra of A and B tones are reversed.

Condition 11: Sine Tones versus Reverberated Sine Tones

In Condition 11A, the A tones are made by passing sine tone pulses through a digital reverberator with a reverberation time RT_{60} equal to 1 sec. The B tones are sine tone pulses without reverberation, as in the null standard. Melodies A and B excite the same peripheral channels, but they sound very different. The A tones seem to come from a distance. Further, the A tones are much longer in duration. In this respect, Condition 11 is an exaggerated version of Condition 6. It is the only one of our conditions in which tones overlap in time.

Condition 12: Sine Tones versus Rough Tones

In Condition 12A, the A tones are made by multiplying the melody sine tone by a 12.5-Hz sine, using a balanced modulator. The resulting spectrum has two components, separated by 25 Hz. For example if the melody frequency is 440 Hz, then the spectrum of the tone heard by the listener has one component at 452.5 Hz and the other component at 427.5 Hz. The pitch of the complex is approximately 440 Hz, as expected, but the two components beat at rate of 25 Hz, which is the rate for maximum roughness for the frequency range of the melodies (Terhardt, 1974). Melody B is a sine tone, as in the null standard. Melodies A and B excite the same peripheral channels, but they sound very different. Stream segregation is expected according to the source-grouping principle because of this large difference in tone color.

In addition to the three new conditions, Experiment 2 included two conditions from Experiment 1. These were the null standard and octave shift, respectively Conditions 0 and 1. These conditions were included to give a common framework for Experiments 1 and 2. The null standard is the hardest condition, the octave shift is one of the easiest. There were, therefore, a total of five conditions in Experiment 2.

PROCEDURE

The procedure in this experiment was identical to that in Experiment 1. Listeners were selected and trained in the same way, and each did A and B runs on each of two passes. Again there were 30 trials per run. Therefore, each listener made 600 melody pair identifications. There were six listeners in the experiment. Listeners 2 and 4 were available from Experiment 1; the other four are identified below by symbols 8, 9, A, and B.

RESULTS

Data for Experiment 2 were given the same analysis as the data in Experiment 1. Figure 5 shows the percentage of correct pair identifications for each listener and each condition. The data are averaged over the A and B conditions and over the two passes. The error bars have a length that is two standard deviations ($N-1 = 3$ weight). Figure 6 shows the same data with each listener's mean subtracted from the scores. The figures show that best performance occurred with the octave shift (Condition 1) followed by the odd/even spectral difference and the reverberated/dry difference, respectively Conditions 10 and 11. The rough/smooth difference in Condition 12 appears to promote segregation no better than the null standard (Condition 0) does.

The data were submitted to a within-subjects analysis of variance. There was a significant effect of condition ($F(4, 20) = 16.47$, $p < .0001$]. The results of a Newman-Keul post-hoc test ($\alpha = 0.05$) are shown in Figure 7. According to this test, Experiment 2 does not distinguish between the three best conditions. According to Fisher and Duncan tests, however, the octave shift gives significantly better performance than the reverberated/dry condition.

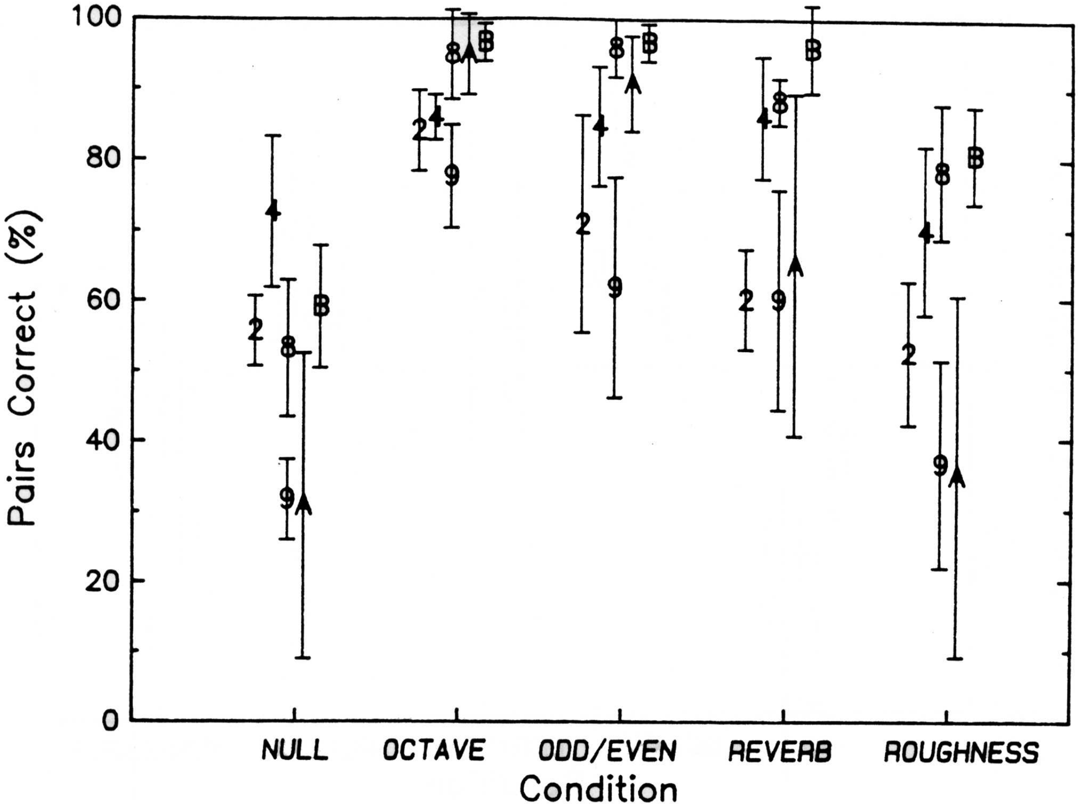

Fig. 5. Percentage of correct pair identifications for the five conditions of Experiment 2, for six listeners, labeled 2, 4, 8, 9, A, and B. Data points show averages over four runs, and error bars have a length that is two standard deviations ($N-1 = 3$ weight).

As in Experiment 1, Experiment 2 finds that those conditions that segregate A and B tones into different peripheral channels are the conditions that lead to good performance in stream segregation. Condition 10 shows that channeling occurs for interleaved spectra. We suspect that low-order harmonics, with frequencies less than 2 or 3 kHz, are responsible, because odd and even harmonics in this range fall into different critical bands.

Discussion

We performed an interleaved melody identification experiment with the goal of better understanding the cause of stream segregation. At the outset we had in mind two models for the stream segregation effect. One model, suggested by the work of van Noorden, is that stream segregation is caused by peripheral channeling. Peripheral channels are either tonotopic (frequency) or lateral (left ear–right ear). The other model, eloquently expounded by Bregman and his colleagues (e.g., McAdams, 1984a, 1984b),

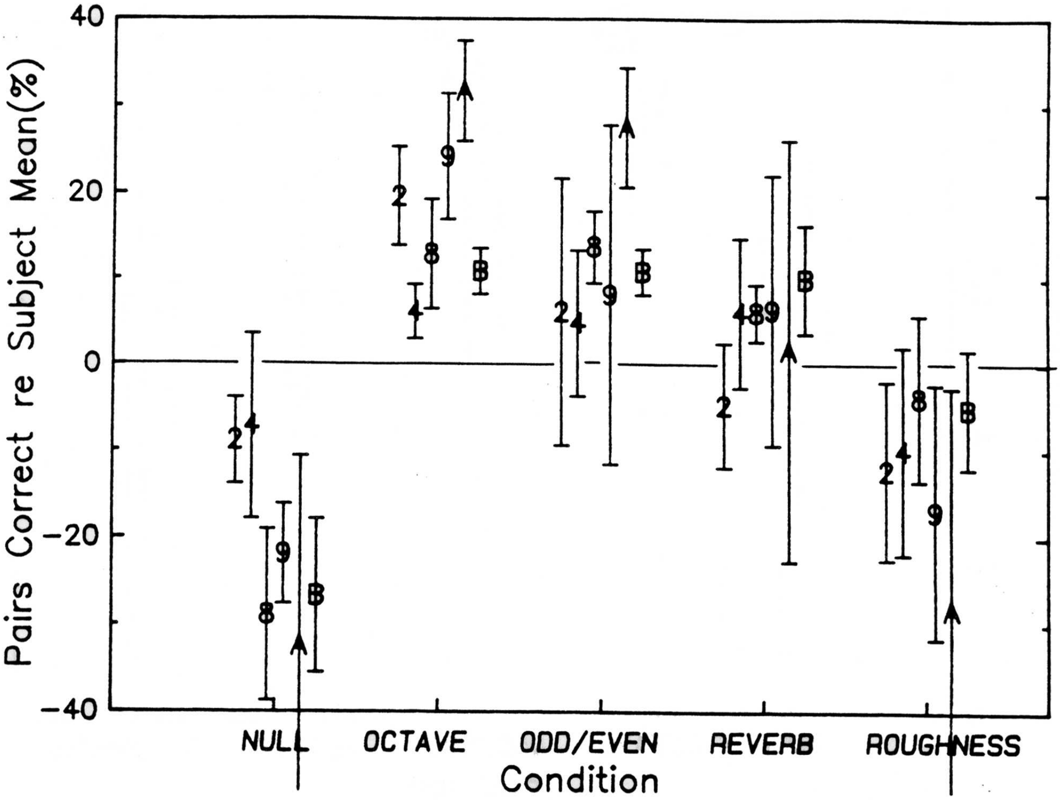

Fig. 6. The data of Figure 5 are replotted with respect to each listener's average over all five conditions.

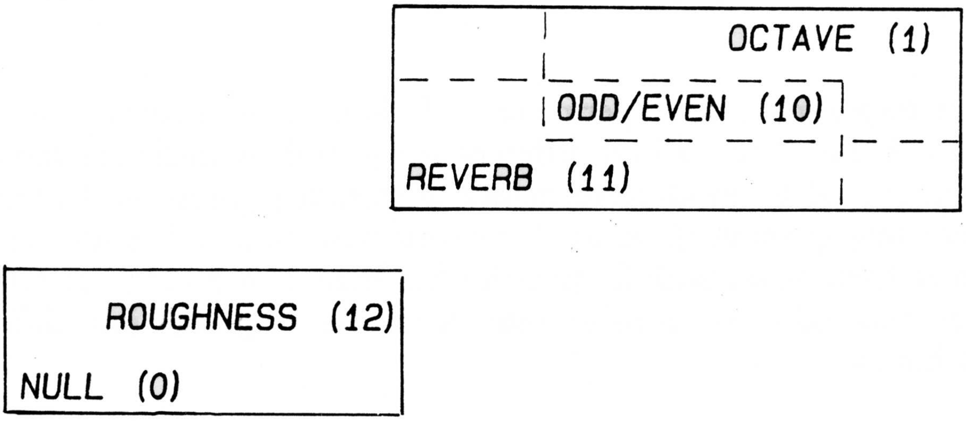

Fig. 7. Newman Keul's range test for Experiment 2, similar to Figure 4. Dashed lines show that Fisher and Duncan tests distinguish between conditions 1 and 11, reinforcing the notion that peripheral channeling (as in 1 but not in 11) is important for stream segregation.

holds that two tones are separately streamed if they are identified by the listener as originating from different sources.

Our melody identification experiment employed a dozen different tone conditions. In some of these, stream segregation could be expected on the basis of both peripheral channeling and source grouping. For instance, the tones of A and B melodies might differ in frequency range, or ear of presentation, or spectrum, or (expanding the definition of the periphery somewhat) interaural time delay. These conditions do not distinguish between the models; they serve as baseline measurements for comparison.

In other tone conditions, the tones of A and B melodies excite the same peripheral channels but are otherwise clearly distinguishable as different sources. Therefore, stream segregation may be expected only if segregation is determined by source grouping. These conditions are those in which tones differ in duration, or intensity, or amplitude envelope shape, or rhythmic position, or roughness. Our experiments found that all of these conditions lead to segregation performance that is significantly inferior to the performance obtained when peripheral channeling is operative. Most telling are the conditions of amplitude envelope shape and roughness, where performance cannot be distinguished from performance on the null standard. Amplitude envelope and roughness are clear identifiers of tone sources, and yet they appear to make no significant contribution to stream segregation.

The results of the experiments lead to the practical conclusion that if a composer, or orchestrator, or gambler needs to make a prediction as to whether two melodies will be segregated or not, then the best strategy is to base the prediction on whether the two melodies excite different peripheral channels or not. To apply these results to a careful study of the competition between peripheral-channeling and source-grouping models requires that some logical problems be solved.

First of the logical problems is that the peripheral-channeling model and the source-grouping model are not alternatives on the same level of description. The source-grouping model includes peripheral channeling as a possible component in the identification of sources, but does not stop there. The model includes the entire gestalt that makes one particular tone distinguishable from another (cf. Moore, 1982). The question then is not which of the two alternative models is correct. The question is whether anything is gained by adding the framework of source grouping to a parsimonious peripheral channeling model. If it should turn out that all experimental data on segregation can be understood in terms of peripheral channeling, then one would conclude that nothing is gained. This would not mean that the source-grouping model is wrong; it would mean only that the model is irrelevant in explaining the data.

The data from our experiments suggest that peripheral channeling is almost an adequate explanation of the stream segregation effect. The qualification comes from the observation that there is a small enhancement in segregation when the two melodies involve tones of significantly different durations (Conditions 6 and 11). Condition 11, however, was outside the usual rules for an interleaved-melody experiment, because each long-duration tone was so long that it overlapped following tones. We are inclined to discount the evidence for segregation mediated by intensity differences (Condition 4). Had the effect been large, a source-grouping explanation based on loudness might have been supported. But the effect of intensity is small enough that it seems possible that it could be caused by increased spread on a peripheral tonotopic axis with increased intensity (8 dB). In the end, therefore, our experiments do not suggest an important role for source-grouping processes based on gestalt rules so far as stream segregation itself is concerned. It is possible that there are other experimental conditions, not unexplored by us, for which the complex principles of source grouping become important. Segregation based on different temporal evolutions of the spectra would seem to be a likely candidate.

Given the, admittedly unconventional, importance that we attach to peripheral channeling in this article, there is a need for a more complete description of the meaning of peripheral channeling in the context of stream segregation. It certainly cannot mean that tones that are assigned to one auditory stream all have to be in the same peripheral channel. The tones of a musical scale do not break into separate streams as the scale crosses a critical band boundary (or any other boundary). Instead, the stream segregation effect is observed when there are large and rapid jumps along some dimensions(s). To solve this problem, Van Noorden's view is helpful. He regards the integration of tones into a single stream as a default process. Therefore, there is no burden of explanation if a listener hears an integrated stream. What needs to be explained is the segregation of streams, where successive tones are alternately assigned by the listener into separate patterns. In this context, the peripheral channeling model says that the segregation will occur if successive tones are in different peripheral channels. The model denies that segregation will occur if successive tones differ along some dimension(s) that do not correspond to different peripheral channels. The emphasis on successive tones, ignoring the overall range of an extended melody, appears also in the crossing and contact counting model of Appendix 2.

Although our data point up the importance of peripheral channeling in stream segregation, it cannot be supposed that the process of stream segregation actually takes place in the peripheral auditory system. A peripheral origin would seem to be excluded by either of our tone-color experiments (Conditions 3 or 10). Here, tones with fundamental frequencies in the same range are segregated because they have harmonics that do not share the same peripheral channels. A possible explanation for this result is that the harmonics are first peripherally segregated and then somehow carry the melody pitches along with them. Evidence against this explanation is that noise bands that are separately segregated cannot carry melody pitches with them (Condition 8). It appears, therefore, that the harmonics of individual complex tones in our tone color experiments are integrated to form individual tone entities before or in parallel with the process that segregates the individual complex tones into streams. The process whereby the harmonics of individual tones are integrated must be a central one, combining the outputs of different peripheral tonotopic channels (Hartmann, 1988). Just how peripheral channeling makes its presence felt in such a central process is unclear, although it is known that tonotopic organization, established in the periphery, is maintained all the way through to the auditory cortex (Romani, Williamson, & Kaufman, 1982). What does seem clear is that peripheral channeling is the dominant characteristic that is seen in stream segregation.[2, 3]

Appendix 1

Common Key Experiment

In the experiments reported in the body of this paper, the melodies all had the same mean semitone number. Transposing melodies to a common mean note maximizes the number of crossings between melody pairs, but puts the different melodies in different keys. We adopted this procedure because we conjectured that maximizing the number of melodic crossings would make the task most difficult, a desirable goal given our highly trained listeners. This conjecture might have been wrong. Possibly the task would have been more difficult if all the melodies had been in the same key, even though melodic crossings were not maximized. The experiment described in this appendix checks our conjecture.

The experiment was run after the completion of the Experiment 1. Five of the seven subjects from Experiment 1 participated; their melody sets were unchanged. The condition was the null standard, Condition 0. Six runs with all melodies in the same key were randomly interleaved with six runs with equal mean semitone number for all melodies. The results are shown in Table 3. The table shows that for all five listeners, performance was better with melodies in the same key than with melodies having equal mean semitone number. The difference is significant at the .001 level for two of the five and not significant for the others. We conclude that our conjecture is modestly supported by the data. In any case there is no support for the contrary view. Therefore, it appears that the procedure used in the experiment, whereby mean note numbers were equated, is the correct way to make melody identification difficult.

---

[2] Part of this work formed Mr. Johnson's senior honors project in the Department of Psychology at Michigan State University.

[3] Helpful advice on the experiments and their interpretation was given by Drs. A. Bregman, J. Dowling, and S. McAdams. Ms. Anne Bloomquist lent us a large stack of scores of children's songs that contributed to the melody table. Mr. Tom Lemense designed and built the digital circuits in the response box. Dr. Brad Rakerd of the Michigan State Audiology and Speech Department assisted greatly in the analysis of the data. This work was supported by NIDCD grant R01 DC00181.

TABLE 3

Average Number of Correct Pairs in 30 Trials and Standard Deviation ($N-1$ Weight) for $N = 6$ Runs

Listener	Equal Mean Semitone Number	Same Key
1	20.7 (3.3)	27.7 (1.6)
2	17.4 (2.8)	19.0 (3.1)
4	11.6 (2.0)	18.8 (1.9)
6	25.6 (3.3)	27.6 (1.7)
7	22.0 (2.7)	24.6 (2.0)

Appendix 2

Melody Crossing and Contact Statistics

In order to identify two interleaved melodies, a listener must track each melody for some span of time. It is reasonable to assume that a crossing of the melodic lines of the two melodies makes it more difficult to track either melody. It also seems likely that a contact between the melodic lines, where the two melodies touch but do not cross, contributes to the difficulty, perhaps not as much as a crossing. This appendix is concerned with the calculation of crossing and contact statistics for a set of melodies.

The first problem is to define the terms *crossing* and *contact*. There are actually several different definitions that might be used, with no obvious way to choose among them. This appendix introduces one such definition with several desirable properties: The definition is unambiguous. It is easily implemented with a numerical algorithm. It counts crossings and contacts in a way that generally agrees with the impression that one gets from looking at a plot of the interleaved melodic lines on graph paper or on a musical staff.

It is assumed that the two melodies A and B have the same number of notes, namely N. Therefore successive notes of melody A, expressed in semitone number, are $A(1)$, $A(2)$, $A(3)$, . . . $A(N)$. The notes of melody B are $B(1)$, $B(2)$, $B(3)$, . . . $B(N)$. By definition, the odd-numbered notes of the interleaved string of notes are those of the A melody, and therefore the string of semitone numbers, as presented to the listener is

$$S = S(1)S(2)S(3)\ldots S(2N) = A(1)B(1)A(2)B(2)A(3)B(3)\ldots A(N)B(N) \qquad (1)$$

For present purposes we will assume that the string is presented only once and does not cycle back to the start.

THE ALGORITHM

The algorithm begins with the sequence of successive differences, defined as

$$d(i) = S(i+1) - S(i), \qquad (i = 1, 2N-1) \qquad (2)$$

We next define sequence D, which is made from sequence d by changing the signs of all the even members of sequence d.

$$D(i) = [S(i+1) - S(i)](-1)^{(i-1)}, \qquad (i = 1, 2N-1) \qquad (3)$$

The algorithm asserts that the number of melodic crossings is equal to the number of sign changes in D, as defined by Eq. (3), and that the number of contacts is equal to the number of zeros in D. This rule is thought to be adequate because it gives plausible results in a number of simple but general test cases, illustrated in Figure 8.

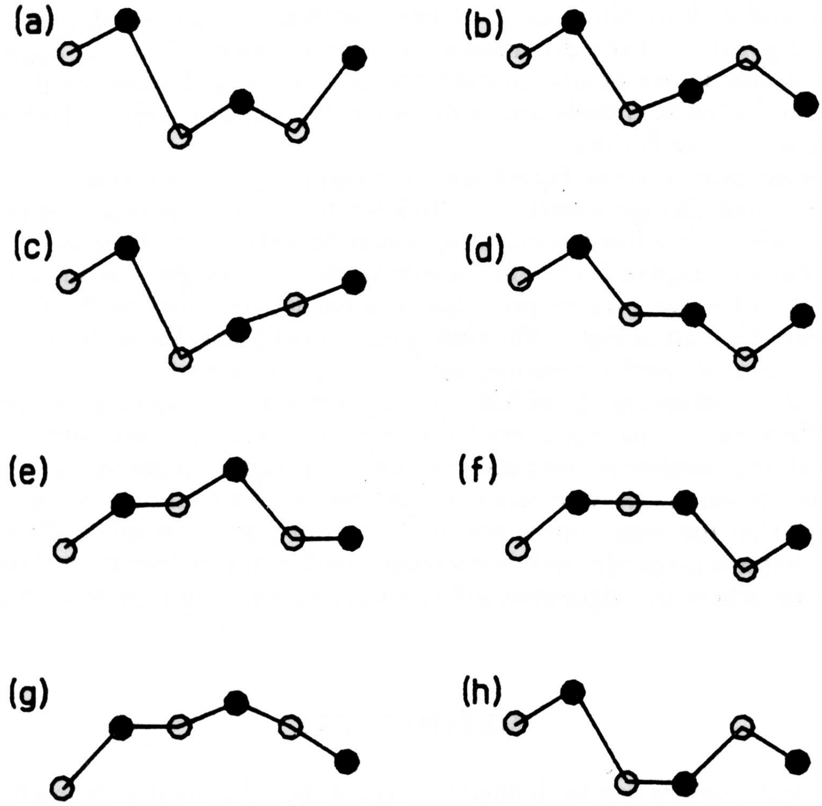

Fig. 8. Interleaved melodies, A with open circles, B with closed circles, to illustrate the operation of the crossing and contact counting algorithm of Appendix 2.

SIMPLE CROSSINGS

The algorithm is most easily tested in cases where the string *S* contains no repeated notes. There are, therefore, no contacts. There may or may not be some crossings.

Figure 8a shows a case that is central to the explanation of the algorithm. It contains no crossings. The B melody is always above the A melody, which means that there is no B note with a height between the heights of successive A notes, and there is no A note with a height between the heights of successive B notes. As a result, the sequence of successive differences in equation (2) is an alternating sequence, and the modified sequence *D* in equation (3) has no sign changes, as indicated by the signs below the symbols in Figure 8a. Clearly it is not necessary that all the B notes lie above (or below) all the A notes in order for there to be a case of no crossings. The overall frequency range of the melodies can migrate up or down without creating a crossing. Instead, a crossing depends on the local environment.

Figures 8b and 8c show cases where there are one and two crossings, respectively. In Figure 8c, the B notes are above the A notes until the fourth note of the string, where a B note intervenes between two A notes. Then on the fifth note of the string, an A note intervenes between two B notes. There are, therefore, two crossings.

CONTACTS

Figure 8d shows a single contact, where an element of the successive difference sequence *D* is zero. There are no sign changes in the sequence, so there are no crossings. Notes B lie entirely above notes A except at the contact point where the A note string and the B note string touch.

Figures 8e and 8f both show cases of two contacts and no crossings. No account is taken in this algorithm of the distribution of contact points. They all count the same.

Figure 8g shows a case of one contact and one crossing. Except for the contact, the B melody can be heard as entirely above the A melody until the very end where an A note falls between successive B notes.

An interesting case arises in Figure 8h. Like Figure 8g, there is one zero and one sign change, but the sign change immediately follows the zero. This indicates that there is a crossing, as in Figure 8b, but the crossing occurs at a flat part of the graph where there is a contact. Possibly Figure 8h should be interpreted as a single crossing and no contact (like Figure 8b). However, we suspect that it is harder to associate the first and second B notes in Figure 8h than in Figure 8b. This extra difficulty may be modeled by interpreting Figure 8h as a contact and a crossing, as the algorithm says.

In summary, the counting algorithm is straightforward: It begins with the sequence of successive differences *D*, having every other sign reversed. The number of zeros in the sequence equals the number of contacts; the number of sign changes in the sequence equals the number of crossings. (A zero has no sign and cannot count for a sign change.)

It is evident that the algorithm is not derived from prior principles. What can be said in its favor is that it is plausible, and that testing it with many different interleaved melodies revealed no case where the algorithm led to an ambiguous result or to a counterintuitive result.

MELODY SETS

Crossings and contacts can be defined for sets of melodies by pairing each melody with every other melody. It is necessary to observe the order of the pairs because the crossings and contacts may be different for different orders. For example, Yankee Doodle as A and Peter Pumpkin Eater as B give 4 crossings and 9 contacts whereas the reverse order gives 6 crossings and 10 contacts.

For a set of 12 melodies, as used in our experiments, there are 132 ordered pairs. The number of crossings and contacts in a set can best be described by an average number of crossings and contacts per pair. Such an average is given in Table 4 for the sets of each of the listeners in Experiment 1. A second set of numbers is given for those listeners who heard melodies all in the same key.

It is evident that putting all the melodies in the same key leads to a dramatic reduction in the number of crossings, compared with the case when mean notes are equated for each melody. The reduction, on average from 4.6 crossings to 2.2 crossings, may well be responsible for the improved performance, shown in Appendix 1, from 65% correct pairs to 78%. Also evident is that this change leads to an increase in the number of contacts.

TABLE 4
Average Crossings and Contacts per Ordered Pair of Melodies

Listener	Mean Note = A440		All Melodies in Same Key	
	Cross	Contact	Cross	Contact
1	4.56	2.95	2.43	5.14
2	4.56	3.18	2.15	5.01
3	4.65	3.29		
4	4.88	3.30	2.27	5.24
5	4.63	3.47		
6	4.61	3.76	2.14	4.06
7	4.23	3.74	1.83	4.76

The latter change is not unexpected. When simple melodies are all put into the same key, there is frequent use of the same notes.

Appendix 3

Interaural Time Differences

The interaural time differences for Condition 7 were created by two all-pass networks, one network for each melody (Hartmann, 1979). Therefore, the interaural time differences were not independent of the tone frequency. In this appendix we calculate the variation due to this experimental simplification and show that it is small.

The time delay introduced by an all-pass network is given by

$$T_d = (N/\pi f)\tan^{-1}(\pi f T_c/N)$$

where T_c is the characteristic time for a single all-pass stage, and N is the number of stages, here taken to be identical. The delay depends on the tone frequency f. As N increases, the time delay becomes increasingly insensitive to the frequency (Hartmann, 1978), so that the all-pass response becomes closer to a true time delay. For the narrow range of frequencies used in this experiment, two stages give adequate frequency insensitivity. With $N = 2$ and T_c set to give 500 μsec delay at 440 Hz (MIDI note 69), the delay at the maximum frequency of 622 Hz (MIDI 75) is 482 μsec and the delay at the minimum frequency of 311 Hz (MIDI 63) is 510 μsec. Values of 482 and 510 are close to the nominal value of 500, and the range of 30 μsec is not much larger than the interaural-time-difference JND for a tone of constant frequency (500 Hz), namely 20 μsec (Durlach & Colburn, 1978, p. 418).

References

Anstis, S., & Saida, A. Adaptation of streaming of FM tones. *Perception and Psychophysics*, 1985, ***11***, 257–271.

Bregman, A. S. The formation of auditory streams. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, NJ: Erlbaum, 1978.

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

Bregman, A. S., & Campbell, J. Primary auditory stream segregation and the perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 1971, ***89***, 244–249.

Bregman, A. S., & Dannenbring, G. L. The effect of continuity on auditory stream segregation. *Perception and Psychophysics*, 1973, ***13***, 308–312.

Bregman, A. S., & Pinker, S. Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, 1978, ***32***, 19–31.

Bregman, A. S., & Rudnicky, A. I. Auditory segregation: Stream or streams? *Journal of Experimental Psychology, Human Perception and Performance*, 1975, ***1***, 263–267.

Brokx, J. P. L., & Nooteboom, S. G. Intonation and the perceptual segregation of simultaneous voices. *Journal of Phonetics*, 1982, ***10***, 23–36.

Butler, D. A further study of melodic channeling. *Perception and Psychophysics*, 1979, ***25***, 264–268.

Dannenbring, G. L., & Bregman, A. S. Streaming vs fusion of sinusoidal components of complex tones. *Perception and Psychophysics*, 1978, ***24***, 369–376.

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

Deutsch, D. Binaural integration of melodic patterns. *Perception and Psychophysics*, 1979, ***25***, 399–405.

Deutsch, D. Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music*. New York: Academic Press, 1982, pp. 91–134.

Dowling, W. J. Rhythmic fusion and perceptual organization (abstract). *Journal of the Acoustical Society of America* 1968, ***44***, 369.

Dowling, W. J. The perception of interleaved melodies. *Cognitive Psychology*, 1973, *5*, 322–337.

Dowling, W. J., Mei-tak Lung, K., & Herbold, S. Aiming attention in pitch and time in the perception of interleaved melodies. *Perception and Psychophysics*, 1987, ***41***, 642–656.

Durlach, N. I., & Colburn, H. S. Binaural phenomena. In E. Carterette (Ed.), *Handbook of perception*, vol. IV. New York: Academic Press, 1978.

Handel, S., Weaver, M. S., & Lawson, G. Effect of rhythmic grouping on stream segregation. *Journal of Experimental Psychology, Human Perception and Performance*, 1983, ***9***, 637–651.

Hartmann, W. M. Flanging and phasers. *Journal of the Audio Engineering Society*, 1978, ***26***, 439–443.

Hartmann, W. M. Some psychoacoustical experiments with all-pass networks. *American Journal of Physics*, 1979, ***47***, 29–34.

Hartmann, W. M. Digital waveform generation by fractional addressing. *Journal of the Acoustical Society of America*, 1987, ***82***, 1883–1891.

Hartmann, W. M. Pitch perception and the segregation and integration of auditory entities. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Auditory function*. New York: Wiley, 1988.

Kubovy, M. Concurrent-pitch segregation and the theory of indispensable attributes. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Erlbaum, 1981.

Massaro, D. W. Perceptual grouping in audition. *Perception*, 1977, ***6***, 541–553.

McAdams, S. *Spectral fusion, spectral parsing and the formation of auditory images*. Unpublished doctoral dissertation, Stanford University, Stanford, CA, 1984a.

McAdams, S. The auditory image: A metaphor for musical and psychological research on auditory organization. In R. Crozier & A. Chapman (Eds.), *Cognitive processes in the perception of art*. Amsterdam: North Holland, 1984b, pp. 183–187.

McAdams, S., & Bregman, A. S. Hearing musical streams. *Computer Music Journal*, 1979, ***3***, 26–43.

Miller, G. A., & Heise, G. A. The trill threshold. *Journal of the Acoustical Society of America*, 1950, ***22***, 637–638.

Moore, B. C. J. *An introduction to the psychology of hearing*, 2nd ed. London: Academic Press, 1982.

Noorden, L. P. A. S. van. *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation. Institute for Perception Research, Eindhoven, Holland, 1975.

Noorden, L P. A. S. van. Minimum differences of level and frequency for perceptual fission of tone sequences ABAB. *Journal of the Acoustical Society of America*, 1977, ***61***, 1041–1045.

Romani, G. L., Williamson, S. J., & Kaufman, L. Tonotopic organization of the human auditory cortex. *Science*, 1982, ***216***, 1339–1340.

Singh, P. G. Perceptual organization of complex tone sequences: A tradeoff between pitch and timbre? *Journal of the Acoustical Society of America*, 1987, ***82***, 886–899.

Steiger, H., & Bregman, A. S. Competition among auditory streaming, dichotic fusion, and diotic fusion. *Perception and Psychophysics*, 1982, ***32***, 153–162.

Terhardt, E. On the perception of periodic sounds (roughness). *Acustica*, 1974, ***30***, 202–213.

Tougas, Y., & Bregman, A. S. Crossing of auditory streams. *Journal of Experimental Psychology*, 1985, ***11***, 788–798.

Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. Auditory sequence confusion of patterns other than speech or music. *Science* 1969, **164**, 586–587.

Wessel, D. L. Timbre space as a musical control structure. *Computer Music Journal*, 1979, *3*, 45–52.

Yin, T. C. T., & Kuwada, S. Neuronal mechanisms of binaural interaction. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Dynamic aspects of neocortical function*, New York: Wiley, 1984.