The perceptual integration of auditory onsets and offsets in stimulus patterns of two partly overlapping frequency glides

Remijn, Gerard Bastiaan

https://doi.org/10.15017/458553

出版情報：Kyushu Institute of Design, 2002, 博士（芸術工学），課程博士
バージョン：

CHAPTER 3: Psychophysical investigation of other peripheral processes that may be involved in the perception of the middle tone: combination tones and spectral splatter.

3.1 General purpose

The general purpose of the following two experiments is to investigate whether the middle tone, as perceived in stimulus patterns consisting of two partly overlapping glides, is the result of peripheral processes concerned with combination tones or spectral splatter. Since the overlap of two glides can be viewed as a kind of a two-tone complex, it needs to be investigated whether or not the perception of the middle tone in Experiments 1 and 2 was the result of a combination tone, or tones. Nonlinear distortion in the auditory system can render various combination tones that become audible during the presentation of a two-tone complex, consisting of a lower and higher frequency component. With f1 being the lower and f2 being the higher frequency of a two-tone complex, the most easily audible combination tones have pitches at frequencies f2-f1 (difference tone), and 2f1-f2 and 3f1-2f2 (cubic difference tones). These combination tones can become audible even when the sensation level of both f1 and f2 is 40 dB or lower (Plomp, 1965; Smoorenburg, 1972), although some authors argue that only cubic difference tones are audible at these low sensation levels (Humes, 1985), especially when both f1 and f2 are higher than 1000 Hz (Hall, 1972b).

Next, the stimulus patterns as used in Experiments 1 and 2, and as used by Nakajima et al. (2000), consisted of glides with very short rise and fall times at the overlap. Since the rise and fall times in Experiments 1 and 2 were only 4 ms, and were linearly raised, it is necessary to establish whether or not the appearance of the middle tone is related to spectral splatters. Splatters can be particularly prominent when a pure tone is turned on and off abruptly (Hartmann, 1998), for example, at the onset of the second and the offset of the first glide in the present paradigm.

The matter will be investigated with psychophysical methods. In Experiment 3, the method of adjustment will be used. In this method, the participant is asked to adjust a value of a stimulus, which can be varied continuously, and set it to apparent equality with a standard (Kaufman, 1986). In Experiment 3, participants are asked to adjust the pitch and duration of a comparison tone to those of the middle tone, if perceived. In Experiment 4, a rating scale method is used. In this method, each of the presented stimuli is given an absolute rating in terms of some attribute (Stevens, 1951). In Experiment 4, participants are asked to rate the clarity of the middle tone, if perceived.

3.2 Experiment 3: Matching the duration and the pitch of the middle tone.

3.2.1 Purpose

The purpose of this experiment was to get points of subjective equalities (PSEs) with regard to the duration and the pitch of the middle tone, if perceived, in stimulus patterns consisting of two partly overlapping glides. In the stimulus patterns, the rise and the fall time of the onset and the offset that delimit the overlap was made 20 ms, and the ramps were raised cosine. This strongly reduced the appearance of spectral splatters. If the middle tone, as perceived in Experiments 1 and 2, depended on the spectral splatter at the onset of the second glide, or at the offset of the first glide, it would be difficult for the participants to perceive a middle tone in the present condition.

Information about the pitch of the middle tone can clarify whether the middle tone is caused by a combination tone or not. If the pitch of the middle tone is matched in frequency regions such as f2-f1, 2f1-f2, and 3f1-2f2, with f1 and f2 being the lower and higher frequency of a two-tone complex, then the middle tone may actually be a combination tone. Informatiòn about the duration of the middle tone is also important for understanding the perceptual mechanism behind its appearance. If the matched duration of the middle tone differs clearly from the physical duration of the overlap, then it seems difficult to assume that the middle tone results from the perceptual connection of the onset and the offset that delimit the overlap, as suggested by Nakajima et al. (2000).

3.2.2 Method

Stimulus patterns

A total of 32 stimulus patterns was generated. All the sounds in the stimulus patterns had rise and fall times of 20 ms, with a cosine-shaped ramp, which strongly reduced spectral splatter at the onset or at the offset of the sounds. Twelve of the stimulus patterns were double-glide stimulus patterns (Figure 10A). The glides had a slope of one octave per second, and each glide had a duration of 1000 ms. The double-glide stimulus patterns had variations in overlap duration, instantaneous frequency separation, and in reference frequency (the center of the frequency region, at the temporal midpoint of the stimulus patterns).

The overlap duration in the double-glide stimulus patterns was 100 or 200 ms.

The 400 ms overlap in the stimulus patterns as used in Experiments 1 and 2 gave rise to a middle tone that was perceived as steady-state by most observers. Informal listening indicated that shorter overlaps gave rise to a middle tone that was close to a steady-state tone according to all listeners. The use of relatively short overlaps of 100 or 200 ms, therefore, may enable the participants in the present experiment to assign a global pitch to the pitch of a middle tone with little difficulty. The pitch of the middle tone can be matched with a comparison consisting of a pure steady-state tone. The six stimulus patterns with the 100 ms overlap duration had a total duration of 1900 ms. The other six stimulus patterns, with the 200 ms overlap duration, lasted 1800 ms in total.

The glides in the double-glide stimulus patterns were calculated from a reference frequency at the temporal midpoint of the stimulus patterns, at t = 900 or t = 950 ms from the onset of the first glide (t = 0 ms). The reference frequency was either 800 Hz or 1600 Hz. The instantaneous frequency separation between the overlapping glides was 1, 2, or 3 ERBs of the reference frequency. For each stimulus pattern the frequency separation value was divided by two, and this value was both added to and subtracted from the reference frequency. This resulted in two frequency points, at the temporal middle of the stimulus pattern, from which the first and the second glide were calculated linearly on a logarithmic scale. The second glide started at a frequency higher than that of the first glide at that point in time.

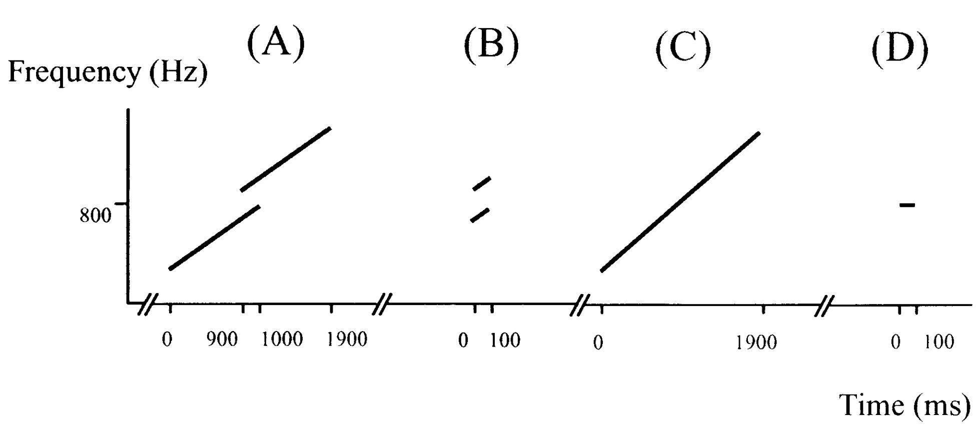


Figure 10. *Stimulus patterns used in Experiment 3 (sample). The x-axis depicts time, and the y-axis frequency on a logarithmic scale.*

Twelve stimulus patterns that consisted of just the overlaps of the double-glide stimulus patterns were generated (Figure 10B). The variations in reference frequency (800 or 1600 Hz), instantaneous frequency separation (1, 2, or 3 ERBs of the reference frequency), and overlap duration (100 or 200 ms) were maintained for these control stimulus patterns.

The next four stimulus patterns were single glides (Figure 10C). The glides moved through either 800 or 1600 Hz, and had a duration of either 1800 or 1900 ms. These values correspond to the reference frequency and the total duration of the double-glide stimulus patterns. Another four stimulus patterns consisted of a short steady-state tone (Figure 10D). These steady-state tones had a frequency of 800 or 1600 Hz and a duration of 100 or 200 ms, similar to the reference frequency and the overlap duration of the double-glide stimulus patterns. The four single-glide stimulus patterns and the four single-tone stimulus patterns were also used as control stimulus patterns. They were used to check for bias in the participant's responding with regard to whether a middle tone was perceived or not, and to check whether or not the participant had any difficulty in operating the buttons on the computer display.

The comparison tone consisted of a steady-state tone with 20 ms rise and fall times, with cosine raised ramps. Its initial duration was randomly chosen from a range of 60 through 80 ms in half of the cases (ascending series, 'A') and from a range of 400 through 600 ms in the other half (descending series, 'D'). The initial durations for the ascending series were sufficiently shorter than the shortest (100 ms) physical duration of the overlap of the double-glide stimulus patterns (and the physical duration of the single-tone stimulus patterns). The initial durations of the descending series were sufficiently longer than the longest (200 ms) physical duration of the overlap of the double-glide stimulus patterns, and the short control stimulus patterns.

The initial frequency of the comparison tone was randomly chosen from a range of 60 Hz through 100 Hz in half of the trials (ascending series) and from a 3600 Hz through 4000 Hz range in the other half (descending series). The low frequency range (60 – 100 Hz) was below the corresponding frequency of a difference tone (f2-f1) that could appear during the overlap in the double-glide stimulus patterns and in the control stimulus patterns that consisted of just the overlaps. The frequency range was also below the corresponding frequency of a cubic difference tone (3f2-2f1) that could appear in the stimulus patterns as used. The lowest combination tone that was possibly audible was a difference tone with a frequency of 103.6 Hz, which could result from the 200 ms overlap of glides or glide components separated by 1 ERB at a reference frequency of 800 Hz. The high frequency range (3600 – 4000 Hz) of the initial frequency of the comparison tone was, just in case, above the corresponding frequency of the highest possible summation tone (f1 + f2) that could appear in the stimulus patterns, although summation tones are generally less clearly audible than (cubic) difference tones (Gelfand, 1998),.

Apparatus

The stimulus patterns were generated by a computer (16 bit, sampling frequency 44100 Hz). Via a DAT recorder (Tascam DA-30 MKII), that was used as a D/A converter, a low-pass filter at 8.3 kHz (NF Electronic Instruments DV-04), a parametric equalizer (Kenwood GE 1001), an amplifier, and headphones (STAX Lambda Nova), the stimuli were monaurally presented to the participant's preferred ear. The sound level averaged 68 dBA for the stimulus patterns with a single glide or steady-state tone, whereas the level for the stimuli with overlapping glides or short, simultaneous glides, averaged 71 dBA. Levels were measured (fast-peak) with a sound level meter (Brüel & Kjær 2209) and an artificial ear (Brüel & Kjær 4153), mounted with a microphone (Aco 7013).

Participants

Four participants, two females and two males, participated in the experiment. They had normal hearing, were studying in the field of auditory perception, and had received basic training in music and training in technical listening for acoustic engineers. They were 22 through 24 years of age.

Procedure

The stimulus patterns were monaurally presented to the participant in a sound proof booth. By clicking a mouse on a "present" button on a computer screen, one of the stimulus patterns was presented, followed by a 3 s silence and a comparison tone. The initial duration and the initial frequency of the comparison tone were chosen from either an ascending (A) or descending series (D). That is, the 'initial duration - initial frequency' values originated from A-A, A-D, D-A, or D-D series, so that every stimulus pattern was presented four times with different initial comparison tones. These 32 x 4 trials were subdivided in 8 blocks, consisting of 16 trials each.

In each block of 16 trials, eight random trials were stimulus patterns of a short tone. The stimulus patterns were randomly chosen from the stimulus patterns of two simultaneous, short glide tones without the glide context, or the single steady-state tones. When a stimulus pattern with a short tone was presented, the background color of the computer screen turned blue and the following instructions appeared on the computer screen: (1) Adjust the duration of the comparison tone to the duration of the short tone; (2) Adjust the pitch of the comparison tone to the global pitch of the short tone. The participant was explained that the pitch of the comparison tone could be adjusted by clicking "higher" and "lower" buttons on the computer screen. With these buttons, small changes in the duration of the comparison tone could be made for precise approximation. By clicking the "higher roughly" and the "lower roughly" buttons, relatively large changes in the duration of the comparison tone could be made. By clicking the "present" button again, the participant could listen to the same stimulus pattern, followed by the comparison tone with its changed pitch.

Changes in the duration of the comparison tone could be made by clicking "longer" and "shorter" buttons, with which small changes in the duration of the comparison tone could be made for precise approximation. Clicking "longer roughly" and "shorter roughly" buttons resulted in relatively large changes in the duration of the comparison tone. By clicking the "present" button again, the stimulus pattern and the manipulated comparison tone could be heard. When the frequency or duration of the comparison tone was decreased, the step-size of the change was limited to be equal to or smaller than half the value of the duration or frequency of the comparison tone in the former presentation. When the frequency or duration of the comparison tone was increased, the maximal step-size was a doubling of the value of the frequency or the duration that the comparison tone had on the presentation before.

The participant was free to adjust the pitch or the duration of the comparison tone in any order and could make as many manipulations as wished. When satisfied with the matching of both global pitch and duration, the participant could click the "finish" button on the computer screen. This left a pop-up menu on the screen, which asked the participant to confirm his/her choice. After confirmation, the next stimulus pattern was presented.

The other eight random trials in one block consisted of long single-glide or double-glide stimulus patterns. When presented, the background color of the computer screen turned white, and the following instructions appeared on the screen (Figure 11): (1) If, after listening to the same stimulus pattern for five times, a short tone near the temporal middle of the stimulus pattern cannot be heard, click the "no middle tone" button. (Clicking this button resulted in the presentation of the next trial). (2) If a short tone can be heard in the temporal middle of the stimulus pattern, adjust the pitch and duration of the comparison tone to that of the tone perceived in the stimulus pattern.

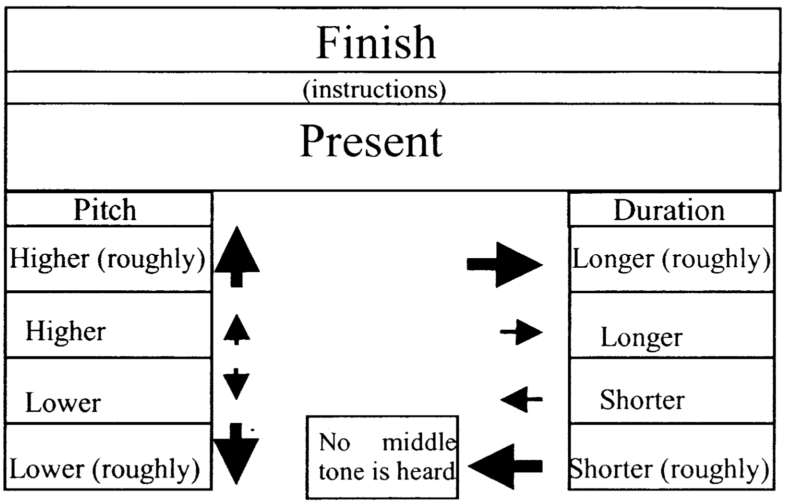


Figure 11. *The computer display used in Experiment 3 for conditions in which a double-glide stimulus pattern was presented. By clicking the buttons next to the arrows on the display, the participant could adjust the pitch and the duration of a comparison tone to those of the middle tones in the standard stimulus patterns, if perceived.*

Each participant received 8 randomized blocks, completing 4 sessions of 32 stimulus patterns. At the start of each block, two warm-up trials were given. None of the participants reported to have difficulty in understanding the task or manipulating the comparison tone via the buttons on the computer screen. The task lasted eight hours in total.

3.2.3 Results and discussion

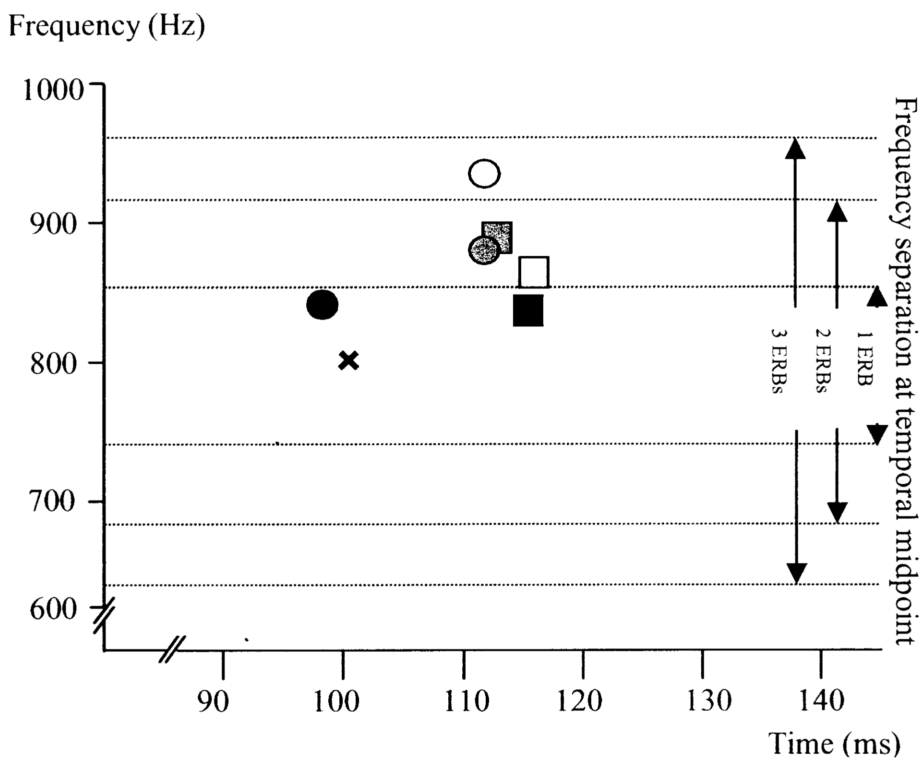
In the stimulus patterns that consisted of a single glide (cf. Figure 10C), none of the participants heard a middle tone, and the data were left out of the analysis. Since one participant did not perceive a middle tone in five trials divided over three of the four sessions, the data of just one session of the participant were analyzed. The participant reported to have difficulty in perceiving an independent middle tone, with a clear onset and a clear offset, in the double-glide stimulus pattern with an instantaneous frequency separation of 3 ERBs. An alternative would be to omit all the data with the 3 ERB conditions of this participant, but in order to perform the statistical analysis (Friedman tests with multiple comparisons), only complete sessions were used. The other three participants reported that they could hear a middle tone in all trials in all of their four sessions. The statistical analysis was therefore done over 13 sessions.

Figures 12A – 12D depict the mean PSEs of the duration and the pitch of the single-tone stimulus patterns (cf. Figure 10D), the 'only overlap' stimulus patterns (cf. Figure 10B), and the middle tone in the double-glide stimulus patterns (cf. Figure 10A). Eight separate Friedman tests were done ($p<0.05$), and multiple comparisons followed. A Friedman test regarding the PSEs of duration and a Friedman test regarding the PSEs of pitch were done for stimulus patterns with a (reference) frequency and (overlap) duration of 800 Hz and 100 ms (Figure 12A), 800 Hz and 200 ms (Figure 12B), 1600 Hz and 100 ms (Figure 12C), or 1600 Hz and 200 ms (Figure 12D). Significant cases of the multiple comparisons are indicated in the figures.

In general, a middle tone was robustly perceived by the participants, except for the participant who had difficulty in hearing a middle tone when the glides were separated by 3 ERBs in a few trials. The finding that a middle tone could still be perceived in stimulus patterns in which the amount of spectral splatter at the onset of the second or the offset of the first was strongly reduced, shows that the middle tone as perceived in Experiments 1 and 2 was not the result of spectral splatters. Moreover, the results of the present experiment confirm that a middle tone can be perceived in double-glide stimulus patterns with an instantaneous frequency separation larger than a critical bandwidth.

Next, the mean PSEs show that the pitch of the middle tone clearly differs from the pitches of possible combination tones. The perceived pitch of the middle tone generally lies in the frequency region between the reference frequency and the frequency of the second glide tone at the temporal midpoint of the double-glide stimulus patterns. The middle tone, therefore, is not likely the result of a combination tone, or tones. The finding that its pitch is close to the onset of the second glide in the double-glide stimulus patterns, however, may be related to one of the perceptual modes found in Experiments 1 and 2.

A



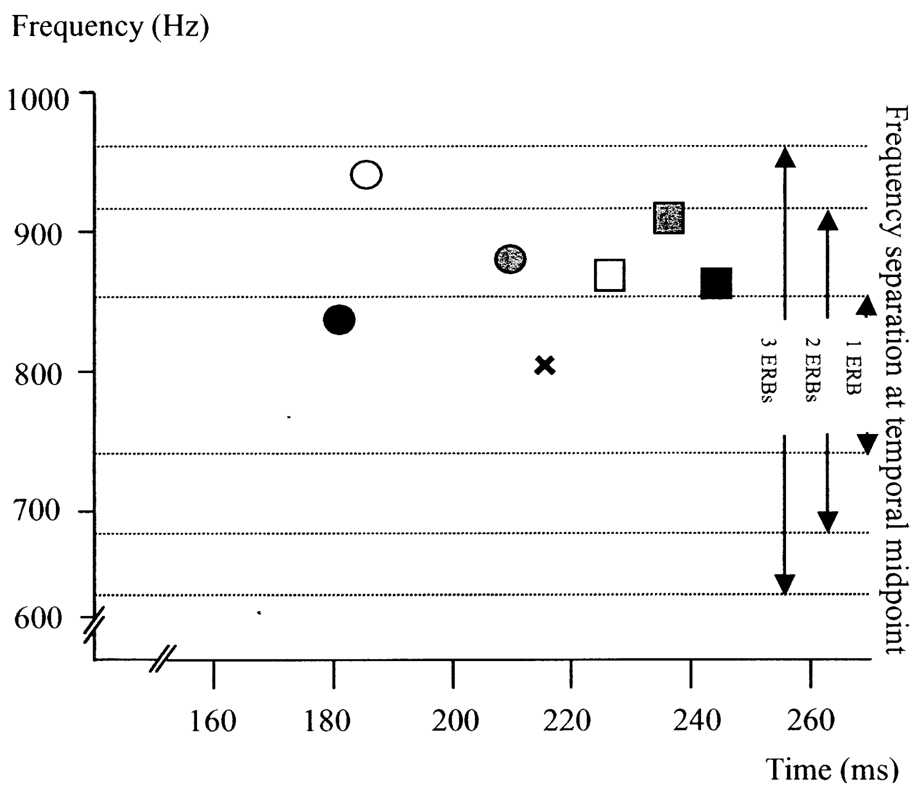
Legend:		Instantaneous frequency sep.
Double glide stimulus patterns:	1 ERB	●
	2 ERBs	◍
	3 ERBs	○
'Just overlap' stimulus patterns:	1 ERB	■
	2 ERBs	▩
	3 ERBs	□
Single tone (800 Hz, 100 ms)		×

Significant differences ($p<0.05$)	
PSE Pitch	*PSE Duration*
○ higher than × ▩ higher than ×	none

Figure 12. *Results of Experiment 3. Mean PSEs of (middle tone) pitch and duration of stimulus patterns with a (reference) frequency of 800 Hz and (overlap) duration of 100 ms.*

Continued

B



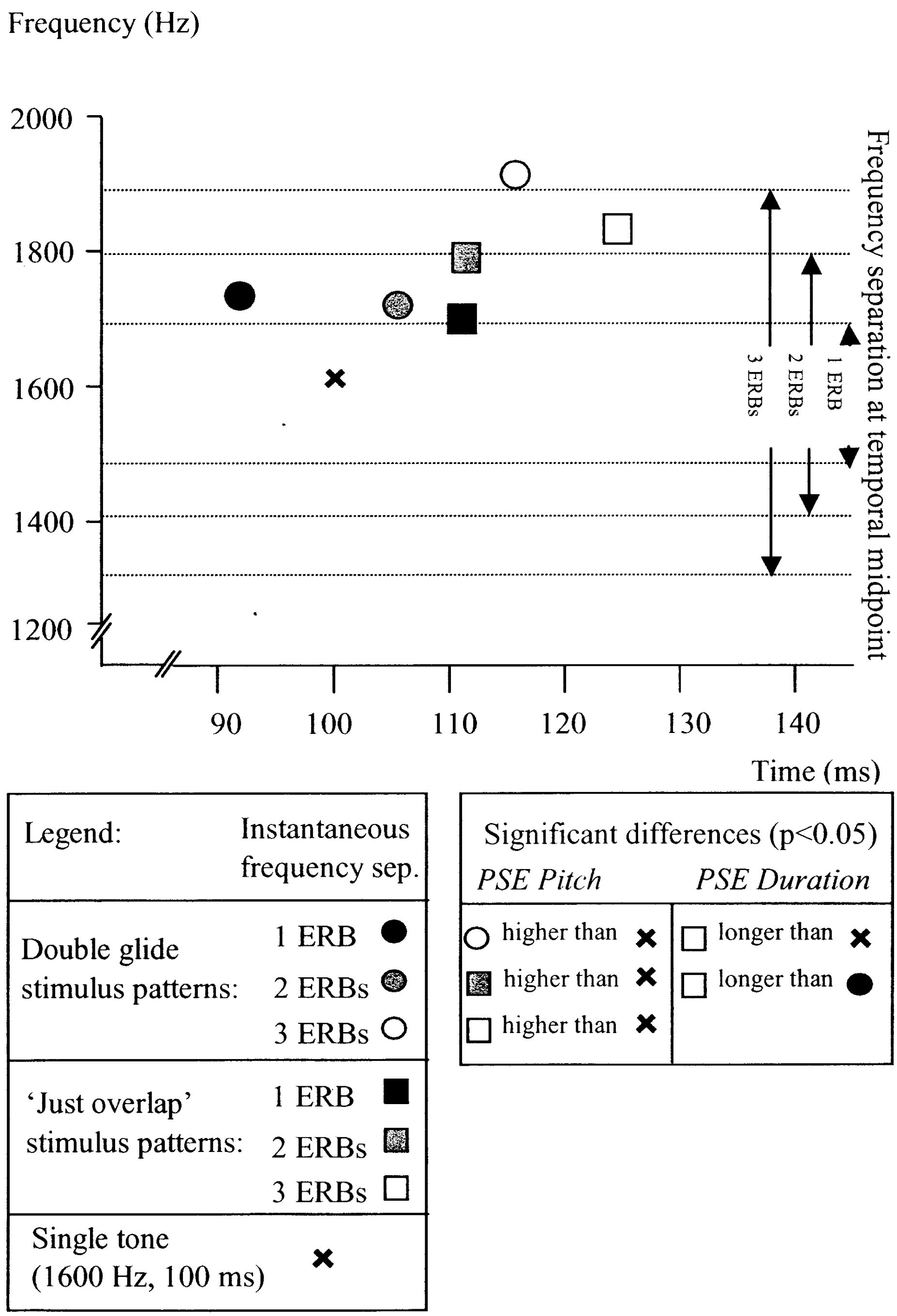
Legend:	Instantaneous frequency sep.	
Double glide stimulus patterns:	1 ERB	●
	2 ERBs	◍
	3 ERBs	○
'Just overlap' stimulus patterns:	1 ERB	■
	2 ERBs	▩
	3 ERBs	□
Single tone (800 Hz, 200 ms)		×

Significant differences (p<0.05)	
PSE Pitch	*PSE Duration*
○ higher than ×	▩ longer than ●
▩ higher than ×	▩ longer than ◍
▩ higher than ◍	▩ longer than ○
	□ longer than ●

Mean PSEs of (middle tone) pitch and duration of stimulus patterns with a (reference) frequency of 800 Hz and (overlap) duration of 200 ms.

Continued

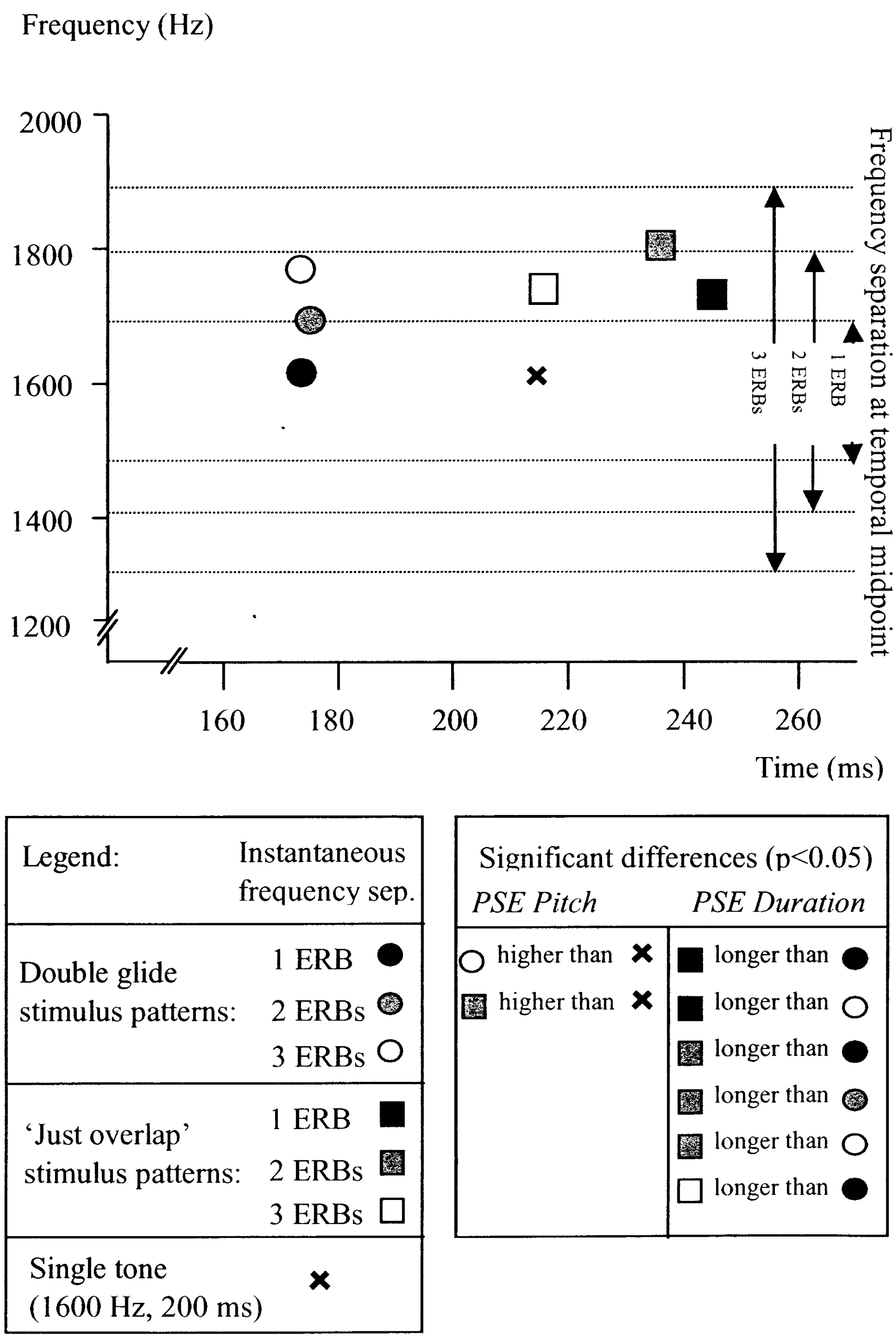
C



Mean PSEs of (middle tone) pitch and duration of stimulus patterns with a (reference) frequency of 1600 Hz and (overlap) duration of 100 ms.

Continued

D



Mean PSEs of (middle tone) pitch and duration of stimulus patterns with a (reference) frequency of 1600 Hz and (overlap) duration of 200 ms.

In Experiments 1 and 2, observers reported a perceptual mode in which the beginning of the second of two pitch trajectories was more salient than the remainder of the percept (Mode C). It was argued that, since the auditory system seems to restart a pitch-computing process when a sound joins an ongoing spectrum with a sudden onset (Bregman & Ahad, 1994), the pitch of the onset of the second glide could have contributed dominantly to the pitch of the percept at that point in time. The present data seem to suggest that the pitch of the onset of the second glide in the double-glide stimulus patterns can also contribute dominantly to the pitch of the middle tone. In view of this, is it possible that the middle tone is simply the (segregated) onset of the second glide? If only the onset of the second glide were important, then this would have implications for the subjective duration of the middle tone.

The subjective duration of the middle tone, however, shows that both the onset of the second *and* the offset of the first glide must play a role in the perception of the middle tone in the double-glide stimulus patterns. The PSE of the middle tone was close to the physical duration of the overlap of the double-glide stimulus patterns, either 100 or 200 ms (Figure 12A-12D). No significant differences were found between the PSE of the middle tone and the PSE of the single-tone stimulus patterns of 100 or 200 ms. This means that the hypothesis of Nakajima et al. (2000), which states that the middle tone results from the perceptual connection of the onset and offset that delimit the overlap, is supported on basis of the subjective duration of the middle tone. In Experiment 4, a different method and paradigm are used to further investigate the role of spectral splatters, combination tones, and the hypothesis of Nakajima et al. (2000) in the perception of the middle tone.

3.3 Experiment 4: Judging the clarity of the middle tone in double-glide stimulus patterns with masking noise bands.

3.3.1 Purpose

In Experiment 4, participants were asked to judge with a rating scale if and how clearly they could perceive a middle tone in the double-glide stimulus patterns. One purpose of the experiment was to confirm the results of Experiment 3, in which it was found that the appearance of the middle tone did not seem to be related to the appearance of a combination tone during the overlap of the glides. In the present experiment, gliding narrow-band noises were presented along with the overlapping glides. The noises were used to 'mask' possible combination tones. Masking refers to the process by which the audibility threshold of one sound is raised by the presence of another (masking) sound. A masking sound influences the audibility threshold of the masked sound best when it is similar in frequency to the masked sound (Moore, 1997). In the present experiment, three masking noise bands crossed the frequency ranges at which combination tones could occur during the overlap of the double-glide stimulus patterns. The noises were used with the intention of making combination tones inaudible. If the middle tone were the result of a combination tone, or tones, then a middle tone could not be heard in stimulus patterns in which possible combination tones are masked. The stimulus patterns in the experiment were presented at low sensation levels. When the sensation level of f1 and f2 is 40 dB or lower, only combination tones at frequencies f1-f2, 2f1-f2, and 3f1-2f2 can become audible (Plomp, 1965; Smoorenburg, 1972), with f1 being the lower and f2 being the higher frequency of a two-tone complex. Each masking noise-band was used to mask one of these three possible combination tones.

In Experiments 1, 2, and 3, it was found that the instantaneous frequency separation between the glides influenced the appearance of a middle tone and a long pitch trajectory. However, when the instantaneous frequency separation changes, the frequency separation between the onset and offset that delimit the overlap also changes. Another purpose of the experiment, therefore, was to investigate whether or not the frequency proximity between the onset and the offset that delimit the overlap in the double-glide stimulus patterns influences the appearance of the middle tone, as suggested by Nakajima et al. (2000). In the present experiment, the effect of frequency proximity between the onset and the offset is investigated under conditions in which the instantaneous frequency separation between the glides is kept constant on the logarithmic scale. Stimulus patterns were used in which the second glide started at a higher and at a lower frequency than that of the first glide at that point in time.

3.3.2 Method

Stimulus patterns

Twelve double-glide stimulus patterns were generated, consisting of two, ascending glides that had a slope of one octave per second (Figure 13). The two glides were 1000 ms each, and overlapped each other for 200 ms. The total duration of the stimulus patterns was therefore 1800 ms. The rise time and the fall time of each glide were 20 ms, with cosine raised ramps. The starting frequency of the second glide was varied. In half of the double-glide stimulus patterns, which were named 'step-up' patterns, the second glide started at a frequency higher than that of the first glide at that point in time, at t = 800 ms from the onset (t = 0) of the first glide (Figure 13, left plane). In the other half of the double-glide stimulus patterns, called 'step-down' patterns, the second glide started at a frequency lower than that of the first (Figure 13, right plane).

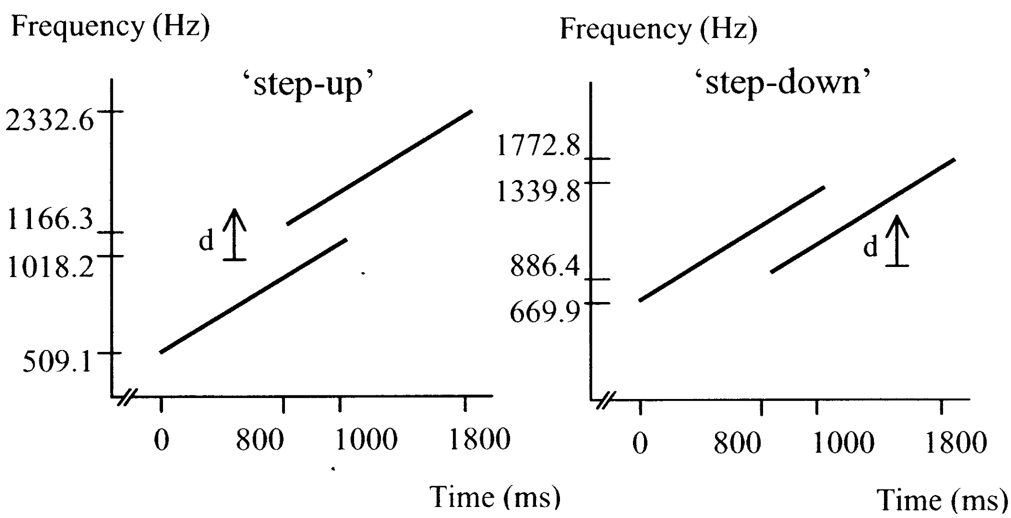


Figure 13. *Examples of a 'step-up' (left) and a 'step-down' stimulus pattern (right) employed in Experiment 4. In both panels, the horizontal axis represents time, whereas the vertical axis represents frequency on a logarithmic scale. The instantaneous frequency separation (d) in both the step-up and the step-down stimulus patterns was 100, 200, or 300 Hz (0.7, 1.4, and 2.1 ERBs, respectively, at the temporal midpoint). The 'step-up' and the 'step-down' stimulus pattern depicted in the figure have an instantaneous frequency separation (d) of 300 Hz (2.1 ERBs).*

The glides were calculated from a reference frequency of 1100 Hz at the temporal midpoint of the stimulus patterns (t = 900 ms from the onset of the first glide). At the temporal midpoint, the lower glide component (f1) and the higher glide component (f2) had frequencies of 1050 - 1150 Hz, 1000 - 1200 Hz, *and* 950 - 1250 Hz, respectively. These three variations in frequency separation of 100, 200, and 300 Hz correspond approximately to frequency separations of 0.7, 1.4, and 2.1 times the ERB of the reference frequency. The ERB at the reference frequency is 143.4 Hz. A frequency separation of 1.4 ERB corresponds roughly to one critical bandwidth of 1100

Hz as described by Zwicker and Fastl (1999). The calculation of the frequency separation between the frequency components was similar to that in the former experiments.

In half of the double-glide stimulus patterns, three noises were presented along with the overlapping glides (Figure 14). The gliding noises were meant to mask the possible combination tones appearing during the overlap of the glides, between t = 800 and t = 1000 ms from the onset of the first glide (t = 0 ms). The gliding noises ascended with a slope of one octave per second, parallel to the glides. Each gliding noise crossed one of three frequency ranges at the temporal midpoint of the stimulus pattern (t = 900 ms). In these frequency ranges, the pitches of combination tones f2-f1, 2f1-f2, and 3f1-2f2 could have become audible, with f1 and f2 being the frequencies of the higher and lower glide component at the temporal midpoint. When the frequency separation between the glides was 100 Hz (0.7 ERB), the three noises crossed f2-f1 = 100 Hz, 2f1-f2 = 950 Hz, and 3f1-2f2 = 850 Hz, respectively, at the temporal midpoint of the stimulus pattern (t = 900 ms). At a frequency separation of 200 Hz (1.4 ERB), the noises crossed f2-f1 = 200 Hz, 2f1-f2 = 800 Hz, and 3f1-2f2 = 600 Hz, whereas at the frequency separation of 300 Hz (2.1 ERBs), the noises covered f2-f1 = 300 Hz, 2f1-f2 = 650 Hz, and 3f1-2f2 = 350 Hz, respectively. The width of the noise bands was 0.1 octave. From t = 800 through t = 1000 ms, each noise covered a frequency range of 0.05 octave higher and 0.05 octave lower than the frequency at which a combination tone could occur.

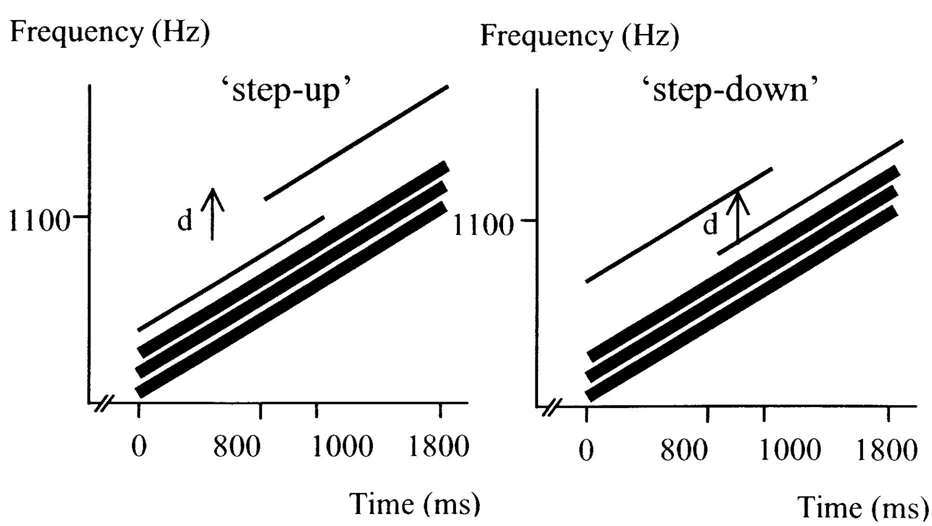


Figure 14. *Examples of a 'step-up' (left) and a 'step-down' (right) stimulus pattern employed in Experiment 4. The three narrow band noises in the stimulus patterns were used to mask possible combination tones.*

Each gliding narrow-band noise consisted of 100 glide components. The frequency range of each noise at the temporal midpoint (t = 900 ms) was divided into 100 equal units, numbered from 1 to 100, on a log scale. For each n from 1 to 99, one spectral component was randomly chosen from either unit n or unit n + 1. One component was chosen randomly from unit 100 or unit 1. All the components had equal amplitudes. They had random initial phases and frequencies. Of the 100 gliding components, 49 to 51 components had random frequencies in the lower half of the logarithmic frequency range, whereas the other 49 to 51 components fell in the upper half of the range. The rise and fall time of the noise was 20 ms, with a cosine shaped ramp.

The level of each of the masking noises was 7 dB lower than that of the glides. When both primaries f1 and f2 of a two-tone complex are presented at 40 dB SPL with frequencies of 1000 and 1100 Hz respectively, the sensation level of each of the three possible combination tones is at least 15 dB lower (Goldstein, 1967; Smoorenburg, 1972). In theory, the noise bands were therefore sufficiently intense to mask possible combination tones. This was also confirmed empirically in a control condition described below. Combining the variations in relative starting frequency of the second glide ('step-up' or 'step-down'), the frequency separation between the glides (0.7, 1.4, and 2.1 ERB), and the presence or absence of masking noise bands, a total of 2 x 3 x 2 = 12 double-glide stimulus patterns was generated.

Two control stimulus patterns were generated, in which 'dummy combination tones' were used to test whether or not the noise bands indeed could function as maskers of 'real' combination tones (Figure 15). One stimulus pattern (Figure 15, left) consisted of a single, ascending glide of 1800 ms. Three pure frequency components of 200 ms were presented simultaneously with the single glide, between t = 800 and t = 1000 ms from the onset of the single glide. The three short tones were ascending and had a slope of one octave per second. In the temporal middle of the stimulus patterns, at t = 900 ms, the three tones had frequencies of, respectively, 200, 600, and 800 Hz. The frequencies and the duration of the three short tones therefore corresponded to the pitches and the duration of the combination tones f2-f1, 2f1-f2, and 3f1-2f2, that could become audible in the double-glide stimulus pattern in which the glides were separated by 1.4 ERB. The level of each of the dummy combination tones was 15 dB lower than that of the long, single glide. At the temporal midpoint of the stimulus pattern, at t = 900 ms, the single glide moved through 1100 Hz, equal to the reference frequency of the double-glide stimulus patterns. In the second control pattern (Figure 15, right), three narrow-band noises were added to the single glide and the three dummy combination tones in such a way that each noise, respectively, covered the frequency range of one of the three dummy combination tones. The noise bands were generated as described before with regard to the double-glide patterns with masking noise bands. The level of each of the noise bands was 8 dB higher than that of the dummy combination tones, and 7 dB lower than that of the long, single glide. The rise and fall times of the long glide, the dummy combination tones, and the noise bands were 20 ms, with cosine shaped ramps. Combining the twelve double-glide patterns and the two control patterns, a total of 14 stimulus patterns was generated.

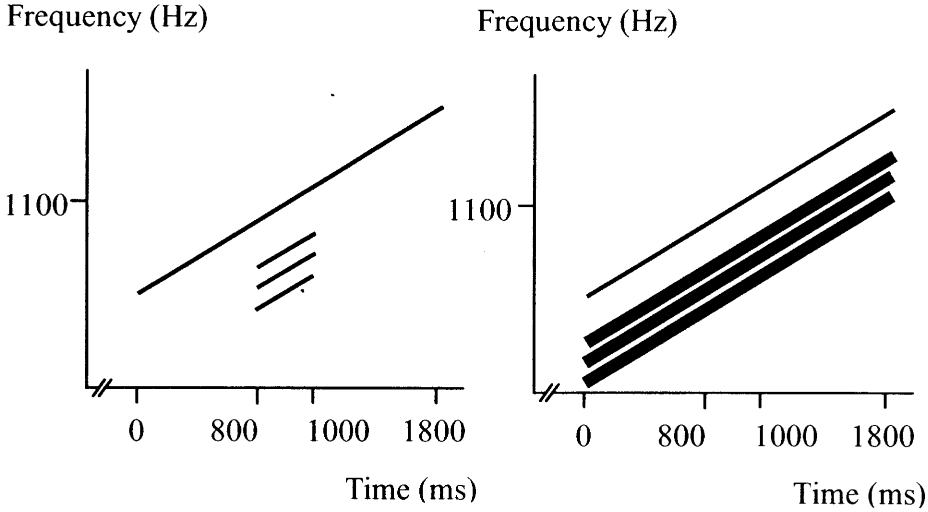


Figure 15. *Control stimulus patterns used in Experiment 4. In order to examine whether or not the noise bands could mask possible combination tones in the double-glide stimulus patterns, a stimulus pattern was used in which a single glide was presented with three dummy combination tones (left). The same stimulus pattern was made with three masking noises superimposed on the dummy tones (right).*

Apparatus

The stimulus patterns were generated by a computer (16 bit, sampling frequency 44100 Hz). Via a DAT recorder (Tascam DA-30 MKII), that was used as a D/A converter, a low-pass filter at 7 kHz (NF Electronic Instruments DV-04), an amplifier (Sansui 405α), and headphones (Sennheiser HAD 200), the stimuli were monaurally presented to the participant's preferred ear in a sound proof booth. The sound level of the stimulus patterns averaged 39 dBA. The levels were measured (fast-peak) with a sound level meter (Node 2075), and an artificial ear (Brüel & Kjær 4153), mounted with a microphone (Brüel & Kjær 4134).

Participants

Eight students of auditory perception, two females and six males, participated. They were 22 through 26 years of age. They all had normal hearing. All had received basic training in music, and training in technical listening for acoustic engineers.

Procedure

In a sound proof booth, the 14 stimulus patterns were presented to the participant in a randomized way. By clicking the mouse in a plane on a computer screen, each stimulus patterns was presented after 4 s silence. The participant could listen to the same stimulus pattern as many times as he/she wanted. The task of the participant was to decide if and how clear he/she could hear an auditory event, different from the ongoing sound, in the temporal middle of the presented stimulus pattern. The participant was required to make this judgment after listening to the same stimulus pattern for at least five times. The judgment was made on a 7-point rating scale, by clicking one of seven buttons (marked '1' through '7') on the computer screen. The participant was informed that the right extreme of the scale, indicated by the button marked with the number '7', represented the complete clarity of an auditory event deviating from the ongoing sound. The participant was instructed to press this button only if he/she could explicitly identify the auditory event as a different sound appearing along with the ongoing sound. The participant was asked to choose the left extreme of the scale, indicated by the button with the number '1', when no auditory event in the temporal middle of the ongoing sound could be heard.

The participant was instructed to click the mouse in a plane on the screen, marked 'finish', after making the judgment. This left a pop-up window on the screen, in which the participant had to confirm his/her choice. The pop-up window enabled the participant to listen to the same stimulus pattern again, if wanted, and to change his/her rating. After confirmation of the judgment in the pop-up window, the next stimulus pattern was presented. Each participant received three randomized sessions, consisting of the 14 randomized trials, with three warm-up trials before each session. The participant was required to take a break after each session. The experiment lasted about

45 minutes.

3.3.3 Results and discussion

The results of Experiment 4 are depicted in Figure 16. A short auditory event, in or near the temporal middle of the ongoing sound, was perceived in the control pattern that consisted of the long, single glide and the three dummy combination tones. The perception of this auditory event can be attributed to the temporary presence of the dummy tones, which together made up a short harmonic glide complex. When the three louder noise bands were superimposed on the dummy tones, however, the auditory event in the temporal middle could not be heard. The narrow-band noises, therefore, successfully masked the three dummy tones with frequencies and intensities that were similar to those of combination tones that could become audible during the overlapping part of the double-glide stimulus patterns. In order to mask possible combination tones in these stimulus patterns, the masking potential of the narrow band noises thus turned out to be sufficient. These two control patterns were left out of the statistical analysis.

The raw data of the twelve double-glide stimulus patterns were normalized by using a square-root transformation, and analyzed by using the multivariate analysis of variance (MANOVA) model in a three-way, within-subject design. The independent measures were the relative starting frequency of the second glide (2 levels), the instantaneous frequency separation between the tones (3 levels), and the presence or absence of the masking noises (2 levels). As can be seen from Figure 16, a middle tone could be perceived in the temporal middle of the double-glide stimulus patterns, even though the spectral splatter at the onset of the second glide tone was strongly reduced by using cosine shaped ramps of 20 ms. Spectral splatters, therefore, cannot be the vital cause of the perception of the middle tone.

The main effect of 'masking noise bands' was not significant. The middle tone could be perceived regardless of the presence or absence of noise bands with the potential to mask the possible combination tones. This result corresponds to the results of Sasaki and Nakajima (1996). They used stimulus patterns in which two partly overlapping glides held a harmonic frequency relationship during the overlap. As a consequence, the difference tones and the cubic difference tones should have had identical frequencies. A masking glide that ran from the onset of the first to the offset of the second of the overlapping glides was added with a frequency identical to the (cubic) difference tone. With the possible combination tone masked by the extra glide, listeners could still hear a middle tone. The results of the present experiment, and those of

Experiment 3, show that the middle tone is not the result of a combination tone, or tones.

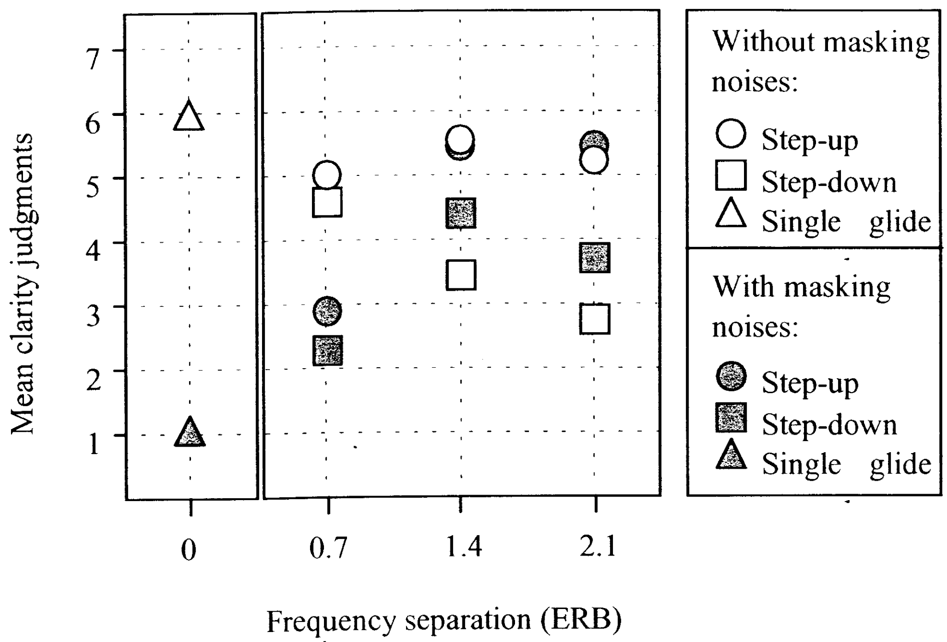


Figure 16. *Mean clarity judgments of the middle tone as heard in double-glide stimulus patterns, with masking noise bands (gray squares and circles) and without masking noise bands (white squares and circles). The triangles in the small cadre on the left show the judgments of the audibility of the dummy tones in the control patterns with masking noise bands (gray triangle) and without masking noise bands (white triangle).*

Figure 16 shows that when the glides were separated by 0.7 ERB of the reference frequency, a rather large difference appeared in the clarity of the middle tone between the stimulus patterns in which the noises were present, and those in which they were absent. This difference was far less prominent in conditions with larger values of frequency separation. Although the interaction effect between 'masking noise bands' and 'frequency separation' was not significant, two factors may have caused the faintness of the middle tone in double-glide stimulus patterns with the 0.7 ERB frequency separation. First, in these conditions, two of the three masking noises had frequency ranges close to the overlap. Two noises centered around 850 Hz and 950 Hz, and had a width of 0.1 octave. Especially the noise band centered around 950 Hz was rather close to the 1050 Hz frequency of the lower glide component at the temporal middle of the stimulus pattern. The masking potential of these noises might have been too good: since low-frequency tones mask higher ones better than vice versa (Moore, 1997), this so-called upward spread of masking from the masking noises could have heightened the audibility threshold of the lower glide component during the overlap. This could have influenced the perception of a middle tone. In the stimulus patterns with the larger frequency separations, the frequency distance between the noise bands and the lower overlapping component was larger, so that the masking influence of the noises on the lower glide must have been smaller.

Another factor that could have caused the faintness of the middle tone in the 0.7 ERB concerns the timbre of the middle tone in this condition. When the two glides were separated by less than a critical bandwidth, or one ERB, the interaction between the two overlapping components could have given a rough quality to the middle tone. Some of the phenomenological descriptions in Experiments 1 and 2 confirmed that the middle tone could have a rough quality when the instantaneous frequency separation between the glides was smaller than a critical bandwidth. The roughness of the middle tone could have facilitated its perceptual integration with the noise bands. Some of the reports of Experiment 1 also indicated that when the instantaneous frequency separation between the overlapping glides became larger than a critical bandwidth, the pitch quality of the middle tone became closer to that of a pure tone. This difference in timbre may have facilitated the perceptual segregation of the middle tone and the noises. In general, a difference in timbre facilitates the perceptual segregation of sounds (Bregman, 1990).

The main effect of 'instantaneous frequency separation' was not significant. Figure 16 shows that the double-glide patterns also yielded a middle tone at frequency separations of 1.4 and 2.1 ERB. This confirms the finding that the perception of the middle tone does not depend solely on the interaction of the overlapping components within a critical band, or one ERB. The 'step-up' patterns yielded a more prominent middle tone than the 'step-down' patterns. This tendency was significant, [$F(1,4) = 9.4142$, $p<0.0374$], and may have some important implications. First, it once more confirms that the middle tone is not the result of a combination tone. When both the 'step-up' and 'step-down' stimulus patterns had the same instantaneous frequency separation, the overlapping components in both stimulus sets were identical with respect to duration, frequency and intensity. Both the 'step-up' and 'step-down' patterns, in conditions of equal frequency separations, should have rendered similar combination tones. If the perceived middle tone were in fact a combination tone, the perceived clarity of the middle tone in both sets of stimulus patterns would have been more or less similar, and not significantly different.

Second, the difference in the clarity of the middle tone between 'step-up' and 'step-down' supports the hypothesis that proximity between the onset and the offset that delimit the overlap facilitates the appearance of the middle tone. In conditions where the instantaneous frequency separation between the glides was equal, the main difference between the 'step-up' and the 'step-down' stimulus patterns concerned the frequencies of the onset and the offset that delimit the overlapping part. In the 'step-up' patterns, the frequency difference between the onset and the offset was far smaller than in the 'step-down' patterns, under conditions of equal frequency separation. When the glides were separated by 1.4 ERB, for example, the difference between the onset and the offset that delimited the overlap in the 'step up' stimulus pattern was |1119.64 - 1071.77| = 48.87 Hz. In the 'step-down' stimulus pattern, this difference was as much as |933.03 – 1286.13| = 353.1 Hz. Proximity in frequency between the onset and the offset that delimited the overlap in the 'step-up' stimulus patterns thus may have facilitated the perception of the middle tone in the present experiment.

3.4 Summary

Experiments 3 and 4 showed that a middle tone can be heard in stimulus patterns consisting of two partly overlapping glides by different participants and with the use of psychophysical methods. The middle tone could be perceived in both experiments when the instantaneous frequency separation between the glides was larger than a critical bandwidth, or an ERB, similar to the results of the Experiments 1 and 2. In both Experiments 3 and 4, though, double-glide stimulus patterns with cosine shaped rise and fall times of 20 ms were used, in order to reduce spectral splatter at the onset or the offset that delimit the overlap. Because a middle tone could still be heard, it can be concluded that spectral splatters did not underlie the appearance of the middle tone in Experiments 1 and 2. The PSEs of the pitch of the middle tone (Experiment 3) and the appearance of the middle tone in stimulus patterns with masking noise bands (Experiment 4) showed that the middle tone is not the result of a combination tone.

The PSEs of the duration of the middle tone (Experiment 3) showed that the duration of the middle tone is comparable to the physical duration of the overlap of the double-glide stimulus patterns. This indicates that both the onset of the second glide and the offset of the first glide, delimiting the overlap, are important to the perception of the middle tone. The significant difference in the clarity of the middle tone between 'step-up' stimulus patterns and 'step-down' stimulus patterns (Experiment 4) indicates that the frequency relationship between the onset and offset that delimit the overlap is important for the perception of the middle tone. Compared with the ‘step-down’ stimulus patterns, the relatively small frequency separation between the stimulus edges that delimited the overlap in the ‘step-up’ stimulus patterns may have resulted in a significantly more salient middle tone in these stimulus patterns. In the Chapter 4, appearance of the middle tone in relation with the proximity between the onset and the offset that delimit the overlap is further investigated.