4

Timing and synchronization in ensemble performance

RUDOLF A. RASCH

Introduction

For the last six centuries Western music has been largely polyphonic. This means that a piece of music consists of two or more simultaneous parts or 'voices' (either vocal or instrumental). In order to make performance possible, the temporal relations between the various voices as well as those within a single voice have to be defined exactly and unambiguously. These relations must be clearly stated in the score of the music and appropriately realized in a performance. In its initial stage polyphonic music was often called 'measured music' (*musica mensurata*) as opposed to 'plain music' (*musica plana*), that is, plainchant in which there was no requirement of strict time structure.

Observations from such widely different fields as timbre experiments, instrumental sounds, and notation lead to the conclusion that the beginning or onset of a tone is the decisive point in time for that tone. Timbre experiments have shown that recognition of musical instrument tones is severely impaired by taking away the starting transient portion (Clark *et al.* 163; Berger 1964; Saldanha and Corso 1964). Widely used instruments such as piano, harpsichord, and guitar (in general, all plucked or struck instruments) produce tones without steady-state parts, which begin to decay immediately after the onset. Traditional music notation puts notes at the onset moment of the horizontal time scale, and places those with simultaneous onsets together irrespective of their lengths. Investigation of the temporal structure of music is therefore predominantly a study of onset moments.

In this chapter we deal with the problem of synchronization in musical performance, that is, how musicians manage to co-ordinate their own temporal (onset) structures with those of the other performers in such a way that the temporal structures actually match each other and fuse into one common temporal structure. This means that the tones which are prescribed in the score as simultaneous must be played at the same moment of time, especially with regard to their onsets. In reality, perfect synchronization is never realized and there will always be some asynchronization. Onsets meant to be simultaneous but in performance not exactly simultaneous will be called quasi-simultaneous onsets.

In the first and second sections of this chapter, we will present a model for the description of synchronization of ensemble performances (Rasch 1979, 1981). It is based on the standard deviations of relative onset times and onset time differences. The third section is devoted to the question of how to define the onset times of musical tones in a synchronization situation (Vos and Rasch 1981). Finally, we will speculate a little about the role of the conductor in the synchronization of musical performances.

The description of synchronization

When musicians play together in an ensemble, they will try to synchronize as much as possible the tones meant to be simultaneous. For a number of reasons, such as the restricted accuracy of human motor performance and time perception, the relative ease of tone production within or between instruments, and the time lag between the production of a player's own tones and the perception of the tones produced by others, a perfect synchronization is not possible in a live performance. There will always be some degree of asynchronization.

Onsets of tones meant to be simultaneous will, in reality, scatter a little in time. Assume an ensemble of *n* players. Figure 4.1, illustrates some of

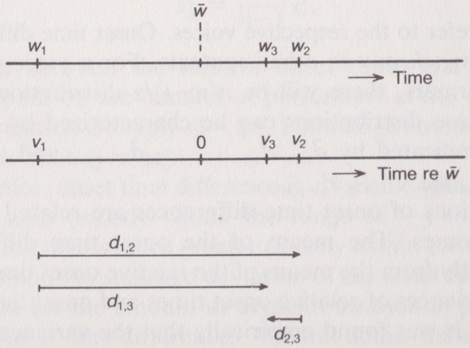

Fig. 4.1. Illustration of the concepts used for synchronization analysis. w_1, w_2 and w_3 are three onsets meant to be simultaneous. $\bar{w}$ is the mean onset time. v_1, v_2, and v_3 are the onset times expressed relative to $\bar{w}$. The onset time differences are denoted by d_{12}, d_{23}, d_{31}.

the concepts used for an ensemble of three performers. The *absolute onset times* of the tones meant to be simultaneous will be denoted by $w_1, w_2, \ldots, w_n$. For each group of such onsets there will be a mean onset time:

$$\bar{w} = \frac{w_1 + w_2 \ldots + w_n}{n}.$$

Now, the absolute onset times can be converted into *relative onset times*:

$$\begin{aligned} v_1 &= w_1 - \bar{w}, \\ v_2 &= w_2 - \bar{w}, \\ &\ldots, \text{ and} \\ v_n &= w_n - \bar{w}. \end{aligned}$$

The relative onset times are relative to the mean onset time, not to other onset times. Below, we will as a rule use the relative onset times. The absolute onset times of the tones in a piece of performed music can be measured and transformed into relative onset times. This will result in a set of n distributions of relative onset times, one per voice. These distributions can be characterized by their means, which will be indicated by $\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n$, and their variances, which will be denoted by $s_1^2, s_2^2, \ldots, s_n^2$. The means indicate which instruments lead or lag in their onsets relative to the other instruments in a performance.

Time differences between onsets are particularly significant for perception. The *onset time differences* will be denoted by the letter d:

$$\begin{aligned} d_{12} &= v_2 - v_1, \\ d_{13} &= v_3 - v_1, \\ &\ldots, \text{ and} \\ d_{n-1,n} &= v_n - v_{n-1}. \end{aligned}$$

The subscripts refer to the respective voices. Onset time difference has been termed *onset asynchrony* in other contexts. For a piece of music with n voices or performers, there will be $n(n-1)/2$ distributions of onset time differences. These distributions can be characterized by their respective means, to be indicated by $\bar{d}_{12}, \bar{d}_{13}, \ldots, \bar{d}_{n-1,n}$, and variances, to be indicated by $s_{12}^2, s_{13}^2, \ldots, s_{n-1,n}^2$.

The distributions of onset time differences are related to those of the relative onset times. The means of the onset time differences can be calculated directly from the means of the relative onset times. The relation between the variances of relative onset times and onset time differences is more complex. It was found empirically that the variances of the several distributions of relative onset times for one specific piece of music do not differ very much from each other. For simplicity we will assume that they are equal, and denote their value by s_v^2. In addition we will assume that the covariances between distributions are equal. Their value will be denoted by c_v. With these assumptions it is possible to derive a simple formula that relates the variances of the relative onset times and the onset time differences.

Each relative onset time can be written as a linear sum of all other relative onset times, for example:

$$v_1 = (v_2 + v_2 + v_3 + \ldots + v_n).$$

The variances of such a sum equals the sum of the variances of the terms of the sum plus twice the sum of all covariances among the terms:

$$s_v^2 = \sum_{i=2}^{i=N} s_{v_i}^2 + 2 \sum_{i=2}^{i=n} \sum_{j=i+1}^{i=n} c_{v_{ij}} \quad (j \neq i)$$

in which $c_{v_{ij}}$ is the covariance of v_i and v_j.

Since, by assumption, $s_{v_1}^2 = s_{v_i}^2 = s_v^2$ and $s_{v_{ij}}^2 = c_v$, the variance of the relative onset times in general can be expressed in the following equation:

$$s_v^2 = (n-1)s_v^2 + (n-1)\,(n-2)c_v.$$

This leads to:

$$c_v = \frac{s_v^2}{1-n}.$$

Since an onset time difference can be considered to be a linear combination of the form $d_{12} = v_2 - v_1$, its variance is given by:

$$s_d^2 = 2s_v^2 - 2c_v.$$

Substituting the value of c_v in the last equation leads to:

$$s_d^2 = \frac{2n}{n-1} s_v^2.$$

So the relation between the relative onset times and the onset time differences depends on the number of performers in the ensemble. For a duo (n=2) the relation is simply $s_d^2 = 4s_v^2$. For a large ensemble, the relation converges to the limit $s_d^2 = 2s_v^2$.

In actual practice, onset time difference is a variable with a mean not very different from zero (mostly within the range of −5 to +5 ms) and a standard deviation that is much greater (mostly within the range of 30 to 50 ms). For this reason the standard deviation of the onset time differences is a better measure for the amount of asynchronization in performed music than the mean onset time differences. We will define the *asynchronization of a pair of voices* as the standard deviation of the onset time differences of simultaneous tones of those voice parts. In formula, the asynchronization will be denoted by the capital A: $A_{ij} = s_{ij}$.

The standard deviations of the various distributions of onset time differences in one piece of music do not differ very much from each other. We do not lose very much information if the asychronizations of the various voice pairs are averaged, by taking the root-mean-square. This averaging has the advantage that the asynchronization of a piece of music can be expressed as a single value. We will define the *asynchronization of a piece of performed music* as the root-mean-square of the standard deviations of the onset time differences for all pairs of voice parts. The asynchronization of a piece of music will be denoted by the capital A without subscripts.

It should be noted that the asynchronization defined quantitatively in this way indicates ranges in which onset time differences will fall with certain probabilities. The range from $-A$ to $+A$ includes 68 per cent of onset time differences, the range from $-2A$ to $+2A$ 95 per cent. If we discard the sign of the onset time differences, it can be stated that the median unsigned onset time difference is $0.68A$, the lower quartile $0.32A$, and the higher quartile $1.15A$.

The synchronization of simultaneous tones is only one of the temporal tasks of performing musicians. Within one voice, the succession of tones also has to be strictly timed. We will call the timing of tones of equal duration in one voice the *isochronization* of the tones. The isochronization of a voice will be defined as the standard deviation of tone durations meant to be equal. In actual practice, there will be a number of sources that contribute to the variance of equal tone durations, like tempo trends and fluctuations. Also, the asynchronization of simultaneous tones will be mirrored in isochronization. If we exclude all other sources of the variance of 'equal' tone durations, there is a certain relation between asynchronization and isochronization (actually, a-isochronization). We consider two consecutive tones with absolute onset times w_i and w_{i+1} (see Fig. 4.2). The duration of the first tone is

$$h_i = w_{i+1} - w_i.$$

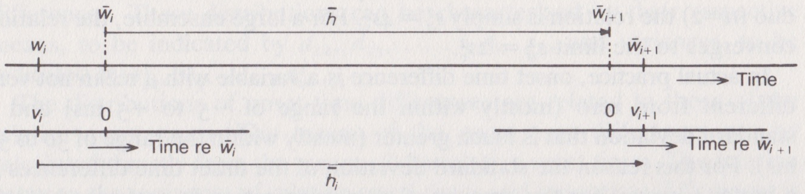

Fig. 4.2. Illustration of the concepts used in determining the relation between asynchronization and the standard deviation of tone durations. $\bar{w}_i$ and $\bar{w}_{i+1}$ are the mean onset times of two successive triplets of tones. w_i and w_{i+1} are the onsets of two successive tones in one voice part; v_i and v_{i+1} are the same onsets but expressed relative to the mean onset times. $\bar{h}$ and h_i are the time intervals between the mean onsets and the onsets in a single voice, respectively.

The absolute onset time is the sum of the mean onset time and the relative onset time. We denote the mean onset time of tones simultaneous to w_i as $\bar{w}_i$, that of tones simultaneous to $\bar{w}_{i+1}$. We assume that the successive mean onset times $(\ldots, \bar{w}_i, \bar{w}_{i+1}, \ldots)$ are perfectly isochronous. Therefore, mean tone duration $\bar{h}$ is

$$\bar{h} = \bar{w}_{i+1} - \bar{w}_i.$$

The relative onset times of the two tones in succession will be denoted by v_i and v_{i+1}. Now we can rewrite the actual tone duration as follows:

$$h_i = \bar{h} + (v_{i+1} - v_i).$$

The variance of this last expression is

$$s_h^2 = 2(1-r)s_x^2.$$

in which r is the correlation between v_i and v_{i+1}, or, the autocorrelation of relative onset times. The above equation can also be written as

$$s_h^2 = \frac{n-1}{n}(1-r)s_d^2.$$

The relation between isochronization and synchronization can be used in the comparison of the results of our measurements with the results of experiments on sequential timing behaviour which exist in the literature.

The measurement of synchronization

The descriptive model given will now be illustrated with a set of measurements of live musical performances specifically recorded for this purpose (Rasch 1979, 1981).

In order to study synchronization in musical performances it is necessary to know the exact temporal structure of each individual part of the composition performed. Attempts have been made to analyse the compound sound of a piece into its individual parts, for instance by Tove *et al.* (1967) at the Royal Institute of Technology in Stockholm, and by Moorer (1975) at the Center for Computer Research in Music and Acoustics of Stanford University. Up to now, these attempts have been only partially successful. Several serious problems prevent application of the methods employed to larger, not specially selected samples of music. The main problem areas in this field are the storage and processing of the prohibitively large amounts of data, and the separation of fused harmonics in the cases of musical intervals with simple frequency ratios.

A more feasible approach is the one in which the acoustics signals from each instrument are picked up and recorded before they are mixed into the ensemble sound. This can be done by using either contact microphones under normal acoustic conditions or directional microphones in an anechoic chamber. The latter method was chosen for our recordings because the recorded sound most resembled the sound in normal listening conditions. The recorded signals passed a logarithmic envelope detector with an output voltage proportional to the sound-pressure level of the temporal envelope of the signals. The sampling rate was 200 Hz, the time interval between two successive samples being 5 ms. The envelopes could be displayed on the graphic terminal of the computer. Peaks in the enveloped were matched with the notes of the score played by the musicians. The onset time of a peak was defined as the moment that the envelope reached or surpassed a certain, previously defined threshold level. This level was taken about 15 or 20 dB below the maximum levels of the signals. Figure 4.3 shows the envelopes of a fragment of 9 s of music, with the peaks numbered according to the computer output and the score of the music inserted for comparison and identification purposes.

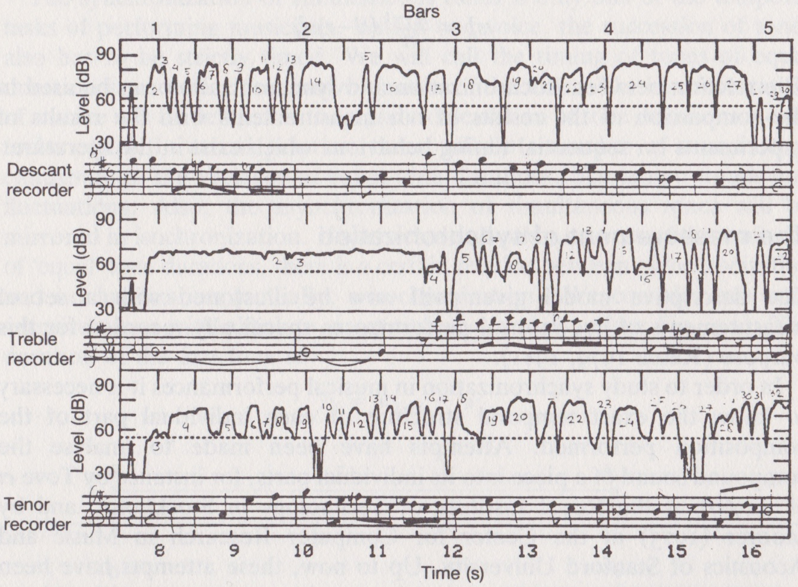

Fig. 4.3. The envelopes of three simultaneous voices of a fragment of 9 s. The peaks have been numbered, and the notes of the score have been inserted.

The following step was to determine which peaks in the three envelopes corresponded to a triplet of simultaneous notes in the musical score. Table 4.1 shows the simultaneous onsets from Fig. 4.3, the relative onset times, and the onset time differences. Note that the temporal resolution of the processing system was 5 ms. As a matter of fact, not all the triplets of simultaneous notes found in the score had a corresponding triplet of peaks in the envelopes. If the initial part of a tone did not correspond to a peak in the envelope, no corresponding onset time was available for the calculations.

Table 4.1 *Peaks corresponding to simultaneous notes from the music fragment of Fig. 4.3, with absolute onset times, relative onset times, and onset time differences.*

Instrument 1	2	3	absolute onset time (ms)			relative onset time (ms)			onset time differences (ms)		
Peak no.			w_1	w_2	w_3	v_1	v_2	v_3	d_{12}	d_{23}	d_{31}
18	5	18	11970	11995	11955	−3	21	−18	25	−40	15
19	8	20	12670	12675	12695	−10	−5	15	5	20	−25
22	12	22	13405	13390	13410	3	−12	8	−15	20	−5
24	15	24	14075	14105	14125	−27	3	23	30	20	−50
25	17	26	14795	14800	14795	−2	2	−2	5	−5	0
27	20	18	15500	15510	15470	7	17	−23	10	−40	30

The statistical computations were mostly done per composition, sometimes per movement of a composition. First, means, standard deviations, and root-mean-squares of standard deviations of relative onset times and onset time differences were calculated. These figures make up the basic data set of our analysis. Further computations included the autocorrelations of relative onset times, correlations with tempo measures, and conversion of synchronization into isochronization values.

Recordings were made of music played by three professional trio ensembles: a recorder trio, a reed trio (oboe, clarinet, bassoon), and a string trio. Each ensemble played two compositions that were typical for the repertoire.

Means and standard deviations of relative onset times were calculated for each recorded composition. The mean relative onset times are graphically presented in Fig. 4(a). They indicate to what extent certain instruments tend to lead or lag in simultaneous onsets relative to the other instruments. Certain tendencies can be discerned. In the string and wind trios the main melody instruments (violin, oboe) tend to lead relative to the others, with the bass instruments (cello, bassoon) in the second place, and the middle voice (viola, clarinet) in the third place. This corresponds to our intuitive notion that in most of the music for these ensembles the melody part is the leading, 'first' voice. Recorder ensemble music is more polyphonic as a rule, and the bass is the most fundamental voice. This is confirmed by the observation that the tenor recorder—the lowest voice—is the leading voice, on the average.

There are three distributions of onset time differences for each

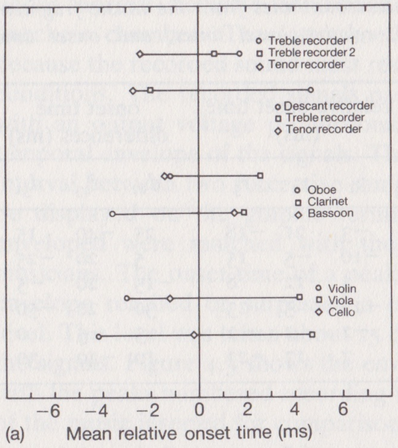

Fig. 4.4(a). Mean relative onset times of the instruments per composition.

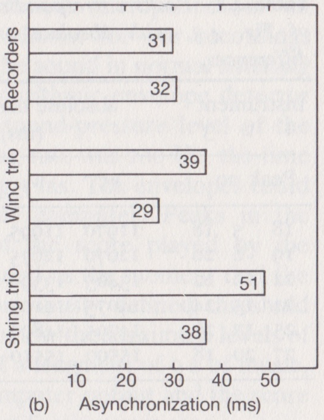

Fig. 4.4(b). Mean asynchronization per composition, averaged over the instruments.

composition. The root-mean-squares of the standard deviations are used as measures for asynchronization. Asynchronization per composition is graphically illustrated in Fig. 4(b). Asycnchronization values fall largely within the range of 30 to 50 ms. Mean asynchronization of this set of data is 36 ms. There are some systematic differences between the ensembles. The recorder ensemble shows a relatively small asynchronization of about 30 ms. The two examples of the wind trio have different amounts of asynchronization: 27 and 37 ms respectively. The largest asynchronizations are found in the string trio examples: 37 and 49 ms. These differences may be related to the rise times of the tones of the instruments in question. Recorders have short rise times, so that the beginnings of the tones are clearly marked. These short rise times make good synchronization both possible and necessary. The conventional woodwinds of the wind trio also have relatively short rise times. The onset of string tones is a more gradual process resulting in long rise times (30 to 100 ms). These longer rise times permit a greater leniency as to synchronization.

The relation between asynchronizatoin and tempo was also investigated (see Fig. 4.5). All data in this table concern single movements of the compositions. As tempo measure we used the time between two successive counting beats, the 'inter-beat time', determined as 60 divided by the metronome value (the number of counting beats per minute). This measure includes a subjective element in the choice of the counting unit. The correlation between the asynchronization of the movements and the tempo measure of 0.80. Evidently, the conclusion is justified that faster

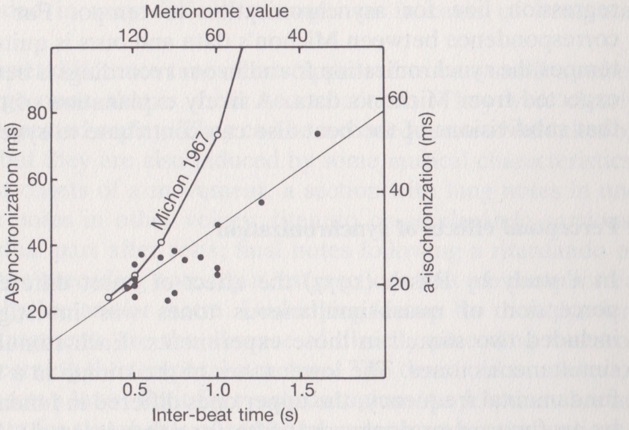

Fig. 4.5. Relation between tempo and asynchronization. Inter-beat time is used as a measure of tempo. The vertical scales indicate asynchronization, defined on the left as the standard deviation of onset time differences, averaged over instruments. On the right, a-isochronization is indicated. The dots are data points for movements. The regression line of tempo on asynchronization is based on the correlation between inter-beat time and asynchronization. The results of Michon's experiments (1967, p. 31) are inserted for comparison.

tempo goes with more synchronization, and slower tempo goes with less synchronization.

As already noted, asynchronization of simultaneous onsets can be the cause of a-isochronization, that is, the existence of deviations from equal durations of successive tones within single voices. It is possible to make an estimate of this a-isochronization from asynchronization. For this estimate the correlation between subsequent relative onset times must be known. This correlation is the autocorrelation of a series of relative onset times. The autocorrelations were first calculated per composition and then algebraically averaged over voices and compositions. The resulting value proved to be 0.21. This indicates that there is a slight tendency of relative onset times to be preserved in the subsequent onsets. With $r = 0.21$, we find a ratio of 0.73 between a-isochronization and asynchronization. Apart from the autocorrelation, the asynchronization patterns seem to be more or less random.

Now we can compare our results with those of experiments conducted by Michon (1967). His subjects had to tap synchronously with a pulse train presented through headphones. His results (p. 31, Table 2B) are standard deviations of time intervals meant to be equal, and as such indicate a-isochronization. Tapping rates were 60, 90, 120, and 180 beats per minute. Figure 4.5 includes both the results of Michon's experiments and the regression line for asynchronization to tempo. For high tempos the correspondence between Michon's data and ours is quite good. For lower tempos the synchronization found in our recordings is better than would be expected from Michon's data. A likely explanation of this discrepancy is that subdivisions of the beat also can contribute to synchronization.

Perceptual effects of synchronization

In a study by Rasch (1977) the effect of onset difference times on the perception of quasi-simultaneous tones was investigated. Each trial included two stimuli in those experiments. Each stimulus contained two simultaneous tones. The lower tones of the stimuli in a trial had the same fundamental frequency, the upper ones differed in fundamental frequency by an (upward or downward) fifth (or other interval). The threshold for perceiving the upper tone of the stimulus was defined as the level resulting in a 75 per cent correct score in a 2AFC paradigm of the judgement of the direction of the pitch jump. For synchronous notes (onset difference time 0 ms) the threshold was between 0 and −20 dB depending on the frequency ratio and phase relation of the simultaneous tones. The threshold could be decreased drastically, down to about −60 dB, by introducing an onset difference time of, say, 30 ms. In the latter case, the threshold was largely independent of factors other than temporal ones. This condition can be compared directly with asynchronization in performed music. In the experiments there was no sensation of order. In supra-threshold conditions asynchronization contributes to the apparent transparency of the compound sound multi-tone stimuli.

When listening to music, onset differences mostly go unnoticed. Actually, the performances of professional ensembles give the impression of perfect synchronization. In the now classic investigation by Hirsh (1959) it was found that, roughly speaking, a time interval of 20 ms between two sounds (tones, noises, clicks) was sufficient for judging which one came first. Several reasons can be mentioned why onset difference times that are considerably longer are not perceived in a musical listening situation. Firstly, the qualities of the simultaneous tones differ, especially as to pitch and timbre (instrument); this impairs an exact judgement of the temporal structures. Secondly, rise times of musical tones are much longer (for example 20–80 ms) than the stimuli used in psychophysical experiments (often 5 or 10 ms). The smoother rise curves also impair exact judgements. Thirdly, the distribution of onset differences is practically random. Fourthly, attention of the listener is not directed to onset differences between voices ('vertical') but rather to melody lines within one voice ('horizontal'). The musicians themselves are, in general, also unaware of the amount of asynchronization, except when unwanted asynchronization cannot be avoided in very difficult passages or in cases of 'temporal mistakes'.

Among the onset difference times there appeared to be a small number of relatively large ones, that is, from 100 to 200 ms. They were clearly audible during repeated replay. They can be a consequence of mistakes by the performer, but they are also induced by some musical characteristics, such as: the first onsets of a movement, a section with long notes in one voice and short notes in other voices; ritenuto or accelerando sections; onsets of individual part after rests; final notes following a ritardando or separated from the preceding notes by a rest; sections that are complicated with regard to rhythm and/or metre. Underlying factors for these harder-to-synchronize fragments are the absence of directly preceding tones or uncertainty concerning the temporal structure. The threshold for order of auditory stimuli sets a lower limit for synchronization. As yet, there are no experimental data for thresholds of temporal order in musical situations.

As a final remark, mention must be made of Seashore's theory that 'beauty in music largely lies in the artistic deviation from the exact or rigid' (1938, p. 249). This postulate leaves open the question of how to define artistic deviation. However, its negative formulation ('there is no beauty in music if the performers adhere exactly and rigidly to the instructions of the score') is certainly true. In every artistic musical performance there is constant deviation from what is prescribed exactly in the score, as to time, frequency, duration, level, etc. These deviations have strong statistical components without being purely random. They are of primary importance for the 'live' character of music performed by human beings. The asynchronization of simultaneous tones should be regarded as one of the vital deviations in the performance of music.

The moment of synchronization

In the preceding section, the decisive time moment of the tone, used as a basis for the calculation of synchronization, was taken where the temporal envelope of the tones passed a threshold, quasi-arbitrarily defined as the level 15 or 20 dB below the average maximum level of the respective voice part. Since this choice is fairly central in the description of synchronization behaviour, the threshold model concerning the perceptual onset of musical tones will be considered here in some detail. The model assumes that music listeners (including music performers) base their timing on the passing of the temporal envelope through a certain threshold level, which is defined in relation to the average maximum level of the stream of musical tones (Vos and Rasch 1981). The physical temporal envelope of a musical tone can be roughly divided into three successive portions: the rise, the steady-state, and the decay portions (see Fig. 4.6).

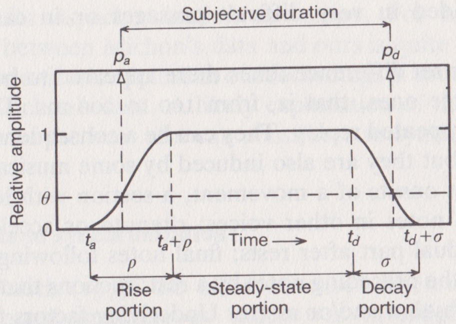

Fig. 4.6. Temporal envelope of a musical tone, divided into three successive portions. The perceptual onset (p_a) and offset (p_d) of the tone are defined as the moment at which the temporal envelope passes the relative threshold amplitude, θ.

If the rise function, which describes the envelope during the rise, is monotonously increasing, then the inverse function is unambiguous. Throughout this chapter, we will regard rise functions as monotonously increasing functions. In our model, the perceptual onset of a tone is the moment at which the temporal envelope during the rise portion passes a certain relative threshold level. If the moment of the perceptual onset is known, the threshold level can be calculated. If the threshold level is known, the perceptual onset can be calculated. The subjective duration of a tone is defined as the time interval between the perceptual onset and offset of a tone. The following paragraphs, describing an extension of our model, will deal only with the perceptual onset.

The model can be extended to groups of tones, either simultaneous or successive ones. Two tones are called *perceptually synchronous* when their perceptual onsets coincide in time (see Fig. 4.7).

Tone sequences are defined as *perceptually isochronous* if the time intervals between successive perceptual onsets are all equal to each other. Figure 4.8(c) shows the temporal envelopes of the tones A, B, and A', which are perceptually isochronous. An experimental paradigm with successive perceptually isochronous tones with different rise functions and/or rise times can be used to determine the threshold amplitude for the perceptual onset (Vos and Rasch 1981). If tones are physically isochronous, that is, when the time intervals between successive physical onsets are all equal to each other but have different rise times and/or functions, the perceptual onsets will not, as a rule, be isochronous.

In the experiments exploring the applicability of the model a paradigm was used in which a sequence of tones had to be isochronized, that is, the tones had to be adjusted in such a way that the onsets were perceived isochronously. Each trial started with a tone sequence that was decidedly nonisochronous. The starting sequence consisted of successive pairs of tones, A and B, with a different interval between the physical onsets of A and the following B and between B and the following A (see Fig. 4.8(a)). The onset times of tones B, relative to those of A, could be adjusted by the subjects, by turning a knob. The experimental task was to adjust the onset times of tones B in such a way that the sequence ABABAB . . . was perceptually isochronous, that is, that the perceived onsets of the tones followed each other after strictly the same time interval. This is illustrated in Fig. 4.8(b). Because the tones A were repeated every 800 ms, the subjective repetition time t of the tones in the entirely isochronized sequence is 400 ms. Rise times were varied independently. The time between the successive physical onsets was derived from the position of the turning knob at which the subject judged the tone sequence to be isochronous. Now all variables that are necessary for computing the threshold amplitude for the perceptual onset were known.

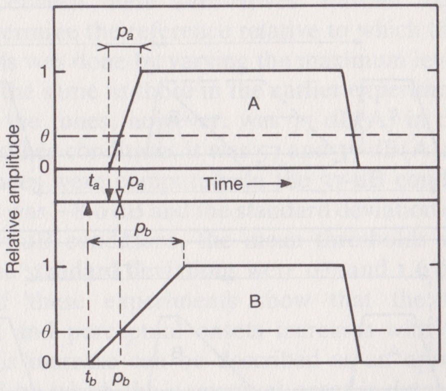

Fig. 4.7. Temporal structure of a perceptually synchronous stimulus, containing tones A and B with physical onsets t_a and t_b, rise times p_a and p_b, and maximum amplitude 1. The perceptual onsets, p_a and p_b, coincide in time and are by definition located at the moment at which the temporal envelopes pass threshold amplitude, θ.

Relative threshold amplitudes were computed for a number of combinations of unequal rise times. Mean threshold level was −15.4 dB, and the standard deviation equalled 1.6 dB. Considering this consistent result—that is, that about the *same* relative threshold level was found in ten physically different conditions—it is justified, for the time being, to define the perceptual onset moment of a tone as the time at which its envelope passes a certain threshold level. Moreover, from these experimental data, this threshold level can be estimated as −15 dB, relative to the maximum level of the tones.

The perceptual onset threshold has up to now been given relative to the

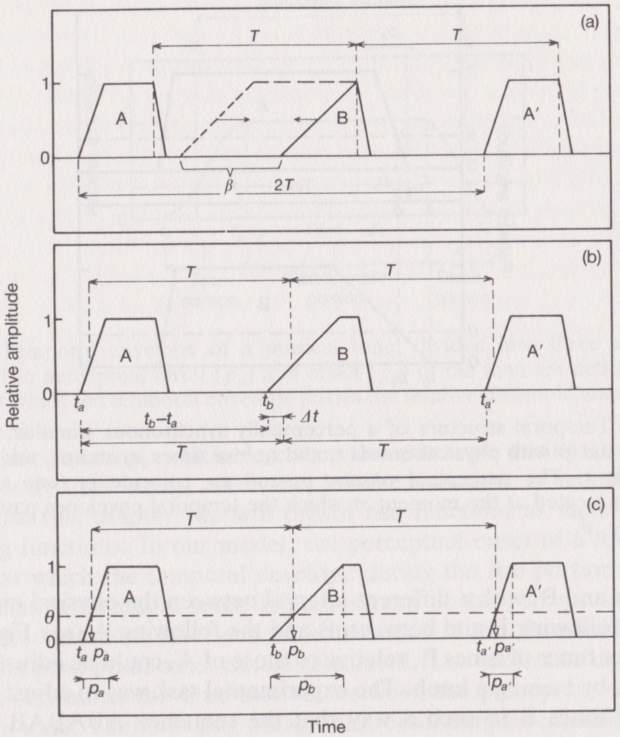

Fig 4.8. (a) Illustration of a perceptually nonisochronous starting sequence. The physical onsets of tones B could be adjusted by the subject within time interval β. At the start of a trial, the physical onset of tone B was either at the beginning or at the end of β. The physical offset of tones A and B, were fixed. (b) Temporal structure of a perceptually isochronous tone sequence in which the time intervals between successive perceptual onsets are all equal to each other. The physical onsets of tones A and B are t_a and t_b, respectively. The repetition time of tone A equals $2T$, so that T is the perceptual repetition time. The dependent variable is denoted by the time interval $\Delta t = T-(t_b-t_a)$. (c) Temporal structure of tones A, B, and A′, which are perceptually isochronous. The time interval between successive perceptual onsets is denoted by T. The perceptual onsets p_a, p_b, and $p_a{}'$ are defined as the moment at which the temporal envelopes pass relative threshold amplitude, θ.

maximum level of the tones. But it is evident that the threshold can also be defined as a level above background noise or hearing threshold. In short, the following question can be asked: Is the threshold fixed, with respect to maximum level, to background, or to some other criterion? Our first experiments were designed to test the threshold hypothesis in general, not to discriminate between these alternatives. Further experiments were carried out to determine the reference relative to which the threshold had to be defined. This was done by varying the maximum levels of the tones. The stimuli were the same as those in the earlier experiments. The sound-pressure level of the tones, however, was 77 dB(A) in the highest level condition; in the other conditions it was 57 and 37 dB(A). Again, relative threshold amplitudes were computed. In the 37-dB condition, the mean relative threshold was −8.0 dB and the standard deviation equalled 0.9 dB. In the 57- and 77-dB conditions, the mean thresholds were −11.7 and −13.0 dB, and the standard deviations were 0.7 and 1.6 dB, respectively.

The results of these experiments show that the time difference between physical and perceptual onsets increases with decreasing tone intensity, that this increase can be described as an upward shift in the relative threshold by which the perceptual onset is determined, and that the shift in threshold is small relative to the shift in stimulus level. Therefore, the threshold can be most conveniently described relative to the maximum level of the stimulus.

From the experiments, it may be concluded that the perceptual onsets of musical tones can be defined as the times at which the envelopes pass a *relative* threshold of about 6 to 15 dB below the maximum level of the tones. In a number of experiments, we have shown that the level of the relative threshold depends on the tone level above masked or absolute threshold. The data from the present experiments seem to suggest adaptation to a certain constant stimulus level. At the time the adaptation threshold is passed by the stimulus level presented, the onset of the stimulus is perceived.

In studies on the temporal structure of performed music, our threshold model can be applied to determine the perceptual onsets of musical tones. When music is performed on instruments with very short rise times, like the piano, harpsichord, and drums (Gabrielsson 1974; Povel 1977; Sundberg and Verrillo 1980), the difference between the physical onset and the perceptual onset is very small. In these cases, level above threshold, too, does not have a great impact on this difference. However, when ensemble music is performed on instruments producing tones with relatively long rise times, such as bowed string instruments, the perceptual onset heavily depends on the relative threshold.

In such musical practice in which dynamic differences are not very large, perceptual onset is clearly affected only by the rise times and rise functions of the different instruments. This, however, is a variable with which the respective musicians can cope by adjusting their physical onset times in order to establish the appropriate timing of the perceptual onsets of their tones.

Future research should be focused on the perceptual onset of musical tones in synchronously perceived tone pairs. The sensation levels of simultaneously presented tones, especially, are dependent on the amount of auditory masking (Zwislocki 1978). To apply our model to simultaneously produced tones, experimental results of binaural masking experiments with complex tones are needed. In addition, it would be interesting to see if our model also works in cases of complex tones consisting of partials with unequal rise times and unequal physical onsets (Freedman 1967; Grey and Moorer 1977) and of tones with substantially differing amplitude envelopes (Strong and Clark 1967).

Synchronization and the conductor

Small ensembles usually perform without a conductor, while larger ensembles and orchestras are guided by the gestures of a conductor. Since one of the tasks of the conductor is to keep the performing musicians synchronized, it is interesting to comment upon the present or absence of a conductor in terms of our descriptive model of synchronization.

When there is no conductor, the musicians will have to synchronize upon each others' sound outputs. For this condition, we consider the standard deviation of the onset time differences as being of some fixed, constant magnitude. In that case, the standard deviation of the relative onset times varies as a function of the number of performers in such a way that it becomes larger when that number is increasing (since $s_v^2 = (½(n-1)/n)s_d^2$). If we use our mean value of the standard deviation of onset time differences (36 ms), then the standard deviation of relative onset times is 18 ms for an ensemble of two players, 20.8 ms for three players, 22 for four, 22.8 for five, 23.2 for six, 24 for nine. For ensembles with more than nine players it converges to its maximum value of 36/2½ or 25.5 ms. Evidently, synchronization quality is less in larger ensembles.

When a conductor is leading an ensemble or an orchestra, it may be supposed that the performers will try to synchronize not with each other, but with the movements of the conductor. For that condition, it may be supposed that the standard deviation of the relative onset times will be of some fixed, constant magnitude, since all musicians have the same task in this respect. We cannot extract an estimate for this quantity from our experiments, but via a detour a guess can be made. When the standard deviation of relative onset times is a constant quantity, the standard deviation of onset time differences is diminishing when a number of performers is increasing (since $s_d^2 = (2n/(n-1))s_v^2$), or, put the other way around, the standard deviation of onset time differences is increasing when the number of performers is increasing. With this relationship in mind, it is easy to understand that a large orchestra will always need a conductor, or, at least, will do better with a conductor. If the ensemble shrinks, however, synchronization quality is diminishing when the conductor stays in charge of the synchronization.

A survey of recorded chamber music with five to 15 performances showed that ensembles with up to and including nine performers did, more often than not, without a conductor, while ensembles with 10 or more players almost always had a conductor (see Fig. 4.9). This must mean that the standard deviation of relative onset times of an imaginary ensemble of 9.5 players marks, on the average, the maximum allowable value of that standard deviation (in our average case 24.1 ms). It may be assumed that in this imaginary situation synchronization is as good with conductor as without, which gives us at the same time the assumed constant value of the standard devation of relative onset times in a situation with conductor. When there are those 9.5 performers, the standard deviation of onset time differences (or the asynchronization) is just 36 ms. But when there were only two performers, asynchronization would have been 48 ms; with three, 42 ms; with four, 39 ms; etc. When the number of performers gets substantially larger than 10, asynchronization converges to its limit of 34.1 ms.

Our line of reasoning concerning the usefulness of a conductor when it comes to synchronization is summarized in Fig. 4.10. It amounts to the

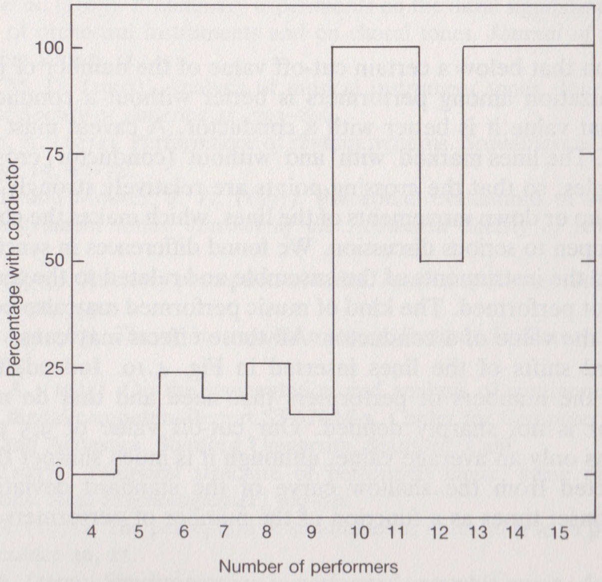

Fig. 4.9. Percentage of performances with a conductor, taken from a representative set of recorded quartets, quintets, sextets, etc. Only two compositions for 12 players could be located, of which one was with, one without conductor.

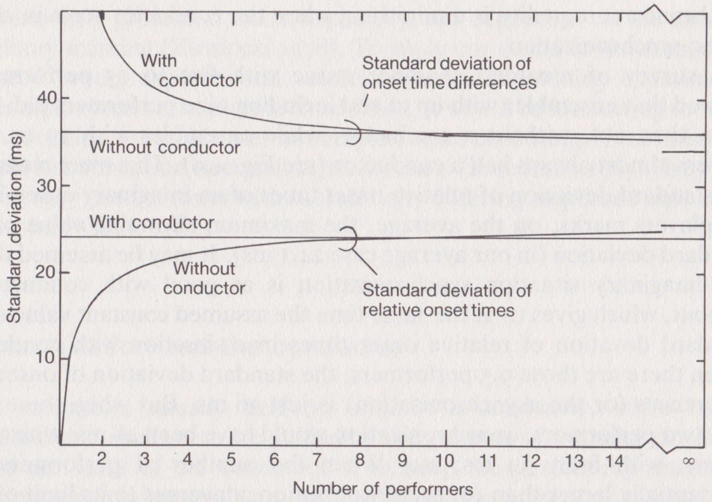

Fig. 4.10. Idealized standard deviations of relative onset times and onset times differences of performances with and without conductor, as a function of the number of performers. Numerical values are anchored to a mean asynchronization of 36 ms.

conclusion that below a certain cut-off value of the number of performers synchronization among performers is better without a conductor, while above that value it is better with a conductor. A caveat must be added, however. The lines marked 'with' and 'without' (conductor) cross at rather small angles, so that the crossing points are relatively strongly influenced by small up or down movements of the lines, which makes the choice of the crossing open to serious discussion. We found differences in synchronization related to the instruments of the ensemble and related to the tempo of the movement performed. The kind of music performed may also be expected to affect the value of a conductor. All these effects may cause upward or downward shifts of the lines inserted in Fig. 4.10. Indeed, the cut-off between the numbers of performers that need and that do not need a conductor is not sharply defined. Our cut-off value of 9.5 performers represents only an average value, although it is much sharper than was to be expected from the shallow curve of the standard deviation of the relative onset times as a function of the number of performers.

Conclusion

In this chapter, we have presented some thoughts and data about timing and synchronization in ensemble performance. The position taken in this chapter has to do with model making and acoustical measurement rather than with processes of the psychology of music. However, before any such psychological interpretation comes into play, a sound and reliable way of measuring the relevant characteristics of musical behaviour should exist, and that is what this chapter aims to provide.

Measures of musical behaviour, including musical performance, are essentially statistical measures. For this reason we developed our statistical model of synchronization, based on relations between variances of relative time moments. We showed that with this framework it is possible to describe in a sensible way synchronization in ensemble performance. The research we have done was not initiated by the wish to uncover the psychological processes that are behind synchronization behaviour. Eventually, however, the descriptive part should be supplemented by a psychological, interpretative part.

References

Berger, K. (1964). Some factors in the recognition of timbre. *Journal of the Acoustical Society of America* **36**, 1888.

Clark, M., *et al.* (1963). Preliminary experiments on the aural significance of parts on tones of orchestral instruments and on choral tones. *Journal of the Audio Engineering Society* **11**, 45.

Freedman, M. D. (1967). Analysis of musical instrument tones. *Journal of the Acoustical Society of America* **41**, 793–806.

Gabrielsson, A. (1974). Performance of rhythm patterns. *Scandinavian Journal of Psychology* **15**, 63–72.

Grey, J. M. and Moorer, J. A. (1977). Perceptual evaluations of synthesized musical instrument tones. *Journal of the Acoustical Society of America* **62**, 454–62.

Hirsh, I. J. (1959). Auditory perception of order. *Journal of the Acoustical Society of America* **31**, 759.

Michon, J. A. (1967). *Timing in temporal tracking*. Institute for Perception TNO, Soesterberg.

Moorer, J. A. (1975). On the segmentation and analysis of continuous musical sound by digital computer. Report STAN-M-3, Center for Computer Research in Music and Acoustics, Stanford University, Stanford, Calif.

Povel, D. (1977). Temporal structure of performed music, some preliminary observations. *Acta Psychologica* **41**, 309–20.

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

Rasch, R. A. (1979). Synchronization in performed ensemble music. *Acustica* **43**, 121–31.

Rasch, R. A. (1981). Aspects of the perception and performance of polyphonic music. Unpublished thesis. Groningen.

Saldanha, E. and Corso, J. (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America* **36**, 2021.

Seashore, C. E. (1938). *Psychology of music.* McGraw-Hill, New York. (Reissued in 1967 by Dover, New York).

Strong, W. and Clark, M. (1967). Synthesis of wind-instrument tones. *Journal of the Acoustical Society of America* **41**, 39–52.

Sundberg, J. and Verrillo, V. (1980). On the anatomy of the retard: a study of timing in music. *Journal of the Acoustical Society of America 68*, 772–9.

Tove, P. A., Ejdesjö, L., and Svärdström, A. (1967). Frequency and time analysis of polyphonic music. *Journal of the Acoustical Society of America* **41**, 1265.

Vos, J. and Rasch, R. (1981). The perceptual onset of musical tones. *Perception and Psychophysics* **29**, 323–35.

Zwislocki, J. J. (1978). Masking: experimental and theoretical aspects of simultaneous, forward, backward and central masking. In *Handbook of perception* (ed. E. Carterette and M. Friedman), Vol. 4. Academic Press, New York.

5

Rehearsal skill and musical competence: does practice make perfect?*

LINDA M. GRUSON

In recent years, research on the acquisition of skills has focused upon the study of a variety of complex skills including, for example, competence in games such as chess (de Groot 1965; Chase and Simon 1973; Simon and Chase 1973), bridge (Charness 1979) and the board game, Go (Reitman 1976); in the academic domain, acquisition of skills in such subjects as mathematics (Krutetskii 1976), physics (Simon and Simon 1978; 1980; Chi *et al.* 1980; Larkin *et al.* in press), compositional writing (Flower and Hayes 1979), and engineering thermodynamics (Bhaskar and Simon 1977); and a number of cognitive-motor skills such as typewriting (Thomas and Jones 1970; Shaffer 1975, 1978) and, most pertinent to the present study, musical skills (Bean 1938; Weaver 1943; Wolf 1976; Shaffer 1980, 1981; Sloboda 1974, 1977; MacKenzie *et al.* 1983).

While the research paradigms and target behaviours studied have been quite diverse, one of the most consistent findings that emerges from the research on skills is that as an individual learns a skill, he or she acquires the ability to process increasingly larger and more complex units of meaningful, skill-related information. Chase and Simon's (1973) work with chess players illustrates this process of 'chunking'. They have demonstrated that with increasing mastery, chess players are able to reconstruct significantly more pieces of meaningful chess positions exposed briefly to them. The reconstruction of randomly placed pieces did not improve with increased mastery. Chase and Simon hypothesized that chess masters' superiority over less expert players was due to their ability to encode

---

* This study was carried out in partial fulfilment of the Ph.D. degree. An earlier version of this paper was presented at the eighty-ninth Annual Convention of the American Psychological Association, Los Angeles, California, August 1981