acta acustica 3 (1995) 405-416 October 1995, page 405

Perceptual Separation of Simultaneous Complex Tones: the Effect of Slightly Asynchronous Onsets

Joos Vos

TNO Human Factors Research Institute, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands

(Received 16 December 1994; revised 22 June 1995; accepted 25 August 1995)

Abstract. — Onset asynchrony has been put forward as one of the relevant factors in auditory scene analysis (for references, see A.S. Bregman, MIT Press, Cambridge, 1990). In the present study this issue was further investigated using a new paradigm: musically trained subjects were presented with two successive pairs of simultaneous complex tones and they had to decide whether the two pairs contained tones with the same pitch. If this was the case they had to indicate which tones had been equal in pitch (the lower in each pair, the higher in each pair, the higher in the first pair and the lower in the second pair, or *vice versa*). In Experiment 1, the frequency interval between the lower and the higher tones of each pair varied between 150 and 750 cents, and in the asynchronous conditions the onset of the higher tone was delayed by 20 ms. In Experiment 2 various musical intervals were used which were either pure or slightly mistuned, and in the asynchronous conditions the onset of the higher tone preceded that of the lower tone by 25 ms. The results showed that for the degree of onset asynchrony investigated, a facilitative effect of onset asynchrony on perceptual separation was only found in a small subset of conditions.

Pacs numbers: 43.66Mk.

1. Introduction

In the perception of simultaneous complex tones, many factors may affect the ability of the auditory system to parse individual tones and uncover the individual acoustic sources that gave rise to them (Bregman, 1990). If components have arisen from the same physical event, they will have relationships between them that are unlikely to have occurred by chance.

In the *frequency domain*, for example, both spectral and harmonic relations may be relevant clues for segregation. Since usually higher partials are weaker than the lower ones, the auditory system may prefer to accept partials that fall off in intensity with increasing frequency as parts of the same sound source (Bregman, 1990, Chapter 3). Likewise, a discontinuity in the spectral envelope, such as a jump in the level of a number of partials, may indicate that there are two separate tones. In addition (Bregman, 1990, Chapter 3), harmonically related partials tend to be assigned to the same source: if the calculation of the pitch of the fundamental fails to yield one single pitch, as is the case with two different simultaneous complex tones, this may be an indication that there are two unitary sounds. Our auditory system is capable of perceptually separating simultaneous sources. Using simultaneous complex tones, Plomp et al. (1973) showed that trained listeners were able to categorize the 12 musical intervals between C4 and C5 with an overall mean accuracy of about 85% at a tone duration of 120 ms. In a similar study, Killam et al. (1975) found that their subjects reached a mean score of about 65% correct. Even with missing fundamentals, the intervals between two simultaneous complex tones can be successfully recognized (i.e., the pitches correctly identified) as long as spectral interference between the components is not too strong (Houtsma & Canning, 1983; Houtsma et al. 1984; Beerends & Houtsma, 1989). The identification of the direction of mistuning of musical intervals is another example of our ability to process fundamental pitches in the sound complex. Stretched and compressed versions of harmonic fifths and major thirds were consistently identified by musically trained subjects at mistunings of only 25 cents (Vos, 1982).

In the *time domain*, asynchronous onset times may provide extra clues for segregation. For simultaneous sinusoidal or complex tones, a beneficial effect of onset asynchrony on perceptual separation was reported by Bregman & Pinker (1978), Dannenbring & Bregman (1978), and by Rasch (1978, 1981). In the experiments of Bregman & Pinker (1978) and Dannenbring & Bregman (1978), the subjects had, among other things, to judge the richness or the rate of the tones. In the experiments of Rasch, the subjects had to identify the direction of pitch jumps in a masking paradigm (Rasch, 1978) or they had to identify musical intervals consisting of tones with sound pressure levels that differed by 20 dB at the maximum (Rasch, 1981, Chapter 3). Hall & Grose (1991) showed that the interference of a modulated masker in the detection of amplitude modulation at a target frequency was reduced by a delay of the onset of the target relative to that of the masker. In a profile analysis task Green & Dai (1992) showed that onset asynchrony between target and nontarget components impaired the performance.

The effect of onset asynchrony on perceptual separation was also demonstrated in a paradigm in which the pitch of a complex tone was manipulated by the mistuning of a specific harmonic (Darwin & Ciocca, 1992; Ciocca & Darwin, 1993). The quality of a vowel can be changed in the expected direction by an additional tone if the on- and offsets are presented in synchrony. If the onset of the added tone preceded that of the vowel, the quality of the vowel changed back toward that of the original vowel (Darwin, 1984a, 1984b; Roberts & Moore, 1991).

2. Experiment 1

2.1. *Rationale of the Experiment*

The literature referred to above suggests that the perceptual separation of simultaneous tones is already enhanced by relatively small differences in onset times of the tones: conditions with onset asynchronies of about 30-60 ms (Bregman & Pinker, 1978; Dannenbring & Bregman, 1978), and even asynchronies as small as 10 ms (Rasch, 1978) yielded results which were significantly different from conditions with strictly synchronous onsets.

In the study of Rasch (1978), the strong effects of small asynchronies were obtained in a masking paradigm in which the sound pressure levels of the target tones were much lower than the sound level of the masker. However, in real-life, auditory events are usually about equal in level. Since perceptual effects obtained in threshold conditions do not necessarily represent perception in suprathreshold conditions (Scheffers, 1979; Rasch, 1981), it is worthwhile to investigate the effect of asynchronous onsets in conditions in which both tones are about equal in level.

Furthermore, the nature of the stimuli adopted by Bregman & Pinker (1978) and by Dannenbring & Bregman (1978) was rather cyclical and therefore highly predictable, as a result of which the benefit of asynchrony may have been built up during the long lasting stimulus sequence (Bregman, 1978; Bregman, 1990, Chapter 3; Dannenbring & Bregman, 1978).

The present experiment was designed to verify whether the effect of a small degree of asynchrony, as obtained in the paradigms described above, is relevant to conditions with more realistic and less predictable auditory events. If this were the case, potential utilization of differences in onset time, such as in automatic recognition of speech and non-speech by computers (e.g., see Brown & Cooke, 1994), could also be extended to the enhancement of the identification of warning signals by human beings in background noise and to limits set to the temporal presentation of concurrent auditory signals in multiple alarm systems (Doll & Folds, 1986).

Furthermore, since small differences in onset times, up to about 25 ms, are hardly noticeable by naive listeners, the perceptual separation of simultaneous tones could be enhanced by implementing slightly asynchronous onsets in computer generated polyphonic music. Knowledge about the perceptual significance of small onset asynchronies is also relevant to computer models of human auditory perception (Williams, 1989; Williams et al. 1990; Brown & Cooke, 1994).

In the present experiment, the perceptual relevance of small onset asynchronies was determined by the use of a new paradigm in which subjects were presented with two successive pairs of simultaneous complex tones and had to decide whether the pairs contained tones with the same pitch. This task can be performed only if the two complex tones in each pair are perceptually separated and properly encoded. Only the effect of onset asynchrony was investigated, i.e., the offsets were always kept in strict synchrony. The experiments of Darwin (1984a, 1984b), of Roberts & Moore (1991), and of Green & Dai (1992) have shown that onset asynchrony has a greater effect on perception than offset asynchrony has. Excluding the large asynchronies of 240 ms, the latter results were also confirmed by Darwin & Sutherland (1984). Although Dannenbring & Bregman (1978) found that for a positive effect on segregation to occur, the target tone had either to start before the other components or to end after them, they believed that the potentially smaller effect of offset asynchrony could have been built up due to the cyclical nature of their stimuli. Rasch (1978), finally, did not systematically vary offset asynchrony. From a comparison of his stimuli with complete and short overlap, however, it may be concluded that perceptual separation was not enhanced by an earlier decay of the target relative to that of the masker, which is consistent with the results of Dannenbring & Bregman (1978). The second variable of interest was the interval size between the lower and the higher tones within each pair. We supposed that especially for relatively small interval sizes, asynchrony between the onsets of the tones would have a positive effect on the perceptual separation of the tones.

2.2. *Method*

2.2.1. *Stimuli*

The stimuli consisted of pairs of two simultaneously presented complex tones. These tones consisted of 12 harmonics with amplitude a_n proportional to $1/n$. For each tone, the spectral-envelope slope was therefore -6 dB/oct. The phase of the individual harmonics was chosen at random. The overall sound pressure level of each tone was 73 dB(A). This level was measured by means of an Artificial Ear (Brüel & Kjaer, type 4152). The tone bursts had a 225 ms steady-state portion and rise/fall times of 25 ms, each shaped with the appropriate half-cycle of a raised-cosine function. The fundamental frequency of the lower and the higher tone within an interval was selected from a set of six logarithmically equally distant frequencies centered around 500 Hz. From low to high, the fundamental frequencies of the six tones were 402.6, 439.1, 478.8, 522.1, 569.4, and 620.9 Hz. The interval size between these successive tones is equal to 150 cents; the interval size of an octave equals 1200 cents.

2.2.2. *Apparatus*

The experiment was run under the control of a PDP-11/34 computer. The complex tones were generated with a digital signal processor based on the TMS-32010. One period of the complex was stored in the external memory of the signal processor. The resolution of the digital-to-analog converter was 16 bit. The sampling rate was 33.33 kHz. The tones were low-pass filtered with a cutoff frequency of 8 kHz and with a slope of −48 dB/oct. The signals were presented diotically by means of Beyer DT 48S headphones. Subjects were seated in a sound-proof room.

2.2.3. *Subjects*

Four musically trained subjects with normal hearing were tested over two half-day sessions. They were paid for their participation.

2.2.4. *Experimental Variables*

The independent variables were (1) the frequency interval between the lower and the higher tones (150, 300, 450, 600, and 750 cents), and (2) the time interval between the physical onsets of the lower and the higher tones (0 and 20 ms). The percentage of correct responses (see below) was the dependent variable.

2.2.5. *Procedure*

In the experiment we used the six (= n) complex tones mentioned above. Out of those stimuli $n(n-1)/2 = 15$ different frequency intervals of two simultaneous tones were composed. Each experimental trial consisted of the successive presentation of two of these tone intervals. An example is given in Figure 1. The detailed temporal structure of the experimental trials is shown in Figure 2. Note that the silent interval between the first and the second two-tone stimulus was 1 s.

It was the task of the subject to decide whether the first and the second two-tone stimulus in a trial contained tones with the same pitch, and if this was the case to indicate whether equal pitch applied to the lower tones

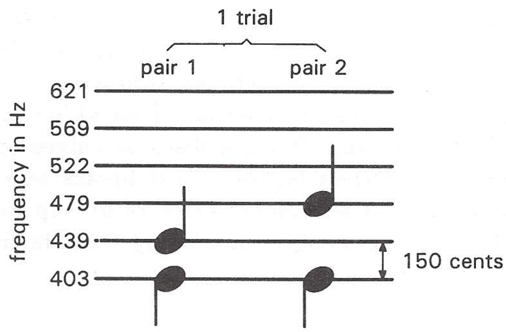

Figure 1. Example of a trial presented in experiment 1.

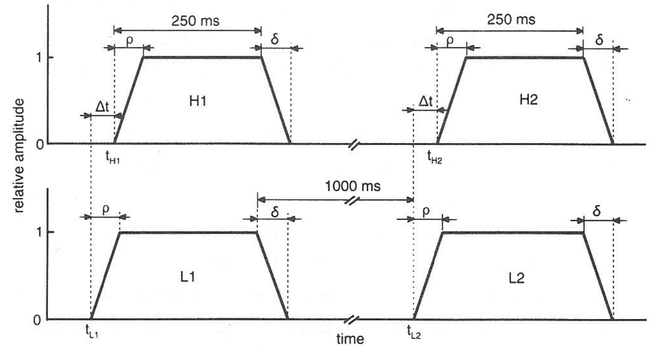

Figure 2. Temporal structure of the stimuli within one experimental trial, containing tones L1 and H1 (first pair) with physical onsets t_{L1} and t_{H1}, and tones L2 and H2 (second pair) with physical onsets t_{L2} and t_{H2}. Rise and decay times of the tones were all equal. The onsets of the higher tones were delayed by Δt ms.

of the successive stimuli (LL, as in Figure 1), the higher tones (HH), the higher tone of the first and the lower tone of the second two-tone stimulus (HL), or *vice versa* (LH). In the experiment $n(n-1)/2[n(n-1)/2-1]/2 =$ 105 pairs were compared: 20 different pairs for each of the LL, HH, and HL stimulus categories and 45 different pairs in which none of the four tones within a pair had a fundamental frequency in common (NC). A possible effect of the presentation order of the two-tone stimuli within each pair on performance was not considered. Since the presentation order was randomized, a portion of the 20 HL conditions in fact were LH conditions.

A series of trials consisted of two sets of 105 pairs, one set in which the physical onsets of the two simultaneous tones were perfectly synchronized, and one set in which the onset time of the higher tone was delayed relative to that of the lower tone. The subjects were not told that a series consisted of both synchronous and slightly asynchronous onsets of the tones. None of the subjects ever reported unequal onsets of the tones, which is consistent with observations made by Rasch (1978). (According to Zera & Green (1993) this relatively low sensitivity to onset asynchrony is typical for simultaneous intervals or chords; sensitivity is much higher for partials in a single harmonic complex tone). Presentation order of the $2 \times 105 = 210$ different pairs was randomized. The subjects were presented with 10-13 series in total.

Responses were given by means of one of four buttons (LL, HH, HL/LH, and NC). Feedback was given by a red light above the correct button. Short breaks were given after each series. A session consisted of five up to seven series. The subjects received a training series in the first session.

2.3. *Results*

In the LL and the HH conditions, there were 10 combinations that were different with respect to the interval size between the tones within each pair (see Table 1). In the HL/LH conditions, there were six different combinations, and the number of different combinations included in the NC conditions was 13. For each subject separately, we computed the percentage of correct responses for all these relevant conditions. The mean percentages are given in Table 1. In all conditions included in the LL and HH stimulus categories, and in most conditions of the HL/LH and NC categories, the smaller interval within a pair is labeled as interval A, and the large interval is labeled as interval B. In a number of HL/LH and NC conditions, however, intervals A and B are equal in size.

LL. Averaged across the four subjects, high percentage correct values between 86% and 100% were obtained. The percentages were subjected to an analysis of variance (ANOVA) (4(subjects) × 2(synchrony) ×10 (interval size combinations), all repeated measures). The analysis showed that (1) the mean values of the 10 different conditions were not significantly different ($p > 0.18$), (2) the performance in the asynchronous conditions was not significantly different from that in the synchronous conditions ($p > 0.36$), and (3) there was no significant interaction effect between interval size combination and synchrony ($p > 0.09$).

HH. Again, all HH conditions yielded high mean percentage correct values between 78% and 98%. An ANOVA performed on the data showed that neither the main effect of interval size combination ($p > 0.13$) nor that of synchrony ($p > 0.14$) was statistically significant. There was, however, a small interaction effect of interval size combination and synchrony ($F(9,27) = 2.17$; $p < 0.06$). Inspection of the data in Table 1 reveals that the interaction effect may be mainly addressed to the relatively low performance in the synchronous condition for the interval size combination of 150 and 450 cents.

HL/LH. In general, the mean percentages of correct responses were considerably lower than those in the LL and HH conditions (Table 1). The combinations in which the size of the two intervals A and B in a pair was equal (150/150 and 300/300) yielded higher percentages of correct responses than the remaining combinations ($F(5,15) = 7.65$; $p < 0.001$). There was no main effect of synchrony ($p > 0.62$) nor an interaction effect of synchrony and interval size combination ($p > 0.56$) on performance level.

NC. In general, the mean percentage of correct responses in the NC conditions (correct = no common pitch) was closer to those obtained in the LL and HH than to those obtained in the HL/LH conditions. Neither the main effect of synchrony ($p > 0.52$) nor the interaction effect of synchrony and interval combination ($p > 0.23$) was significant. The size of the intervals within a pair had an effect on the performance ($F(12,36) = 3.45$; $p < 0.005$). A Newman-Keuls paired comparison test (Winer, 1970) showed that the combinations with the interval sizes of 150/300, 300/450, and 450/600 cents were significantly different from the combinations that included intervals with a size of 750 cents ($\alpha = 0.05$). In these latter combinations, the frequency jumps of the lower and higher tones were always in the opposite direction, i.e., either converging or diverging.

2.4. *Discussion*

2.4.1. *Perceptual Separation and Synchrony*

The present results showed that for the degree of onset asynchrony investigated, no general evidence for a facilitative effect of asynchrony on perceptual separation could be found. These results may be compared to those of Rasch (1978, 1981) who used simultaneous complex tones with spectral contents that were similar to those in the present experiment.

In the pertinent experiment of Rasch (1978), the fundamental frequency of the lower tone was fixed at 250 Hz. The fundamental frequencies of the higher tones within each pair were different: they were set to 500 or 750 Hz. The subjects had to indicate whether the pitch of the higher tones jumped upward or downward. Performance level (75% correct responses) was expressed as the sound pressure level of the higher tone relative to that of the 250 Hz masking tone. The delay of the onset of the masking tone was varied from 0 to 30 ms. The results showed that this delay had a very strong effect on the threshold: each 10 ms delay resulted in about a 12 dB threshold decrement. Further explorations suggested that the effect of asynchrony on the threshold was (1) independent of tone duration (50-200 ms), (2) relatively stable for identical musical intervals in frequency ranges that are one or two octaves higher, and (3) not greatly modified by the frequency of the lower tone (100-455 Hz).

The perceptual significance of small degrees of asynchrony, as found by Rasch (1978) could not be confirmed in the present experiment. According to Rasch, the strong effect of asynchrony on the threshold of the higher tone is closely related to backward masking. In studies in which noise bands short in duration masked even shorter target clicks or pure tones, however, the backward masking curve was different from the data obtained by Rasch (see Zwislocki, 1978, for a review). For example, in the exper-

Table 1. *Mean percentage of correct responses of the four subjects for various conditions. Stimulus categories are described in the text; syn and asyn represent the synchronous and asynchronous conditions, respectively.*

Interval Size Between Simultaneous Tones in Cents		Stimulus Category							
		LL		HH		HL/LH		NC	
Interval A	Interval B	Syn	Asyn	Syn	Asyn	Syn	Asyn	Syn	Asyn
150	150	...	...	...	...	76	71	91	89
150	300	90	86	89	91	49	51	81	82
150	450	87	91	78	94	52	54	89	88
150	600	91	93	84	87	45	46	97	96
150	750	86	95	88	94	...	...	97	98
300	300	...	...	...	...	80	72	88	93
300	450	94	92	88	91	55	58	79	78
300	600	98	100	92	94	...	...	88	89
300	750	93	100	94	94	...	...	99	100
450	450	...	...	...	...	...	...	86	90
450	600	96	97	94	98	...	...	74	86
450	750	98	100	96	96	...	...	100	100
600	600	...	...	...	...	...	...	94	91
600	750	100	98	98	98	...	...	...	...

iments from Elliott (1962a, 1962b), the major portion of the masking effect disappeared in about 10 ms. Wright (1964a, 1964b) used a masker which consisted of 600 ms narrow-band noise and sinusoidal test tones which for the greater part overlapped the masker in time. Temporally, his stimuli were therefore closer to those of Rasch. In contrast to the experimental results obtained by Rasch, Wright found that in about the first 25 ms immediately preceding the masker, the threshold of the test tone was nearly constant and approximately equal to simultaneous masking. This plateau was observed for various masker levels and for various frequencies of the sinusoidal test tone.

Although the literature just described suggests that the effect of asynchrony, as found by Rasch, cannot be completely explained by backward masking, it is precluded that masking has played a significant role in the conditions of the present experiment: the overall sound pressure levels of the tones were all equal.

The discrepancy between the present results and those of Rasch might be explained by differences in the dissonance of the musical intervals. In the present experiment the size of the intervals between the tones was equal to multiples of 150 cents. Since the tones used consisted of many harmonics, and these tones were all presented at the same relatively high sound pressure level, the degree of interference between the harmonics of the simultaneous tones was very high. As a result, these intervals were perceived to be very rough and dissonant.

It was hypothesized that in this dissonant context, the tendency of the tones to fuse was already low and that the attempt to lower this even further by the introduction of onset asynchrony was predestined to fail. In consonant intervals, such as the octave and the twelfth used by Rasch (1978), the perceptual separation of the tones might be more difficult and onset asynchrony might be more effective to enhance this separation. For supra-threshold conditions, this will be tested in Experiment 2 by presenting both pure and mistuned musical intervals with synchronous and asynchronous onsets of the constituent tones.

In the recognition experiments reported by Rasch (1981, Chapter 3), both perfectly consonant intervals (with fundamental frequency ratios, $p : q$, of 2:3 and 3:4), imperfectly consonant intervals (with $p : q$ ranging between 3:5 and 5:8), and dissonant intervals ($p : q$ equal to 8:9 or 9:16) were used. Since Rasch collapsed the identification scores for all these intervals, it is not possible to estimate from his study to which degree the effect of onset asynchrony depends on the consonance of the intervals.

2.4.2. *Perceptual Separation and Interval Size*

In the LL, HH, and NC conditions perceptual separation, expressed as the percentage of correct responses, did not systematically increase with interval size. Even for interval sizes as small as 150 and 300 cents, with the distance between the fundamental frequencies equal to about 50% and 100% of the critical bandwidth, respectively, performance level was already very high.

The failure to find an effect of interval size is consistent with results from experiments in which the effect of fundamental frequency differences on the identification of two simultaneous vowels was investigated: the performance was higher when the fundamentals of the vowels were different, but almost all the improvement was already obtained by changing the separation of the fundamentals from 0 to 50 cents (Chalikia & Bregman, 1989, 1993) or from 0 to 100 cents (Scheffers, 1983, Chapter 4). However, since the vowels in each pair were spectrally different, whereas the spectral content of the tones in our experiment was equal, it may be questioned to which extent this comparison is significant.

One could object that the small effect of onset asynchrony, as found in the interval recognition experiments reported by Rasch (1981), had not been obtained because of a ceiling effect. Since the performance level in the LL, HH, and NC conditions was relatively high (in the conditions specified in Table 1, the mean percentages of correct responses ranged between 74% and 100%; the overall percentages in the LL, HH, and NC conditions ranged between 90% and 94%), it is not excluded that those conditions were less appropriate to test the hypothesis of a facilitative effect of onset asynchrony on perceptual separation. In the HL/LH conditions, however, performance was much lower. At least in the latter conditions, the failure to find a positive effect of asynchrony cannot be attributed to the ceiling effect described above.

3. Experiment 2

Discussing the results of Experiment 1, it was hypothesized that 1) the perceptual separation of tones in dissonant intervals might be easier than that of tones in consonant intervals, and 2) onset asynchrony might therefore be especially effective in the latter intervals. Experiment 2 provides a test of these hypotheses by including both pure and mistuned intervals with synchronous and asynchronous onsets of the constituent tones.

3.1. Method

3.1.1. *Stimuli and Apparatus*

Again, the stimuli were pairs of two simultaneously presented complex tones. These tones consisted of 12 harmonics with amplitude a_n proportional to $1/n^{1/2}$. For each tone, the spectral-envelope slope was therefore −3 dB/oct. By using a slope of −3 dB/oct instead of the slope of −6 dB/oct applied in Experiment 1, we intended to increase the roughness of the mistuned intervals and consequently enhance the difference between the pure and the mistuned intervals. The overall sound pressure level of each tone was about 74 dB(A). Tone duration and the rise/decay functions were identical to those in Experiment 1. The rise and decay times, however, were 15 ms and therefore 10 ms smaller than those used in the previous experiment. The fundamental frequencies of the tones ranged between 390 and 650 Hz. The apparatus was identical to that in Experiment 1.

3.1.2. *Subjects*

Five musically trained subjects with normal hearing were tested over three or four half-day sessions. They were paid for their participation.

3.1.3. *Experimental Variables*

The independent variables were (1) musical interval ($p:q$ equal to 11:12, 5:6, 3:4, 2:3, and 3:5), (2) degree of mistuning (0 and ±15 cents), and (3) time interval between the physical onsets of the higher and the lower tones (0 and 25 ms).

3.1.4. *Procedure*

In general, the procedure was similar to that in Experiment 1. The computation of the fundamental frequencies f_1 and f_2 of the lower and the higher tones was more complex than that in Experiment 1. As a first step, f_1 or f_2 was again selected from a set of six logarithmically equally distant frequencies centered around 500 Hz, but now with interval sizes between these tones equal to 176.8 cents, yielding a maximum interval size between the lower and higher tones of 884 cents (size of a major sixth with $p : q = 3 : 5$). Next, dependent upon which tones in the successive pairs of two-tone intervals had to be equal in pitch (LL, HH, or HL/LH), f_1 or f_2 was determined.

With pure intervals in the LL and NC conditions, for example, f_2 equaled $(q/p)f_1$; with intervals that were mistuned by c cents, f_2 equaled $(q/p)f_1(2^{c/1200})$. With pure intervals in the HH conditions, f_1 equaled $(p/q)f_2$; with the mistuned intervals f_1 equaled $(p/q)f_2(2^{-c/1200})$. To maximize the strength of the beats and roughness in the mistuned intervals, the higher tones were attenuated by 1 dB for intervals with ratios of 5:6 and 3:4, and by 2 dB for the intervals with ratios of 2:3 and 3:5 (Vos & van Vianen, 1985). A series of trials consisted of four sets of 105 pairs, two sets with *pure* intervals in which the onsets of the two simultaneous tones were either perfectly synchronized or in which the onset of the lower tone was delayed relative to that of the higher tone, and two sets with *mistuned* intervals, again one with synchronous and one with asynchronous onsets of the tones. Presentation order of the 4 × 105 = 420 different pairs, as well as the presentation order of the two-tone stimuli within each pair, was randomized. The direction of mistuning (stretched or compressed) of the intervals was also randomized. A session consisted of three or four series. Again, the subjects received a training series in the first session.

3.2. *Results*

LL. For the LL conditions, high mean percentage correct values between 74% and 98% were obtained (see Table 2). The percentages were subjected to an ANOVA (5(subjects) × 2 (synchrony) × 2 (purity) × 10 (interval size combinations), all repeated measures). The analysis showed that interval size combination was the sole stimulus variable that had a significant effect on performance ($F(9,36) = 2.65$; $p < 0.02$). The combinations which consisted of intervals with sizes greater than or equal to 316 cents yielded higher percentages than the combinations in which one of the intervals had a size of 151 cents. Neither the main effects of the remaining stimulus variables, nor the interaction effects with the stimulus variables, were significant.

HH. For the HH conditions, mean percentage correct values between 44% and 96% were obtained. An ANOVA performed on the data showed that the performance in the asynchronous conditions was significantly higher than that in the synchronous conditions ($F(1,4) = 38.1$; $p < 0.005$). Figure 3, in which the results are plotted as a function of the size of the larger interval (B) in a pair and with the size of the smaller interval (A) and the kind of temporal presentation as parameters, shows that the synchrony effect was independent of interval size combination.

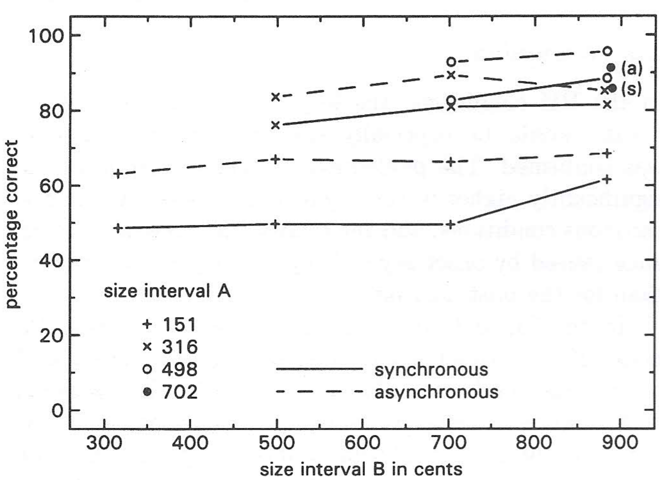

Figure 3. Mean percentage of correct responses in the HH conditions as a function of the size of the larger interval (B) in a pair, and with the size of the smaller interval (A) and the kind of temporal presentation as parameters.

Figure 4 shows that the difference between the synchronous and asynchronous conditions was almost twice as high for the pure intervals than it was for the mistuned intervals ($F(1,4) = 7.31$; $p = 0.05$). This interaction effect was independent of the interval combinations ($p > 0.80$). In addition, the ANOVA demonstrated a highly significant effect of interval size combination ($F(9,36) = 14.0$; $p < 0.000001$). The data in Figure 3 show that, in general, higher percentage correct values

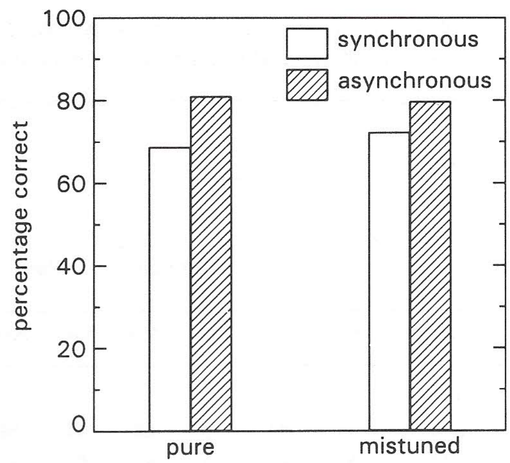

Figure 4. Mean percentage of correct responses in the HH conditions with pure and mistuned intervals, separately shown for the synchronous and asynchronous onsets.

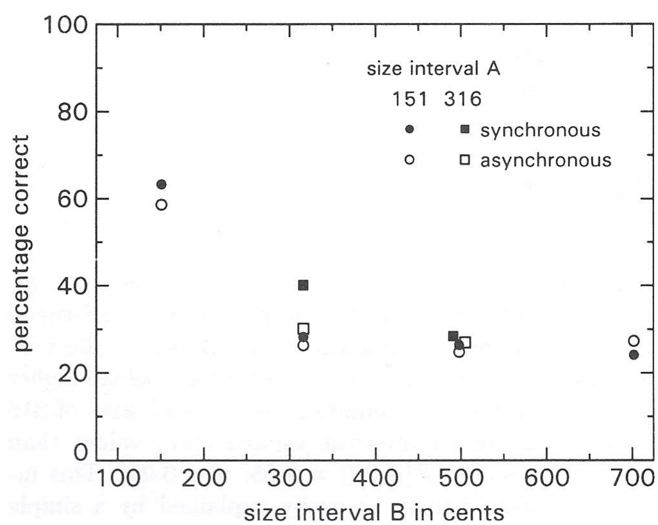

Figure 5. Mean percentage of correct responses in the HL/LH conditions as a function of the size of interval B, and with the size of interval A and kind of temporal presentation as parameters.

were obtained for pairs in which the size of each interval was about 300 cents at the minimum.

HL/LH. Averaged across subjects and the pure and mistuned intervals, the mean percentages are given in Figure 5. In general, these values were considerably lower than those in the LL and HH conditions. The combinations in which both intervals had a size of 151 cents only, yielded significantly higher percentages than the combinations in which one of the intervals was greater in size ($F(5,20) = 5.47$; $p < 0.005$). In five of the six different interval combinations, the performance was (slightly) lower in the asynchronous conditions than in the synchronous conditions ($F(1,4) = 11.5$; $p < 0.05$).

Table 2. *Mean percentage of correct responses for various conditions, including both pure and mistuned intervals. Stimulus categories described in text; syn and asyn represent the synchronous and asynchronous conditions, respectively.*

Pure Interval Size Between Simultaneous Tones in Cents		Stimulus Category															
		LL				HH				HL/LH				NC			
		Pure		Mistuned		Pure		Mistuned		Pure		Mistuned		Pure		Mistuned	
Int. A	Int. B	Syn	Asyn	Syn	Asyn	Syn	Asyn	Syn	Asyn	Syn	Asyn	Syn	Asyn	Syn	Asyn	Syn	Asyn
151	151	...	...	...	...	...	...	...	...	59	55	67	62	90	86	90	86
151	316	77	84	75	80	44	64	53	62	28	28	28	25	83	83	83	84
151	498	82	83	81	83	51	64	49	70	29	25	24	24	75	78	82	81
151	702	84	85	82	86	51	70	48	63	24	26	24	29	78	85	80	86
151	884	80	74	80	83	57	68	66	69	...	...	...	...	89	89	87	89
316	316	...	...	...	...	...	...	...	...	44	34	36	26	87	91	92	88
316	498	91	95	89	94	74	88	78	79	27	32	30	22	81	83	81	87
316	702	96	94	91	91	78	89	84	89	...	...	...	...	83	90	89	91
316	884	91	97	87	89	82	86	81	84	...	...	...	...	93	96	98	97
498	498	...	...	...	...	...	...	...	...	...	...	...	...	90	92	86	86
498	702	96	96	97	98	84	94	81	92	...	...	...	...	84	85	89	90
498	884	97	93	93	98	84	95	93	96	...	...	...	...	96	97	90	95
702	702	...	...	...	...	...	...	...	...	...	...	...	...	87	88	90	89
702	884	98	96	91	95	82	91	89	91	...	...	...	...	...	...	...	...

In four of the six interval combinations, the performance was about the same for the pure and the mistuned intervals. For the combinations with the two smallest intervals of 151 cents, the mistuned intervals yielded higher values, and for the combinations with equal sizes of 316 cents, the mistuned intervals yielded lower values than the pure intervals ($F(5,20) = 2.95$; $p < 0.05$). This interaction effect cannot be easily explained by a simple underlying mechanism.

NC. Averaged across the five subjects, high percentage correct values between 75% and 98% were obtained (see Table 2). An ANOVA performed on the data showed that the main effect of synchrony was not significant ($p > 0.13$). Overall, the percentages for the mistuned intervals were slightly higher than those for the pure intervals ($F(1,4) = 7.55$; $p = 0.05$). The size of the intervals within a pair, however, had a highly significant effect on the performance ($F(12,48) = 3.57$; $p < 0.001$). A Newman-Keuls paired comparison test showed that the combinations with the interval sizes of 151/498 and 151/702 cents yielded significantly lower percentages than the combinations with interval sizes of 316/884 and 498/884 cents ($\alpha \leq 0.05$). The percentages for the combinations 151/316 and 316/498 cents were significantly lower than the combination with 316/884 cents only ($\alpha = 0.05$). The overall effect of interval size is comparable to that obtained in Experiment 1.

3.3. *Discussion*

In the HH conditions, the hypothesis that onset asynchrony would be especially effective for pure intervals was confirmed. The percentage of correct responses was significantly higher in the asynchronous than in the synchronous conditions, and the relative increase of performance caused by onset asynchrony was higher for the pure than for the mistuned intervals.

In the LL and NC conditions, however, neither the main effect of onset asynchrony nor its interaction with purity was significant, whereas in the HL/LH conditions performance was slightly lower in the asynchronous than in the synchronous conditions and this opposite effect was independent of the purity of the intervals. As a result, our hypothesis about the interaction effect of onset asynchrony and purity could not be confirmed in general.

In the present experiment, performance level was lower than that in Experiment 1. Especially for the LL and NC conditions, it is therefore less plausible to attribute the lack of finding a facilitative effect of onset asynchrony to the ceiling effect discussed in Experiment 1. In a different experimental paradigm and with sinusoidal instead of complex tones, Bregman & Pinker (1978, Experiment 3) had also varied synchrony and purity in an orthogonal way. Neither purity nor the interaction of purity and synchrony had any significant effect on the perceptual separation of the two more or less simultaneous tones. In spite of considerable spectral differences that are relevant to the discriminability and subjective purity of musical intervals (Vos, 1986; Vos & van Vianen, 1985), the present overall result with respect to the effect of purity on perceptual separation is therefore similar to that obtained by Bregman and Pinker.

4. General Discussion

4.1. *The Perceptual Relevance of Small Degrees of Onset Asynchrony*

In the current study, onset asynchronies of 20-25 ms generally did not enhance perceptual separation of the simultaneous tones. The perceptual relevance of asynchronies of 10 ms, as found by Rasch (1978) and by Hall & Grose (1991) could not be confirmed.

It is not clear to which extent the latter results are related to the threshold paradigms adopted. For simultaneous vowels presented at a clearly supra-threshold level, conditions with asynchronous onsets (and offsets) did not yield higher identification scores than those with synchronous onsets (Scheffers, 1979). The effect of onset asynchrony was also studied by Rasch (1981, Chapter 3); in his study the identification of musical intervals was the dependent variable, and one of the additional independent variables was the level difference between the constituting tones. The results indicated that the percentage of correct responses with asynchronous onsets (about 67%) was just significantly higher than that with strictly synchronous onsets (61%). Unexpectedly, however, the degree of asynchrony, which ranged between 5 and 40 ms, had no significant effect on performance level. It is unfortunate that the nature of the significant interaction effect of asynchrony and level difference, as found in Rasch's study, was not clarified.

One might wonder whether perceptual separation, as it was investigated in the present experiments, had already been strongly facilitated by presenting the tones twice. In such a paradigm a beneficial effect of onset asynchrony might not show up.

The mere repetition, however, was neither a sufficient nor a necessary condition for perceptual separation: firstly, the repetition of the tones in the HL/LH conditions did not yield high performance levels (overall percentages of correct responses were 59% and 34% in Experiments 1 and 2, respectively). Secondly, in the NC conditions, all four tones were presented only once. In spite of this, perceptual separation was adequate and at about the same performance level as found in the LL and HH conditions (overall percentages of correct responses in the NC conditions were 90% and 87% in Experiments 1 and 2, respectively).

Another issue that might be put forward in an attempt to explain why the perceptual relevance of small degrees of onset asynchrony could not be confirmed is related to the rise times of the tones: the effect of asynchrony may depend not just on the asynchrony itself, but on the degree of non-overlap of the rise portions of the two signals (see Figure 2). In the asynchronous conditions of Experiment 1, the non-overlap was 80%, whereas in Experiment 2, it was 167% of the rise time. For the conditions with the smaller asynchrony value included in, for example, the study of Bregman & Pinker (1978), the non-overlap was even 240% of the rise times of the tones.

The relevance of the degree of non-overlap to perceptual separation may be questioned, because for conditions in which Rasch (1978) already found significant perceptual effects, the degree of non-overlap was considerably lower than the non-overlap that occurred in Experiment 1 of the present study. In Rasch's experiments the rise time was 20 ms, but, in contrast to the definition used in the present study, this was the time between 10% and 90% of the maximum amplitude, implying that between 0% and 100% of the maximum amplitude, his rise time was $1.69(20) = 33.8$ ms. Consequently, in the conditions in which he set the onset asynchrony to 10 or 20 ms, the degree of non-overlap was as small as 30% or 59% of the rise time, respectively.

It is not difficult to understand why the degree of non-overlap fails to explain the discrepancy between the results obtained in the various studies. Basically, the *perceptual onsets* of the two slightly asynchronous tones with identical rise portions were shifted by the same amount, without affecting the degree of asynchrony that was intended to be used as the independent variable (see Vos & Rasch, 1981). The uncertainty about the exact point in time at which the onset of the tones is perceived will increase with increasing rise time of the tones, but this effect is estimated to be small. Moreover, from experiments in which the rise times of the slightly asynchronous tones were different, it was concluded that different rise times had effects comparable to different onset times without increasing onset uncertainty (Rasch, 1978, §4.3, Figure 18).

There is no real conflict between the results of the present study and those of Dannenbring & Bregman (1978) and Bregman & Pinker (1978). In the study of Dannenbring & Bregman (1978), the smaller degree of onset asynchrony investigated was equal to 35 ms and therefore greater than that used in the present study. Bregman & Pinker (1978) presented a pure tone A and a more or less simultaneous pair of lower frequency pure tones B and C in a repeatedly alternating train. The frequency of C was always lower than that of B and the onset (and offset) time delay of C relative to B was varied in about 30 ms steps between −60 and 60 ms. The 30 ms conditions yielded a significant decrease of the richness of the timbre of the tone with the lowest pitch (tone C) only if the frequency separation between A and B was small (about 1 semitone). In this specific condition the overall result may have been a combination of the effect of the asynchrony between C and B and the effect of temporal coherence between A and B. In the conditions in which the frequency interval between A and B was much greater (about 11 or 20 semitones), as a result of which temporal coherence between A and B was very difficult or even impossible to perceive (van Noorden, 1975, Chapter 2), a significant reduction of the richness of tone C's timbre was only obtained at the asynchrony value of 60 ms. The absence of a significant reduction at 30 ms is compatible with the results of the current study.

Darwin (1984a) showed that the quality of a vowel could be changed in the expected direction by an additional tone if the on- and offsets of the vowel and the tone were presented in synchrony. If the onset of the added tone preceded that of the vowel, the quality of the vowel changed back toward that of the original vowel. Especially in conditions in which perception was only mildly affected by masking, significant changes toward the original vowel were already obtained with onset asynchronies as small as about 25 or 40 ms. The latter finding was confirmed in a subsequent experiment with slightly different vowels (Darwin, 1984b). In related experiments, Roberts & Moore (1991) provided a more detailed range of onset asynchronies than Darwin did. Their results showed that asynchronies as small as 10 and 20 ms already resulted in small, but statistically significant changes toward the original vowel.

Our experimental results suggest that the perceptual effect of small onset asynchronies should not be overestimated. This is consistent with results from Darwin & Ciocca (1992). In their experiments they presented two successive complex tones with 12 harmonics each. In the first complex tone, however, the fourth harmonic could be mistuned and it could start earlier than the other 11 partials. Their subjects adjusted the pitch of the second tone to match the pitch of the first tone. The results showed that for onset asynchronies of 20 and 40 ms, the pitch shifts were not significantly different from those obtained in the synchronous conditions, whereas they were smaller for onset asynchronies of 80 and 160 ms, and completely absent for onset asynchronies of 320 and 640 ms. The results of Darwin & Ciocca (1992) imply that the contribution of a resolved frequency component to the pitch of a complex tone can be significantly diminished if onset time differences are greater than 40 ms. The results from subsequent experiments showed that for comparable conditions, such significant effects were obtained for onset asynchronies greater than or equal to 80 ms (Ciocca & Darwin, 1993).

4.2. *Asynchronization in Polyphonic Compositions*

In polyphonic compositions at least a significant portion of the tones of the different voices are prescribed to have synchronous onsets. In real performance, however, perfect synchronization is never realized. For professional trio ensembles, the standard deviation of the differences in onset time of simultaneous tones typically ranged between 30 and 50 ms (Rasch, 1979). In the latter study, the asynchronization may be considered to be non-intentional, which is supported by the stochastic nature of the asynchronization patterns. According to Rasch (1979, 1981), these findings suggested that in practical musical performance asynchronization plays an important role in the separate perception of simultaneous tones and voices. It should be emphasized, however, that this notion has not been verified in independent experimental studies. It might be that for significant effects on perceptual separation to occur, asynchronies even greater than 50 ms are needed. This would be in line with additional ways to asynchronize tone onsets in polyphonic compositions. Intentional asynchronization is involved in musical ornamentation such as the arpeggio and the French suspension; these and many other examples are given in Rasch (1981, Chapter 5). Furthermore, the various voices in polyphonic compositions may for the same reason be asynchronized at a macro-level. Diminution and alternation are well-known examples of macro-asynchronization. For a number of vocal polyphonic works composed by Praetorius, Rasch (1981) showed that macro-asynchronization increased with the number of voices. Since the signal-to-noise ratio of a single voice decreases with the number of the other voices, macro-asynchronization may be considered to be a tool to maintain the perceptibility of the separate voices at an acceptable level (Rasch, 1981). Explicit compositional efforts to avoid occurrences of synchronous tone onsets have also been found in Bach's two-part inventions (Huron, 1993).

5. Conclusion

In the present experiments, the effect of onset asynchrony on perceptual separation of simultaneous complex tones was investigated. In contrast with a number of previous studies reported in the literature, we used a paradigm in which the tones were presented in a noncyclical way and at sound pressure levels which were all clearly above threshold. Furthermore, we decided to restrict our study to onset time differences not greater than 20-25 ms. According to several studies, this degree of asynchrony should already have a significant effect on perception while, interestingly, the unequal onsets are hardly noticeable by naive listeners.

Our results showed that for the degree of onset asynchrony investigated, a facilitative effect of asynchrony on perceptual separation was only found in a small subset of conditions. This finding suggests that with respect to application in conditions in which the predictability of auditory events is relatively low, the effect of small differences in onset time on perceptual separation should not be overestimated. In this respect our results are in line with those of Darwin (1981), Darwin & Ciocca (1992), Ciocca & Darwin (1993), Scheffers (1979), and also with those of Rasch (1981) and, basically, those of Bregman & Pinker (1978).

Acknowledgements

This research was funded by the Netherlands Ministry of Defence. The author is grateful to Reinier Plomp for his stimulating discussions which led to the initiation of these experiments, and to Willem Heiser and Berrie Zielman for their advice on data analysis.

References

BEERENDS J.G., HOUTSMA A.J.M. (1989) "Pitch identification of simultaneous diotic and dichotic two-tone complexes," J. Acoust. Soc. Am., 85, 813-819

BREGMAN A.S. (1978) "Auditory streaming is cumulative," J. Exp. Psychol.: Human Percept. Perform., 4, 380-387

BREGMAN A.S. (1990) "Auditory Scene Analysis: the Perceptual Organization of Sound", MIT Press, Cambridge, MA

BREGMAN A.S., PINKER S. (1978) "Auditory streaming and the building of timbre," Can. J. Psychol., 32, 19-31

BROWN G.J., COOKE M. (1994) "Perceptual grouping of musical sounds: a computational model", J. New Music Research, 23, 107-132

CHALIKIA M.H., BREGMAN A.S. (1989) "The perceptual segregation of simultaneous auditory signals: pulse trains segregation and vowel segregation," Percept. Psychophys., 46, 487-496

CHALIKIA M.H., BREGMAN A.S. (1993) "The perceptual segregation of simultaneous vowels with harmonic, shifted, or random components," Percept. Psychophys., 53, 125-133

CIOCCA V., DARWIN C.J. (1993) "Effects of onset asynchrony on pitch perception: adaptation or grouping?," J. Acoust. Soc. Am., 93, 2870-2878

DANNENBRING G.L., BREGMANN A.S. (1978) "Streaming vs. fusion of sinusoidal components of complex tones," Percept. Psychophys., 24, 369-376

DARWIN C.J. (1981) "Perceptual grouping of speech components differing in fundamental frequency and onset-time," Q. J. Exp. Psychol., 33A, 185-207

DARWIN C.J. (1984a) "Auditory processing and speech perception," in "Attention and Performance X: Control of Language Processes", H. Bouma and D.G. Bouwhuis Eds., Erlbaum, Hillsdale, NJ, pp. 197-210

DARWIN C.J. (1984b) "Perceiving vowels in the presence of another sound: constraints on formant perception," J. Acoust. Soc. Am., 76, 1636-1647

DARWIN C.J., CIOCCA V. (1992) "Grouping in pitch perception: effects of onset asynchrony and ear of presentation of a mistuned component," J. Acoust. Soc. Am., 91, 3381-3390

DARWIN C.J., SUTHERLAND N.S. (1984) "Grouping frequency components of vowels: when is a harmonic not a harmonic?," Q. J. Exp. Psychol., 36A, 193-208

DOLL T.J., FOLDS D.J. (1986) "Auditory signals in military aircraft: ergonomics principles *versus* practice," Appl. Ergonomics, 17, 257-264

ELLIOTT L.L. (1962a) "Backward masking: monotic and dichotic conditions," J. Acoust. Soc. Am., 34, 1108-1115

ELLIOTT L.L. (1962b) "Backward and forward masking of probe tones of different frequencies," J. Acoust. Soc. Am., 34, 1116-1117

GREEN D.M., DAI H. (1992) "Temporal relations in profile comparisons," in "Advances in the Biosciences", Y.Cazals, K. Horner, L. Demany Eds., Vol. 83, Auditory Physiology and Perception, Proc. of the 9th International Symposium on Hearing, Pergamon Press, Oxford, pp. 471-478

HALL J.W., GROSE J.H. (1991) "Some effects of auditory grouping factors on modulation detection interference (MDI)," J. Acoust. Soc. Am., 90, 3028-3035

HOUTSMA A.J.M., CANNING J.M. (1983) "Pitch Perception of Simultaneous Complex Tones", IPO Annual Progress Report 18, Institute for Perception Research, Eindhoven, The Netherlands, pp. 20-25

HOUTSMA A.J.M., CANNING J.M., BEERENDS J. (1984) "A preliminary study of identification of harmonic intervals made by simultaneous complex tones," in "Proceedings of a symposium on computational models of hearing and vision", Academy of Sciences of the Estonian S.S.R., Tallinn, Estonia, pp. 19-23

HURON D. (1993) "Note-onset asynchrony in J.S. Bach's two-part inventions," Music Perception, 10, 435-444

KILLAM R.N., LORTON P.V., SCHUBERT E.D. (1975) "Interval recognition: identification of harmonic and melodic intervals," J. Music Theory, 19, 212-235

PLOMP R., WAGENAAR W.A., MIMPEN A.M. (1973) "Musical interval recognition with simultaneous tones," Acustica, 29, 101-109

RASCH R.A. (1978) "The perception of simultaneous notes such as in polyphonic music," Acustica, 40, 21-33

RASCH R.A. (1979) "Synchronization in performed ensemble music," Acustica, 43, 121-131

RASCH R.A. (1981) "Aspects of the Perception and Performance of Polyphonic Music", Elinkwijk BV, Utrecht, The Netherlands, Doctoral Dissertation

ROBERTS B., MOORE B.C.J. (1991) "The influence of extraneous sounds on the perceptual estimation of first-formant frequency in vowels under conditions of asynchrony," J. Acoust. Soc. Am., 89, 2922-2932

SCHEFFERS M.T.M. (1979) "The role of pitch in perceptual separation of simultaneous vowels," IPO Ann. Progr. Rep., 14, 51-54, Institute for Perception Research, Eindhoven, The Netherlands

SCHEFFERS M.T.M. (1983) "Sifting Vowels: Auditory Pitch Analysis and Sound Segregation", Rijksuniversiteit Groningen, Groningen, The Netherlands, Doctoral Dissertation

Van NOORDEN L.P.A.S. (1975) "Temporal Coherence in the Perception of Tone Sequences", Technical University Eindhoven, Eindhoven, The Netherlands, Doctoral Dissertation

VOS J. (1982) "The perception of pure and mistuned musical fifths and major thirds: thresholds for discrimination, beats, and identification," Percept. Psychophys., 32, 297-313

VOS J. (1986) "Purity ratings of tempered fifths and major thirds," Music Perception, 3, 221-258

VOS J., RASCH R.A. (1981) "The perceptual onset of musical tones," Percept. Psychophys., 29, 323-335

VOS J., VIANEN B.G. Van (1985) "Thresholds for discrimination between pure and tempered intervals: the relevance of nearly coinciding harmonics," J. Acoust. Soc. Am., 77, 176-187

WILLIAMS S.M. (1989) "STREAMER: A Prototype Tool for Computational Modelling of Auditory Grouping Effects", Department of Computer Science Research Report N°98-31, University of Sheffield, United Kingdom

WILLIAMS S.M., NICOLSON R.I., GREEN P.D. (1990) "STREAMER: Mapping the auditory scene," in Proc. of the Institute of Acoustics, Autumn Conference, Windemer, Vol. 12, Part 10, pp. 567-575

WINER B.J. (1970) "Statistical Principles in Experimental Design", McGraw-Hill, London

WRIGHT H.N. (1964a) "Temporal summation and backward masking," J. Acoust. Soc. Am., 36, 927-932

WRIGHT H.N. (1964b) "Backward masking for tones in narrow-band noise," J. Acoust. Soc. Am., 36, 2217-2221

ZERA J., GREEN D.M. (1993) "Detecting temporal asynchrony with asynchronous standards," J. Acoust. Soc. Am., 93, 1571-1579

ZWISLOCKI J.J. (1978) "Masking: experimental and theoretical aspects of simultaneous, forward, backward, and central masking," in Handbook of Perception Vol. IV, E.C. Carterette and M.P. Friedman Eds., Academic Press, New York, Chapter 8, pp. 283-336