Perception & Psychophysics
1989, 46 (5), 487-496

The perceptual segregation of simultaneous auditory signals: Pulse train segregation and vowel segregation

MAGDALENE H. CHALIKIA and ALBERT S. BREGMAN
McGill University, Montreal, Quebec, Canada

In the experiments reported here, we attempted to find out more about how the auditory system is able to separate two simultaneous harmonic sounds. Previous research (Halikia & Bregman, 1984a, 1984b; Scheffers, 1983a) had indicated that a difference in fundamental frequency (F0) between two simultaneous vowel sounds improves their separate identification. In the present experiments, we looked at the effect of F0s that changed as a function of time. In Experiment 1, pairs of unfiltered or filtered pulse trains were used. Some were steady-state, and others had gliding F0s; different F0 separations were also used. The subjects had to indicate whether they had heard one or two sounds. The results showed that increased F0 differences and gliding F0s facilitated the perceptual separation of simultaneous sounds. In Experiments 2 and 3, simultaneous synthesized vowels were used on frequency contours that were steady-state, gliding in parallel (parallel glides), or gliding in opposite directions (crossing glides). The results showed that crossing glides led to significantly better vowel identification than did steady-state F0s. Also, in certain cases, crossing glides were more effective than parallel glides. The superior effect of the crossing glides could be due to the common frequency modulation of the harmonics within each component of the vowel pair and the consequent decorrelation of the harmonics between the two simultaneous vowels.

In most natural listening situations, at any given moment, the vibrations of our eardrums are the result of several sound sources active at the same time. In such cases, the auditory system is faced with the problem of separating the pattern of superimposed sounds into individual subsets of components that correspond to the separate sound sources. Otherwise, nonveridical percepts will be formed, in each of which some of the properties of the perceived sound will derive from one acoustic source, while others will derive from other sources.

The present experiments were designed to examine the effects of two types of grouping cues on the perceptual fusing of certain parts of a spectrum with one another. One of these was the "F0" cue: The spectrum contains partials that can be grouped into two subsets by virtue of the fact that each subset contains the harmonics of a different fundamental (F0). The second was the "common fate" cue: When a set of harmonics of the same F0 is glided, the harmonics change in frequency on parallel paths, and thus the impression of harmonicity is reinforced.

A number of different experiments have been conducted to study the ability to attend to one speech (or speech-like) signal in a mixture of natural or LPC-monotonized (resynthesized) continuous speech signals (Broadbent, 1952; Brokx & Nooteboom, 1982; Brokx, Nooteboom, & Cohen, 1979; Cherry, 1953; Darwin, 1981; Egan, Carterette, & Thwing, 1954; Treisman, 1960). These studies have indicated that perceptual separation can be facilitated when the signals have different pitches (resulting from F0 differences). The results of these experiments were confirmed by Scheffers (1983a), who investigated the effects of F0 differences on the identification of two simultaneous steady-state synthetic vowels. Identification scores improved significantly when the F0s differed by more than 1-2 semitones.

All these studies suggest that the components of a harmonic series—associated with a particular fundamental frequency—tend to fuse together and thus perceptually separate from another simultaneous set of harmonics, if the latter set can be attributed to another fundamental sufficiently different from the first.

The precursor for the present experiments on the common fate cue was a set of experiments done in our laboratory to investigate the role of parallel gliding in binding frequency components together (Bregman, McAdams, & Halpern, 1978, reported in McAdams & Bregman, 1979; Halikia, Bregman, & Halpern, 1982). It was found that

This article is based on research that formed part of a PhD dissertation, submitted to McGill University by M. H. Halikia (now spelled Chalikia) in 1985. The research was supported by the Natural Sciences and Engineering Research Council of Canada, through a grant awarded to A. S. Bregman. We wish to thank S. McAdams, V. Ciocca, B. Roberts, and two anonymous reviewers for their helpful comments on an earlier version of the manuscript. M. H. Chalikia is now in the Department of Psychology, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201. Requests for reprints should be sent to Albert S. Bregman, Department of Psychology, McGill University, 1205 Dr. Penfield Ave., Montreal, Quebec H3A 1B1, Canada.

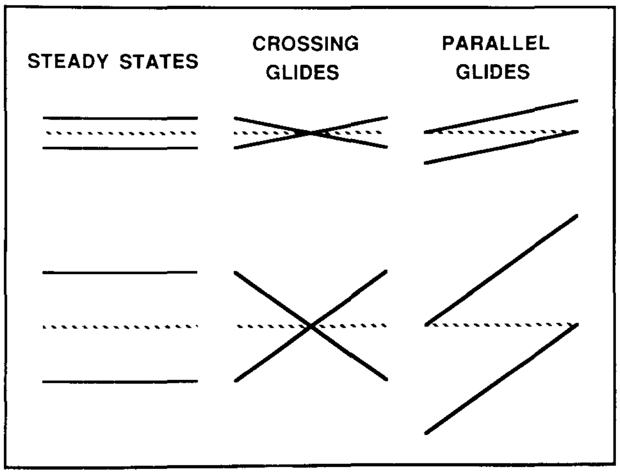

Figure 1. An illustration of two kinds of pitch differences for steady states, crossing glides, or parallel glides. Dashed lines represent the steady-state 140-Hz F0.

fusion would take place among any subset of partials that was moving in parallel in the frequency domain. For example, if three partials ascended in parallel and three partials descended in parallel, the listener heard two tones, one gliding up and one gliding down in pitch.

With the similar goal of binding partials together by changing their frequencies over time in a correlated fashion, McAdams (1982, 1984a, 1984b), following a demonstration by Chowning (1980), was able to split a complex tone into two sounds, each with its own pitch, by imposing different random frequency modulations on different subsets of harmonics (even and odd ones).

In other studies (Halikia & Bregman, 1984a, 1984b), it has been reported that the perceptual segregation of two superimposed vowels is facilitated when the vowels have moving pitch contours (moving F0s), as compared with when they have steady contours (steady-state F0s).

Although vowel sounds are the kinds of pitch-carrying sounds that are most interesting to study, their use as stimuli presents special problems: vowels are familiar, and they have sharp peaks in their spectra. It is not clear how much of the segregation that occurs is due to these two factors. Accordingly, in Experiment 1, we employed only steady-state and gliding pulse trains as stimuli.

In Experiments 2 and 3, synthetic vowels were used as stimuli, because it was hard to tell, using the method that we had employed in the pulse-train experiment, how clearly the two subsets of harmonics were perceptually segregated. In the two later experiments, the subsets formed two vowels whose identities could be reported. These studies also formed an extension of Scheffers's (1983a) work by determining the effects of gliding F0s on the segregation of simultaneous vowels.

Figure 1 illustrates some of the pitch contours that were used in these experiments, for vowel pairs consisting of steady-states, crossing glides, or parallel glides. Similar relations were used in Experiment 1.

EXPERIMENT 1

The aim in Experiment 1 was to examine the perceptual segregation of harmonic nonspeech sounds. Pulse trains were used as stimuli because they have a flat spectrum, and because, by virtue of their periodicity, they evoke a pitch sensation. The pairs used contained two superimposed pulse trains—gliding or steady-state—with either the same or different F0s. In the gliding pulse trains, the harmonics of each member of the pair swept through a frequency range, moving in the same direction and maintaining constant frequency ratios between them. With such stimuli, the task of the subject becomes one of segregating concurrent pitches.

In Experiment 1, band-pass filtered pulse trains were also used. It has been claimed (Moore, 1973; Plomp, 1964, 1967) that pitch perception must be based on the presence of low-numbered frequency harmonics, which are resolved by the peripheral auditory system. Moore (1982) has suggested that a clear perception of pitch may not be evoked when only a small group of high unresolvable harmonics is present (e.g., above the 15th, depending on the F0), because "the time interval corresponding to the F0 will fall outside the range which can be analysed in the channel responding to those harmonics" (p. 142). We wanted to see whether high harmonics could provide the information necessary to segregate two concurrent sounds. We examined this issue by gradually decreasing the range of frequencies available for discrimination.

All signals were band-pass filtered so that neither signal within a pulse-train pair would contain harmonics in a region where the other did not.

Also, the passband was moved through a range of frequencies so that progressively more of the lower harmonics were removed. We wanted to see, assuming that discrimination on the basis of periodicity information is possible, whether there were any differences in the ability to segregate the sounds, given the numbers and frequencies of harmonics available.

Method

Stimuli. The standard stimulus was a pulse train with an F0 of 140 Hz and 32 harmonics. The duration of the pulse train was 1 sec, including 200 msec of rise/fall. This served as a reference (standard) stimulus on the basis of which the other signals were synthesized. Additional pulse trains were created, using either steady-states (SSs, with flat F0s) or crossing glides (CGs, with changing F0s), such that their F0s belonged in a "high" or "low" frequency range (above or below the original F0 of 140 Hz). Combinations of those pulse trains, taken two at a time, always with one high and one low (both components being either SSs or CGs), yielded pulse-train pairs with F0 differences of 0, ½, and 2 semitones. No separations between 0 and ½ semitones were tested, because Scheffers (1983a) had found no significant effect in that range. Table 1 shows

Table 1
Values of the F0s for the Different Semitone Frequency Separations for All the Experiments

Middle Value	High F0	Low F0	Separation of F0s
	Steady States and Crossing Glides		
140	140	140	0
140	142.10	137.90	0.5
140	148.40	132.07	2
140	152.06	128.40	3
140	166.51	117.70	6
140	181.71	107.80	9
140	197.90	98.90	12
	Parallel Glides		
140	140	140	0
140	144.2	135.90	0.5
140	166.50	117.70	3
140	197.90	98.90	6
140	280	70	12

Note—Middle value, high F0, and low F0 are in Hz; separation of F0s is in semitones.

the values of the F0s for the different semitone frequency separations used in all the studies. The two pulse trains were positioned symmetrically around 140 Hz on log-frequency coordinates. For example, in the case where the F0 difference was ½ a semitone, for SSs the high pulse train had an F0 of 142.1 Hz and the low pulse train had an F0 of 137.93 Hz, which yielded a ratio of 1.03. For CGs, the F0 of the low started at 137.93 Hz and ended at 142.1 Hz, and the F0 of the high started at 142.1 Hz and ended at 137.93 Hz. In this case, "high" and "low" define the point at which the glide started. Both glides swept through the same range of frequencies, one gliding up and the other gliding down. F0 separations indicated constant separations for the SSs, but only maximum separations for the CGs. That is, for SSs, a difference of, say, 2 semitones between the F0s in the mixture refers to a *constant frequency separation* of that magnitude maintained throughout the duration of the signal. For the CGs, the difference of 2 semitones refers to the *maximum frequency separation* obtained only at the beginning and the end points, and to less than that separation at all the points in between. Therefore, the given nominal frequency separation of the CGs overestimates their F0 separation and makes it harder for them to be more segregated than the F0-matched SS condition. All glides were linear on log-frequency coordinates (gliding a constant number of octaves per second).

Once the original wideband (WB) six pairs were synthesized (three frequency separations with steady-state and gliding pitch contours), additional pairs were created by filtering. There were three sets of filtered pairs, each based on the preceding six stimuli. The first set (referred to as FA) contained stimuli band-passed in the range 1500-4000 Hz (with a 40-dB drop from 1600 to 1400 Hz on the lower spectral edge, and from 3800 to 4200 Hz on the higher edge). The second set (FB) contained stimuli band-passed in the range 2100-4000 Hz (with a 40-dB drop from 1900 to 1250 Hz on the low edge and from 3800 to 4200 Hz on the high edge). Finally, the third set (FC) contained stimuli band-passed in the range 2600-4000 Hz (with a 40-dB drop from 2750 to 2550 on the low edge and from 3800 to 4200 Hz on the high edge). The levels of all pulse trains were equated and measured with a General Radio 1551-C sound pressure level meter (A weighting) with a flat-plate coupler.

Two test files were prepared, each with two repetitions of the 24 pairs, in a random order, so that each tape contained a total of 48 stimulus pairs. The listeners were assigned to one of these files in a counterbalanced fashion.

Procedure. There were three independent variables: filtering (WB, FA, FB, and FC), pitch contour (SS vs. CG), and F0 separation (0, ½, and 2 semitones). The dependent variable was the segregation score.

The experiment started with a pretest session, during which the listeners were seated individually in a test chamber where they listened to a file that contained the six unfiltered stimuli in a random order. On each trial, the stimulus pair was repeated six times. Facing the subjects on a table was an answer sheet, on which they had to give a response of "1" or "2," depending on whether they had heard one or two different sounds in the pair. In the latter case, one would be heard as a high tone and the other would be heard as a low tone for the SSs; for the CGs, one would be heard as gliding up and the other as gliding down. No feedback was given. After having completed this familiarization task, the listeners proceeded with the testing session, which was similar to the training session.

Subjects. Twenty paid McGill undergraduate students served as subjects.

Apparatus. All the stimuli were synthesized on a PDP-11/34 computer (Digital Equipment Corporation), using the MITSYN signal-processing software (Henke, 1980). The original (wideband) signals were synthesized with 32 harmonics by means of additive synthesis. The band-pass filtering was done computationally, using a 127-point FIR filter with a Hamming window, and its effect was verified by spectral analysis. The signals were output via a 12-bit digital-to-analog converter, at a sampling rate of 12 kHz, low-pass filtered at 5.5 kHz (antialiasing) with a Rockland 851 filter and presented binaurally over TDH-49P headphones at 65 dBA. The subject was seated in an Industrial Acoustics 1202 audiometric chamber.

Results and Discussion

Scoring. Each subject received a score of 1 or 2, depending on whether he or she had reported hearing one or two sounds within each pair. After all the entries had been recorded, the two scores for each pair were averaged, and these averaged entries were used in the statistical analysis.

Analysis. A three-way analysis of variance for repeated measures gave a significant effect for filtering [$F(3,57) = 9.87, p < .0001$], for pitch contour [$F(1,19) = 4.99$,

Table 2
Mean Scores for Each of the F0 Separations (in Semitones) for Each Level of the Filtering and Pitch Contour Variables

		F0 Separations		
		0	.5	2
		Filtering		
WB	*M*	1.22	1.90	1.94
	SE	0.02	0.03	0.04
FA	*M*	1.10	1.83	1.85
	SE	0.04	0.01	0.02
FB	*M*	1.06	1.71	1.81
	SE	0.01	0.03	0.03
FC	*M*	1.07	1.80	1.71
	SE	0.01	0.03	0.00
		Pitch Contour		
Steady States	*M*	1.09	1.79	1.81
	SE	0.04	0.04	0.04
Crossing Glides	*M*	1.13	1.83	1.86
	SE	0.04	0.04	0.05

Note—WB = wideband. FA, FB, and FC = filtered.

$p < .03$], and for F0 separation [$F(2,38) = 114.91$, $p < .0001$], as well as for the filtering × F0 separation interaction [$F(6,114) = 2.49$, $p < .02$].

The mean scores for each of the fundamental frequency separations and for each of the four filtering conditions averaged across pitch contours are given under Filtering in Table 2.

The frequency separation of the fundamentals affected all filtering conditions [$F(2,69) = 151.99$, $p < .01$ for WB; $F(2,69) = 166.71$, $p < .01$ for FA; $F(2,69) = 181.42$, $p < .01$ for FB; $F(2,69) = 147.09$, $p < .01$ for FC], as indicated by tests of simple effects. In all cases, the ½- and 2-semitone differences gave higher scores than did the 0-semitone difference ($p < .01$, Newman-Keuls tests). There were no differences in scores between the ½- and 2-semitone conditions.

Filtering had an effect for all frequency separations [$F(3,142) = 36.04$, $p < .001$, for the 0-semitone difference; $F(3,142) = 24.1$, $p < .01$, for the ½-semitone difference; $F(3,142) = 36.04$, $p < .01$, for the 2-semitone difference], as indicated by tests of simple effects. However, Newman-Keuls tests found only one pairwise comparison significant. For the 2-semitone difference, the score in WB was higher than the score in FC ($p < .05$). The scores for the CGs were higher overall than those for the SSs ($p < .05$, Newman-Keuls). The mean values are shown under Pitch Contour in Table 2.

The results from this study on pulse trains indicate, in agreement with previous findings (Halikia & Bregman, 1984a, 1984b; Scheffers, 1983a; Zwicker, 1984), that increased F0 separations and the use of CGs can facilitate the perceptual separation of simultaneous sounds.

The superiority of the CGs, under conditions where it is found, is probably due to two mechanisms. It should be recalled that the different directions of motion destroy any harmonic relations that exist accidentally between the harmonics of two subsets of partials when two complex tones are played as SSs. Therefore, the effect can be viewed as just another result of the mechanism that groups partials by their harmonic relations. The second mechanism may be one that groups partials if they are moving in parallel, regardless of their harmonic relations. McAdams has reported observations that support the idea that parallel motion of a set of partials on log-frequency coordinates can bind that subset together into a single fused sound and can segregate that subset from other concurrent partials (McAdams, 1984b, 1985, personal communication).

The presence of CGs facilitated segregation in the case of band-pass filtered signals just as in the case of unfiltered signals. Apparently the number of harmonics present in all three cases (FA, FB, and FC) is still large enough to provide sufficient information for the perception of pitch by means of some periodicity-detection mechanism. It is possible that the two pulse trains were segregated on that basis.

The one significant drop in performance between the unfiltered and one of the filtered conditions (FC, at the 2-semitone difference) could be attributed to loss of information, since, in the case of the filtered signals, there were fewer frequency channels available from which to extract periodicity information. It is possible that if the passband had been still narrower, a significant decrease in performance would have been found at the ½-semitone difference as well.

The segregation of the pulse trains from one another did not involve identification, since the listeners only had to decide whether one or two sounds had been present. Consequently, we did not know on what basis they were making their judgments—an increase in beats or a sensation of roughness, for example, rather than really hearing out two tones. Therefore, we turned to the use of synthesized vowels in an identification task.

EXPERIMENT 2

In Experiments 2 and 3, we further explored the findings of Experiment 1 by employing vowels instead of pulse trains as stimuli.

In the perception of vowels, small modulations in pitch may facilitate the identification of a vowel with a method that has nothing to do with the grouping of partials on the basis of harmonic relations. McAdams and Rodet (1988) have found that the ambiguity of the identity of a vowel decreases when a small amount of frequency jitter is applied to the fundamental (and hence to all the harmonics). They assumed that frequency modulation had this effect, because all the harmonics that belonged to a particular F0 moved up and down in frequency with it. By increasing or decreasing in amplitude as they approached or receded from the nearest resonance (formant) peak, the harmonics gave information about the position of the peak, thereby improving the identification of the vowel. In effect, the changes in the amplitudes of the harmonics "traced out" the formants, so we can refer to this as the "formant tracing" cue. In the case of SS vowels, such information was not available. It should be mentioned, at this point, that Sundberg's (1982) work on high-pitched vowel identification in the presence of frequency vibrato has shown slight or detrimental effects, compared with the case in which there is no vibrato.

In Experiment 2, we compared vowel pairs with steady-state or gliding pitches. Crossing glides were used.

Table 3
Formant Frequencies of the Vowels

Formant	a	ʊ	i	ɛ
F_1	730	440	270	530
F_2	1090	1020	2290	1840
F_3	2440	2240	3010	2480

Note—Values are given in hertz.

Method

Stimuli. Four vowel sounds were synthesized, using a serial three-formant method. The glottal pulse was created by additive synthesis of 36 harmonics, with an intensity drop-off of 12 dB/octave, modified by a subsequent radiation characteristic imposed by a first-order difference filter, yielding a net spectral slope of −6 dB/octave. The formants were imposed by a series of three formant filters. The vowels were /a/, /ʊ/, /i/, and /ɛ/, each with a duration of 1 sec, including 200 msec rise/fall. Each contained 36 harmonics and a fundamental frequency at 140 Hz.

The formant frequencies, set equal to the ones found in Peterson and Barney (1952) for a male voice, are shown in Table 3. The formant frequencies remained constant at these values, regardless of any changes in the F0s. The bandwidths of the formant filters were set at 100, 120, and 140 Hz for F_1, F_2, and F_3, respectively.

The vowels were presented as pairs. All possible combinations of these vowels, taken two at a time, resulted in six pairs. Further pairs were created by altering the F0 of each vowel so that the F0s of the two vowels were separated by ½-, 3-, 6-, 9-, and 12-semitone differences with the two F0s placed symmetrically around 140 Hz (on log-frequency coordinates). The frequency values corresponding to these separations are indicated in Table 1.

The pitch contour (change in F0 over time) used for the synthesis was either a steady state or a crossing glide. All the pairs were synthesized using both types of pitch contours for all the semitone differences. In addition, a second series of pairs was synthesized similarly to those above but opposite in assignment of F0 to vowel. For example, if in the pair /a/ and /ʊ/, /a/ had the high F0 (or downwardly gliding one) and /ʊ/ the low F0 (or upwardly gliding one), in the opposite pair, /a/ had the low F0 and /ʊ/ the high F0. The total number of pairs was 126.

The vowels were synthesized so as to equate them subjectively for equal loudness before mixing.

Two test tapes were prepared, each with the 126 pairs in a random order. The listeners were assigned to one of these two tapes in a counterbalanced fashion.

Procedure. There were three independent variables, vowel pair (six vowel pairs), pitch contour (SS or CG), and F0 separation (0, ½, 3, 6, 9, or 12 semitones). The dependent variable was the identification score.

The experiment started with a pretest session. The listeners were seated individually in a test chamber. At first they were presented with a tape that contained 20 of the pairs that were included on the actual test tapes. Each pair repeated itself 40 times, to make the task easy. Facing the subject on a table was an answer sheet with four words on each line, each of which contained one of the four original vowels. The four words were *beat* (for /i/), *card* (for /a/), *put* (for /ʊ/), and *bed* (for /ɛ/). The subject was instructed to circle the words that contained each of the vowels in the mixture. There was no feedback to the subjects. They were asked to circle only one word if they had heard only one vowel. The listeners proceeded with the actual experiment if they had identified correctly both vowels in at least 16 of the 20 pairs. All listeners reached this criterion. The testing session was similar to the training one, except that each pair was repeated 10 times. The repetitions were employed to raise the task to a suitable level of ease of performance.

Subjects. The subjects were 20 paid male and female McGill undergraduate students.

Apparatus. All the sequences were digitally synthesized in the manner described in Experiment 1. The signals were output via a 12-bit digital-to-analog converter at a sampling rate of 15 kHz, low-pass filtered at 6 kHz by a Rockland 851 filter, and tape-recorded on a Sony TC-654/4 tape recorder. At the time of playback, the signal was presented binaurally over TDH-49P headphones at 70 dBA.

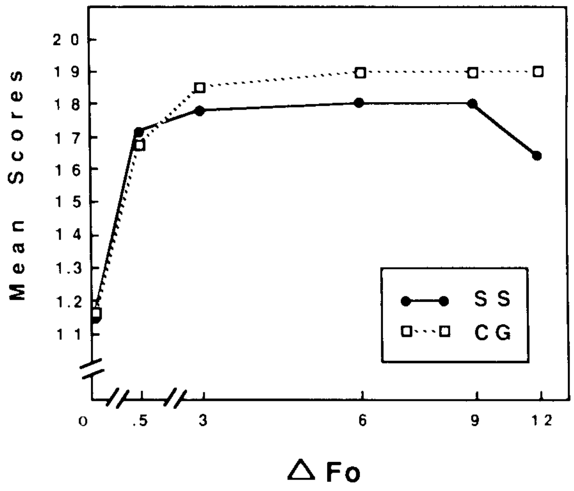

Figure 2. Mean identification scores for steady states (SS) and crossing glides (CG) for each of the F0 separations (log scale) used in Experiment 2. ΔF0 stands for F0 separation.

Results

Scoring. Each subject received a score of 0, 1, or 2, depending on whether none, one, or both of the vowels in the pair had been identified correctly. After all the entries had been recorded, the two scores for every pair (one for the original and one for the opposite pair) were averaged. These averaged entries were used in the statistical analysis. Since the original six mixtures with 0-semitone F0 difference occurred only once, and were not repeated as a baseline set of stimuli for the CG or the opposite pairs, their scores were not included in the analysis.

Analysis. A three-way analysis of variance for repeated measures gave a significant effect for F0 separation [$F(4,76) = 8.64$, $p < .001$], for pitch contour (SS vs. CG) [$F(1,19) = 30.29$, $p < .001$], and for their interaction [$F(4,76) = 9.16$, $p < .001$]. Differences between

Table 4
Mean Scores for Each of the F0 Separations (in Semitones) for Each Level of the Pitch Contour Variable in Experiments 2 and 3

		F0 Separations				
Pitch Contour		.5	3	6	9	12
		Experiment 2				
Steady States	*M*	1.71	1.78	1.80	1.80	1.64
	SE	0.04	0.04	0.03	0.02	0.03
Crossing Glides	*M*	1.68	1.85	1.90	1.90	1.90
	SE	0.05	0.02	0.01	0.02	0 02
		Experiment 3				
		0	.5	3	6	12
Steady States	*M*	1.15	1.65	1.63	1.71	1.38
	SE	0.03	0.03	0.03	0.05	0.03
Parallel Glides	*M*	1.10	1.71	1.71	1.76	1.57
	SE	0.03	0.03	0.05	0.06	0.05
Crossing Glides	*M*	1.11	1.69	1.79	1.81	1.80
	SE	0.03	0.05	0.04	0.05	0 05

particular pairs of vowels were not significant. Figure 2 shows the mean identification scores across vowel pairs for the SSs and CGs for each of the frequency separations. Scores for the 0-semitone difference are shown for reference. The mean values are given under Experiment 2 in Table 4. The frequency separation of the F0 affected the identification scores for both SS [$F(4,144) = 28.74$, $p < .001$] and CG [$F(4,144) = 45.85$, $p < .001$], as indicated by tests of simple effects, but not in the same manner.

In the SS condition, the scores seemed to be very similar for most of the separations (i.e., identification improved as soon as the ½-semitone difference was introduced and remained the same for the other differences)—except for the octave difference, where there was a significant drop in performance, as compared with the 3-, 6-, and 9-semitone intervals ($p < .01, p < .01$, and $p < .01$, respectively, all intervals compared to the octave separation, using Newman-Keuls tests). On the other hand, this did not happen in the CG condition where performance increased significantly from the ½- to the 3-semitone difference ($p < .01$, Newman-Keuls) and remained high without dropping at the octave separation.

Generally, scores were higher when the pitch contours were crossing glides than when they were steady-states, for all separations except the ½-semitone one, as was confirmed by tests of simple effects ($p < .03$, $p < .005$, $p < .009$, and $p < .001$, for the 3-, 6-, 9-, and 12-semitone separations, respectively).

Discussion

The results from the SS condition confirm Scheffers's (1983a) finding that two simultaneous vowels are heard separately when their F0s differ by a certain amount. In spite of differences in the specific F0 values used, it seems that identification performance improves up to a certain level and then remains at that level until the octave separation, where it drops again. The octave case is special, because, due to a great amount of overlap between the harmonics of the two vowels (the harmonics of the vowel with the higher F0 overlap with every other harmonic of the vowel with the lower F0), there is a decrease in the subject's ability to separate the two vowels when the harmonic relations are fixed between the two sounds (as in the case of SSs). The results from the CG condition were generally better, as was expected. It was also apparent that a minimum F0 separation was necessary before the use of CGs could have an additional effect, since the effect was not found at the ½-semitone separation. That this effect was especially strong at the octave separation, where identification performance decreased in the steady-state case, is not surprising. It is obvious that this is due to the fact that, since the harmonics of the ascending and descending glides were moving with respect to one another, they did not maintain the fixed harmonic coincidence at the octave responsible for the drop in the performance on the SS vowels.

An examination of the curves relating vowel identification to separation of F0s shows that, just as with the pairs of pulse trains, the greatest increment that results from separating the two F0s appears in the first half-semitone, a frequency ratio of about 1.03. These data relate to those of other researchers (Darwin & Gardner, 1986; Moore, Glasberg, & Peters, 1985, 1986; Moore, Peters, & Glasberg, 1985), which show that a partial that is part of a harmonic series but is gradually being mistuned starts to no longer fuse with the other harmonics. Apparently, this happens because a mistuned partial or group of partials that surpasses about 3% mistuning no longer falls within the tolerance of the "harmonic sieve" (Duifhuis, Willems, & Sluyter, 1982; Scheffers, 1983a, 1983b). Our experiment, although it involved a mistuned group of harmonics, rather than a single mistuned harmonic, shows similar effects. It is also possible that our results and those of McAdams (1984b) indicate that the sieve can track moving harmonics.

At this point, it should be clear that the use of CGs has an effect that is additional to the effect of F0 separation. However, it should be noted that even with SSs, performance is still high once a difference in F0 is introduced.

It is not clear whether the direction of the glides is important. The glides can be designed so as to cross one another, moving in opposite directions, or to be parallel, moving in the same direction. Would the kind of glides used influence the segregation of the vowels? We have described how the use of glides may improve the information about the positions of the resonances in the sound source. This factor may also facilitate the identification of two superimposed vowels. However, we also assume that there is a second cue available in gliding sets of harmonics. In parallel glides, when the F0s of two vowels, and therefore all their harmonics, glide coher-

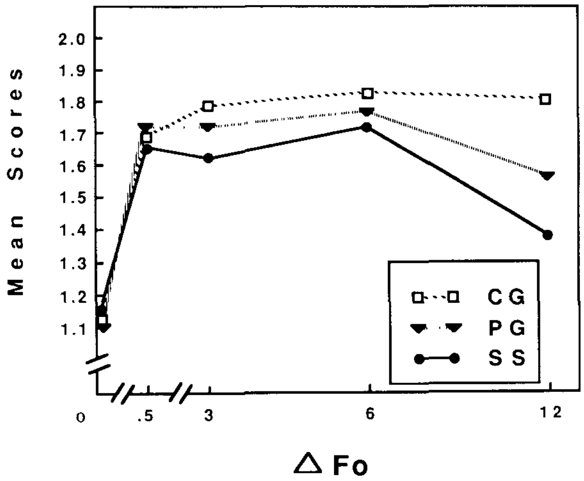

Figure 3. Mean identification scores for the three types of pitch contour for each of the F0 separations (log scale) used in Experiment 3. PG refers to parallel glides, CG to crossing glides, SS to steady-states.

ently, the harmonics of the two vowels retain the same simple frequency-ratio relationships to one another across vowels at every instant in time. However, when the F0s move in opposite directions, as in CGs, carrying their harmonics with them, only the members of the subset of harmonics that are related to the same F0 retain their frequency relationships to one another over time; this incoherence across subsets should assist in the grouping of the members of each subset and their segregation from the other set.

Two hypotheses, then, could explain any improvement that may result from F0s that change as a function of time. First, the changing F0 might give clearer information about the position of a formant peak. McAdams (1984b) has indeed shown that modulating harmonics make a vowel more prominent than nonmodulating ones, possibly due to formant tracing. Coherence across separate harmonic series (analogous to our parallel glides) did not affect his results. If formant tracing is the primary factor, then change itself should be important, but not the relative direction of change of the two F0s. Consequently, it would make no difference whether parallel or nonparallel glides were used. Second, if formant tracing is not the only factor, the direction of change might help different groups of harmonics to be grouped selectively. In such a case, vowels on F0s that moved in different directions should be easier to identify than those on F0s that moved in the same direction. Experiment 3, in addition to SSs and CGs, also contained parallel glides (PGs).

EXPERIMENT 3

In a pilot study for Experiment 3, we noticed that the effects of an increasing F0 difference (combined with, say, CGs) on identification performance were smaller than one would expect from our perceived (subjective) ability to identify the sounds. Scheffers (personal communication, September 1983), and later McAdams (personal communication, December 1988), had also noticed this difference between the measured effect and the phenomenological impression of the sounds in their experiments. It was decided to run the present study including pairs of identical vowels for all the F0s' separations and the different inflections used. We thought it would be interesting to see if two identical vowels could be heard separately on the basis of pitch difference.

Method

Stimuli. For the SSs and the CGs, the stimuli were similar to the ones that were used in Experiment 2, but they were resynthesized at an increased sampling rate of 22 kHz. In addition, more pairs were added by combining each vowel with itself for all the F0 separations and the different pitch contours used. In the parallel glide (PG) pairs, the F0s of the two glides maintained a constant semitone separation as they glided upward or downward. In all cases, the ratios of the F0s represented 0-, ½-, 3-, 6-, and 12-semitone differences, maintaining 140 Hz as the value around which the glides were symmetrically placed (on log frequency coordinates). The 9-semitone difference was omitted, because it had previously produced results similar to those produced by the 6-semitone differences. The frequency values of the F0s for the various ratios are shown in Table 1.

All the vowel pairs were synthesized using the three types of pitch contours for all the F0 separations The total number of pairs was 150. All the vowels were equated subjectively for equal loudness, as in Experiment 2. Two test tapes were prepared, each with the 150 pairs in a random order.

Procedure. The procedure was identical to the one used in Experiment 2. The only differences consisted of the number of vowel pairs, which were 10 in this case, and the number of pitch contours, which were three (SS, PGs, and CGs). The dependent variable was the identification score.

The experiment started with a pretest session. All stimuli were repeated 10 times in both the training and the testing sessions. There was no feedback to the subjects.

Subjects. Twenty male and female McGill undergraduate students, who were paid, participated in the study.

Apparatus. The synthesis and playback conditions were similar to those described in Experiment 2. The signals were output via a 12-bit digital-to-analog converter at a sampling rate of 22 kHz, low-pass filtered at 10 kHz by a Rockland 851 filter, and recorded on tape. At the time of playback, the signal was presented binaurally over TDH-49P headphones at 70 dBA.

Results

Scoring. The scoring was similar to that in Experiment 2.

Analysis. A three-way analysis of variance for repeated measures gave a significant effect of pitch contour [$F(2,38) = 17.16, p < .001$] and F0 separation [$F(4,76) = 61.95, p < .001$], of their interaction [$F(8,152) = 8.74, p < .001$], and of vowel pairs [$F(9,171) = 3.56, p < .004$]. A closer examination of the latter main effect revealed that the mixture /ʊ/ + /ɛ/ produced overall poorer results than did any of the other mixtures, over all conditions (Newman-Keuls tests). This was probably because the subjects often confused /ʊ/ + /ɛ/ when these vowels were combined. We have no explanation for this confusion, because /ʊ/ and /ɛ/ are spectrally quite different; one is a back and the other a front vowel. Theoretically, there should have been no confusion between them. However, it is possible that our computer synthesis resulted in some vowels that were perceived better than others. We also have no explanation for why we found no vowel differences in Experiment 2.

Figure 3 shows the mean identification scores for the three types of pitch contour for each of the F0 separations. Tests of simple effects showed significant effects of the frequency separation of the F0s in the SS [$F(4,158) = 132.58, p < .001$], the PG [$F(4,158) = 172.63, p < .001$], and the CG [$F(4,158) = 216.87, p < .001$] conditions, but inspection shows that the patterns of results are different. The mean values are shown under Experiment 3 in Table 4.

In all three conditions, there is an improvement in performance from the 0-semitone separation to the ½-semitone separation ($p < .01$, Newman-Keuls tests). The scores remain high for the other frequency separations ($p < .01$ for all of them compared to the 0-semitone one), and then they drop at the octave separation for the SS and

PG conditions. In the case of the SSs, the octave-separation results are significantly lower than in the ½-, 3-, and 6-semitone conditions ($p < .01$). In the case of the PGs, the differences are significant at $p < .05$ for the ½- and 3-semitone separations, and at $p < .01$ for the 6-semitone separation (all compared to the 12-semitone separation).

Even though the scores for the PGs appeared to fall in between the scores for the CGs and the SSs (for the 3-, 6-, and 12-semitone separations; see Figure 3), the only case where these differences are significant is at the octave separation ($p < .01$ for all three comparisons). At the 3- and 6-semitone separation, only the scores for the CGs are better than the scores for the SSs ($p < .01$ and $p < .05$, respectively). Planned comparisons indicated that the average score for the CGs is higher than that for the SSs [$F(1,152) = 7.35, p < .01$]. No other differences were significant.

Discussion

Experiment 3 was expected to provide results that clearly showed the differences among the three types of pitch contour. However, the results seem to support the notion that the only important difference is the one between CGs and SSs. The contribution of the PGs is clearly evident only at the octave separation, where performance is better than that with the SSs. This is a meaningful difference, since both PGs and SSs suffer equally from the octave effect due to the harmonic coincidence that occurs there.

The finding that there was no signrificant difference in the identification of vowels between the CGs and the PGs came as a surprise, since we expected some difference (for the reasons stated earlier). As we mentioned in the Method section of Experiment 1 (see under Stimuli), the F0 separations are defined differently for the CGs and the PGs (or the SSs). To see this, an inspection of Figure 1 may prove helpful. The average log-frequency separation of the CGs is smaller than that of the PGs (and the SSs). Yet the CGs always showed themselves superior to the PGs, even though these differences failed to reach statistical significance. At this point, we may need to find a more sensitive measure.

GENERAL DISCUSSION

The results from Experiment 1, in which pulse trains were used, showed that F0 differences and the existence of CGs can facilitate the perceptual separation of simultaneous sounds. These results were extended in Experiments 2 and 3, in which vowel mixtures were used.

Generally, a pitch difference of a half-semitone was sufficient to produce a dramatic improvement in the subject's ability to hear two sounds in a mixture or to identify two vowels, in agreement with Scheffers's findings (1983a, 1983b). To explain how a listener recognizes a vowel mixed with another one, Scheffers has suggested two similar mechanisms: template matching (see Klatt, 1980), and a profile analysis (Green, Kidd, & Picardi, 1983) operating on the spectral envelope of the stimulus. A vowel is identified on the basis of its spectral envelope (profile or template). According to Scheffers, a listener can recognize two vowels in a mixture easily if their respective spectral envelopes are dissimilar. If the spectral envelopes are similar, separation depends on pitch differences, and it is carried out in conformity with the "harmonic sieve" model of pitch perception (Duifhuis et al., 1982). This model suggests that the process of finding the F0 of a set of harmonics is analogous to the use of a sieve that has holes at the harmonic frequencies of a particular F0. When a periodic sound such as a vowel is mixed with another one that has a different F0, the sieve will allow the harmonics of the first sound to fall through (and thus be grouped together) and will block the harmonics of the second sound. The latter set of harmonics may fall through another sieve and constitute a different group. For Scheffers, then, spectral dissimilarity is a very important factor in the separation of two vowels and need not depend on F0 differences. Pitch separation is a secondary factor that aids in the case of spectral similarity of the two vowels by contributing to the fusion of related harmonics.

The present studies indicate that the use of crossing glides improves the identification of the vowels in a mixture. Two possible mechanisms could account for this result:

1. The first could be a "spectral peak picker." Such a mechanism would look for spectral peaks and parse the general spectrum along the lines proposed by Scheffers (1983a). If glides were present (any glides), it would be easier to parse the spectrum, because, due to "formant tracing" the two spectra would be better defined. On the basis of our results, except in the octave case (where both glide conditions produced better results than did the SS one, and CGs were better than the PGs), we cannot conclude that the relative direction of change is important.

The work of McAdams (1984b) seems to support the idea of "formant tracing." However, recent work (Marin, 1987; Marin & McAdams, 1987) does not support the hypothesis that vowel separation may be due to spectral envelope tracing. In this work, two types of vowel synthesis were used: In the first type, harmonics whose frequencies were coherently modulated traced out the spectral envelope. In the second type, the amplitudes of the harmonics remained fixed during frequency modulation and did not trace out the spectral envelope. In this case, because the harmonics retained a constant intensity during frequency modulation, the entire spectral envelope shifted. The results indicated that spectral tracing had no effect on the rating of the prominence of vowels in a mixture (three different vowels at pitch intervals of a perfect fourth). Frequency modulation, even without spectral tracing, seemed to contribute to the perceived prominence of a vowel. The effect may be due to a "grouping" mechanism, if one assumes that there is something special about a harmonic series in coherent motion. Such a notion could also explain our results with the glides, which leads to the postulation of a mechanism different from the "spectral peak picker" one.

2. A grouping mechanism could use two possible kinds of information to parse the overall spectrum: (1) the F0 cue, membership in a particular harmonic series (this method would work with both SSs and glides); (2) the common fate cue, the gliding of a harmonic set with the same temporal contour, which would result in the reinforcement of harmonicity and contribute to the fusion of the harmonics in the set. Common fate could explain the superiority of the CGs over SSs (or PGs at the octave). In a vowel mixture, a common temporal contour for all harmonics within each subset will hold the harmonics together. The existence of two different temporal contours (incoherence of motion across subsets) would contribute to the segregation of the subsets. In the case of the vowels, once segregation into two sets has taken place, identification could be performed on the basis of formant peak information.

Our results are consistent with a grouping mechanism. One way to explain the significant effect of pitch contour in Experiment 3 is to hypothesize that performance with both kinds of glides was perhaps better than with steady-states. It does not follow, however, that the two types of glides produce their facilitation in the same way. As attractive as this grouping theory is, an experiment by Gardner and Darwin (1986) failed to support it. These investigators found that frequency modulating one harmonic in a vowel-like spectrum did not keep it from contributing to the estimate of the frequency of the peak of the formant in which it was located and therefore to the identity of the vowel. On the basis of the grouping theory, one would expect the modulated harmonic to be excluded from the estimation. However, the auditory system may be able to take better advantage of independent modulation patterns in different subsets of harmonics when the separate modulation patterns are defined by more than a single harmonic.

We observed a strong effect of F0. Gardner and Darwin (1986) consider this effect to be more important than coherent frequency modulation. Nonetheless, our results and those of McAdams (1984b) show that coherent motion of its harmonics improves the identifiability of a vowel.

One issue that the present experiments did not address is the question of whether it is important not only that the harmonics of one subset undergo parallel frequency changes, but that they should be harmonically related, in order for their fusion to occur. The results with the steady-state pairs seem to indicate that the existence of harmonic relations is important. However, McAdams has found that parallel frequency change can make a contribution to fusion even when no good harmonic relations are maintained over time among the partials, as in a series made up of partials "stretched" on log-frequency coordinates (McAdams, 1982, 1984b, p. 263; 1985 personal communication). In contrast, a study by Bregman and Doehring (1984) found that it is not sufficient for partials to glide in parallel in order for fusion to occur. They must also maintain simple harmonic relations. This particular issue is far from resolved and further investigation is necessary.

REFERENCES

Bregman, A. S., & Doehring, P (1984). Fusion of simultaneous tonal glides: The role of parallelness and simple frequency relations. *Perception & Psychophysics*, **36**, 251-256.

Bregman, A S., McAdams, S., & Halpern, L. (1978, November) *Auditory segregation and timbre*. Paper presented at the meeting of the Psychonomic Society, San Antonio, TX.

Broadbent, D. E. (1952). Failures of attention in selective listening. *Journal of Experimental Psychology*, **44**, 428-433.

Brokx, J P L., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, **10**, 23-26.

Brokx, J. P. L., Nooteboom, S. G., & Cohen, A. (1979). Pitch differences and the intelligibility of speech masked by speech. *IPO Annual Progress Report*, **14**, 55-60.

Cherry, E. C. (1953). Some experiments on the recognition of speech with one and two ears. *Journal of the Acoustical Society of America*, **25**, 975-979.

Chowning, J. M. (1980). Computer synthesis of the singing voice In *Sound generation in winds, strings, computers* (Publ. No. 29, pp 4-13). Stockholm: Royal Swedish Academy of Music.

Darwin, C. J. (1981). Perceptual grouping of speech components differing in fundamental frequency and onset time. *Quarterly Journal of Experimental Psychology*, **33A**, 185-207.

Darwin, C. J., & Gardner, R. B. (1986). Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality. *Journal of the Acoustical Society of America*, **79**, 838-845.

Duifhuis, H., Willems, L. F., & Sluyter, R. J. (1982) Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception. *Journal of the Acoustical Society of America*, **71**, 1568-1580.

Egan, J. P., Carterette, E. C , & Thwing, E. J. (1954). Some factors affecting multi-channel listening. *Journal of the Acoustical Society of America*, **26**, 774-782.

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

Green, D. M., Kidd, G., & Picardi, M. C. (1983). Successive vs. simultaneous comparison in auditory intensity discrimination. *Journal of the Acoustical Society of America*, **73**, 639-643.

Halikia, M. H., & Bregman, A. S. (1984a). Perceptual segregation of simultaneous vowels presented as steady states and as parallel and crossing glides. *Journal of the Acoustical Society of America*, **75**, S83. (Abstract)

Halikia, M. H., & Bregman, A. S. (1984b). Perceptual segregation of simultaneous vowels presented as steady states and as glides. *Canadian Psychology*, **25**, 210. (Abstract)

Halikia, M. H., Bregman, A. S., & Halpern, L. (1982, June). *Auditory segregation of simultaneous frequency glides*. Paper presented at the 43rd Annual Convention of the Canadian Psychological Association, Montreal.

Henke, W. L. (1980). MITSYN: *A coherent family of command level utilities for time signal processing* [Computer program]. Belmont, MA: Author.

Klatt, D. H. (1980). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 243-288). Hillsdale, NJ: Erlbaum.

Marin, C. (1987). *Role de l'enveloppe spectrale dans la perception des* *sources sonores*. Unpublished master's thesis, Sorbonne University, Parıs.

Marin, C., & McAdams, S. (1987). Role of the spectral envelope in sound source segregation. *IRCAM Annual Report 1987*, pp. 75-92.

McAdams, S. (1982). Spectral fusion and the creation of auditory images. In M. Clynes (Ed.), *Music, mind, and brain* (pp. 279-298). New York: Plenum.

McAdams, S. (1984a). The auditory image: A metaphor for musical and psychological research. In W. R. Crozier & A. J. Chapman (Eds.), *Cognitive processes in the perception of art* (pp. 289-323). Amsterdam: North-Holland.

McAdams, S. (1984b). *Spectral fusion, spectral parsing and the formation of auditory images*. Unpublished doctoral dissertation, Stanford Unıversıty, Palo Alto, CA.

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

McAdams, S., & Rodet, X. (1988). The role of FM-induced AM in dynamic spectral profile analysis. In H. Duifhuis, J. Horst, & H. Wit (Eds), *Basic issues in hearıng* (pp. 359-369). London: Academic Press.

Moore, B. C. J. (1973). Some experiments relating to the perception of complex tones. *Quarterly Journal of Experimental Psychology*, **25**, 451-475.

Moore, B. C. J. (1982). In *An introduction to the psychology of hearing* (pp. 115-149). London: Academic Press.

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *Journal of the Acoustical Society of America*, **77**, 1853-1860.

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America*, **80**, 479-483.

Moore, B. C. J., Peters, R. W., & Glasberg, B. R. (1985). Thresholds for the detection of inharmonicity in complex tones. *Journal of the Acoustical Society of America*, **77**, 1861-1867.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, **24**, 175-184.

Plomp, R. (1964). The ear as a frequency analyser. *Journal of the Acoustical Society of America*, **36**, 1628-1636.

Plomp, R. (1967). Pitch of complex tones. *Journal of the Acoustical Society of America*, **41**, 1526-1533.

Scheffers, M. T. M. (1983a). *Sifting vowels: Auditory pitch analysis and sound segregation*. Unpublished doctoral dissertation, Universıty of Groningen, Groningen, The Netherlands.

Scheffers, M. T. M. (1983b). Simulation of auditory analysis of pitch: An elaboration on the DWS meter. *Journal of the Acoustical Society of America*, **74**, 1716-1725.

Sundberg, J. (1982). Perception of singing. In D. Deutsch (Ed.), *The psychology of music* (pp. 59-98). Orlando, FL: Academic Press.

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, **12**, 242-248.

Zwicker, U. T. (1984). Auditory recognition of diotic and dichotic vowel pairs. *Speech Communication*, **3**, 265-277.

(Manuscript received December 14, 1988;
revision accepted for publication April 10, 1989.)