S T E L L I N G E N

behorende bij het proefschrift van R.A. Rasch, Aspects of the perception and performance of polyphonic music.

1. De kwantitatieve, cijfermatige aanpak van een psychologische vraagstelling kan een veel te gunstig beeld geven van de zinvolheid en de nauwkeurigheid ervan.

2. Interpretaties van getalsmatige gegevens die slechts na statistische analyses mogelijk zijn, en die niet informeel (bijv. in grafiek of tabel) zichtbaar zijn, zijn slechts zelden van werkelijke betekenis.

3. Het begrip chroma (positie van een toon binnen een octaaf) hoort thuis in de musiektheorie, en niet in de muziekpsychologie.

4. Het 31-toonsstelsel is ongeschikt als stemmingssysteem voor zuivere intonatie, en wel in verband met de grote ontstemming van de kwint (meer dan 5 cents).

5. Het uitwerken van basso continuo uit de Barok-periode (1600-1750) is eerder een onderdeel van de contrapuntleer dan van de harmonieleer.

6. De functionele aanduiding van het dominant-kwartsext-accoord met de code $I-{}^{6}_{4}$ is onjuist en bovendien misleidend.

7. In de beschrijving van amplitudemodulaties kan de term modulatieniveau (te definiëren als $20 \log a_{max}/a_{min}$) goede diensten bewijzen.

8. Het gebruik van stemtoonhoogten anders dan de internationaal overeengekomen A4 (of a') = 440 Hz dient afgeraden te worden, ook in het kader van de historische uitvoeringspraktijk.

9. J.S. Bach heeft bij zijn Wohltemperiertes Clavier wel degelijk aan de evenredig-zwevende stemming gedacht.
Voor afwijkende meningen zie men bij: H. Kelletat, Zur musikalischen Temperatur, insbesondere bei Johann Sebastian Bach (Kassel 1960); H.A. Kellner, Eine Rekonstruktion der Wohltemperierte Stimmung von Johann Sebastian Bach (Das Musikinstrument 26, 1977, 34-35); J. Barnes, Bach's keyboard temperament: Internal evidence from the Well-tempered Clavier (Early Music 7, 1979, 236-249).

10. In tegenstelling tot veel gehoorde en gelezen meningen bestaat er voor de afgestudeerde musicoloog geen werkloosheidsprobleem.
Men zie bijv. de tegenstrijdige opmerkingen in de Utrechtse Universiteitsgids 1979-1980, p. 285.

R.A. Rasch

A S P E C T S

O F T H E

P E R C E P T I O N A N D P E R F O R M A N C E

O F

P O L Y P H O N I C M U S I C

RIJKSUNIVERSITEIT GRONINGEN

A S P E C T S
O F T H E
P E R C E P T I O N A N D P E R F O R M A N C E
O F
P O L Y P H O N I C M U S I C

Proefschrift
ter verkrijging van het doctoraat
in de Sociale Wetenschappen aan de
Rijksuniversiteit te Groningen op
gezag van de Rector Magnificus
Dr. M.R. van Gils in het openbaar te
verdedigen op donderdag 2 april 1981
des namiddags te 2.45 uur precies

door

RUDOLF ALEXANDER RASCH
geboren te Borger (Dr.)

Drukkerij Elinkwijk BV - Utrecht
1981

Promotoren

Prof.dr. J.A. Michon

Prof.dr.ir. R. Plomp

Coreferent

Prof.dr. J.J. van der Werff

P R E F A C E

The research described in this dissertation was carried out as part of the project "The perception of tones in polyphonic music", which was financially supported by a grant from the Netherlands Organization for the Advancement of Pure Research (Nederlandse Organisatie voor Zuiver-Wetenschappelijk Onderzoek ZWO) under nr. 15-29-03. Preparatory and supervisory work was done by the section for Music Perception (Muziekperceptie) of the Dutch Psychonomics Foundation (Nederlandse Stichting voor Psychonomie). The investigations were carried out at the Institute for Perception (Instituut voor Zintuigfysiologie) TNO in Soesterberg (Netherlands) in the period from February 1, 1975 until July 1, 1978, under the direct supervision of prof.dr.ir. R. Plomp, head of the Audiology Division of the Institute for Perception TNO. I would like to express my thanks to the Institute for Perception TNO for making available to me the many facilities of the Institute, like its electronic equipment, experimental rooms, computers, technical, secretarial, and library services. My thanks also go to the staff personnel of the Institute, scientific, technical, and secretarial. Finally, I want to thank Ruth Drijfhout for her assistance in typing and further preparing the final version of this dissertation.

Some parts of this dissertation have been published in earlier versions as articles in journals (Chapter Two: Rasch 1978a; Chapter Four: Rasch 1979). The same parts have been presented at scientific meetings, both in the Netherlands and abroad, in 1976 and 1977. A short version of Chapter Four has been published in Dutch (Rasch 1978b,c).

Rudolf Rasch

TABLE OF CONTENTS

Preface 5

Table of contents 6

Chapter One - INTRODUCTION

1.1. The perception of tones in polyphonic music 9

1.2. Polyphonic music as an auditory stimulus 10
1.2.1. Introduction 10
1.2.2. The spectral approach 10
1.2.3. The temporal approach 12
1.2.4. Integrating the two approaches 13

1.3. Outline of the following chapters 14
1.3.1. Introduction 14
1.3.2. Psychophysical experiments 14
1.3.3. Acoustical measurements 15
1.3.4. Theory and practice of music 16

Chapter Two - THRESHOLD EXPERIMENTS

2.1. Introduction 16

2.2. Method 16
2.2.1. Introduction 16
2.2.2. Apparatus 17
2.2.3. Stimuli, trials, and series 17
2.2.4. Subjects 21
2.2.5. Measurement procedure 21

2.3. Harmonic coincidence 22
2.3.1. Introduction 22
2.3.2. Stimuli 23
2.3.3. Results 24

2.4. Synchronous versus asynchronous masker tones 28
2.4.1. Introduction 28
2.4.2. Stimuli 29
2.4.3. Results and discussion 30

2.5. Simultaneous versus continuous masker tones 32
2.5.1. Introduction 32
2.5.2. Stimuli 32
2.5.3. Results 33

2.6. Deviating frequency ratios 34
2.6.1. Frequency modulation 34
2.6.2. Mistuned intervals 35

2.7. Conclusions and discussion 36

Chapter Three - RECOGNITION EXPERIMENTS

3.1. Introduction 38

3.2. Method 38
3.2.1. Apparatus 38
3.2.2. Stimuli, trials, and series 40
3.2.3. Subjects 41
3.2.4. Measurement procedures 41

3.3. Results 43
3.3.1. General 43
3.3.2. Intervals 43
3.3.3. Onset time difference 43
3.3.4. Level difference 45
3.3.5. Onset time difference versus level difference 46
3.4. Conclusions and discussion 47

Chapter Four - THE MEASUREMENT OF ASYNCHRONIZATION

4.1. Introduction 50
4.2. The statistics of asynchronization 50
4.2.1. Introduction 50
4.2.2. Absolute and relative onset times 51
4.2.3. Onset time difference 52
4.2.4. Asynchronization of voice pairs 54
4.2.5. The asynchronization of a piece of music 55
4.2.6. Isochronization 56
4.3. Method 58
4.3.1. Recording 58
4.3.2. Signal analysis 59
4.3.3. Peak detection 60
4.3.4. Onset time detection 61
4.3.5. Statistical computations 63
4.4. Musicians and music 64
4.5. Results 66
4.5.1. Relative onset times 66
4.5.2. Onset time differences: asynchronization 67
4.5.3. Asynchronization and tempo 69
4.6. Relations with other research 69
4.6.1. Isochronization 69
4.6.2. Synchronization of piano chords 70
4.7. Conclusion 71

Chapter Five - OBSERVATIONS FROM MUSICAL PERFORMANCE AND COMPOSITION

5.1. Polyphonic music in performance 72
5.1.1. Spectral factors 72
5.1.2. Temporal factors 74
5.2. Asynchronization in polyphonic composition 78
5.2.1. Counterpoint 78
5.2.2. The theory of macro-asynchronization 80
5.2.3. Macro-asynchronization in polyphonic composition 82

Chapter Six - CONCLUSIONS 88

Summary 90

Samenvatting 93

References 96

Curriculum vitae 99

Ars non habet osorem nisi ignorantem

CHAPTER ONE

INTRODUCTION

1.1. The perception of tones in polyphonic music

Experience and observation tell us that human listeners are often capable of separately perceiving two simultaneous sounds from different sources. Examples are the ability to understand speech against a background of noise and also to follow individual parts or voices in polyphonic music (in the widest sense of the word). This ability cannot be taken for granted. The various simultaneous signals are superimposed on each other on their way to our ears. Our hearing system obviously performs some kind of analysis to make possible the perception of the original sound stimuli. The issue is particularly evident in music where there is more than one melodic line. This is the case in music that is traditionally called polyphonic (mostly composed before 1750), but often also in symphonic music and ensemble chamber music (mostly composed after 1750).

The harmonics of single tones fuse into a single perceptual image as does the doubling of melodic lines by the same or other instruments, either in unison or in octaves. The simultaneous use of organ stops results in a single sound impression. To a lesser extent this is also true in homophonic music such as harmonized melodies. Only one melodic line is clearly perceived. In polyphonic music, however, the simultaneous components of the auditory stimulus do not fuse into one single melodic line but retain their own musical significance.

Until now, most research in musical acoustics and perception has been devoted to problems concerning single tones and single lines of melody. Polyphonic music poses a great number of questions concerning performance and perception that cannot be answered from research on single lines and tones but need to be investigated separately.

We shall tackle the problem of the perception of polyphonic music within frameworks derived from several fields of science. The production and perception of music is a complicated process. If an aspect of this process is investigated, a multidisciplinary approach may be most fruitful. The present study includes chapters that deal with the psychophysical, acoustical, and theoretical aspects of polyphonic music. Of course, no complete description of any of these aspects can be given. However, the presence of these varied approaches creates opportunities for many interesting interpretations. For instance, it can be interesting to see how important a certain acoustical variable is in the theory of music or in psychophysical theory.

In this study it is tried to shed some light on the human capability of perceiving simultaneous lines separately, by investigating some major factors within the various mentioned fields. During the course of these investigations it was decided to concentrate on the aspects of asynchronization of simultaneous tones. This aspect turned out to be an important variable in psychophysical experiments. Its practical significance could be established by measurements of performed music. Connections were made with the theory of the composition and the performance of polyphonic music.

1.2. Polyphonic music as an auditory stimulus

1.2.1. Introduction

Sound stimuli can be viewed from a spectral as well as from a temporal point of view. Spectral information is usually represented by means of a frequency spectrum. Temporal information is primarily embodied in the waveform of the sound. The hearing system is able to use both spectral and temporal information when analysing sound stimuli, as can be seen, e.g., in the perception of speech. In this study temporal information will be associated with the temporal envelopes of tones and not with the temporal fine-structure of the waveforms. It can be represented in diagrams with time and sound-pressure level as horizontal and vertical dimensions, respectively. That is the sense in which we will use the terms 'spectral approach' and 'temporal approach'. In practical situations the two approaches can never be completely separated from each other. Conceptually, however, they must be clearly distinguished. In the design of experiments, it is necessary to be aware of the features of either approach. We will, in the context of this study, interpret these approaches with respect to the perception of polyphonic music. In order to use polyphonic music as an auditory stimulus in psychophysical experiments, certain simplifications must be made. For experimental purposes, we will use pairs of simultaneous complex tones as first approximations of the stimulus conditions present in polyphonic music.

1.2.2. The spectral approach

The spectral approach is illustrated in Fig. 1. The stimulus consists of the superposition of two series of harmonics. The harmonics are not

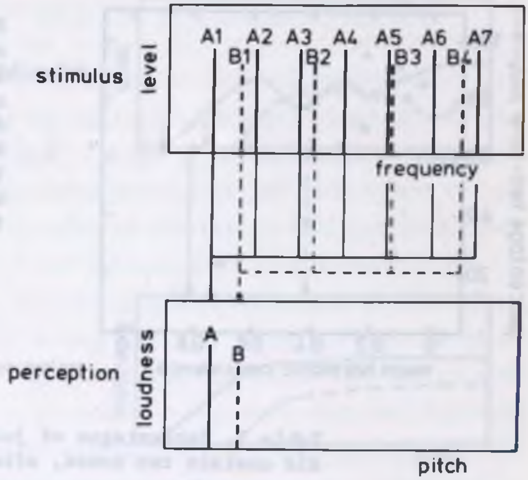


Fig. 1. Schematic diagram of the "spectral approach" to the separate perception of two simultaneous tones. The stimulus consists of two tones A and B, each with many harmonics. In perception, the two harmonic patterns are recognized as such, and give rise to the perception of two distinctive tones with pitches corresponding to the fundamental frequencies of the stimulus tones.

perceived as single entities, but the ones belonging to each tone are grouped together in perception by a process that may be considered to be a form of pattern recognition. Each of the series of harmonics give rise to the perception of a distinctive tone.

In the history of perception research the spectral approach can be traced back to Carl Stumpf, who describes experiments on the perception of two-tone complexes in the second volume of his Tonpsychologie (1890). Stumpf produced his stimuli by playing tones on piano or organ. His main concern was the subjective fusion (Verschmelzung) of musical tones. He asked his subjects whether they heard one or two tones. The percentage of two-tone judgment was his measure of fusion. He related these results to the rather speculative concept of specific perceptual energies. We will relate his fusion measures to our concept of "harmonic coincidence", the percentage of harmonics of a tone that coincide with harmonics from another tone, which will be developed quantitatively in section 2.3. The results of Stumpf's experiments interpreted along this line can be found in Table I and Fig. 2. They show that high coincidence corresponds with high fusion (low percentage of two-tone perception) and vice versa. Recently, there has been research on the perceptual separation of simultaneous harmonic sounds with different fundamental frequencies in the context of speech perception (Darwin 1979; Scheffers 1979).

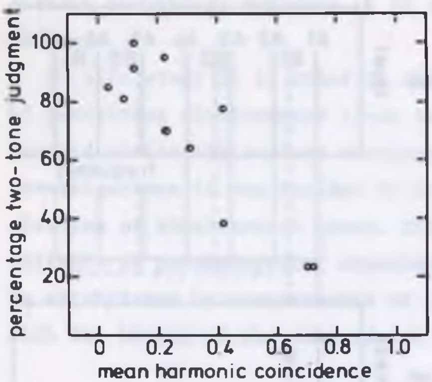


Fig. 2. Percentage of correct judgement of the presence of two tones in a two-tone stimulus, as a function of mean harmonic coincidence. After data from Stumpf (1890). The mean harmonic coincidence is the geometrical mean of the harmonic coincidences of the lower and the higher tone of the interval (see 2.3.1.).

Table I. Percentages of judgement that an interval did contain two tones, after Stumpf (1890).

Interval	p.145	p.148
octave	24	24
fifth	78	38
fourth		64
major third	95	70
major second	100	91
minor seventh		81
tritone		85

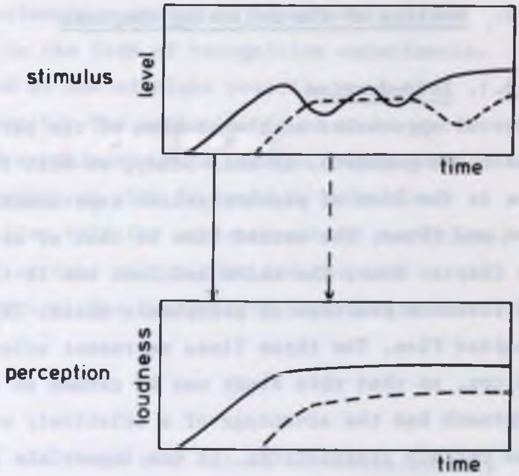
1.2.3. The temporal approach

The temporal approach is illustrated in Fig. 3. The stimulus is viewed as consisting of two tones with different temporal structures as represented by the temporal envelopes. Frequency components with identical temporal structures are grouped together and considered to originate from one tone. The approach may be seen as an example of the Gestalt law of common fate. It can account for the separation of simultaneous tones. Their characterization is dependent on the spectral composition of the tones as well.

Helmholtz (1885, p. 59) already recognized the role of temporal information in the sound stimulus with respect to the separate perception of simultaneous tones:

"Now there are many circumstances which assist us in separating the musical tones arising from different sources, and secondly, in keeping together the partial tones of each separate source. Thus when one musical tone is heard for some time before being joined by the second, and then the second continues after the first has ceased, the separation in sound is facilitated by the succession of time. We have already heard the first musical tone by itself, and hence know immediately

Fig. 3. Schematic diagram of the temporal approach to the perception of two simultaneous tones. The tones are distinguished by their differing temporal envelopes.



> what we have to deduct from the compound effect for the effect of this first tone. Even when several parts proceed in the same rhythm in polyphonic music, the mode in which the tones of different instruments and voices commence, the nature of their increase in force, the certainty with which they are held, and the manner in which they die off, are generally slightly different for each. (...) But besides this, in good part music, especial care is taken to facilitate the separation of parts by the ear. In polyphonic music proper, where each part has its own distinct melody, a principal means of clearly separating the progression of each part has always consisted in making them proceed in different rhythms and on different divisions of the bars."

Recently, this approach has been followed in experiments on auditory stream formation and segregation by Van Noorden (1975, 1977) and by Bregman and co-workers (Bregman & Pinker 1978; Dannenbring & Bregman 1978).

1.2.4. Integrating the two approaches

The two approaches should be seen as complementary and mutually supporting, not as excluding each other. The relative contributions to perception of spectral and temporal information depend on stimulus characteristics. In steady-state sound temporal differentiation is lacking, so that our hearing system can only use spectral cues. Temporal cues predominate in situations with sounds that are richly differentiated from a temporal point of view. Performed music as a stimulus for the hearing system shows differentiation both in the frequency and in the time domain. The steady-state parts of musical tones provide clear spectral patterns. The sequence of tones in time and the temporal characteristics of the individual tones provide a continuous stream of temporal differentiation.

1.3. Outline of the following chapters

1.3.1. Introduction

Several approaches to the problem of the perception of tones in polyphonic music are possible. In this study, we will follow three lines. The first one is the line of psychophysical experimentation, described in Chapters Two and Three. The second line is that of acoustical measurement, followed in Chapter Four. The third and last one is that of musical theory and performance practice of polyphonic music. This aspect is dealt with in Chapter Five. The three lines represent scientific fields of different nature, so that this study may be termed an interdisciplinary study. This approach has the advantage of a relatively wide overview, but it also has certain restrictions. It was impossible to explore all different sides of the problem extensively during the limited time available for the investigations. Instead, important topics, drawn from these three fields, were worked out in greater detail. The choice of the topics was determined by their expected significance for the problem, by the possibility of experimentation and by the possibilities of mutual connections. The topics investigated focus upon the phenomenon of asynchronization in polyphonic music.

1.3.2. Psychophysical experiments

Psychophysical experiments were done with pairs of simultaneous tones as stimuli. We chose two paradigms, leading to threshold experiments and recognition experiments. In the threshold experiments, the pitch threshold of a tone (the test tone) in a two-tone stimulus was to be determined. The other tone (the masker tone) was kept at a constant sound-pressure level. The threshold of the test tone expressed in dB relative to the level of the masker was used as a measure of the separate perception of the test tones. This measure was easy to obtain and had good reliability. It was evaluated in the sense that conditions with low thresholds were considered to represent conditions with good possibilities for separate perception (viewed as an essential feature of the perception of polyphonic music), while conditions with high thresholds were considered to represent conditions with poor separate perception. The use of thresholds as measures of perception has certain restrictions. It cannot be guaranteed that conclusions derived from them also hold in supra-threshold conditions. However, they are very useful as a first approximation. The threshold experiments are described in Chapter Two.

In addition to the threshold experiments a series of experiments under supra-threshold conditions was run, in the form of recognition experiments. Two-tone stimuli had to be recognized as one of eight possible musical intervals (major second to minor seventh). The variables found to be the most important ones in the threshold experiments were used as independent variables in the recognition experiments, so that it could be determined whether the "threshold conclusions" also hold for supra-threshold conditions. In these experiments, percentages of correct recognition and response times were used as measures of perception. The recognition experiments are described in Chapter Three.

1.3.3. Acoustical measurements

In the psychophysical experiments it was found that minute asynchronizations of the simultaneous tones played an important part in their perception. Therefore, is was decided to focus the acoustical measurements of performed music on this aspect. Special recordings were made of ensemble performances so that each voice of the ensemble was accessible on a separate tape channel. This enabled us to measure the degree of asynchronization present in the performances. It turned out that the amount of asynchronization was about the same as the asynchronization used in the psychophysical experiments in the form of onset time differences. This opened the way to extending the interpretation of the psychophysical experiments from laboratory to practical musical conditions. The acoustical measurements are described in Chapter Four.

1.3.4. Theory and performance of music

The problem of the perception of polyphonic music had been handled by composers and performers many centuries before the beginning of scientific research. This experience has led to a large amount of practical knowledge to be found in theoretical writings on the composition and performance of polyphonic music. To give a complete summary of this pre-scientific knowledge is not an easy task. We chose from these fields of information a number of topics that refer to the aspects already dealt with from an experimental point of view. Special emphasis was put on the role of deliberate asynchronization in performance and in the composition of polyphonic music. These aspects of the theory of music are described in Chapter Five.

Finally, Chapter Six brings all the lines followed in the preceding chapters together in a short concluding overview, emphasizing again the role of asynchronization in polyphonic music.

CHAPTER TWO

THRESHOLD EXPERIMENTS

2.1. Introduction

In this chapter a number of threshold experiments will be described in detail. They aim to investigate the conditions under which a tone can be perceived in the presence of another tone. The determined thresholds are masked thresholds. We reasoned that a condition resulting in a low threshold would be a favorable condition for the separate perception of simultaneous tones and that such conditions would be very practical in the performing situation of polyphonic music.

Two series of threshold experiments were run, series A and B. Experimental conditions were replicated but for some minor differences. Series A concentrated on harmonic coincidence and onset time difference as the most important variables, while in series B special attention was paid to the influence of different types of masker tones. The following description of the experimental method applies to series A. Deviations from it in series B are mentioned in parentheses. Since there were differences in apparatus (e.g., another computer), the results of both series of experiments have not been combined. If the results concern the same variables, they are presented separately.

2.2. Method

2.2.1. Introduction

In the threshold experiments trials containing two successive stimuli were used. Each stimulus included a test tone and a masker tone. The masker tone was either simultaneous with the test tone or sounded continuously. The masker tone had the same frequency in both stimuli of a trial. This frequency was constant throughout a series. The test tone, however, jumped in frequency, up or down, using the same two frequencies for all trials of a series. Subjects were asked to indicate whether the pitch of the test tone jumped upward or downward. The level of the test tones relative to that of the masker tones resulting in 75% correct responses was considered to be the threshold for the test tones, and, provisionally, the threshold for the separate perception of the two tones in the experimental condition concerned. With this experimental set-up the influence of many variables could be investigated in a convenient way. The procedure was easily handled by the subjects and resulted in reliable and reproduceable data.

The choice of the pitch jumps as the threshold criterion rather than mere detection was based on the consideration that we were interested in the perception of tones in a pseudo-musical context, not in the detection of whatever there may be in the stimulus in addition to the masker tone.

2.2.2. Apparatus

A flow diagram of the apparatus used is shown in Fig. 4. The experiments were run on-line under the control of a PDP-7 (B: PDP-11/10) computer. The two simultaneous tones were generated as two separate continuous signals. Waveforms were stored in two external revolving memories (recirculators) as 128 (B: 256) discrete samples with 10-bit accuracy. These memories were read out by digital-to-analogue converters with a sample rate determined by a pulse train derived from a programmable frequency generator. Phase-locking of the two signals was possible by connecting both recirculators (at least one of them with more than one period of the waveform) to the same frequency generator. The signals were filtered by a low-pass filter with a cut-off frequency of 4 (B: 5) kHz, in order to remove sample frequencies and to equalize the spectral range of the tones. The signals for the test tones, for which the threshold was to be determined, passed through a programmable attenuator. The temporal structures of the tones were determined by on/off gates with half-sinusoidal rise and decay amplitude envelopes. Rise and decay times were set manually. Reported rise and decay times concern the trajectory from 10 to 90% of the final amplitude and vice versa. Switching the gates on and off was achieved by pulses derived from the computer clock (B: from a programmable timer). After gating, the tones were mixed. The sounds were presented to the subjects through headphones, binaurally, in a soundproof room. Responses were made by the subjects by pushing one of two push-buttons of a response box. Responses were processed immediately by the computer.

2.2.3. Stimuli, trials, and series

The basic trial form is illustrated in Fig. 5, using musical notation. A trial consisted of two successive stimuli, each stimulus including a test tone and a masker tone which were presented simultaneously. The frequency of the masker tones in the two stimuli of a trial was the same.

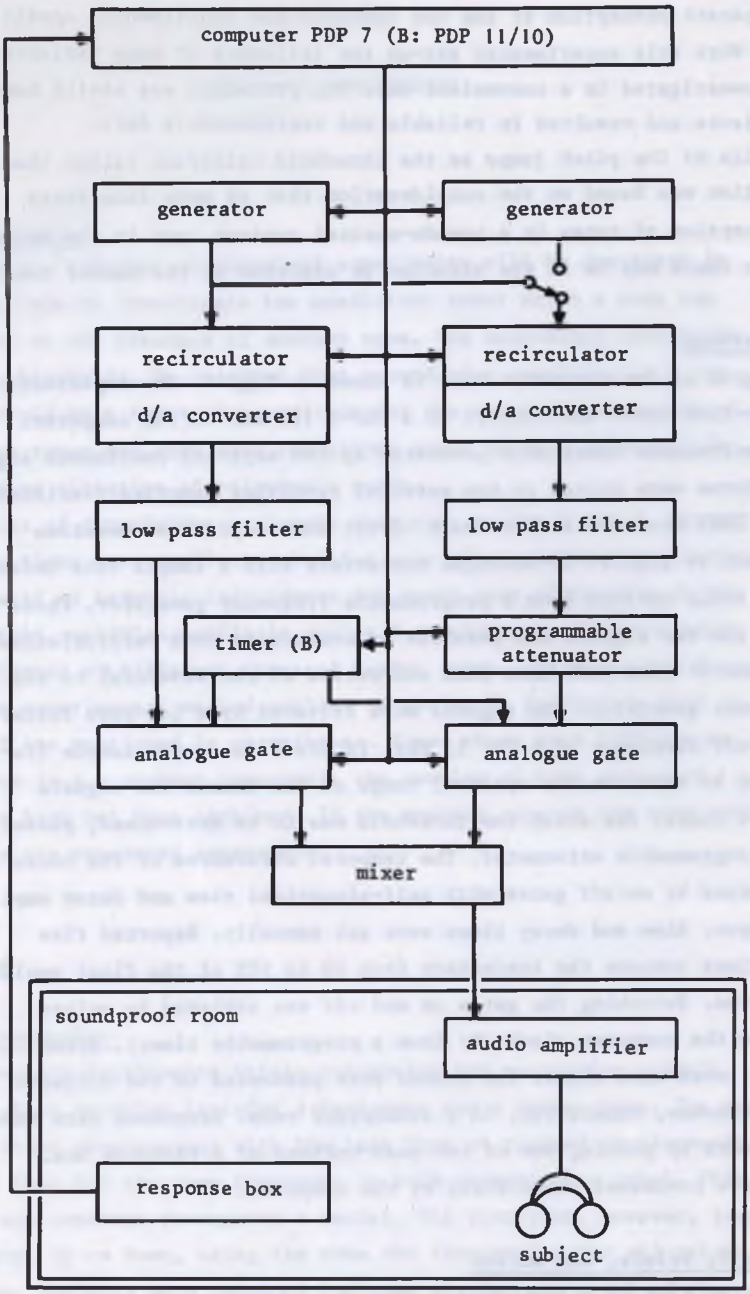


Fig. 4. Basic set-up of the apparatus used in the threshold experiments. The sample frequency of the test-tone recirculator was derived either from the generator of the masker tone (phase-locked generation) or from its own generator (random-phase generation).

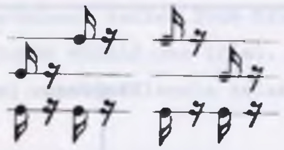


Fig. 5. Musical representation of the two stimulus configurations used in the threshold experiments. Each trial consists of two two-tone stimuli. The masker tone has the same frequency in both stimuli. The test tones have different frequencies, resulting in a pitch jump, either upward (low-high trial) or downward (high-low trial). The lower test tone has a fundamental frequency of about 500 Hz. Tone duration was 200 (B: 250) msec; the silent interval between the two tones of a stimulus was equal to tone duration. This corresponds to a metronome value of 150 (B: 120). A diagram like this figure was shown to the subjects as a part of the experimental instruction.

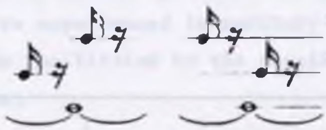


Fig. 6. Musical representation of the trials with a continuous masker tone. For details see the legend of Fig. 5.

In some experiments the masker was a continuously sounding tone. This trial form is illustrated in Fig. 6. The frequencies of the test tones in the two stimuli of a trial were different, so that, perceptually, the pitch of the test tone jumped upward or downward against the background of stationary masker tones.

Waveforms of the tones were calculated with the formula

$$p(t) = \sum_{n=1}^{n=20} \frac{1}{n} \sin (2\pi nft + \phi) \qquad (1)$$

which results in harmonic spectra with an envelope slope of -6 dB/octave. Phases employed were usually $\phi=0$ (sine addition) for the masker tones and $\phi=\pi/2$ (cosine addition) for the test tones (B: also $\phi=0$ for the test tones). The perpendicular phase relation between phase-locked masker and test tones in series A was chosen since it represents the mean phase relation of conditions with random phases. Fundamental frequencies were chosen in such a way that the lower test tone was about 500 Hz. The frequencies of the masker tones were within the range of 250 to 420 Hz, those of the higher test tones within 672 to 835 Hz. Within one series, three frequencies

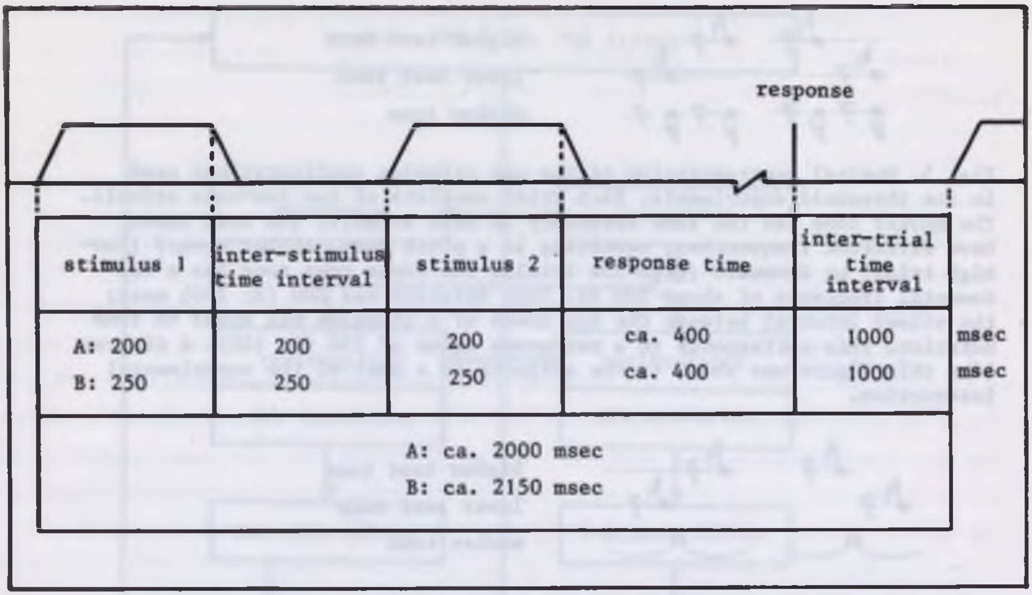


Fig. 7. Temporal structure of the trials in the threshold experiments. The successive time slices are not to scale. During the response time a red light bulb on the response box lit up in order to inform the subject that a response was expected.

were used: one for the masker tones and two for the test tones.

The durations of the tones and the inter-stimulus time intervals were 200 (B: 250) msec. Rise and decay times were 20 msec. After the decay of the second stimulus of a trial the subject could respond by pressing one of two push-buttons. If he heard a downward pitch jump, he had to press down the left-hand button; if an upward jump, the right-hand button. There was no time-restriction for responding. After responding a pause of 1 second followed before the next trial. With a mean response time of 400 msec (estimated) each trial lasted about 2 sec (B: 2150 msec), on the average. Series included 180 (B: 100) trials, lasting about 6 min (B: 3.5 min).

In the experiments of series A the masker tones were always simultaneous with the test tones, either synchronous (on and off at the same moments) or asynchronous (with delayed onset of the masker tone). In series B experiments were also run with the masker sounding as a continuous tone.

The masker was kept at a constant sound pressure level of about 70 dB SPL (re 20 µPa). The levels of the test tones, for which the threshold was to be determined, varied from trial to trial, but was the same for the two test tones within one trial. All levels of test tones are expressed as spectral envelope levels relative to the spectral envelope of the masker tone.

2.2.4. Subjects

The subjects were music students (conservatory or musicology) and musically competent amateurs. They were paid for their services. The author also participated as a subject in a number of experiments. Since the experiments aimed at studying a musical-perceptual competence, the choice of music students or at least musically trained persons as subjects was a deliberate one. It was necessary to have subjects with a musical background. Moreover, they were experienced in auditory tasks. Communication about the experimental task was facilitated by the possibility of using musical terminology and notation.

2.2.5. Measurement procedure

The influence of various experimental conditions on the separate perception of the two tones in the stimuli was evaluated by their effect on the pitch threshold of the test tones in the stimuli. The threshold measurement procedure was uniform in all experiments. In an experimental series the variable under investigation was set at a certain value. The level of the test tones was varied randomly from trial to trial over a range of 25 dB covered in steps of 5 dB. This corresponds to the well-known method of constant stimuli. The direction of frequency jumps in the trials (up or down) was also random. The random alternation of supra- and sub-threshold test tones in a series keeps the attention of the subjects directed towards the perception of tones, and not of mere sounds.

The subjects were informed about the structure of the stimuli and trials by showing them Fig. 5. A two-alternative forced choice (2AFC) procedure was used in which the subjects were asked to indicate whether they perceived an upward or a downward pitch jump. This procedure leads to an expectation of 100% correct responses for unambiguously perceived supra-threshold test tones, and 50% (i.e. chance performance) when the test tones are not perceived at all. The subjects had to guess when they could not decide about the direction of the pitch jump from perception. The level of the test tones that resulted in 75% correct responses was considered as the threshold. As a rule, the threshold was determined by linear interpolation between the nearest supra- and sub-threshold levels used in the series.

If there was more than one subject or series per subject, the threshold values of all series were averaged in order to arrive at a single value. Standard deviations turned out to be in the order of one or a few dB (see Fig. 16 for typical values).

The results of the experiments are presented in diagrams with the variable under investigation as the horizontal axis and the level of the tones as the vertical axis. The curves are threshold curves for the perception of the frequency jump. The subjects and the number of series per subject (if more than one) for each data point are given in the legends of the graphs.

From a psychoacoustical point of view the experiments are masked threshold determinations. They differ from detection experiments by the presence of a test tone in both stimuli and by the pitch jump as criterion for the subjects. Some of the experimental conditions were also investigated with detection paradigm. The subject had to decide whether the test tone was in the first or the second stimulus. The outcomes of these experiments were very similar to those of the masked threshold determinations with the pitch-jump criterion.

2.3. Harmonic coincidence

2.3.1. Introduction

When temporal differentiation of two tones in a stimulus is absent, the hearing system can only make use of spectral information in order to separate the two tones. The use of spectral information was investigated with the help of stimuli containing tones with identical temporal descriptions, that is, with identical onset times, durations, and rise and decay times. Such stimuli are called synchronous stimuli. Fig. 8 illustrates the temporal structure of a synchronous stimulus.

It is to be expected that the extent of coincidence of the harmonics of the two tones in the stimulus is an important factor in whether or not they are perceived separately. If two complex tones sound simultaneously, harmonics of both tones can coincide in frequency, depending on the ratio of their fundamental frequencies. We will denote the coincidence of harmonics as harmonic coincidence and consider it to be a feature of tones in harmonic musical intervals. The coinciding harmonics of tones are vectorially added. We will call this vector sum the common harmonic of the tones. Harmonic coincidence can be defined quantitatively as the fraction of the harmonics of a tone that coincides with the harmonics of one or more other tones.

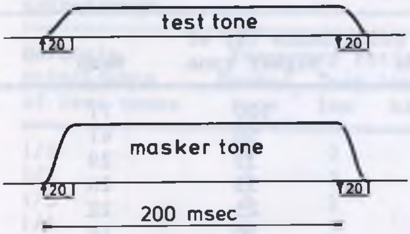


Fig. 8. Temporal structure of a synchronous stimulus, as used in the threshold experiments. The two tones had equal duration, equal rise and decay time and sounded simultaneously.

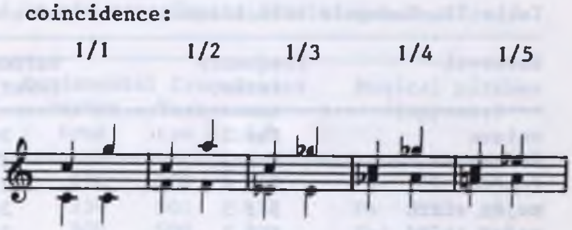


Fig. 9. Approximate musical pitches of the tones used in the stimuli of the threshold experiments. The approximation is based on C5 = 500 Hz.

In the case of two tones, with frequencies of f_a and f_b, the harmonic coincidence C_a is equal to the frequency of the tone divided by the least common multiple of the frequencies of the two tones (denoted $M(f_a,f_b)$):

$$C_a = f_a/M(f_a,f_b). \qquad (2)$$

The calculation of the harmonic coincidence can also be done with the frequency ratio instead of the frequencies themselves. When working with a simple frequency ratio, $f_a : f_b = p : q$, the formula is simplified to:

$$C_a = \frac{1}{q} \text{ and } C_b = \frac{1}{p}. \qquad (3)$$

A coincidence of 1/1 is called complete coincidence.

The harmonic coincidence of a musical interval can be defined as the geometric mean of the harmonic coincidences of the individual tones in the interval:

$$C_{ab} = (C_a C_b)^{\frac{1}{2}} = (pq)^{-\frac{1}{2}}. \qquad (4)$$

Harmonic coincidence is an important musical variable since musical intervals are defined by certain simple frequency ratios.

2.3.2. <u>Stimuli</u>

By choosing the right frequency ratio, stimuli can be constructed with a certain harmonic coincidence of the test tones. In two-tone stimuli the harmonic coincidence of the test tone can only take values that can be described as 1/n, n being an integer number. We have constructed stimuli with a coincidence of 1/1, 1/2, 1/3, 1/4 and 1/5 for the test tones.

Table II. Harmonic coincidences of the most common consonant musical intervals.

Interval	Frequency ratio	Harmonic coincidence (%) of lower tone	higher tone	Mean
octave	1 : 2	50	100	71
fifth	2 : 3	33	50	41
fourth	3 : 4	25	33	29
major sixth	3 : 5	20	33	26
major third	4 : 5	20	25	22
minor third	5 : 6	16	20	18
minor sixth	5 : 8	13	20	16

The two test tones within one series (with different frequencies) had the same harmonic coincidence. The frequencies were chosen in such a way that the lower test tone had a frequency of about 500 Hz. The frequencies of the tones in the various stimuli are listed in Table III. They are illustrated with approximate musical pitches in musical notation in Fig. 9.

As to the phase relations between masker and test tones, two different conditions were used: random phase and controlled phase. The masker tones were generated by sine addition, and test tones by sine (A, B) or cosine (A) addition. In series A experiments were also done with the phases of the harmonics of the test tones varying from zero to 90^{o}, but only with complete harmonic coincidence. Control of the phase relation between coinciding harmonics is important because this relation can considerably affect the amplitude of the common harmonics.

If the two tones of the stimulus are generated with phase-locking, and amplitudes and phases of the harmonics of the tones are known, the amplitudes and phases of the common harmonics can be calculated. The coinciding harmonics from masker and test tones are given by sin $2\pi ft$ and b sin $(2\pi ft + \beta)$ respectively, b being the relative amplitude of the test tone and β the phase of the test tone's harmonics. Amplitude c and phase γ of the common harmonics are given by:

$$c = (b^2 + 2b\cos\beta + 1)^{\frac{1}{2}} \quad (5)$$

$$\gamma = \arcsin(\frac{b}{c}\sin\beta). \quad (6)$$

Values of β used were 0, 30, 60 and 90^{o}. Values of c and γ under these conditions are given in Table IV.

2.3.3. Results

The results of the experiments on the effects of harmonic coincidence on the pitch-jump threshold are illustrated in Fig. 10 (effect of harmonic

Table III. Frequencies of tones used in the experiments involving harmonic coincidence

Harmonic coincidence of test tones	Frequency ratios Masker tone	Test tones low	Test tones high	Fundamental frequencies Masker tone	Test tones low	Test tones high	Musical pitches (approx.)		
1/1	1	2	3	250	500	750	C4	C5	G5
1/2	2	3	5	334	501	835	F4	C5	A5
1/3	3	5	7	300	500	700	E♭4	C5	G♭5
1/4	4	5	7	400	500	700	A♭4	C5	G♭5
1/5	5	6	7	420	504	588	A4	C5	E♭5

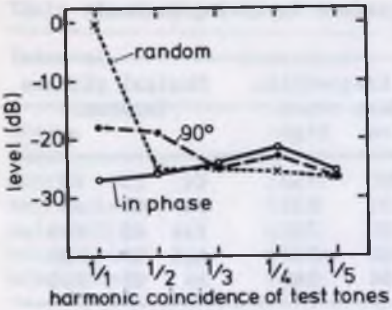
Table IV. Relative amplitudes and phases of the common harmonics when two tones are generated phase-locked for several phase differences between the coinciding harmonics. The amplitude of the masker tone is unity, b is the amplitude of the test tone. The phases of the single harmonics of the masker tones are zero.

Phase difference between coinciding harmonics (degrees)	Relative amplitude of common harmonics	Phase of common harmonics
0	$b + 1$	0
30	$(b^2 + 3^{\frac{1}{2}}b + 1)^{\frac{1}{2}}$	arcsin $(\frac{1}{2}\frac{b}{c})$
60	$(b^2 + b + 1)^{\frac{1}{2}}$	arcsin $(3 - \frac{1}{2}\frac{b}{c})$
90	$(b2 + 1)^{\frac{1}{2}}$	arcsin $(\frac{b}{c})$

coincidence under various phase conditions, A), Fig. 11 (idem, B) and Fig. 12 (effect of phase under complete coincidence). They can be summarized as follows.

If the harmonic coincidence is incomplete, i.e. 1/2 or less, neither the exact amount of coincidence, nor the phase relations between the coinciding harmonics play a role of any importance regarding the threshold of the test tones. Thresholds of -20 to -27 dB were found for these conditions.

If the harmonic coincidence is complete (1/1) the phase relations between the coinciding harmonics do matter. With random phase differences a threshold of about 0 dB was found. When the phase differences between coinciding harmonics were fixed and not greater than 90^o (the amplitudes of the common harmonics are then not smaller than those of the single harmonics), thresholds were found in the same range as for partial coincidence, i.e. about -20 to -25 dB. With complete coincidence, there are no single harmonics of the test tones. The perception of the test tones is based on the level differences between the common harmonics and the masker tone's single components. These level differences are dependent on the phase



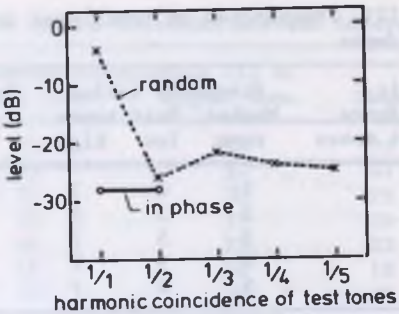


Fig. 10. Pitch-jump threshold of the test tones in synchronous stimuli as a function of harmonic coincidence. The fundamental frequencies used are listed in Table III. Parameter is the phase relation between the coinciding harmonics of masker and test tones. Data-points are means of four subjects: AB, PS, RR, RS (A).

Fig. 11. Threshold of the test tones in synchronous stimuli as a function of harmonic coincidence. For further explanation see Fig. 10. Subjects: JPK, JV, RR (B).

differences between coinciding harmonics. Therefore, the latter variable has a considerable influence on the thresholds under conditions with complete coincidence. With random phase differences, some common components are stronger, other ones are weaker than the single components of the masker tone. Only the stronger components contribute to the perception of the test tone. This results in the relatively high threshold for this condition (0 dB). The constantly varying levels of the common components in a stimulus cause perceivable differences in the timbre of the test tones.

With controlled and equal phase differences between the coinciding harmonics, each common harmonic contributes to the perception of the test tones. The thresholds found can be interpreted as difference limina of the common and single harmonics. The smaller the phase differences between the coinciding harmonics, the lower the level of the test tones required for the common harmonics to reach the threshold level. For this reason it can be expected that the threshold is lower for an in-phase generation of the coinciding harmonics than for a generation with a 90° phase difference. The phase-dependence of the threshold was investigated in experiments in which the phase

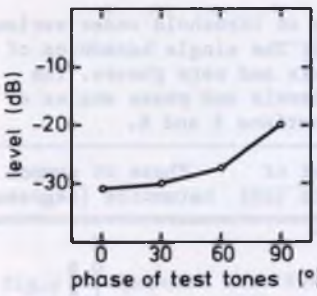


Fig. 12. Threshold of the test tones as a function of phase. The stimuli were synchronous with complete coincidence. Subjects: CE, RR (A).

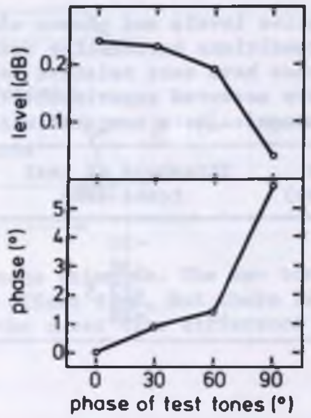


Fig. 13. Level and phase of common harmonics at threshold level as a function of the phase of the test tones. The stimuli were synchronous and with complete coincidence of the test tones.

of the test tones was varied from 0 to 90^{o}. Phase angles greater than 90^{o} were not used, since in that case the level of these common harmonics no longer increases monotonically with the level of the test tones. The results of these experiments are presented in Fig. 12. They show that there is indeed an effect of phase. The effect is in the expected direction. The threshold is lower for in-phase generation than for greater phase angles. With the help of the equations (5) and (6) it is possible to calculate the levels and phases of the common harmonics at threshold. The results of these computations are presented in Table V (Fig. 13). The perceptual separation of the masker and the test tones in synchronous stimuli with complete coincidence must be based on the differences between the common and the single harmonics. The numbers in Table V (Fig. 13) represent the differences between the common and single harmonics at threshold measured for the different phases as used for the test tone's harmonics. It appears that at threshold the level of the common harmonics is higher by only a very small fraction than that of the single harmonics. The relative level at threshold decreases when the test tone's harmonics are generated with non-zero phases. In case of in-phase generation of the test tones the level difference is the only cue for per-

Table V. Relative levels and phases of common harmonics at threshold under various phase-angle conditions between the coinciding harmonics. The single harmonics of the masker tones have zero relative sound-pressure levels and zero phases. The thresholds were measured experimentally. The relative levels and phase angles of the common harmonics were computed with the help of Equations 5 and 6.

Phase of test tones (degrees)	Threshold of test tones (dB)	Relative level of common harmonics (dB)	Phase of common harmonics (degrees)
0	-31	0.24	0
30	-30	0.23	0.9
60	-27.5	0.19	1.4
90	-20	0.04	5.7

ceptual separation. If the test tones are generated with harmonics with non-zero phases, there will also be a phase difference between the common and single harmonics. The data of Table V show that the larger the phase difference at threshold is, the smaller the level difference. From this it must be concluded that the phase difference between common and single harmonics is used as a cue for the perceptual separation of the masker and the test tones. The phase differences involved are very small, up to about 5°.

2.4. Synchronous versus asynchronous masker tone

2.4.1. Introduction

In practical musical situations synchronous tones seldom occur. For this reason experiments were conducted in which temporal differences between the two tones in the stimulus were introduced, in order to study the effect of these differences upon the threshold of the tones and upon the subjective aspects of perception.

The simplest way to remove the synchronization of two tones is to give them different onset times. This is effective only if the weaker (test) tone starts before the stronger (masker) tone, or, in other words, if the onset of the masker tones is delayed relative to the onset of the test tone. The temporal structure of such an asynchronous stimulus is illustrated in Fig. 14. Such a stimulus can be split into two successive parts. During the first part the test tone is the only sound in the stimulus. During the second part, beginning with the onset of the masker tone, both the masker and test tones are present. Because of the exclusive presence of the test tone at the beginning of the stimulus, thresholds lower than in synchronous

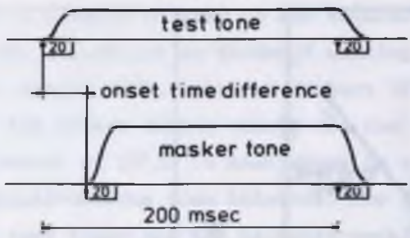


Fig. 14. Temporal structure of an asynchronous stimulus. The two tones have equal rise and decay times, and equal offset time, but there is an onset time difference. In the experiments the onset time difference was at most 30 msec.

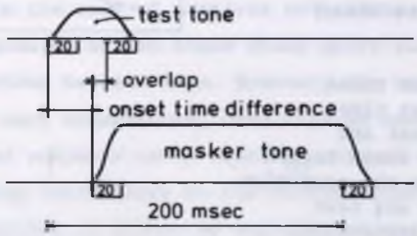


Fig. 15. Temporal structure of an asynchronous stimulus with short overlap.

conditions can be expected. In asynchronous threshold conditions the test tone is masked effectively by the masker during the second part of the stimulus. In order to see whether the perception of the test tone could be accounted for completely by its short exclusive exposition preceding the onset of the masker tone, experiments were done with the test tone ending immediately after the onset of the masker. This condition was labelled "short test tones" and is illustrated in Fig. 15.

2.4.2. Stimuli

The experiments described in this section were performed with a masker tone of 250 Hz, and test tones of 500 and 750 Hz. Phase differences between coinciding harmonics were 90°. The masker tone had its onset either synchronously with the test tones, or 10, 20 or 30 msec delayed. The test tones ended either at the same time as the masker ("long test tones") or 5 msec after the onset of the masker tones ("short test tones").

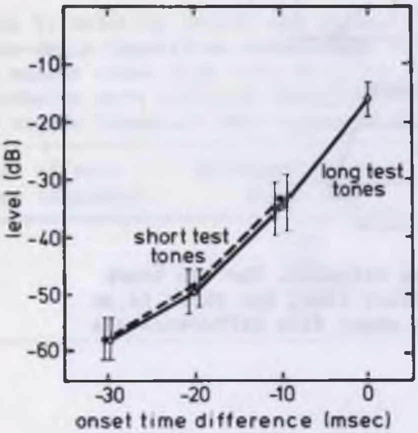


Fig. 16. Threshold of test tones as a function of the onset time difference between the test and masker tones. The masker tones lagged behind the test tones. In the condition "long test tones" masker and test tones ended exactly simultaneously. In the condition "short test tones" the test tones ended 5 msec after the beginning of the masker tones. The duration of the masker tones was always 250 msec. The harmonic coincidence of the test tones was complete. Vertical bars indicate standard deviations.Subjects: AB, CE, FS, PB, PS (A).

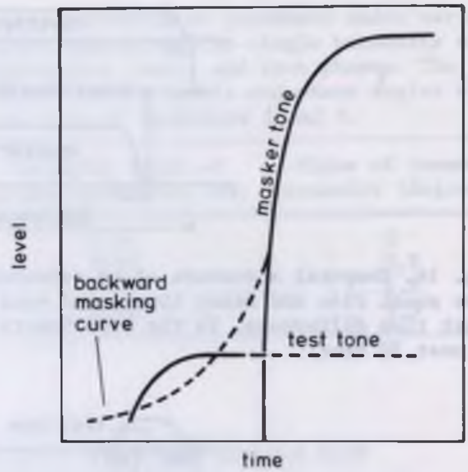


Fig. 17. Illustration of the relation between the backward-masking curve of the masker tone and the thresholds of the test tones. If test tones have lower levels, they must start earlier relative to the masker tone, in order to overcome the backward-masking effect of the masker tone.

2.4.3. Results and discussion

The results of the threshold measurements are presented in Fig. 16. The open points connected by continuous lines depict the thresholds found for long test tones, the solid points connected by dashed lines those found for short test tones. It appears that the threshold is virtually unaffected by the amount of overlap but strongly affected by the onset time difference between test and masker tones. Each 10 msec delay results in a 10 dB downward shift of the threshold. The threshold for an onset time difference of 30 msec is near to the threshold in quiet of the test tones. The onset time difference turned out to be the most powerful variable in the experiments, causing the largest threshold shifts measured.

The effective portion of the test tones occurs as a weak signal just before the onset of the much stronger masker tones. As a result, the threshold curve found is closely related to the backward masking curve of the masker tones. The literature on backward masking is fairly extensive (see e.g. Elliot 1962, Wright 1964, Wilson & Carhart 1971). Although estimates on the range of the effect differ widely for the various experimental conditions, a time interval of 10 to 15 msec seems to be the largest significantly effective backward-masking time interval. The relation between the thresholds of the test tones and the backward-masking curve of the masker tone is illustrated in Fig. 17.

The subjective perception of stimuli with delayed masker tones differs considerably from that of synchronous stimuli. While in synchronous stimuli the tendency of the two tones to fuse into one "sound image" is strong and particularly so in the case of complete coincidence of harmonics, in non-synchronous stimuli the two tones stand apart clearly, independent of the spectral relations between them. However, the unequal onset times of the tones are not very conspicuous, even with an onset time difference of 30 msec. Uninformed subjects never reported unequal onset times of the tones. This finding corresponds to the minimum time interval necessary for the correct perception of order, as reported in the few papers on this subject (e.g. Hirsh 1959). The perception of the order of the two signals is additionally impaired by their different qualities (see Warren 1974) and by their smooth rise curves. All this means that the two tones in the stimulus are perceived as two separate but simultaneously occurring sounds.

Although the perception of the test tones can only be based on 10 to 30 msec of presentation, one hears this tone as having the same duration as the masker tone, and not as a very short tone burst. This is an example of the continuity effect, a phenomenon also found in other auditory conditions. The continuity effect typically operates when weak continuous sounds are masked temporarily by stronger sounds with short durations. Although the weak sound cannot be perceived when being masked, perceptual coding is simplest if its presence is assumed. Since during the masked time interval the weak sound does not contribute any energy to our hearing system, the suggestion of its presence holds even when it is actually absent during the presence of the strong sound. The continuity effect is therefore based on the assumed or possible presence of a weak sound that is or would be masked by a strong sound. The effect is found to play a part in a number of psycho-physical paradigms, especially the rapid alternation of weak and strong sounds. In that case one hears the weak sound continuously through the strong sound bursts (Thurlow & Elfner 1959). The time range of the continuity effect is restricted. Exact data on the range of the effect have not yet been published. The optimal time interval seems to be in the range of 100 to 200 msec. Verschuure & al. (1976) mention 200 msec as upper limit in a pulsation-threshold situation, but elsewhere greater values have been used (Fastl 1975: 300 msec). The maximum time interval for which the effect holds probably lies in the range of 250 to 300 msec, a time interval that corresponds, perhaps not accidentally, to the average duration of a syllable and a musical tone.

Another observation from the stimuli with delayed masker tones is that the subjective loudness of the test tones is more or less the same as that of test tones of 200 msec in isolation, i.e. higher than the subjective loudness of a tone of only a few milliseconds which actually are the basis for perception. Evidently, the continuity effect restores not only the probable temporal course of the test tone, but also its probable loudness. This corresponds with an observation by Pickett (1959, note 5).

2.5. Simultaneous versus continuous masker tone

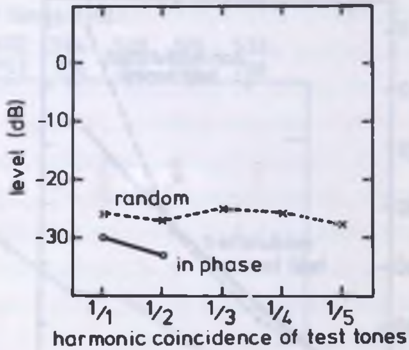
2.5.1. Introduction

In the preceding section it was shown that asynchronization of masker and test tones has a profound influence on the threshold of the test tone. The asynchronization in these experiments was realized by delaying the masker's onset for some msecs. In this section the effects of another form of asynchronization will be dealt with, viz. short test tones against a background of a continuous masker tone. We will see that this condition also results in a decrease of the threshold of the test tones, be it a slight decrease compared to the condition with a delayed simultaneous masker tone.

2.5.2. Stimuli

The comparison of conditions with simultaneous short and continuous masker tones was an important part of the experiments of series B. The stimuli used were largely analogous to those used for synchronous tones in the series B experiments. Harmonic coincidence varied from 1/1 to 1/5. All waveforms were composed by sine addition. The stimuli with coincidence 1/1 or 1/2 were presented in two conditions as to phase relation: with random phase relation; and with phase-locking (in-phase). In the latter

Fig. 18. Threshold of the test tones with continuous masker tones, as a function of coincidence. The fundamental frequencies used are listed in Table III. Parameter is the phase relation between the coinciding harmonics of masker and test tones. Subjects: JPK, JV, RR (B). The data point for random phase relation and complete coincidence is based on RR only.



conditions a click could be heard just before every stimulus as a result of restarting the clock generators. This restart was necessary in order to get the required phase-locking between the waveforms of the masker and the test tones. The clicks were actually short interruptions of the masker tones. Seemingly, they did not influence the subjects' performance on the experimental task. The stimuli with coincidence 1/3, 1/4 or 1/5 were only presented with random phase relations between coinciding harmonics.

2.5.3. Results

The results of the threshold measurements with continuous masker tones are presented in Fig. 18. They should be compared to the results shown in Figs. 10 and 11, which are derived from similar stimuli, but with synchronous masker and test tones instead of a continuous masker tone. The thresholds with continuous masker tones are lower than with simultaneous masker tones. The decrease does not vary very much as a function of coincidence. The average decrease is about 3 dB. As a matter of fact, this is considerably less than the decrease that is found in the conditions with a delayed masker tone.

The condition with complete coincidence and random phase is again an exceptional one. It appeared impossible for the subjects, except for the experimenter himself, to get reliable results. The difficulties of this condition for the subjects are quite understandable. The phase shifts at the onsets of the test tones result in pitch shifts which are randomly distributed over the various common harmonics. The confusing effects of these onset (and offset) phenomena can only be overcome after extensive exposure to the stimuli. The experimenter-subject managed to get a 27 dB decrease of the threshold under the continuous-masker condition compared to the synchronous-masker condition.

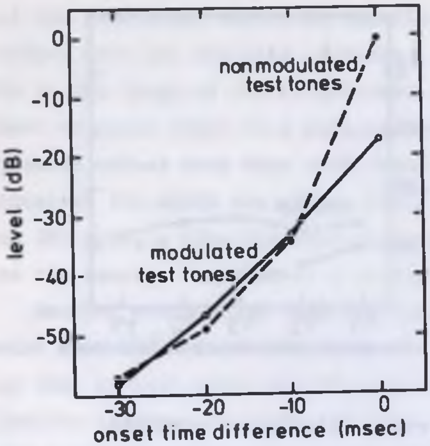


Fig. 19. Threshold of frequency-modulated test tones as a function of onset time difference between masker and test tones. Thresholds of non-modulated tones have been inserted for comparison. They have been borrowed from Fig. 10 (synchronous stimuli with independently generated tones) and Fig. 16 (non-synchronous stimuli with phase-locked generation of tones). Subject: RR (A).

2.6. Deviating frequency ratios

2.6.1. Frequency modulation

In musical performance vibrato is a rather common phenomenon. Vibrato consists of both frequency and amplitude modulation, the relative importance of each type of modulation being dependent on the instrument and playing technique involved. Modulation frequencies mostly fall within the range of 4 to 8 Hz. One of the functions of vibrato is to make the tones more conspicuous. If this is true, the threshold of frequency-modulated test tones may differ from that of non-modulated test tones. Experiments were run in which the test tones were frequency-modulated with a modulation frequency of 5 Hz and a modulation depth of 4% (peak to middle). Only stimuli with complete coincidence (masker tone 250 Hz, test tones 500 and 750 Hz) and random phase relations were used. The stimuli were either synchronous or asynchronous, with a 10 to 30 msec delay of the masker tone. The results are shown in Fig. 19, in which the results of the experiments with non-modulated tones (Fig. 10, 16) have also been inserted.

It turned out that the frequency-modulated test tones of synchronous stimuli with random phase relation had a threshold of -17.5 dB. This must be compared to the 0 dB threshold of the comparable non-modulated test tones. So it can be concluded that for synchronous tones frequency modulation contributes to a rather large extent to the perceptibility of a tone. It can be expected that the same holds in a musical situation. The role

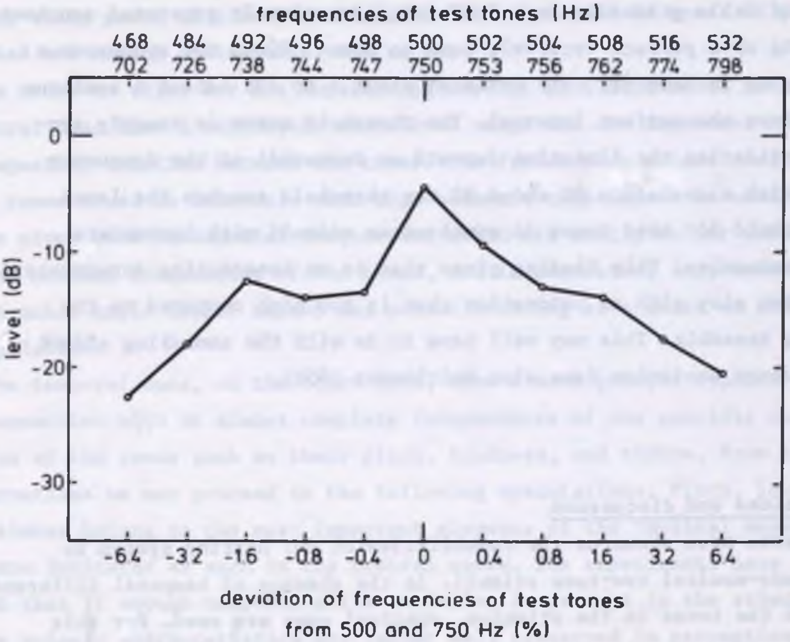


Fig. 20. Threshold of test tones that deviate in frequency from the pure ("just") intervals 250/500 and 250/750 Hz. The stimuli were synchronous. Masker and test tones were independently generated. Subject: RR (A).

of vibrato in the perceptual separation of tones is also illustrated by experiments by McAdams (1980).

The frequency-modulated test tones in asynchronous stimuli had thresholds that come very close to the thresholds of non-modulated test tones. Evidently, in these conditions the onset time difference plays the predominant role in perception.

2.6.2. Mistuned intervals

As part of the series A experiments a short experiment was run in order to see how deviations from the perfect intervals 250/500 and 250/750 Hz would affect the thresholds of the test tones. The experiments were done with synchronous masker and test tones, the masker tones being 250 Hz, the test tones differing ± 0.4, 0.8, 1.6, 3.2, or 6.4% from 500 and 750 Hz. For comparison: a semitone is about 6%. Masker and test tones were generated independently.

The results of the experiments are shown in Fig. 20. Deviation from the perfect intervals did influence the thresholds of the test tones.

The threshold falls gradually from 0 dB for independently generated synchronous tones, making up a perfect interval, down to about -20 dB for synchronous tones making up an interval that deviates about 5 to 10% (about a semitone to a tone) from the perfect interval. The threshold curve is roughly symmetrical considering the direction (upward or downward) of the frequency deviation. With a deviation of about 5% the threshold reaches the level of the threshold for test tones in synchronous stimuli with incomplete overlap of harmonics. This finding gives rise to an interesting interpretation. Soloists often play with an intonation that is too high compared to the accompanying ensemble. This may well have to do with the unmasking effect of the frequency deviation (see also Weitbrecht 1950).

2.7. Conclusions and discussion

The experiments have shown us the capabilities of our hearing system to analyze pseudo-musical two-tone stimuli. In the absence of temporal differentiation between the tones in the stimulus, spectral cues are used. For this type of perception, the levels of the tones must not differ more than about 20 dB. If the hearing system can make use of temporal cues for the separation and perception of the tones, the effective range widens drastically, say, to 50 dB. This is the case when the weaker tone precedes the stronger tone for a very short time interval, e.g., 10 to 30 msec. This difference in onset times is too small to lead to perception of order. In addition, the perception of the weaker tone is facilitated by the continuity effect. The perception of the weaker tone is not impaired by its being masked by the stronger one after the onset of the latter.

It can be assumed that in practical musical circumstances both spectral and temporal characteristics play their parts. Spectral cues predominate in steady-state parts of musical tones. Organ, string and wind instruments produce spectral patterns that are clearly recognizable. But temporal cues are also abundant, first of all because musical parts and voices are composed as a succession of tones, each of which has its own temporal structure. When the musical score prescribes different onset moments for tones of different parts, this is particularly effective in making them stand apart. But even if synchronous onset is required, actual perfect synchronization of the tones is exceptional. Other sources of temporal cues of tones are instrumental and performance features like the widely differing durations and shapes of rise curves, modulation phenomena in steady-state parts of tones, the impulse-like character of tones of percussion, plucked and struck instruments, and so on.

In order to allow for the separate perception of simultaneous tones, spectral cues have a relatively restricted range and are dependent on the spectral relation between the tones. The fusion that arises easily when tones sound in intervals with high harmonic coincidence is used in organ stops with fundamental frequencies which are multiples (or submultiples) of the nominal frequencies of the tones, and in orchestral instrumentation, which makes ample use of unison and octave doubling as a means of timbre manipulation.

The temporal cues, on the other hand, show a much greater dynamic range in connection with an almost complete independence of the specific characteristics of the tones such as their pitch, loudness, and timbre. From these observations we may proceed to the following speculations. Pitch, loudness and timbre belong to the most important elements of the "musical message" and are indicated as such in the musical score. The experiments have shown that if enough temporal differentiation is present in the stimulus these primary characteristics are rather well preserved in perception. Also, the separation of the tones is well preserved. One may conclude that temporal perception processes must have played an important part in the development of polyphonic music as a perceptually justified musical structure. This is an example of the interaction between musical composition techniques on the one hand, and the acoustic features of performed music and the response of the human hearing system to these features, on the other.

CHAPTER THREE

RECOGNITION EXPERIMENTS

3.1. Introduction

In the preceding chapter experiments were described that concerned the measurement of the limits for hearing individual tones against a background of other tones. Implicit in these investigations was the assumption that stimulus conditions with low thresholds would correspond to good separate perception of tones in normal, supra-threshold conditions, and that, vice versa, high threshold conditions would correspond to poor separate supra-threshold perception. In order to test this assumption a series of experiments with simultaneous tones under supra-threshold listening conditions was carried out. The stimuli were two-tone musical harmonic intervals (from major second to minor seventh) which had to be identified by the subject. The percentage of correct identification was used as a measure for the separate perception of the tones. It was assumed that the separate perception of the tones of the interval would result in better identification of the interval between the tones. For this experimental task, it is necessary that the subjects are able to name the intervals. Music students achieve this ability as a part of 'ear training' (solfège), a standard course in all forms of musical education.

The experimental set-up of this musical-interval recognition experiment was as follows. Stimuli were constructed that consisted of two complex tones. There were four independent variables, which were varied randomly and independently from each other: (1) musical interval (from major second to minor seventh, excluding the tritone); (2) onset time difference (from -40 to +40 msec); (3) level difference (from -20 to +20 dB); and (4) the absolute frequency (about the middle range for musical tones). The subject was presented with a random series of 250 stimuli which had to be identified by pressing one of eight buttons of a response box.

3.2. Method

3.2.1. Apparatus

The apparatus used in the experiments is illustrated schematically in the diagram in Fig. 21. The experiments were run on-line under the control of a PDP 11/10 computer. Waveforms were stored as 256 discrete samples in two revolving memories (recirculators) with 10-bit accuracy. The recirculators were read out by digital-to-analogue converters with sample frequencies which were 256 times the required fundamental frequencies. The sample pulse trains were derived from frequency generators. Each of the signals delivered by the d/a converters passed a gate which opened and closed on timer pulses, a low-pass filter with cut-of frequency of 4 kHz, and an attenuator. Then the signals were mixed and presented to the subject by headphones, binaurally. The subject was seated in an acoustically treated room. Responses could be made by way of a response box with eight buttons. In each of the buttons there was a little light. After the response the correct button lit up for 0.5 sec so that the subject had immediate feedback of results. Response times were measured by starting a clock at the beginning of the decay portion of the tones and reading out the time at the moment of response.

The computer was used to calculate waveforms, to determine the composition of stimuli, trials and series, to control apparatus (frequency-generators, recirculators, attenuators, timers, clock), and for the registration and processing of responses and the display of results.

3.2.2. Stimuli, trials, and series

Stimuli were harmonic musical intervals consisting of two complex tones. Waveforms were calculated with Equation (1) (Section 2.2.3.). The duration of the tones was 250 msec. Rise and decay times were 20 msec, with half-sinusoidally shaped rise and decay amplitude envelopes. The stimuli were varied along four parameters, which were independent from each other:

(1) Musical interval: major second (frequency ratio 9/8), minor third (6/5), major third (5/4), perfect fourth (5/4), perfect fifth (3/2), minor sixth (8/5), major sixth (5/3) and minor seventh (16/9).

(2) Onset time difference, the onset time of the higher tone relative to that of the lower tone: ± 40, 20, 10, 5, and 0 msec. A negative onset time difference means that the higher tone came first.

(3) Level difference, the level of the higher tone relative to that of the lower tone: ± 20, 10, 5, 2, and 0 dB. A negative value means that the lower tone had the higher level. Absolute levels were such that the combined level of the two tones was about 70 dB re 20 μPa.

(4) Absolute frequencies. For the first group of four subjects, the fundamental frequency of the lower tone was determined first (250, 300, 350, 400 or 450 Hz) or that of the higher tone (300, 350, 400, 450 or 500 Hz). The frequency of the other tone was determined according to the

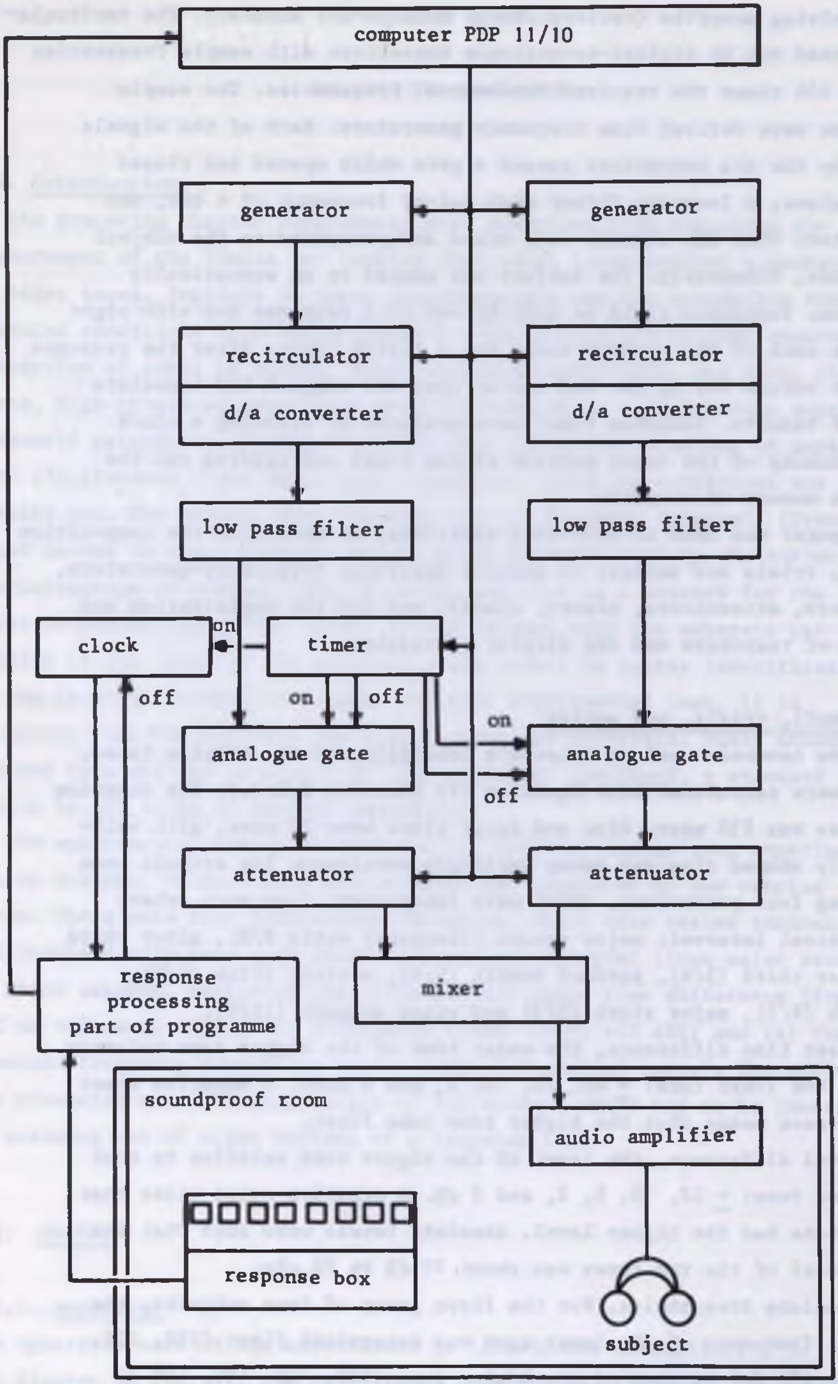


Fig. 21. Set-up of apparatus used in the interval recognition experiments.

interval. With this method the fundamental frequencies of the lower tones varied from 168.75 to 450 Hz (musically from about E3 to A4), those of the higher tones from 281.5 to 800 Hz (about C4 sharp to G5). For the second group of four subjects, absolute frequencies were determined by taking 250, 275, 300 ... 450, or 475 Hz as the arithmetic mean frequency of the interval. With this method the fundamental frequencies of the lower tones varied from 180 to 447 Hz (about F3 to A4), those of the higher tones from 264.7 to 608 Hz (about C4 to D5).

The time structure of the trials is illustrated in Fig. 22. Trials consisted of four consecutive segments: (1) the stimulus, with a duration of 250 msec; (2) the response time, measured from the beginning of the decay of the tones, until the moment of the subject's response; (3) the feedback time, 500 msec, during which the button of the correct response lit up; and (4) the inter-trial time, 500 msec. Mean response time turned out to be about 1750 msec, so that the mean duration of a trial was about 3000 msec.

Series consisted of 250 trials. With a mean trial duration of 3 sec the duration of a series was about 12.5 min. The values of the four independent variables were determined for each trial by random-number procedures. This method results in a series of stimuli that is random in every respect. Actually, there are 8 x 9 x 9 x 10 = 6480 different stimulus conditions. Each series was, in fact, a random sample from this 'stimulus population'.

The subject had to respond by pushing one of the eight buttons of the response box. Each of these buttons corresponded with one of the eight musical intervals in the experiments, which were indicated next to the buttons. If the response time was longer than 4 sec, the response was not processed. This was done because the reliability of such responses could not be expected to be very high, while they would have influenced the mean response times in an uncontrolled way.

3.2.3. Subjects

As in the threshold experiments, musicology and conservatory students served as subjects. There were two groups of four subjects. The experimental programme differed slightly between the groups, both in stimulus construction (absolute frequencies) and in response processing.

3.2.4. Measurement procedure

The subject was asked to identify the musical interval presented by means of pressing one of the buttons of the response box. He was informed about

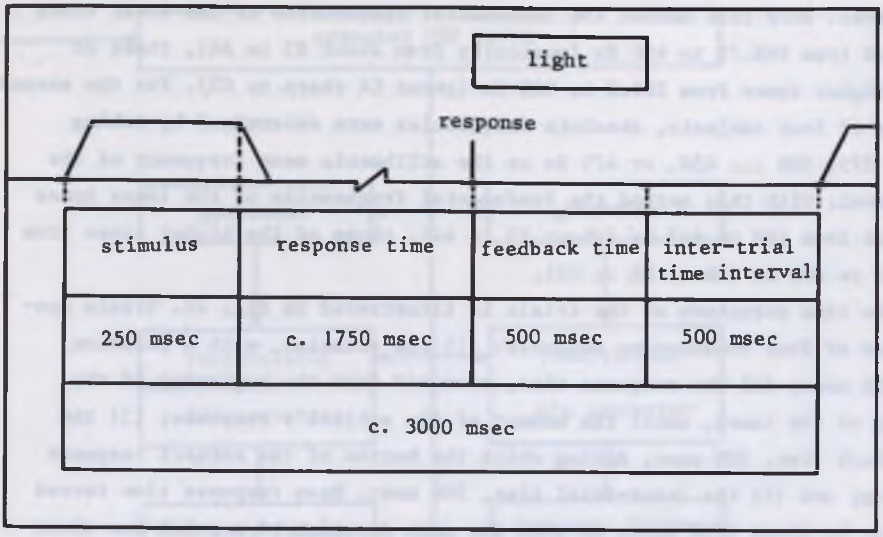


Fig. 22. Temporal structure of the trials in the recognition experiments. The successive time slices are not to scale.

the correctness of his identification by a light in the correct button. This feedback procedure appeared to be an important factor in the motivation to perform the experimenal task as well as possible. The subject was asked to react reasonably quickly. He was not informed that his response times were measured or that his too-long responses were skipped.

The percentages of correct responses and the mean response time were the dependent variables to be determined. The percentages of correct responses were calculated for each value of the independent variables separately, summing over the conditions created by the other variables. For the second group of four subjects, the percentages of correct responses were also calculated for the 'combined conditions', determined by the values of onset time difference and level difference.

Mean response times were computed for all values of onset time difference and level difference separately, and, for the second group of four subjects, also for the 'combined' conditions. The means were calculated for all responses, and for the correct and incorrect responses separately.

Table VI. Correlation coefficients between percentages of correct responses and mean response times, calculated per category of onset time difference and level difference.

variable	mean response time computed over all responses	correct responses	incorrect responses
onset time difference	-0.762	-0.747	-0.442
level difference	-0.709	-0.382	-0.245

3.3. Results

3.3.1. General

The overall percentage of correct responses was 65%. For the first group it was 60%, for the second group 70%.

The mean response time was 1918 msec. The correct responses had a lower mean response time: 1754 msec. The incorrect responses required a longer time on the average: 2304 msec. It must be taken into account that responses with times longer than 4000 msec were discarded.

The correlations between the percentages of correct responses and the response times were calculated for the scores on the variables onset time difference and level difference. These calculations were performed for the mean response times for all responses and for the correct or incorrect responses separately. The resulting correlation coefficients are shown in Table VI. It appears that the correlation is highest when recognition times are taken over all scores, both the correct and the incorrect ones.

3.3.2. Intervals

The recognition of the musical intervals themselves was the experimental task, but not the aim of the experiments. The scores are presented in Table VII. The dissonant intervals (major second, minor seventh) were the easiest ones to recognize. The perfectly consonant intervals (fifth and fourth) were better recognized than the imperfectly consonant ones (the thirds and sixths), in this experimental context.

3.3.3. Onset time difference

Onset time difference was varied from -40 to +40 msec. The recognition score as a function of onset time difference is presented graphically in Fig. 23. This figure includes both the percentages of correct responses and the mean response times, determined for all responses and for the

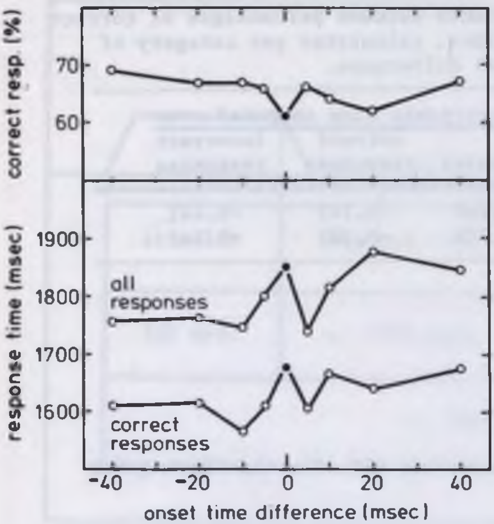


Fig. 23. Percentages of correct responses and mean response times of musical intervals as a function of the onset time difference between the two tones of the interval.

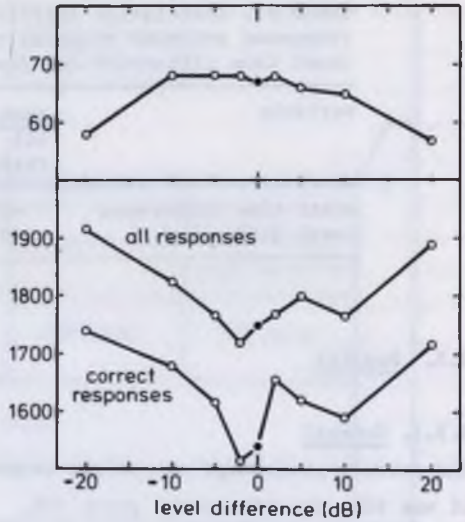


Fig. 24. Percentages of correct responses and mean response times of musical intervals as a function of the level difference between the two tones of the interval.

Table VII. Percentage of correct responses of musical intervals.

interval	percentage of correct responses
major second	92
minor third	56
major third	60
perfect fourth	65
perfect fifth	77
minor sixth	53
major sixth	48
minor seventh	76
mean	66

correct responses only. These data are summed over all the conditions with a particular onset time difference, ignoring the variation in interval, level difference and absolute frequency.

The differences between the scores for the various conditions with respect to onset time difference are not very large. However, each data point is based on a large number of trials (about 800), which varied in other respects. Two trends can be discerned in the results. The condition without onset time difference resulted in the smallest percentage of correct identification (61%). Onset time difference led to better recognition, but the amount of onset time difference is not very influential. The smallest difference value (5 msec) was sufficient to elicit the recognition score that held for asynchronous stimuli in general (ca. 67%). Both positive and negative onset time differences gave rise to better identification than no onset time difference. However, the effect is not strictly symmetrical. Conditions with negative onset time difference (higher tone starts before the lower one) result in better identification than do positive onset time differences. This trend is to be expected from the results of the threshold experiments. The lower tone can easily mask the higher one. The masking can be compensated for if the higher tone starts a little before the lower one, but not if the lower tone leads.

3.3.4. Level difference

The experimental results concerning level difference are presented in Fig. 24. As in the case of onset time difference, the data are summed over variations of the other independent variables. The scores are more or less symmetrical for positive and negative level differences. The recognition score is not influenced by the level difference to any substantial degree unless the difference is larger than 10 dB. For these level differences the percentage of correct responses was about 67%. For a level difference of 20 dB (positive or negative), recognition of the intervals was poorer (about 58%).

In the interpretation of the data concerning level differences one should take into account that levels are expressed in sound-pressure level differences, not in spectral envelope level differences, as was the case with the threshold experiments. If the envelope level difference is taken as a point of departure for studying the recognition scores, a slight asymmetry arises. The envelopes of the tones have slopes of -6 dB/octave. The mean interval between the higher and the lower tones is half an octave. This means that the relative level of the envelopes of the higher tones is, on the average, 3 dB higher than the relative sound-pressure level. So the sound-pressure level difference can be corrected into spectral envelope level by adding +3 dB. Negative envelope level differences (lower tone stronger than higher tone) resulted in poorer identification than positive envelope level differences. This can be explained by the fact that the lower tone can mask the higher one more easily than vice versa.

3.3.5. Onset time difference versus level difference

Both onset time difference and level difference influence the recognition of the intervals. For the second group of four subjects, the scores for the conditions characterized by a specific combination of onset time difference and level difference were determined. With these data, it is possible to investigate the relative importance of onset time difference and level difference in the recognition task. Two methods were used to determine the relative importance of the two factors: a simple form of regression analysis and analysis of variance.

For the first method, we defined a predicted score for a 'combined condition' as the weighted sum of the score according to its onset time

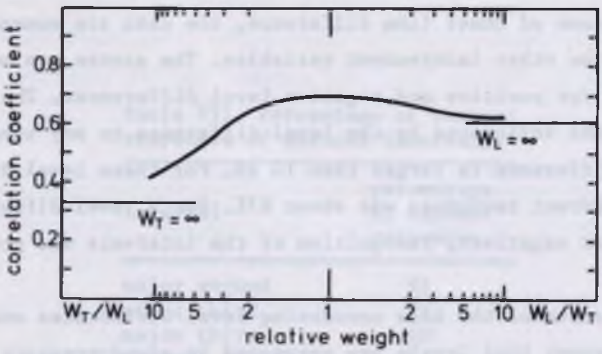


Fig. 25. Correlation coefficient between obtained and predicted scores of interval recognition in various conditions. The obtained scores refer to conditions characterized by an onset time difference and a level difference. The predicted scores are weighted sums of scores for a particular onset time difference (averaged over all other variables) and one for a particular level difference (averaged over all other variables).

difference and the score according to its level difference:

$$\hat{s}_{T,L} = w_T \cdot s_T + w_L \cdot s_L \qquad (7)$$

where $\hat{s}_{T,L}$ - predicted score for the conditions with onset time difference T and level difference L

s_T - mean score for onset time difference T

s_L - mean score for level difference L

w_T - weight for level difference score

w_L - weight for onset time difference score

The correlation between the predicted and the measured scores of the combined conditions was computed, with the ratio of weights w_T/w_L varying from 0.1 to 10. It turned out that the highest correlation (0.69) was found with $w_T/w_L = 0.95$. That means that actually the mean of the "onset time difference score" and the "level difference score" gives a good estimate of the score of a condition characterized by a certain combination of onset time difference and level difference. The correlation between predicted and measured scores as a function of the relative weight is presented graphically in Fig. 25.

Analyses of variance were carried out over the percentages of correct responses and over the response times. The data matrices were three-dimensional, with subject, onset time difference and level difference as variables. The results of the analyses are summarized in Tables VIII and IX. It turns out that the differences between subjects are responsible for more than half of the total variance. Onset time differences and level differences explained a few additional percentage points of variance at a significant (or nearly significant) level, as did the interaction between these two variables. Interaction variances with subject differences were negligible, except for the interaction between subject and level difference, regarding response times.

3.4. Conclusions and discussion

The psychoacoustical experiments described in Chapter Two dealt with the limits of separate audibility of tones. Most practical situations create perceptual conditions that are less extreme. For this reason the interval recognition experiment was carried out with stimulus conditions that more or less reflected musical reality. Level differences between simultaneous

Table XVIII. Analysis of variance of the percentages of correct identifications. The variables are: 1. subject (four levels: random); 2. onset time difference (nine levels: fixed); 3. level difference (nine levels: fixed).

Variable	Degrees of freedom	F-ratio	Probability (%)	Percentage of explained variance
1	(3,192)	153	0.00	61.60
2	(8,24)	2.41	4.55	0.82
3	(8,24)	6.54	0.03	3.82
1-2	(24,192)	0.72	82.58	-0.90
1-3	(24,192)	0.85	66.68	-0.48
2-3	(64,192)	1.36	5.86	2.32

Table IX. Analysis of variance of the response times. The variables are: 1. subject (four levels: random); 2. onset time difference (nine levels: fixed); 3. level difference (nine levels: fixed).

Variable	Degrees of freedom	F-ratio	Probability (%)	Percentage of explained variance
1	(3,192)	187	0.00	65.30
2	(8,24)	4.49	0.22	1.40
3	(8,24)	2.34	5.10	1.59
1-2	(24,192)	0.57	94.71	-1.20
1-3	(24,192)	1.69	2.86	1.93
2-3	(64,192)	1.46	2.64	2.57

tones were never more than 20 dB, onset time differences ranged from zero to 40 msec. The results of the experiments show that these differences influence the recognition of the interval, but not to a very large extent.

Intervals with a level difference between the tones of 20 dB were hardest to recognize. Equal levels made perception easiest. Synchronous intervals were somewhat harder to recognize than asynchronous intervals. But the variance in the data caused by these variables was much less than the variance to be attributed to the differences between subjects.

This experiment was based on the idea that interval recognition would improve if the tones of the interval were better perceived separately. By this reasoning interval recognition could be connected with separate perception and used as a measure of the latter. Experiments by Plomp et al. (1973) gave rise to the conclusion that 'tone distance' (frequency difference) contributed more than 'interval quality' (frequency ratio) to the correct perception of a musical interval. Experiments on the categorical perception of a musical interval by Burns & Ward (1978) point in the same direction. Still, it may be possible that in our experiments the tone quality of the intervals played a part in the recognition of intervals, resulting in a decrease of variance to be explained by onset time and level differences.

From the experiments described in this chapter we can conclude that onset time differences also operate in supra-threshold conditions, in which the level differences between the tones of the intervals are not very large. However, the effect of onset asynchrony in these supra-threshold conditions is definitely smaller than in threshold conditions in which large level differences occur. The effect of level difference could not be investigated in the experiments described in Chapter Two (threshold experiments) because there it was the dependent variable. From the present experiments it can be concluded that in supra-threshold conditions this variable is effective to the same order of magnitude as is onset time difference. The practical consequence of this finding is simple. Tones in simultaneous lines of music should not have very different levels, if they are to be perceived separately.

CHAPTER FOUR

THE MEASUREMENT OF ASYNCHRONIZATION

4.1. Introduction

In the preceding chapters psychoacoustical experiments were described concerning the perception of stimuli consisting of two simultaneous tones. One of the most striking outcomes was that small temporal differences between the tones (which have no effect on the perception of order of the tones) are important for the ability to distinguish the tones in perception. These results were derived from laboratory experiments. In this chapter investigations will be described that were carried out in order to see if the temporal differences used in the experiments also occur in actually performed music. The primary question was what asynchronization existed between tones meant to be simultaneous according to the musical score. To this end a series of special recordings was made in which the various single voice parts of the music were available separately. Analyses of these recordings used specially developed, rather extensive computer programming. The outcomes of these analyses were values of statistical measures which were chosen in accordance with a statistical model developed for the description of asynchronization of simultaneous tones. The calculations showed that a certain extent of asynchronization was inherent in every musical performance. This asynchronization was of the same order of magnitude as that used in the psychoacoustic experiments. The result makes it plausible that the laboratory results concerning the perception of two-tone stimuli can be extended to real polyphonic music.

4.2. The statistics of asynchronization

4.2.1. Introduction

There was no theoretical model available that could serve as a background for synchronization analysis. Therefore, a statistical model for the analysis of synchronization and asynchronization has been specially developed. It is based on the variances of time intervals.

When musicians play together in an ensemble, they will try to synchronize

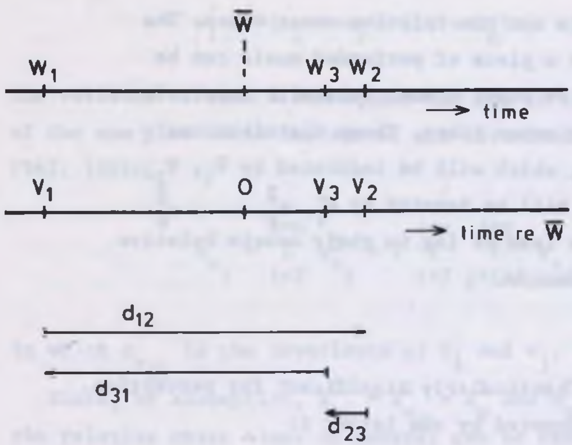


Fig. 26. Illustration of the concepts used for synchronization analysis. w_1, w_2 and w_3 are three onsets meant to be simultaneous. $\bar{w}$ is the mean onset time. v_1, v_2 and v_3 are the onset times expressed relative to $\bar{w}$. The onset time differences are denoted by d_{12}, d_{23} and d_{31}.

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.

4.2.2. Absolute and relative onset times

Onsets of tones meant to be simultaneous will, in reality, scatter a little in time. Assume an ensemble of n players. Fig. 26 illustrates some of the concepts used for an ensemble with three performers. The absolute onset times of the tones meant to be simultaneous will be denoted by w_1, w_2,..., w_n. For each group of such onsets there will be a mean onset time

$$\bar{w} = \frac{w_1 + w_2 + \ldots + w_n}{n}. \quad (8)$$

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

$$\begin{aligned} v_1 &= w_1 - \bar{w} \\ v_2 &= w_2 - \bar{w} \\ &\ldots \\ v_n &= w_n - \bar{w}. \end{aligned} \quad (9)$$

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. These distributions can be characterized by their means, which will be indicated by $\bar{v}_1$, $\bar{v}_2$, ..., $\bar{v}_n$, and their variances, which will be denoted by s_1^2, s_2^2, ..., s_n^2. The means indicate which instruments lead or lag in their onsets relative to the other instruments in a performance.

4.2.3. Onset time difference

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

$$d_{12} = v_2 - v_1$$
$$d_{13} = v_3 - v_1$$
$$\cdots$$
$$d_{n-1,n} = v_n - v_{n-1}.$$

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 $\bar{d}_{12}$, $\bar{d}_{13}$, ..., $\bar{d}_{n-1,n}$, and variances, to be indicated by s_{12}^2, s_{13}^2, ..., $s_{n-1,n}^2$.

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 will be 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, e.g.:

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

The variance 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 (Nunnally 1967, 138):

$$s^2_{v_1} = \sum_{i=2}^{i=n} s^2_{v_i} + 2 \sum_{i=2}^{i=n} \sum_{j=i+1}^{i=n} c_{v_{ij}} \qquad (j \neq i)$$

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

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

$$s^2_v = (n-1)s^2_v + (n-1)(n-2)c_v .$$

This leads to

$$c_v = \frac{s^2_v}{1-n} . \qquad (11)$$

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:

$$s^2_d = 2s^2_v - 2c_v . \qquad (12)$$

Substituting the value of c_v (Eq. 11) in Equation 12 leads to

$$s^2_d = \frac{2n}{n-1} s^2_v . \qquad (13)$$

So it is shown that 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 $s^2_d = 4s^2_v$. For a large ensemble, the relation converges to the limit $s^2_d = 2s^2_v$.

Interpreting these relations for small and large groups of performers depends on whether the variance of the relative onset times or that of the onset time differences is constant irrespective of the number of performers. For the moment, we will assume that the variance of relative onset times is constant when the ensemble is headed by a conductor. Every player will try to synchronize with the conductor's gestures independent of the onsets of the other players. From this assumption it can be concluded that the variances of onset time differences are greater for small than for larger groups of players. When there is no conductor, we assume that the onset time differences are the constant elements. In this case, the variances of relative onset times will be greater for larger than for smaller groups. If this interpretation is valid, it can be concluded that larger groups do better with a conductor with respect to synchronization, while smaller groups do better without. In order to know the critical number of performers in deciding whether to have a conductor or not, it is necessary to know the relative variances of synchronizing with an auditory and a visual signal.

4.2.4. Asynchronization of voice pairs

In actual practice (see Sections 4.5.1 and 4.5.2), onset time difference is a variable with a mean not very different from zero (mostly within the range of -5 to +5 msec) and a standard deviation that is much greater (mostly within the range of 30 to 50 msec). For this reason the standard deviation of the onset time difference is a better measure for the amount of asynchronization in performed music than the mean onset time difference. 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: $A_{ij} = s_{ij}$.

Since only n-1 out of n onset time differences can be chosen independently, the asynchronization of a voice pair can be derived from the asynchronizations of the other voice pairs and the correlation between onset time differences. Take, as an example, an ensemble with three musicians. Since $d_{12} = -(d_{23} + d_{31})$:

$$A_{12}^2 = A_{23}^2 + A_{31}^2 + 2A_{23/31}$$

in which $A_{23/31}$ is the covariance between d_{23} and d_{31}. The correlation between d_{23} and d_{31} equals

$$r_3 = \frac{A_{23/31}}{A_{23}A_{31}}$$

so that

$$A_{23/31} = r_3 A_{23} A_{31}$$

and

$$A_{12}^2 = A_{23}^2 + A_{31}^2 + 2r_3 A_{23} A_{31}. \qquad (15)$$

The correlations between the various pairs of onset time differences are not entirely independent either. If n=3, it is possible to calculate two correlation coefficients from the asynchronization values and one correlation coefficient. If r_2 (the correlation between d_{12} and d_{23}) is known, r_1 and r_3 can be calculated by combining the equation given above for A_{12} with similar ones for A_{23} and for A_{31}

$$r_1 = -\frac{A_{12} + r_2 A_{23}}{A_{31}} \qquad (16)$$

$$r_3 = -\frac{A_{23} + r_2 A_{12}}{A_{31}}.$$

If we assume all asynchronizations of voice parts to be equal, this implies a fixed value for the correlation between the onset time differences, dependent on the value of n. For n=3, e.g., the correlation between the onset time differences of voice pairs can be derived from Equation 16 by assuming $A_{12}^2 = A_{23}^2 = A_{31}^2$ and $r_1 = r_2 = r_3 = r$. The result is $r = -0.5$.

Significance testing is possible for the mean onset time differences, by the t-test for differences. The testing magnitude is (Spitz 1968, 302):

$$t_{12} = \frac{\bar{d}_{12}(N-1)^{\frac{1}{2}}}{A_{12}} \qquad (17)$$

in which N is the number of onset time differences.

4.2.5. The asynchronization of a piece of music

In practice, 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 asynchronizations 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:

$$A = \left(\frac{\sum_{i=1}^{i=n-1} \sum_{j=i+1}^{j=n} A_{ij}^2}{\frac{1}{2}n(n-1)} \right)^{\frac{1}{2}} . \qquad (17)$$

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% of onset time difference, the range from -2A to +2A 95%. 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.

4.2.6. Isochronization

The synchronization of simultaneous tones is only one of the temporal tasks of performing musicians. Within one voice part, the succession of tones also has to be strictly timed. We will call the timing of tones of equal duration in one voice part the isochronization of the tones. The isochronization of a voice part 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}. The duration of the first tone is

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

The absolute onset time is the sum of the mean onset time and the relative onset time (Equation 9). We denote the mean onset time of tones simultaneous to w_i as $\bar{\bar{w}}_i$, that of tones simultaneous to w_{i+1} as $\bar{\bar{w}}_{i+1}$. We assume that the successive mean onset times (..., $\bar{\bar{w}}_i$, $\bar{\bar{w}}_{i+1}$, ...) 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:

$$h_i = \bar{h} + (v_{i+1} - v_i). \qquad (19)$$

The variance of this last expression is

$$s_h^2 = 2(1-r)s_v^2. \qquad (20)$$

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

$$s_h^2 = \frac{n-1}{n}(1-r)s_d^2. \qquad (21)$$

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 literature (see Section 4.6.1.).

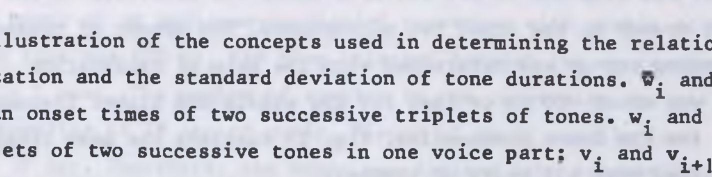


Fig. 27. Illustration of the concepts used in determining the relation between asynchronization and the standard deviation of tone durations. $\bar{w}_i$ and $\bar{w}_{i+1}$ are the mean onset times of two successive triplets of tones. w_i and w_{i+1} are the onsets of two successive tones in one voice 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.

4.3. Method

4.3.1. Recording

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 analyze 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 acoustical 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.

Recordings were made in the anechoic room of the Institute for Perception TNO in Soesterberg. The angle of least sensitivity of the directional microphones used was about $\pm 120^{\circ}$ relative to front. This is especially suitable for ensembles with three musicians. The three musicians were seated in a circle with the microphones in the middle directed towards them. The distance between instrument and microphone was about 1 m. (See Fig. 28). The sounds were recorded on three tracks of a four-channel tape recorder, with a tape speed of 38 cm/sec. (The fourth channel was used to record the sound from an omnidirectional microphone positioned at the midpoint of the set-up.) There was crosstalk of the instrumental sounds to the other two microphones. The amount of crosstalk of the recording set-up was determined with the help of calibration signals. It was about -20 dB or less for the middle and higher frequencies, but greater for the lower frequencies. Fig. 29 displays the mean crosstalk levels, averaged over all pairs of channels.

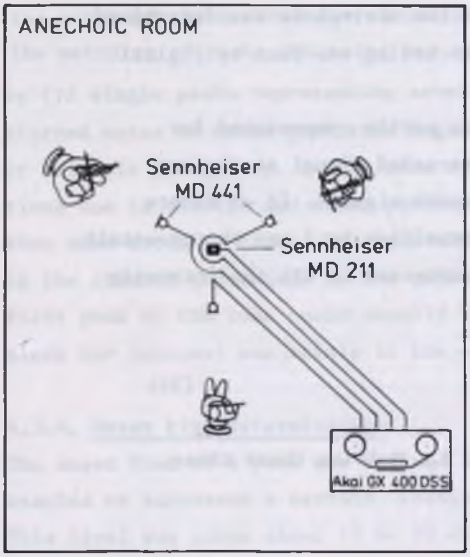


Fig. 28. Set-up for music recordings. By using directional microphones in an anechoic room each instrument is recorded on a separate channel of the tape recorder.

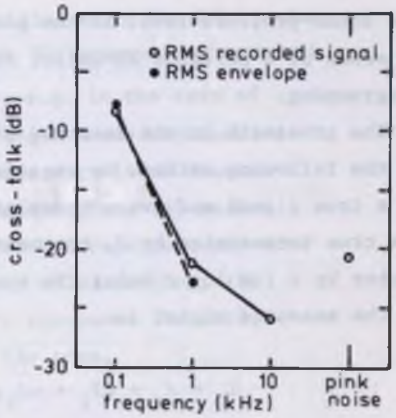


Fig. 29. Cross-talk as measured with calibration signals and averaged over all pairs of channels. The rms of the envelopes was determined by means of the same envelope detector as was used for the music analysis.

4.3.2. Signal analysis

All 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 dynamic range of the envelope detectors was about 60 dB. The three envelope signals were fed into the PDP 11/10 computer by means of an analogue-to-digital converter with 10-bit accuracy. The sampling rate was 200 Hz, the time interval between two successive samples being 5 msec. The number of samples in one conversion run was 3600, corresponding to 18 sec of music. Pieces of music usually last longer than 18 sec. Therefore, the music was divided into successive parts of 18 sec each.

The absolute level of the envelopes was determined by comparing it to the level of a specially recorded calibration signal of which the sound-pressure level at the position of the microphone was determined by means of a portable dB-meter. Further processing was done by digital programming.

The crosstalk in the recorded signals was partly compensated for by the following method. We regarded each recorded signal as the sum of a true signal and two uncorrelated crosstalk signals. If we denote the true intensities by J, the measured intensities by I and the crosstalk factor by c (10 log c being the crosstalk expressed in dB) the intensity of the measured signal is

$$I_1 = J_1 + cJ_2 + cJ_3. \qquad (22)$$

Similar equations can be written for I_2 and I_3. Solving these three equations with three unknowns, one finds

$$J_1 = \frac{I_1 + c(I_1 - I_2 - I_3)}{(1 + 2c)(1 - c)} \qquad (23)$$

and, by analogy, for J_2 and J_3. Strict application of this correction formula is impossible since the crosstalk is frequency-dependent. Choosing crosstalk values conservatively, it is possible to obtain a moderate compensation for the crosstalk in the signals.

4.3.3. Peak identification

The envelopes, both uncorrected and corrected for crosstalk, could be displayed on the graphic terminal of the computer. Paper copies of these display-screen presentations were made, which were used for further inspection of the envelopes. Fig. 30 gives an example of such graphic displays.

The next step in the analysis was to match the peaks in the envelopes with the notes of the score as played by the musicians. Peaks were defined as the regions between two successive minima in the envelopes. Minima that differed 3 dB or less from the surrounding maxima were excluded in the peak detection procedure. The peak detection in the sampled envelopes was performed by a computer. A number was assigned to each successive peak and its onset time (the determination of this feature is discussed in the next paragraph), offset time, duration, maximum level and mean level (rms) were determined. With the help of these data, the notes of the musical scores could be identified with specific peaks. The matching of peaks and notes was quite good. Differences were caused by (1) single peaks representing several notes, e.g. in the case of slurred notes or notes performed legato, (2) single notes represented by multiple peaks, e.g. in the case of long tones with level fluctuations due to vibrato and other performing techniques, and (3) peaks that were actually caused by crosstalk which did not represent any note in the channel concerned. In the second category of mismatches, the first peak of the tone could usually be used to represent the note, since our interest was mainly in the onset of the tone.

4.3.4. <u>Onset time determination</u>

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. This is actually an operational definition. As yet there has been little psycho-acoustical research as to the relevant perceptual beginning of a musical tone (see Vos & Rasch, in press). From the various a priori alternatives the one chosen seems to be the most reasonable. Other, rejected, alternatives were the moment the envelope started to rise and the -3 dB threshold relative to the maximum level. The moment the envelope started to rise was not suitable as a perceptual onset measure because the levels involved were too low. Moreover, measurements at these low levels are more seriously affected than those at higher signal levels. The moment of reaching -3 dB below maximum level of a peak, which is often employed as a definition of the end of the rise portion of an envelope, is not very stable relative to the internal temporal structure of an envelope. Using the passing of a threshold level halfway between minimum and maximum levels turned out to be a feasible compromise. Later experimental research by Vos & Rasch (in press) justified the choice of a threshold level of about 15 dB below maximum level for a definition of onset time.

Fig. 31 shows the envelopes of a fragment of 9 secs of music, with the peaks numbered according to the computer output and the score of the music inserted for comparison and identification purposes. Table X lists the onset times of the peaks in the fragment together with other information about the peaks.

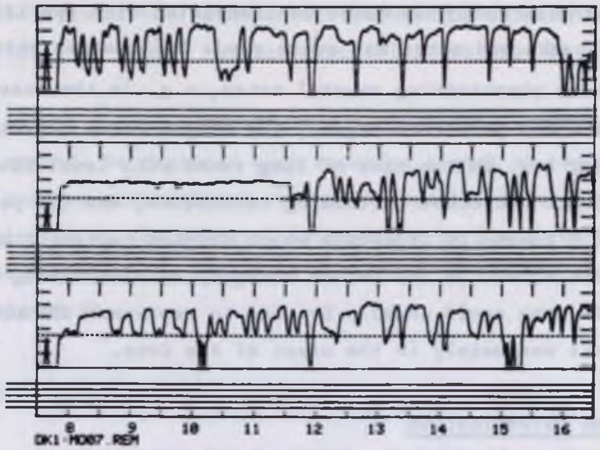


Fig. 30. "Unprocessed" envelopes of the three simultaneous voices of a fragment of music of 9 sec as displayed automatically on the graphic terminal of the computer.

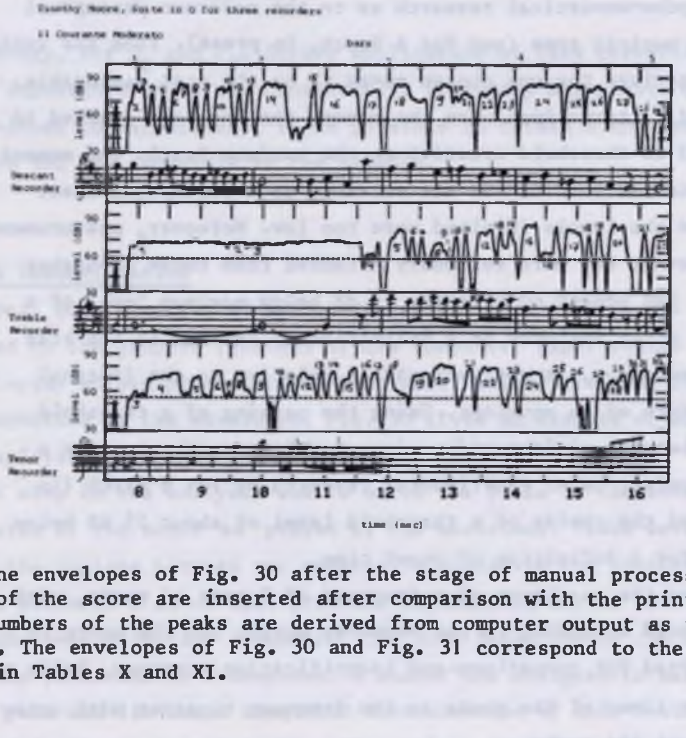


Fig. 31. The envelopes of Fig. 30 after the stage of manual processing. The notes of the score are inserted after comparison with the printed score. The rank numbers of the peaks are derived from computer output as illustrated in Table X. The envelopes of Fig. 30 and Fig. 31 correspond to the data presented in Tables X and XI.

Table X. Fragment of the output from the peak detecting programme. The peaks of each channel have been numbered and their onsets (passing of a threshold; absolute onset time), durations (from the onset to the downward passing of the threshold, or to the onset of the next peak) and maximum levels have been determined. The figures correspond to the envelopes displayed in Fig. 30. Table XI illustrates further statistical processing.

	Channel 1			Channel 2			Channel 3		
no.	onset (msec)	duration (msec)	level (dB)	onset (msec)	duration (msec)	level (dB)	onset (msec)	duration (msec)	level (dB)
1	7610	40	67	7855	1585	70			
2	7825	180	83	9440	305	68			
3	8005	125	82	9745	1835	70	7905	235	60
4	8175	80	76	11630	235	68	8140	695	74
5	8350	90	74	11995	315	79	8835	275	72
6	8525	90	79	12325	110	72	9120	300	76
7	8615	220	85	12485	100	72	9435	85	60
8	8875	80	74	12675	285	68	9520	325	75
9	9040	100	77	12995	10	57	9865	220	76
10	9225	170	81	13030	85	73	10265	125	71
11	9395	120	83	13190	80	75	10390	155	68
12	9570	95	77	13390	315	78	10555	335	75
13	9720	110	82	13720	165	77	10890	150	77
14	9905	490	85	13885	150	76	11040	185	77
15	10745	110	83	14105	310	84	11225	340	79
16	10915	620	83	14435	315	81	11565	150	76
17	11600	265	79	14800	305	79	11715	150	75
18	11970	640	84	15160	90	74	11955	355	77
19	12670	355	84	15315	135	77	12350	345	75
20	13025	90	74	15510	320	83	12695	405	82
21	13115	230	77	16035	105	76	13165	205	79
22	13405	305	81	16210	285	77	13410	405	80
23	13720	315	82				13915	135	73
24	14075	685	85				14125	475	76
25	14795	320	84				14615	180	74
26	15130	340	83				14795	220	72
27	15500	455	82				15350	120	72
28	16065	35	60				15470	205	70
29	16315	10	55				15675	155	74
30	16380	100	62				15830	155	79
31							15985	190	79
32							16200	270	79
33							16470	395	79

4.3.5. Statistical computations

The following step was to determine which peaks in the three envelopes corresponded to a triplet of simultaneous notes in the musical score. This had to be done by hand. The numbers of the quasi-simultaneous peaks were fed into the computer so that the relative onset times of the peak-triplets

Table XI. Peaks corresponding to simultaneous notes from the music fragment of Table X with absolute onset times, relative onset times and onset time differences.

Channel 1	2	3	absolute onset time (msec)			relative onset time (msec)			onset time difference (msec)		
Peak no.			w_1	w_2	w_3	v_1	v_2	v_3	$d_{1,2}$	$d_{2,3}$	$d_{3,1}$
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

and their onset time differences could be computed. Table XI shows the simultaneous onsets from Table X, the relative onset times and the onset time differences. Note that the temporal resolution of the processing system was 5 msec. 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. An onset time that was clearly affected by crosstalk was also excluded from further processing.

The statistical computations were mostly done per composition, sometimes per movement of a composition. First, means, standard deviations and rms's 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 the conversion of synchronization into isochronization values.

4.4. Musicians and music

Recordings were made of music played by three professional trio ensembles: (1) Ensemble Saratoga, consisting of Johanna Middelhof, Marjan van den Borren and Maria Wüst, recorders; recording date May 6, 1976. (2) Trio di Fiati, consisting of Cor Coppens, oboe, Gijs Karten, clarinet, and Piet Scheers, bassoon; recording date May 7, 1976. (3) A string trio consisting

Table XII. Data concerning the performed compositions that were analyzed for asynchronization.

Ensemble	Composer	Composition	Instruments	Duration (sec)
Saratoga	H.U. Staeps	Trio	1. treble recorder 1 2. treble recorder 2 3. tenor recorder	415
	T. Moore	Suite	1. descant recorder 2. treble recorder 3. tenor recorder	323
Trio di Fiati	A. Nudera	Divertimento	1. oboe 2. clarinet 3. bassoon	506
	Fr. Devienne	Sonatine	same as Nudera	155
Trio Van Helden	L. v. Beethoven	Andante con Variazione and Marcia from Trio, Op.8	1. violin 2. viola 3. cello	568
	E. von Dohnányi	Marcia from Serenade, Op.10	same as Beethoven	121

of Henriette van Helden-Verrijn Stuart, violin, Jo van Helden, viola, and F. Olivier, cello; recording date February 23, 1977. Each ensemble played several compositions, with a total duration of about 45 min. Two compositions per ensemble were chosen for synchronization analysis: (1) Saratoga: H.U. Staeps, Trio (Edition Doblinger Stp 308) with movements Allegro moderato, Adagio con moto, Allegretto grazioso; T. Moore, Suite in G (Edition Schott 10554) with movements Andante, Courante-Moderato, Sarabande-Maestoso, Allegro molto. (2) Trio di Fiati: A. Nudera, Divertimento C major with movements Allegro moderato, Menuetto, Andante, Allegro; F. Devienne, Sonatine D minor with movements Allegro, Allegretto. (3) String trio: L. v. Beethoven, from Serenade Opus 8: Andante quasi Allegretto con Var. 1-4, Allegro, Tempo I, Marcia-Allegro; E. v. Dohnányi, from Serenade Opus 10: Marcia. For instruments and durations, see also Table XII. After an initial period of getting used to the extreme acoustical conditions, the performers were not disturbed by those conditions to any significant degree. The recordings sound quite convincing as examples of well-performed ensemble music.

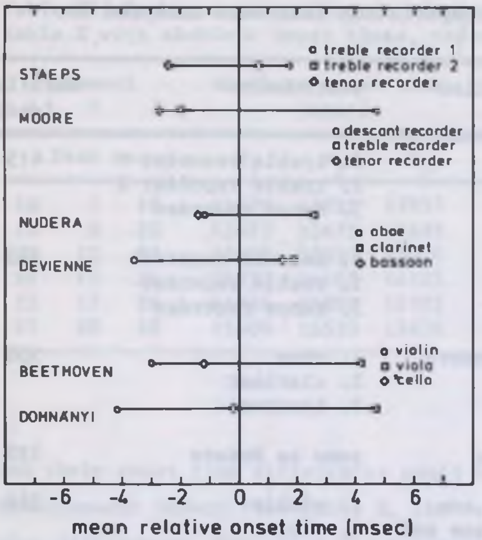


Fig. 32. Mean relative onset times of the instruments per composition. The values in this diagram correspond to Table XIII.

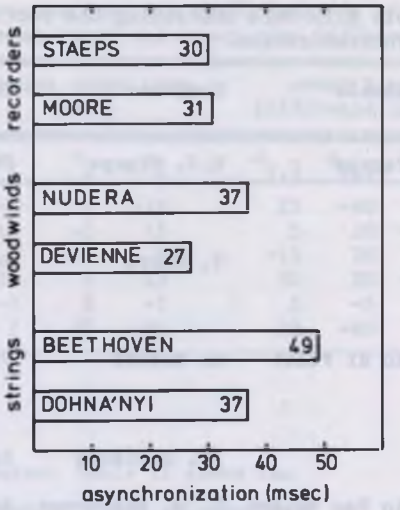


Fig. 33. Mean asynchronization per composition, averaged over the instruments. The values in this diagram correspond to Table XIV.

4.5. Results

4.5.1. Relative onset times

Means and standard deviations of relative onset times were calculated for each recorded composition (see Table XIII). The mean relative onset times are graphically presented in Fig. 32. 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, directing, 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.

Statistical significance of the mean relative onset times (as to their deviation from their mean, that is, zero) can be judged from the significance

Table XIII. Means ($\bar{v}_1$, $\bar{v}_2$, $\bar{v}_3$), standard deviations (s_1, s_2, s_3) and rms of standard deviations (S) of relative onset times in msec calculated for N triplets of onsets per composition. Subscripts refer to channels; see Table XII for instruments.

Composition	N	$\bar{v}_1$	$\bar{v}_2$	$\bar{v}_3$	s_1	s_2	s_3	S
Staeps	372	2	1	-2	18	16	18	17
Moore	215	5	2	-3	18	14	19	17
Nudera	419	-1	3	-1	20	23	19	21
Devienne	162	-3	2	2	15	18	13	15
Beethoven	361	-3	4	-1	30	26	28	28
Dohnãnyi	111	-4	5	0	23	21	21	21

Table XIV. Means ($\bar{d}_{1,2}$, $\bar{d}_{2,3}$, $\bar{d}_{3,1}$), standard deviations ($s_{1,2}$, $s_{2,3}$, $s_{3,1}$) and rms of standard deviations (A) of onset time differences (asynchronizations) in msec calculated for N triplets of onsets per composition. In parentheses is the median unsigned onset time difference, estimated as 0.68A. Subscripts refer to channels; see Table XII for instruments. Significance of the mean differences as determined by the t-test is indicated by asterisks (*: $p \leq 5\%$; **: $p \leq 1\%$).

Composition	N	$\bar{d}_{1,2}$	$\bar{d}_{2,3}$	$\bar{d}_{3,1}$	$s_{1,2}$	$s_{2,3}$	$s_{3,1}$	A
Staeps	372	-1	-3*	4*	29	32	29	30 (20)
Moore	215	-7**	-1	7**	28	35	28	31 (21)
Nudera	419	4	-4	0	40	32	38	37 (25)
Devienne	162	5*	0	-5**	30	23	27	27 (18)
Beethoven	361	7**	-5*	-2	50	52	45	49 (33)
Dohnãnyi	111	9*	-5	-4	39	38	35	37 (25)

of the respective onset time differences as determined with the t-test for differences. These significances are indicated in Table XIV. A number of mean onset time differences are significant. However, it is with great reservation that we accept this significance and give interpretations like the ones just presented, because the means are small compared to the standard deviations, and they are of the same order of magnitude as the time unit of the measurement (5 msec). On the other hand, the means are based on fairly large numbers of samples.

4.5.2. Onset time differences: asynchronization

There are three distributions of onset time differences for each composition. The root-mean-squares of the standard deviations are used as measures for asynchronization (see Table XIV). Asynchronization per composition is graphically illustrated in Fig. 33. Asynchronization values fall largely within the range of 30 to 50 msec. There are some systematic differences between the ensembles. The recorder ensemble shows a relatively small asynchronization

Table XV. Asynchronization and tempo per movement. Inter-beat time (the reciprocal of metronome value) is used as a measure for tempo. Metronome indications stated in the score are given in parentheses after the tempo indication.

Composition	Tempo indication		Asynchronization (msec)	Metronome value (beats/min)	Inter-beat time (sec)
Staeps	I	Allegro moderato (𝅗𝅥 = 116)	31	𝅗𝅥 = 120	0.50
	II	Adagio con moto (♩ = 60)	36	♩ = 66	0.91
	III	Allegro grazioso (𝅗𝅥 = 132)	27	𝅗𝅥 = 130	0.46
Moore	I	Andante (𝅗𝅥 = 72)	31	𝅗𝅥 = 60	1.00
	II	Moderato (Courante) (𝅗𝅥 = 80-84)	23	𝅗𝅥 = 85	0.85
	III	Maestoso (Sarabande) (♩ = 66)	35	♩ = 75	0.75
	IV	Allegro molto (♩ = c.144)	30	♩ = 120	1.20
Nudera	I	Allegro moderato	32	𝅗𝅥 = 60	1.00
	II	(Menuetto)	36	♩ = 90	0.67
	III	Andante	46	𝅗𝅥 = 60	1.00
	IV	Allegro	24	♩ = 120	0.50
Devienne	I	Allegro	25	𝅗𝅥 = 80	0.75
	II	Allegretto	28	♩ = 120	0.50
Beethoven	I	Andante quasi Allegretto	53	♩ = 48	1.25
	II	Allegro	38	♩ = 80	0.75
	III	Andante quasi Allegretto	73	♩ = 40	1.50
	IV	Allegro	34	♩ = 120	0.50
Dohnányi		(Marcia)	37	♩ = 110	0.54

of about 30 msec. The two examples of the wind trio have different amounts of asynchronization: 27 and 37 msec respectively. The largest asynchronizations are found in the string trio examples: 37 and 49 msec. 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 msec). These longer rise times permit a greater leniency as to synchronization.

The mean ratio of the standard deviation of the onset time differences (asynchronizations) and the standard deviation of the relative onset times was 1.7765, which is very close to the theoretically expected value of $3^{\frac{1}{2}} = 1.7321$ (see Equation 13).

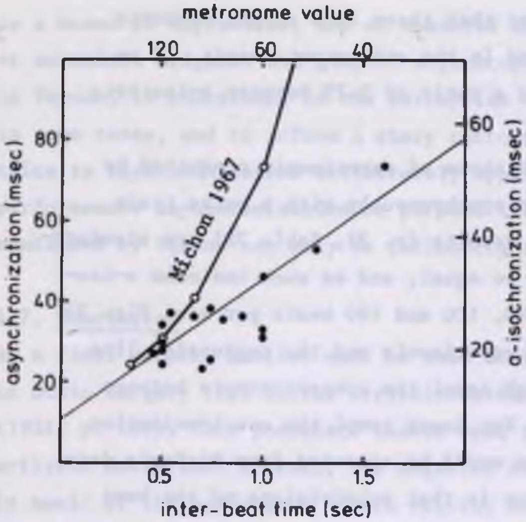


Fig. 34. Relation between tempo and asynchronization. Inter-beat time is used as a measure of tempo. The vertical scales indicate asynchronization, defined on the left as the standard deviation of onset time differences, averaged over instruments. On the right, a-isochronization is indicated. The dots are data points for movements. The regression line of tempo on asynchronization is based on the correlation between inter-beat time and asynchronization (see Table XV). The results of Michon's experiments (1967, p. 31) are inserted for comparison.

4.5.3. Asynchronization and tempo

The relation between asynchronization and tempo was also investigated. The data used in determining this relation are given in Table XV (see Fig. 34). 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 was 0.80. Evidently, the conclusion is justified that faster tempo goes with more synchronization, and slower tempo goes with less synchronization.

4.6. Relations with other research

4.6.1. Isochronization

As already noted in section 4.2, 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. If we insert r = 0.21 in Equation 21 we find a ratio of 0.73 between a-isochronization and asynchronization.

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 by means of 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 min. Fig. 34 includes both the results of Michon's experiments and the regression line for asynchronization to tempo. For high tempi the correspondence between Michon's data and ours is quite good. For lower tempi the synchronization found in our recordings is better than would be expected from Michon's data. A likely explanation of this difference is that subdivisions of the beat also can contribute to synchronization.

4.6.2. Synchronization of piano chords

The only synchronization measurements available in the literature are those of Vernon (see Seashore 1938, p. 248-253, the original publication was not available for inspection). Vernon investigated the synchronization of piano performances. He used rolls for player pianos. By measuring the positions of the holes on the rolls a time resolution of 10 msec could be reached. In Seashore's description the quantitative measure used, called "deviation", is not properly defined. Therefore, a direct comparison of Vernon's results with ours is not possible. However, the order of magnitude of the asynchronization of piano chords seems to be about the same as that of simultaneous tones in ensemble music.

There are a number of differences between piano and ensemble performances that must be kept in mind when interpreting or comparing synchronization data for both kinds of musical performance. Firstly, motor coordination between simultaneous tones is possible for a piano player, but impossible for ensemble players. (Four-handed piano playing would be an interesting case for research.) Secondly, intentional asynchronization is quite possible in piano playing, but difficult and risky for an ensemble. Thirdly, the playing techniques and the piano mechanism can be causes of unwanted asynchronization. Ensemble players can usually concentrate on one tone at a time.

According to Vernon, asynchronization in piano playing is the effect of wanted asynchronization applied either deliberately or subconsciously as a means of expression, and of unwanted asynchronization from mechanical or motorical origins. The goal of asynchronizations is, still according to Vernon, to contribute to the perception of single tones in certain chords in some cases, and to soften a sharp contour in other cases. All this pertains to asynchronization deliberately applied or permitted. In ensemble performances asynchronization on purpose is a rare phenomenon. The goals mentioned by Vernon can only be indirectly relevant for ensemble playing.

4.7. Conclusion

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.

CHAPTER FIVE

OBSERVATIONS FROM MUSICAL PERFORMANCE AND COMPOSITION

5.1. Polyphonic music in performance

5.1.1. Spectral factors

A single part in a polyphonic composition is in a rather unfavourable situation with respect to mean signal-to-noise ratio in performance. If we assign equal sound-pressure levels to the different voice parts in a performance, we can calculate a signal-to-noise ratio for a single voice, in such a way that the voice is considered the signal, the other voices the noise. In a two-voice composition the S/N ratio of each voice is 0 dB, in a three-voice composition -3 dB, in a four-voice composition -4.8 dB, etc. Fig. 35 depicts the S/N ratios of a single voice in an ensemble of equally loud voices.

If a voice is to be perceived separately, this unfavourable S/N ratio has to be overcome in one way or another. Actually, it is not known what S/N ratio is needed for a single musical voice to be perceived separately against a background of other voices. There are several ways for a musician to make his voice conspicuous within the context of other voices. The simplest way is to play at higher sound-pressure levels. The need for solo instruments that could be heard clearly over an accompanying orchestra has led to the development of so-called "concert instruments" with a powerful tone. This development is especially evident in the history of piano, violin and flute in the 19th century. The late-19th-century examples of these instruments are capable of producing sound-pressure levels that are much higher than their predecessors of about a century earlier. In an ensemble, a player can give extra weight to his voice by playing louder. In performing polyphonic, contrapuntal compositions (e.g. fugues by Bach) on a piano, frequent use of dynamic accents is made in order to make it possible to hear thematic material in the middle voices.

Usually there is a lot of mutual masking of components of different tones played simultaneously. However, since in general the higher harmonics of a tone have lower levels than the lower harmonics, the tones of the highest voice are in a favourable position. There is a fair chance that its harmonics are a few dB stronger than those of lower voices in the same frequency regions, so that the highest tones can overcome the masking by the other tones. On the other hand, the middle voices are very likely to be masked by each other,

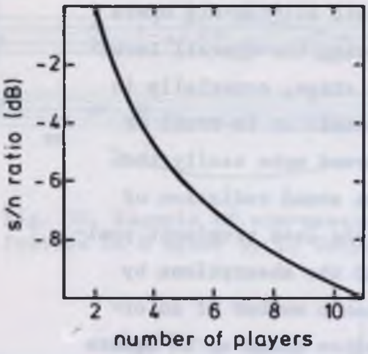


Fig. 35. Signal-to-noise ratio of a single player in an ensemble of n equally loud players.

by the top voices and perhaps also by lower voices with relatively strong upper partials. The lower partials of the tones of the lowest voices are, again, more or less free from masking effects. This situation accounts for the compositional experience that the top voice is most suitable for melody playing, that the bass line should also be treated carefully, but that the middle voices function mainly as supporting, filling-up voices.

Low instruments (cello, bassoon, bass voice) are in a disadvantageous starting position as solo instruments. In the case of the cello as a solo instrument, very frequent use is made of its top register, not lower than the violin's middle register. In the case of a bass voice as a solo part with orchestral accompaniment, such high fundamental frequencies are impossible, of course. It has been noted, however, that good male solo singers make use of an extra formant, the so-called singing formant (Sundberg 1977). This formant lies in the region of 2000 to 3000 Hz and provides strong harmonics in this region, harmonics that are stronger than the harmonics of tones from the accompanying ensemble in the same frequency region. This singing formant can help the singer to be heard over the accompaniment.

When music consisting of solo and accompanying parts is recorded on tape or reproduced via some kind of electro-acoustical system, the S/N ratio of the soloist can be manipulated at will by the placement of microphones and the use of adjusted amplifier settings for the various components of the musical sound. In popular music this is an essential means of making the vocalist (who has often a limited vocal technique from a classical point of view) heard against his/her accompanying background.

If a soloist is not aided by electro-acoustical means, the relative spatial position of soloist and accompaniment can be used to influence the S/N ratio of the soloist in a positive way. Wagner hid his all-too-big opera orchestra more or less below the stage, thus decreasing the overall level of its sound relative to that of the singers on the stage, especially in the higher frequency regions. If a solo violinist stands up in front of the orchestra, the sound of his instrument can be heard more easily than that of his sitting colleagues in the orchestra. The sound radiation of the orchestra violins is less efficient because of the less prominent position of the instrument with regard to the public and the absorptions by music stands, music papers, clothes, seats, etc. When a member of an orchestra plays an obligato solo part, he will very often stand up to ensure an efficient sound radiation. In some scores players of reed wind instruments are instructed to hold their instruments high, probably also to help the tones reach the ears of the public.

Because we have two ears, a different localization of the musicians leads to a different directionality of the sounds they play, which can also be helpful in distinguishing simultaneous sounds. This feature can be and is very successfully applied in stereo recording and reproducing.

It is possible that the often noted sharp intonation of, say, solo violinists, has to do with the better perception of their tones because of their intonation technique. Vibrato playing can have the same function (see section 2.6.)

5.1.2. Temporal factors

To end this section, we will point to a special technique of prominence that gets strong support from our psycho-acoustical experiments described in Chapter 2. It is the use of minute temporal differences, especially onset time difference between the various simultaneous parts (tones), in order to improve separate perception. These onset time differences were found to be a very strong factor in laboratory experiments. A number of observations from musical practice point in the same direction.

In today's "authentic" performance of early music on the harpsichord, asynchronization of simultaneous tones is a fundamental aspect of playing. Since harpsichord music is very often polyphonic/contrapuntal, this technique is of special significance, because it facilitates the perception of the polyphonic musical structure. Striking the keys exactly simultaneously results in a heavy sound which obscures the single tones. We give two examples of harpsichord music in which de-synchronization is led to the extreme. The first example has been taken from a Prélude non mesuré by Louis Couperin (Fig. 36), the other from a keyboard piece by Forqueray (Fig. 37) in which the asynchronization of both hands has been indicated

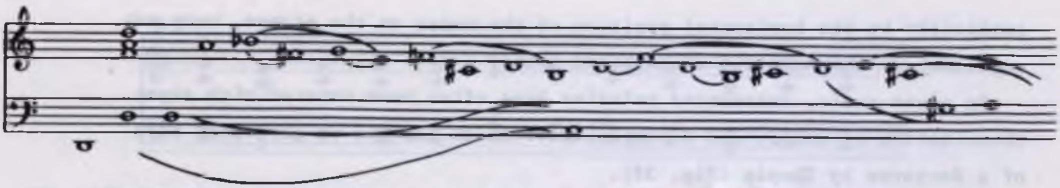


Fig. 36. Example of non-measured, desynchronized music: the beginning of a Prélude in D minor by L. Couperin (17th century; ed. 1959, p. 1; ed. 1972, p. 72).

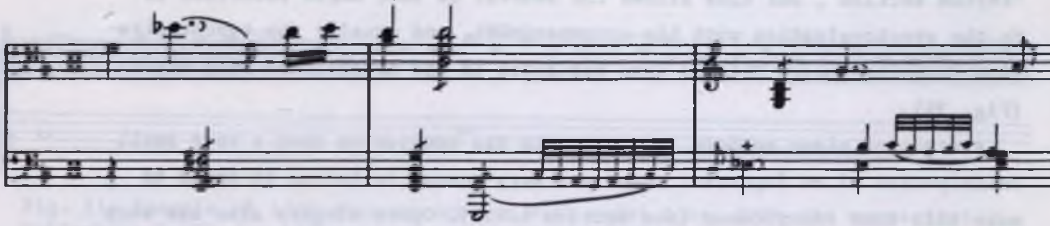


Fig. 37. Example of asynchronization brought into notation: the beginning of the Sarabande La d'Aubonne by A. Forqueray (18th century; ed. 1974, p. 74).

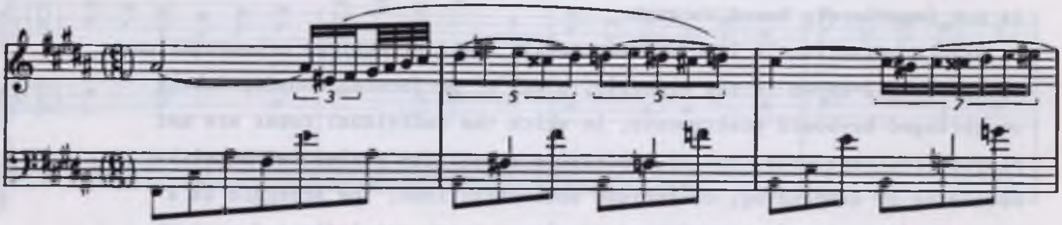


Fig. 38. Example of asynchronization between melody and accompaniment: a fragment from the Nocturne Op. 9, no. 3 by F. Chopin (1830-1831; ed. 1974, p. 74).

Fig. 39. Example of asynchronization in popular music: the playing of the final chord before the beat and here given in notation. The example shows the final measures of Scott Joplin's rag The Entertainer (1902).

explicitly by the horizontal position of the notes on the staves, very exceptional in the notation of keyboard music.

In piano music, ornamented melodies have often been notated with asynchronization of melody and accompaniment. As an example we give some bars of a Nocturne by Chopin (Fig. 38).

In blues, jazz, soul and pop music, melody, accompaniment and rhythm have been divided among different instruments or groups of instruments. The accompaniment usually follows closely the temporal structure given by the "rhythm section", but this allows the soloist to take ample liberties as to the synchronization with his accompaniment, and usually the synchronization is quite loose. This is even reflected in the notation of this music (Fig. 39).

A study of piano methods led Vernon to the conclusion that a very small anticipation of an important tone is a very common technique in order to make this tone conspicuous (see section 4.6.2). Opera singers also use such anticipations, singing just before the beats, for the same purposes. It is to be expected that instrumental soloists (violinists, oboists, etc.) also make use of this simple but refined technique, refined because it is not immediately heard as such.

Asynchronization is involved in certain forms of musical ornamentation. The most well-known is the arpeggio, a way of performing chords, mostly on stringed keyboard instruments, in which the individual tones are not played simultaneously, but as successive tones. The series can be either ascending or descending, or include both directions. The arpeggio is a very common device in harpsichord playing, though not indicated as such in the score. In piano playing it is also used quite often. A less well-known temporal ornament is the 18th-century French suspension (to be distinguished from the suspension as a dissonance on a strong beat that has to be resolved on the following weak beat). A suspension implies a clearly audible retardation of the tone. It is used to put a special stress on that tone. The suspension can also be used in combination with other ornaments such as trills and mordents (see Brunold 1965, p. 37-40).

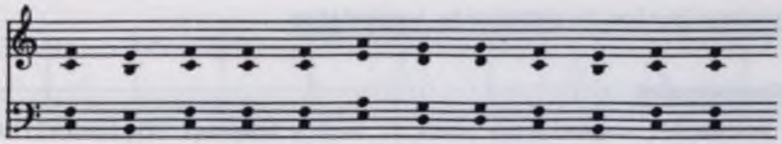


Fig. 40. Example of medieval parallel organum. From the theoretical treatise Musica Enchiriadis by Hucbald (9th century; ed. Gerbert 1764, p. 24).

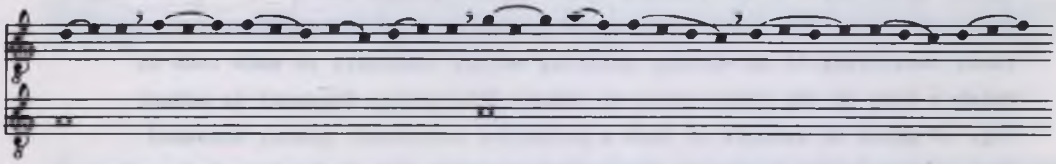


Fig. 41. Example of 13th-century organum with pedal tone. From the manuscript Wolfenbüttel 677 (or 628; ed. Baxter 1931), fol. 39v.

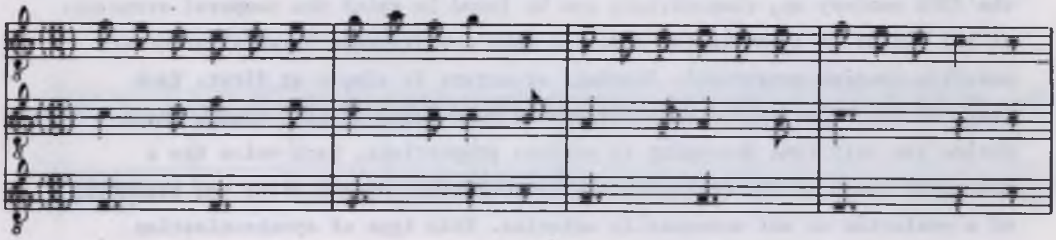


Fig. 42. Example of 13th-century polyphony with three voices with their own rhythmic structures. From the Montpellier manuscript (ed. Rockseth II, 1936, p.123).

Fig. 43. Example of polyphony showing the principle of alternation. From J. Tinctoris' Liber de arte contrapuncti, 1477 (ed. 1961, p. 111).

5.2. Asynchronization in polyphonic composition

5.2.1. Counterpoint

In many polyphonic compositions there is such a temporal (rhythmical) structure that the onsets of the tones in the different voices coincide only partly or not at all. The various voices are asynchronized at macro-level, the level of composition and notation. We call this type of asynchronization macro-asynchronization. In polyphonic music macro-asynchronization is a common procedure. In the oldest, medieval, polyphony, macro-asynchronization does not occur. The first examples of polyphony show note-against-note composition, consisting of an already existing melody (Cantus), to each tone of which a tone of the other voice or voices (Discantus, Organum) is added. Fig. 40 gives an example of such a primitive polyphonic piece, borrowed from a theoretical treatise. The oldest examples of polyphonic writing with macro-asynchronization are the organa of the 13th century. The Cantus consists of long, unmeasured tones, the Organum voice of a sequence of shorter tones (melisma) that does not have a strict rhythmical structure (Fig. 41). From the 13th century on, compositions can be found in which the temporal structure of all voices is specified in order to make a polyphonic temporal structure possible (musica mensurata). Rhythmic structure is simple at first. Each temporal unit (perfectio) contains one or more tones in each voice, which divide the unit time according to various proportions. Each voice has a tone with onset at the beginning of the perfectio. Onsets after the beginning of a perfectio do not necessarily coincide. This type of synchronization and asynchronization is a characteristic of much polyphonic music from the 13th and 14th centuries (Fig. 42). This type of macro-asynchronization, consisting of non-coinciding onsets in single voices preceded and followed by coinciding onsets, will be called diminution. Diminution is an asymmetrical form of asynchronization. All onsets of one voice coincide with an onset of the other voice(s). Only some of the onsets in the other voice(s) coincide with onsets in the first voice.

During the 14th century a new type of asynchronization arises, which will be called alternation. Alternation is a succession of non-coinciding onsets in several voices without intervening coinciding onsets. This type of asynchronization is related to the concept of syncope in music theory. An example of alternation in polyphonic composition is given in Fig. 43. Alternation is a symmetrical form of asynchronization.

Diminution and alternation suffice as descriptive categories of macro-asynchronization in theory and practice, both for the early and the later

Fig. 44. Two-voice examples of the five species of polyphonic music writing, as given by J.J. Fux in his Gradus ad Parnassum (1725). The examples occur on p. 47, 59, 66, 74, and 77, respectively.

polyphony. Contrapuntal-didactic theory makes use, from the 17th century on, of five illustrative species, as prototypes of the rhythmical (temporal) relations between voices. The first species is note-against-note counterpoint. There is no macro-asynchronization. The second and third species include structures with two or more tones in one voice against one tone in another voice. In these species diminution is specifically applied. The fourth species is syncopated counterpoint, with the alternation type of macro-asynchronization. The fifth species, called contrapunctus floridus, comprises all kinds of rhythmical organization in an artistic way. Both diminution and alternation can be found as ways of macro-asynchronization. The classical formulation of the five species is the one by Joannes Josephus Fux. Fig. 44 gives examples of all the species, borrowed from Fux's famous Gradus ad Parnassum (1725).

5.2.2. Theory of macro-asynchronization

In the following paragraphs we will develop a quantitative measure for macro-asynchronization, called the asynchronization index B. We will use the following terminology. A voice has a certain number of onsets. Some of these coincide with onsets in other voices. These are the coinciding onsets or common onsets. The onsets that do not coincide with onsets in other voices will be called single onsets. The number of onsets of a voice will be denoted by p, the number of coinciding or common onsets by c, the number of non-coinciding or single onsets by s. The total number of onsets of a piece or of a group of voices will be denoted by t. In this total the coinciding onsets are counted as one onset.

The (macro-)asynchronization of a voice is defined as the ratio of the number of non-coinciding onsets to the total number of onsets of that voice:

$$B = \frac{s}{p} = \frac{p - c}{p} = 1 - \frac{c}{p} \qquad (24)$$

in which p - number of onsets of the voice v
s - number of non-coinciding onsets
c - number of coinciding onsets
B - asynchronization of voice v

For this asynchronization index it is necessary to indicate for which and relative to which voices asynchronization has to be determined. The index becomes zero if all onsets of the voice coincide with the onsets in the other voice or other voices. It becomes one if no onsets coincide with onsets in other voices. The asynchronization of a pair of voices can be defined as the ratio of the sum of the non-coinciding onsets of the voices to the sum of the numbers of onsets of the voices:

$$B_{1,2} = \frac{s_1 + s_2}{p_1 + p_2} = 1 - \frac{2c}{p_1 + p_2} = \frac{t - c}{t + c} \qquad (25)$$

in which $B_{1,2}$ - macro-asynchronization of a voice pair
p_1 - number of onsets in voice 1
p_2 - number of onsets in voice 2
c - number of coinciding onsets
t - total number of onsets of voices 1 and 2 ($t = p_1 + p_2 - c$).

If the number of onsets in both voices is equal ($p_1 = p_2 = p$), the equation can be simplified to

$$B_{1,2} = 1 - \frac{c}{p}. \qquad (26)$$

If the macro-asynchronization is purely diminutive, with a number of tones in one voice against one tone in the other voice, asynchronization is equal to

$$B_{dim} = \frac{q - 1}{q + 1} \qquad (27)$$

in which q is the number of tones in one voice against one tone in the other voice(s). The asynchronization of a piece of music can be defined as the arithmetical mean of the asynchronizations per voice pair:

$$B = \frac{B_{1,2} + B_{1,3} + \ldots + B_{n-1,n}}{\frac{1}{2}n(n-1)}$$

$$= \frac{2}{n(n-1)} \sum_{i=1}^{i=n-1} \sum_{j=i+1}^{j=n} \left(1 - \frac{2c_{ij}}{p_i + p_j}\right)$$

$$= 1 - \frac{4}{n(n-1)} \sum_{i=1}^{i=n-1} \sum_{j=i+1}^{j=n} \frac{c_{ij}}{p_i + p_j}. \qquad (28)$$

When the numbers of onsets per voice and the number of common onsets per voice pair are not too different, c_{ij}, p_i and p_j can be replaced by their average values $\bar{c}$ and $\bar{p}$:

$$B = 1 - \frac{\bar{c}}{\bar{p}}. \quad (29)$$

The use of equation 28 requires a lot of statistical computation. An estimate of the asynchronization of a piece of music, which requires much less counting and tallying, can be carried out in the following way. If we assume that the macro-asynchronization is equally and evenly distributed over all voices of a composition, it is possible to determine the relation between the ratio of the mean number of onsets per voice to the total number of onsets of a composition (the coinciding onsets counted as one onset) and macro-asynchronization. We assume equal numbers of onsets per voice, $\bar{p}$. One voice has $\bar{p}$ onsets. A second voice adds $B\bar{p}$ onsets that are not coincident with the first voice. A third voice adds $B^2\bar{p}$ "new" onsets, and so on. In formula:

$$t = \bar{p} + \bar{p}B + \bar{p}B^2 + \ldots + \bar{p}B^{n-1}$$

or

$$t = \bar{p} \sum_{i=0}^{i=n-1} B^i.$$

Now, the ratio of the mean number of onsets per voice and the total number of onsets is equal to:

$$\frac{\bar{p}}{t} = 1/ \sum_{i=0}^{i=n-1} B^i. \quad (30)$$

Fig. 45 presents this ratio for n = 2 to 6 (two to six-voice composition) as a function of macro-asynchronization. If this ratio is known, if it possible to estimate the mean macro-asynchronization. This estimation is much easier than exact calculation. In practice, it turns out that the estimation of macro-asynchronization from the ratio $\bar{p}/t$ results in an overestimation.

5.2.3. Macro-asynchronization in polyphonic composition

As sample compositions for the numerical analysis of macro-asynchronization we chose a number of vocal polyphonic compositions from the Musae Sioniae, Fünfter Theil (1607), and Sechster Theil (1609) by Michael Praetorius.

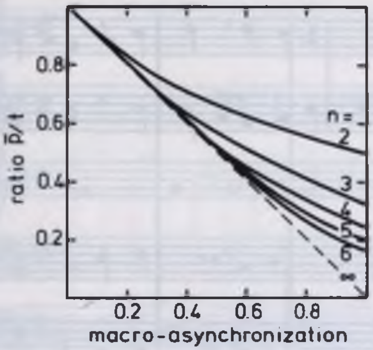


Fig. 45. The ratio of the mean number of onsets per voice in a composition to the total number of onsets in the composition. The curves are based on the assumption that the macro-asynchronization is spread equally over the voices. Parameter is the number of voices in the composition.

Praetorius's works are fairly representative examples of the eclectic ecclesiastical polyphonic style that developed in all European countries toward the turn of the seventeenth century. The Musae Sioniae volumes contain settings of Lutheran hymn tunes. Very often a single tune is present in many different settings, for a different number of voices and with different voice textures. The number of voices varies from two to six. The settings can be roughly divided into homophonic settings and polyphonic settings. Tables XVI-XVII contain the numerical data which result from the analysis of macro-asynchronization. Table XVI refers to the thirteen settings of Nun bitten wir den heiligen Geist (Fig. 46), Table XVII to the nine settings of Allein Gott in der Höh sei Ehr. The distinction between homophonic and polyphonic settings is clearly visible in the asynchronization indices. Homophonic compositions have asynchronizations of about 10% or less, polyphonic compositions have asynchronizations from 30 to 60%. If we average the asynchronizations of all two-voice, all three-voice, etc. polyphonic compositions, we obtain the following figures: two voices 34%, three voices 34%, four voices 36%, five voices 43%, and six voices 49%. This means that in polyphonic compositions the asynchronization increases with the number of voices. We may consider macro-asynchronization as a method of emphasizing the presence of different simultaneous voices. Since the signal-to-noise ratio of a single voice in a polyphonic composition becomes less with an increasing number of other voices (Fig. 35) the increasing macro-asynchronization can be viewed as a technique to keep the single voices as perceptible as possible.

A comparison of the averages of asynchronization per voice within a composition gives an insight into the evenness of spread of asynchronization over the various voices of a composition. In some compositions, like the six-part Nun bitten wir (Musae Sioniae V, 6) and Allein Gott

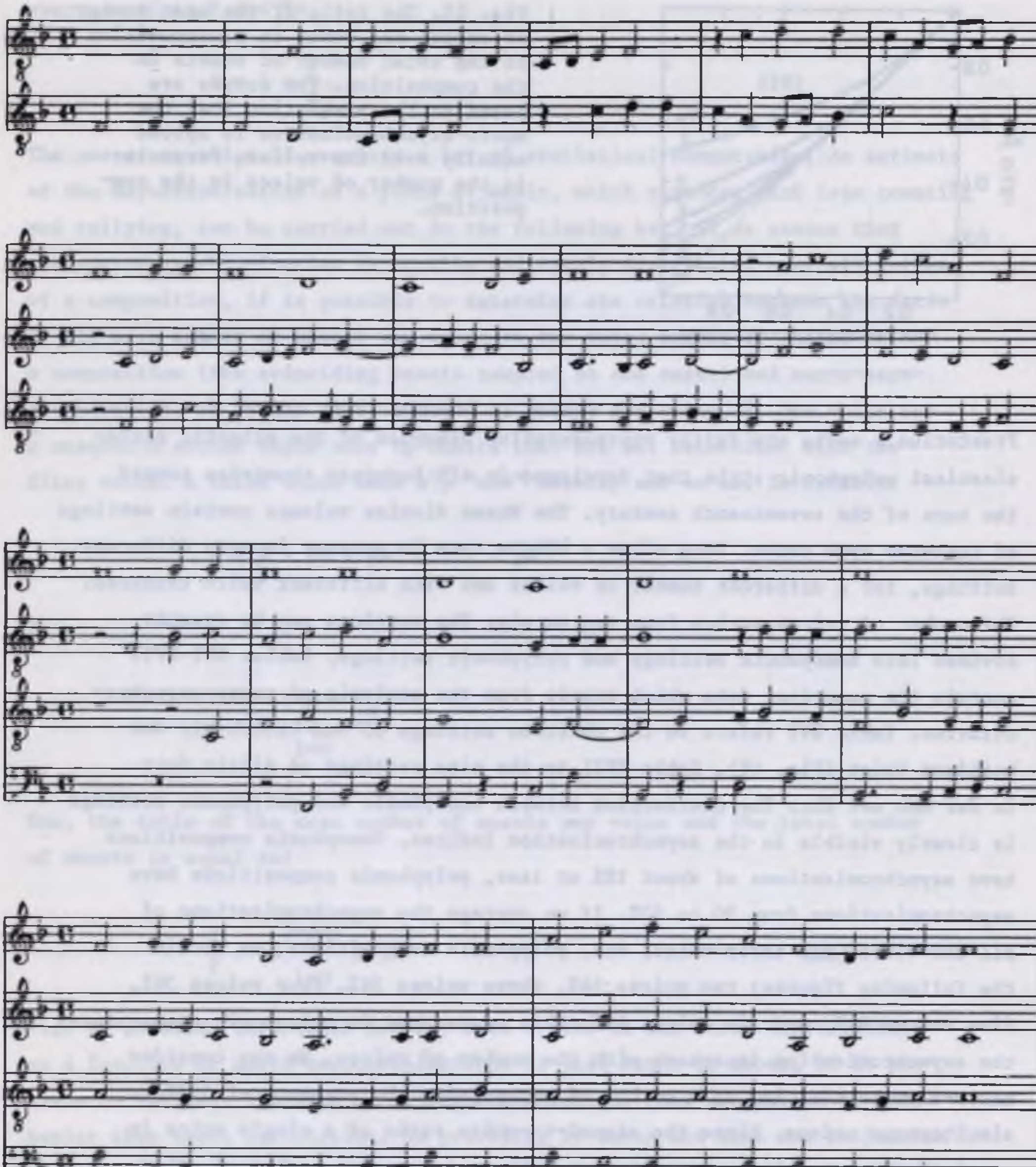


Fig. 46. Incipits and fragment of a number of settings of Nun bitten wir den Heiligen Geist, from M. Praetorius' Musae Sioniae, V and VI (1607, 1609), which are analyzed for macro-asynchronization. The examples given are the incipits from nos. 4, 5, 7 (from Vol. V), no. 157 (from Vol. VI; homophonic) and nos. 8, 9 (fragment) and 11 (from Vol. V).

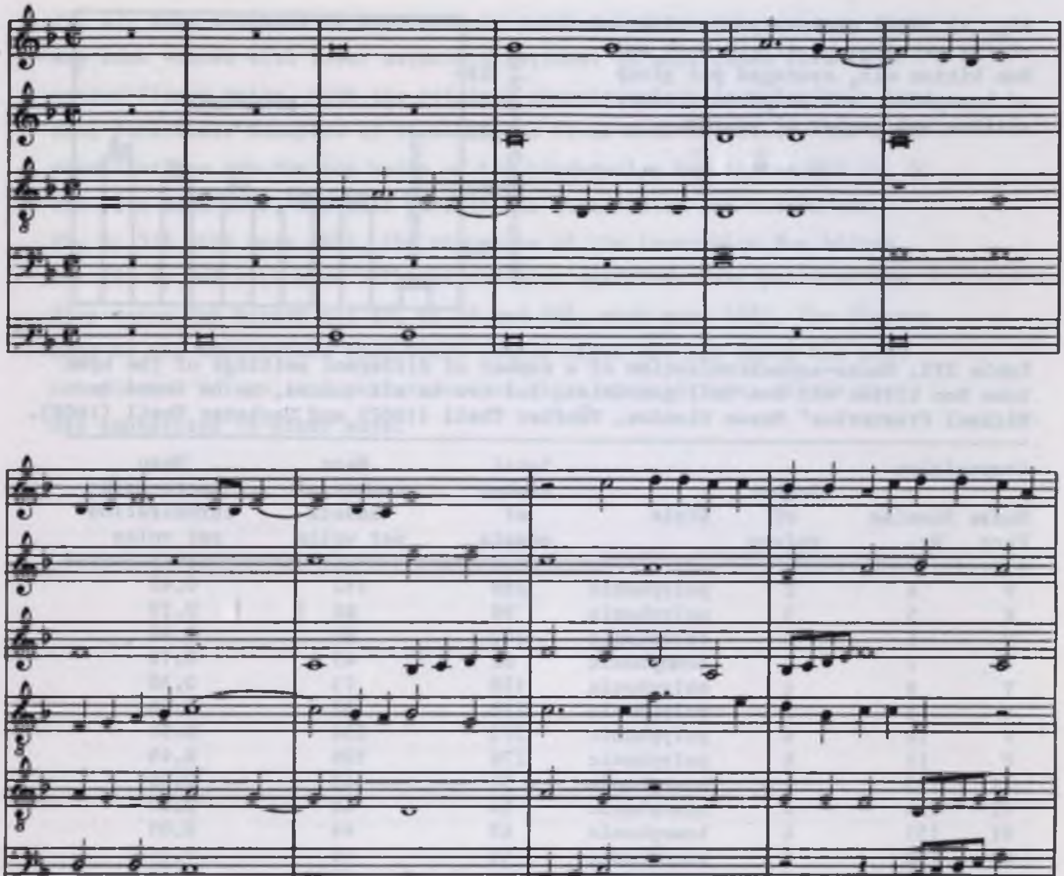


Fig. 47. Mean macro-asynchronizations of the settings of Allein Gott and Nun bitten wir, averaged per group of homophonic and polyphonic compositions and number of voices.

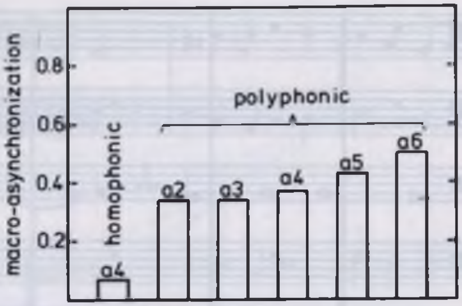


Table XVI. Macro-asynchronization of a number of different settings of the hymn tune Nun bitten wir den heiligen Geist, for two to six voices, to be found in Michael Praetorius' Musae Sioniae, Fünfter Theil (1607) and Sechster Theil (1609).

Composition Musae Sioniae Part	Nr.	Number of voices	Style	Total number of onsets	Mean number of onsets per voice	Mean macro-asynchronization per voice
V	4	2	polyphonic	269	192	0.40
V	5	3	polyphonic	96	66	0.29
V	6	3	polyphonic	140	80	0.46
V	7	4	homophonic	62	47	0.16
V	8	4	polyphonic	128	73	0.38
V	9	5	polyphonic	120	82	0.46
V	10	6	polyphonic	575	250	0.51
V	11	6	polyphonic	276	106	0.49
VI	155	4	homophonic	52	48	0.05
VI	156	4	homophonic	54	48	0.07
VI	157	4	homophonic	45	44	0.01
VI	158	4	homophonic	46	46	0.01
VI	159	4	homophonic	42	42	0.00

Table XVII. Macro-asynchronization of a number of different settings of Allein Gott in der Höh in Praetorius' Musae Sioniae.

Composition Musae Sioniae Part	Nr.	Number of voices	Style	Total number of onsets	Mean number of onsets per voice	Mean macro-asynchronization per voice
V	18	2	polyphonic	194	150	0.28
V	19	3	polyphonic	153	107	0.29
V	20	4	polyphonic	227	144	0.33
V	21	4	homophonic	68	62	0.06
V	22	5	polyphonic	294	157	0.40
V	23	6	polyphonic	387	185	0.47
VI	168	4	homophonic	72	63	0.09
VI	169	4	homophonic	81	65	0.11
VI	170	4	homophonic	79	70	0.07

(V, 23) this evenness is surprisingly good. In other compositions, there are some voices with lower asynchronizations. In most cases this is a cantus-firmus voice, with the original chorale melody in relatively long tone durations. Examples of these cantus firmi with relatively low asynchronizations are the top voice of the three-voice Nun bitten wir (V, 5; 16% with mean 29%), the bass part of the three-voice Nun bitten wir (V, 6; 31% with mean 46%), the top voice of the four-voice Nun bitten wir (V, 8; 26% with mean 38%), and the alto and first bass parts of the five-voice Nun bitten wir (V, 9; 29 and 26%, with mean 46%). The figures lead to the expectation that these cantus firmi are not among the most conspicuous voices in a performance of these compositions, unless they are emphasized in other ways.

CHAPTER SIX

CONCLUSIONS

The perception of music is a complex process. The complexities are even greater in the case of polyphonic music than in the case of homophonic music. It cannot be expected that a single factor or a simple combination of factors will suffice for the description of the perceptual processes involved. Auditory processes can be roughly divided into two groups. One group refers to the spectral structure of the sound: frequency analysis, masking, pitch, timbre, consonance, etc. The other group refers to the temporal structure of the sound: duration, onset, rise, order, modulation, etc. All these processes play their parts in the perception of music.

With the help of psychophysical experiments, we have tried to determine the relative importance of the several stimulus properties which reflect these perceptual processes. We can now give the following tentative conclusions. Complete harmonic coincidence (overlap of harmonics) leads to fused perception of tones. This complete coincidence is present particularly in intervals with frequency ratio 1:n, such as the octave and the twelfth. Therefore, in polyphonic music in which voices and tones have to be perceived separately, these intervals should not be used exclusively. We have found that if two simultaneous tones with incomplete coincidence of partials (less perfect consonances and dissonances) are to be heard separately, their levels should not differ by more than 20 to 25 dB. The same was found when complete coincidence was made incomplete by frequency deviations caused by frequency modulation (vibrato) and mistuning (musical temperament, non-exact intonation). Asynchronization of tones leads to very clear separate perception. If the onset time difference was not greater than 30 msec, no temporal order effect was perceived. This means that this asynchronization range can be used effectively in having "simultaneous" tones perceived separately. Moreover, the effect of asynchronization is largely independent of the musical intervals involved. This finding points to the conclusion that temporal factors are of greater importance than spectral factors.

Our musical-interval recognition experiments confirm the role of asynchronization of simultaneous tones. They also show that the level differences between tones that should both be perceived should not be too large.

In view of the psychophysical experiments, an interesting question was what extent of asynchronization was typical of actually performed music. The results of the measurements, described in Chapter Four, clearly indicate that this asynchronization is of the same order of magnitude as the asynchronizations used in the laboratory experiments. From this correspondence we conclude that asynchronization in the performance of music is an important factor in the separate perception of simultaneous tones and voices. In some respects, this is a rather surprising conclusion. If musicians play together, they always try to synchronize as well as possible. Of course, perceptual and motor skills set a limit to the amount of synchronization that is possible. It is surprising that the asynchronization left over is functional in the perception of the performed music. Another aspect is that this asynchronization is not an element of musical composition but of performance. Because of this, it applies to all types of polyphonic compositions, from the Middle Ages to recent times.

The line of asynchronization was followed further in the fifth chapter. In this chapter we defined the concept of macro-asynchronization, the asynchronization arising from tones that must have onsets at different moments in time. The theory of polyphonic composition is traditionally called "counterpoint". Although this term is derived from "punctus contra punctum" (note-against-note), macro-asynchronization plays an important part and is included in the principles of the contrapunctus diminutus, syncopatus, and floridus. Apart from polyphonic music, there is also homophonic music, in which there is one melodically important voice, together with other supporting voices that do not have to be perceived separately. This style is characterized by a consistent synchronization of the various voices. So the theory of counterpoint gives further support to the idea that asynchronization is an important factor in the separate perception of single voices in polyphonic music.

SUMMARY

In this study research is described that deals with perceptual aspects of polyphonic music. Music is polyphonic if more than one voice part is melodically significant. As a result of this there are simultaneous tones that should be perceived separately most of the time. The problem of the perception of polyphonic music can thus be restated, in a simpler form, as the problem of the perception of simultaneous tones. The perception of tones has a dualistic character. Both spectral properties (frequency structure) and temporal properties (duration, onset, etc.) play a part. In the literature both of these lines have been explored.

The second and third chapters are devoted to psychophysical experiments concerning the separate perception of simultaneous tones. The second chapter describes a series of threshold experiments. The use of threshold experiments was based on the idea that conditions showing a low threshold for the perception of a weak tone in the presence of another tone would ensure easy separate perception of two tones in comparable supra-threshold conditions. The threshold in these experiments was related to the ability to detect a pitch jump of two successive tones in the presence of a masker tone with constant frequency. Independent variables investigated were: harmonic coincidence (overlap of harmonics), onset time difference, simultaneous versus continuous masker tone, and deviation from exact frequency ratio (modulation, mistuning). The main results can be summarized as follows: 1. If tones have complete harmonic coincidence (frequency ratio 1:n) it is difficult to separate them in perception, unless phase can be controlled; 2. Synchronous tones with incomplete harmonic coincidence can be heard separately if their levels do not differ by more than 20 to 25 dB; 3. If the masker is continuous, it is easier to hear the weak test tone than when the masker tone is strictly synchronous; 4. Onset time difference has a profound influence on perception. If the higher, weaker tone starts 30 msec before the lower masking tone, the weaker tone can be heard very clearly. In these conditions the onset time difference is not perceived as temporal order. 5. Deviation from simple frequency ratios makes separate perception easier. From these observations two general conclusions can be drawn: A. Temporal differences between tones

facilitate separate perception, and B. Simple frequency relations lead to fused perception, or, conversely, less simple frequency ratios give better separate perception.

A second series of experiments consisted of interval recognition experiments under supra-threshold conditions. These experiments are described in Chapter Three. Eight different musical intervals within the octave were used. Independent variables of the presentation were onset time difference and level difference. Dependent variables were the percentage of correct identification and the time needed to make a response. There was a strong negative correlation between the two dependent variables. It turned out that synchronous intervals (onsets and offsets simultaneous) had slightly lower percentages of correct responses than asynchronous intervals. Also, the level difference played a part. Recognition was best when the levels of the two tones in the interval were not too different. An analysis of the scores of the conditions defined by both an onset time difference and a level difference showed that both independent variables had about equal weight in recognition.

If psychophysical laboratory experiments point to a certain variable as being important for perception, the question arises how this variable is present in actually performed music. From the experiments we concluded that onset time difference was an important factor. Therefore we chose this variable for further investigation, which is described in Chapter Four. We designed a set of statistical measures concerning the asynchronization of onsets. The most important concepts are: relative onset time (onset time relative to the mean onset time of a group of "simultaneous" tones), onset time difference, and asynchronization, the latter concept defined as the root mean square of the standard deviations of onset time differences of all pairs of voices. Recordings of music played by trio ensembles were made in an anechoic room with directional microphones. The three separate tracks passed an envelope detector, the output of which was fed into a computer by a/d conversion. Further analyses were done by digital programming. Onset time was defined as the moment at which the envelope passed a threshold about 15 of 20 dB below maximum level. The analyses showed asynchronizations to be in the range of 30 to 50 msec. There were slight tendencies for some instruments to lead or lag at the average. The relations between asynchronization and tempo and between synchronization and isochronization were taken into account. Synchronization is better with higher tempi. There is agreement between our data and the data about isochronization of Michon (1967). The findings suggest that in practical musical performance asynchronization plays an important part in the separate perception of simultaneous tones and voices.

In the final chapter of this study (Chapter V) we looked at the theory of music and also at aspects of musical performance, in order to see whether a connection with the psychophysical and physical measurements was possible. Musicians make use of a number of particular techniques to make their parts audible in a polyphonic context. These techniques have to do partly with the spectral aspects of tone production and perception (singing formant), partly with the temporal aspects (asynchronization), but also with level differences (concert instruments) and spatial positioning. On the other hand, the composition of polyphonic music as codified in the theory of counterpoint puts a special emphasis on asynchronization in composition (which we call macro-asynchronization), the extent to which the score prescribes that tones must start at different moments in time. We investigated macro-asynchronization quantitatively in a number of hymn-tune settings by Praetorius, for two to six voices, both in homophonic and in polyphonic styles. It turned out that compositions with more voices had a larger amount of macro-asynchronization than compositions with fewer voices. This finding is another illustration of the general idea that asynchronization is one of the most important factors in the perception of polyphonic music.

S A M E N V A T T I N G

In dit proefschrift wordt onderzoek beschreven dat zich bezighoudt met de perceptieve aspecten van polyfone muziek. Muziek is polyfoon als er verschillende melodisch significante partijen tegelijk voorkomen. Hierdoor zijn er dikwijls gelijktijdige tonen die apart waargenomen moeten worden. We kunnen het probleem van de waarneming van meerstemmige muziek daarom ook eenvoudiger formuleren als het probleem van de waarneming van gelijktijdige tonen.

Toonperceptie heeft een tweezijdig karakter. Aan de ene kant spelen spectrale eigenschappen een rol (harmonischen) aan de andere kant temporele eigenschappen (duur, inzet, enz.). Beide lijnen zijn in de literatuur aan te wijzen.

Hoofdstuk Twee beschrijft een reeks van drempelexperimenten. We gingen er van uit dat condities met een lage drempel voor de aparte waarneming van twee gelijktijdige tonen de weg zouden wijzen naar bovendrempelige condities met gemakkelijke gescheiden waarneming van tonen. De drempel werd gedefinieerd aan de hand van de waarneming van een toonhoogte-sprong in aanwezigheid van een maskerende toon met constante frekwentie. Onafhankelijke variabelen waren: harmonische overlap, inzetverschil, gelijktijdige versus continue maskeerder, en afwijking van precieze frekwentieverhouding (modulatie, ontstemming). De resultaten kunnen als volgt samengevat worden: 1. Tonen met volledige overlap (frekwentieverhouding 1:n) worden moeilijk apart waargenomen, tenzij er een vast faseverschil is. 2. Synchrone tonen met onvolledige overlap kunnen waargenomen worden als de niveauverschillen niet groter dan 20 tot 25 dB zijn. 3. Tegen de achtergrond van een continue maskeerder worden tonen iets gemakkelijker waargenomen dan met een synchrone maskeerder. 4. Inzetverschil heeft een grote invloed op de waarneming. Als een hoge, zachte toon 30 msec eerder begint dan een lage, maskerende toon, klinkt de zachte toon zeer duidelijk. In deze conditie lijken de tonen gelijk te beginnen. 5. Afwijking van eenvoudige frekwentieverhoudingen (bijv. door vibrato of ontstemming) leidt tot betere waarneming. Uit deze observaties kunnen de volgende globale conclusies getrokken worden. A. Temporele verschillen tussen tonen maken de gescheiden waarneming eenvoudiger. B. Eenvoudige frekwentieverhoudingen leiden tot versmelting in de waarneming, of, andersom, afwijking daarvan werkt gescheiden waarneming in de hand.

Een tweede serie van experimenten bestond uit herkennings-experimenten (Hoofdstuk Drie). De proefpersoon moest een interval identificeren uit een verzameling van acht intervallen, alle kleiner dan een octaaf. Onafhankelijke variabelen waren inzetverschil en geluidniveauverschil. Afhankelijke variabelen waren het percentage correcte identificaties en de responstijd. Er was een sterke negatieve correlatie tussen deze twee variabelen. Het bleek dat synchrone intervallen (inzet gelijk) iets moeilijker te herkennen waren dan asynchrone. Ook het niveauverschil speelde mee. De beste herkenning vond plaats bij ruwweg gelijk niveau van de tonen. Verdere analyses wezen uit dat inzet- en niveauverschillen ongeveer in even sterke mate de herkenbaarheid van de intervallen beïnvloedden.

Als psychofysische laboratorium-experimenten een bepaalde variabele aanwijzen als zijnde belangrijk voor de waarneming, komt de vraag naar voren hoe deze variabele zich in de praktijk gedraagt. De drempelexperimenten hadden inzetverschil aangewezen als een belangrijke factor. Daarom besloten we deze variabele uit te kiezen voor verder onderzoek (zoals beschreven in Hoofdstuk Vier). Voor de analyse van de asynchronisatie van inzetten in uitgevoerde muziek ontwierpen we een verzameling van statistische maten. De belangrijkste begrippen in deze analyse zijn: relatieve inzettijd (inzettijd ten opzichte van de gemiddelde inzettijd van een groep gelijktijdige tonen), inzettijdverschil, en asynchronisatie gedefinieerd als het kwadratisch gemiddelde van de standaarddeviaties van inzettijdverschillen van alle paren van stemmen. Opnamen werden gemaakt van muziek gespeeld door verschillende trio's, met behulp van richtinggevoelige microfoons in een echovrije ruimte. De drie afzonderlijke signalen doorliepen een omhullende-detector waarvan de output via a/d conversie de computer inging. Het overige deel van de analyse werd uitgevoerd met behulp van digitale programmering. Inzettijd werd gedefinieerd als het moment waarop de omhullende een bepaalde drempel passeerde, die geplaatst was op -15 tot -20 dB onder het maximum niveau van de tonen. De analyses wezen uit dat asynchronisatie in uitgevoerde muziek gewoonlijk 30 tot 50 msec bedraagt. Sommige instrumenten hadden de tendens om, over het algemeen, iets voor of achter te komen. Ook de relatie tussen synchronisatie en tempo werd onderzocht (resultaat: hogere tempi gaan samen met betere synchronisatie) en die tussen asynchronisatie en isochronisatie (resultaat: overeenstemming met Michon's (1967) resultaten in dit opzicht). De resultaten suggereren dat de asynchronisatie die gevonden wordt in uitgevoerde muziek van belang kan zijn in de aparte waarneming van gelijktijdig klinkende tonen en stemmen in meerstemmige muziek.

Hoofdstuk Vijf richt de blik op de theorie van de muziek en ook op de uitvoeringspraktijk, om te zien of een verbinding tussen de psychofysische experimenten en de fysische metingen enerzijds en de muziektheorie en uitvoeringspraktijk anderzijds mogelijk was. Musici gebruiken een aantal technieken die gericht zijn op het hoorbaar maken van hun partij in een polyfone context. Sommige van deze technieken sluiten aan bij de spectrale aspecten van toonproductie en -perceptie (zangformant), andere bij temporele aspecten (asynchronisatie). Daarnaast zijn niveauverschillen (concertinstrumenten) van belang, en de ruimtelijke opstelling. Aan de andere kant, in de compositie van polyfone muziek (zoals beschreven in de theorie van het contrapunt) wordt speciale nadruk gelegd op het voorschrijven van ongelijke inzetten van tonen. Wij noemden dit aspect macro-asynchronisatie. Ook hiervoor kan een kwantitatieve formulering gegeven worden, die toegepast werd op een reeks vergelijkbare meerstemmige bewerkingen van koralen door Michael Praetorius, voor twee tot zes stemmen, zowel polyfoon als homofoon. Het bleek dat de composities met een groter aantal stemmen een grotere asynchronisatie van inzetten hadden. Deze uitkomst confronteert ons opnieuw met de idee dat in de waarneming van meerstemmige muziek asynchronisatie een belangrijk element is.

R E F E R E N C E S

Baxter, J.H. An old St. Andrews music book. London, Oxford University Press, 1931.

Bregman, A.S. & Pinker, S. Auditory streaming and the building of timbre. Canad.J.Psychol. 32 (1978), 19-31.

Brunold, P. Traité des signes et agréments employés par les clavecinistes français des XVIIe et XVIIIe siècles. Nice, Delrieu, 1965.

Burns, E.M. & Ward, W.D. Categorical perception: phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. J.Acoust.Soc.Amer. 63 (1978), 456-468.

Chopin, F. Nocturnes. Ed. I.J. Paderewski. Warschau, Polnischer Musikverlag, 1959. (Sämtliche Werke 7)

Couperin, L. Pièces de clavecin. Ed. P. Brunold & Th. Dart. Monaco, l'Oiseau-Lyre, 1959. Ed. A. Curtis. Paris, Heugel, 1972. (Le Pupitre 18)

Dannenbring, G.L. and Bregman, A.S. Streaming vs. fusion of sinusoidal components of complex tones. Perc.Psychophys. 24 (1978), 369-376.

Darwin, C.J. Perceptual grouping of speech components. In: Creutzfeld, O., Scheich, H., & Schreiner, Chr. (eds.), Hearing mechanisms and speech. Berlin, Springer, 1979, 333-340.

Elliott, L.L. Backward masking: monotic and dichotic conditions. J.Acoust.Soc.Amer. 34 (1962), 1108-1115.

Fastl, H. Pulsation patterns of sinusoids vs. critical band noise. Percept. Psychophys. 18 (1975), 95-97.

Forqueray, A. Pièces de clavecin. Ed. C. Tinley. Paris, Heugel, 1974.

Fux, J.J. Gradus ad Parnassum. Vienna Austriae, 1725. (Facs.ed. A. Mann. Graz, Akademische Druck u. Verlag, 1967 = J.J. Fux, Sämtliche Werke, Serie VII, Band 1)

Gerbert, M. (ed.). Scriptores ecclesiastici de musica. Tomus I. S. Blasius, 1764. (Reprint Milano, Bollettino, 1931)

Helmholtz, H.L.F. On the sensations of tone as a physiological basis for the theory of music. Transl. A.J. Ellis, London, 1885. (Facs.ed. New York, Dover, 1954)

Hirsh, I. Auditory perception of order. J.Acoust.Soc.Amer. 31 (1959), 759-767.

McAdams, S. Pitch as a perceptual attribute assigned to a source. Paper presented at the Third Workshop on Physical and Neurological Foundations of Music, Ossiach, Austria, August 8 - 12, 1980.

Michon, J. Timing in temporal tracking. Soesterberg, Institute for Perception TNO, 1967.

Moorer, J.A. On the segmentation and analysis of continuous musical sound by digital computer. Stanford University, 1975. Report STAN-M-3, Center for computer research in music and acoustics.

Noorden, L.P.A.S. van. Temporal coherence in the perception of tone sequences. Dissertation Eindhoven, 1975.

Noorden, L.P.A.S. van. Minimum differences of level and frequency for perceptual fission of tone sequences ABAB. J.Acoust.Soc.Amer. 61 (1977), 1041-1045.

Pickett, J.M. Backward masking. J.Acoust.Soc.Amer. 31 (1959), 1613-1615.

Plomp, R., Wagenaar, W.A. and Mimpen, A.M. Musical interval recognition with simultaneous tones. Acustica 29 (1973), 101-109.

Praetorius, M. Musae Sioniae, Teil V (1607) and Teil VI (1609). Ed. F. Blume & H. Költzsch, resp. F. Reusch. Wolfenbüttel, Möseler, n.d. (Gesamtausgabe der musikalischen Werke von Michael Praetorius, Band V, resp. Band VI).

Rasch, R.A. The perception of simultaneous notes such as in polyphonic music. Acustica 40 (1978a), 22-33.

Rasch, R.A. Synchronisatie bij ensemblespel. Mens en Melodie 32 (1978b), 41-47.

Rasch, R.A. Synchronisatie in uitgevoerde ensemblemuziek. Nederlands Akoestisch Genootschap, Publikatie 44, 1978c, 10-22.

Rasch, R.A. Synchronization in performed ensemble music. Acustica 42 (1979), 121-131.

Rokseth, Y. Polyphonies du XIIIe siècle. Tome II. Paris. l'Oiseau-Lyre, 1936.

Scheffers, M.T.M. The role of pitch in perceptual grouping of simultaneous vowels. Institute for Perception Research (Eindhoven), Annual Progress Report 14, 1979, 51-60.

Seashore, S.E. Psychology of music. New York, McGraw-Hill, 1938. (New York, Dover, 1967)

Siegel, J.A. and Siegel, W. Categorical perception of tonal intervals: Musicians can't tell sharp from flat. Perc.Psychophys. 21 (1977), 399-407.

Spitz, J. C. Statistiek voor psychologen, pedagogen, sociologen. Amsterdam, Noord-Hollandsche, 1968.

Stumpf, Tonpsychologie, Zweiter Band. Leipzig, Hirsch, 1890.

Sundberg, J. The acoustics of the singing voice. Sc.Amer. 236 (1977), 82-91.

Thurlow, W.R. and Elfner, L.F. Continuity effects with alternatively sounding tones. J.Acoust.Soc.Amer. 31 (1959), 1337-1339.

Tinctoris, J. The art of counterpoint (Liber de arte contrapuncti). Transl. & ed. A. Seay. Amer.Inst.Musicol., 1961. (Musicol. Studies and Documents 5)

Tove, P.A., Ejdesjö, L. & Svärtstrom, A. Frequency and time analysis of polyphonic music. J.Acoust.Soc.Amer. 41 (1967), 1265-1271.

Verschuure, J., Roodenburg, M. & Maas, A.J.J. Presentation conditions of the pulsation-threshold method. Acustica 35 (1976), 47-54.

Vos, J. & Rasch, R.A. The perceptual onset of musical tones. Accepted for publication in Percept.Psychophys. (1981).

Vos, P.G.M.M. Waarneming van metrische toonreeksen. Dissertatie Nijmegen, 1973.

Warren, R.M. Auditory pattern recognition by untrained listeners. Percept. Psychophys. 15 (1974), 195-500.

Weitbrecht, W. Über den Einfluss nichtlinearer Verzerrungen auf die Hörbarkeit von Verstimmungen musikalischer Intervalle. Fernmeldetechnisches Zeitschrift 3 (1950), 336-345.

Wilson, R.H. & Carhart, R. Forward and backward masking: interactions and additivity. J.Acoust.Soc.Amer. 49 (1971), 1254-1263.

Wright, H.N. Backward masking for tones in narrow-band noise. J.Acoust.Soc. Amer. 36 (1964), 2217-2221.

CURRICULUM VITAE

Rudolf Alexander Rasch was born in Borger (Drente, Netherlands) on December 15, 1945. He received his secondary school education at the Gemeentelijk Lyceum in Den Helder (Noord-Holland, Netherlands), from which he graduated in 1963 (Gymnasium-B). He studied musicology and experimental psychology at the University of Amsterdam, graduating in 1968 (musicology) and 1974 (psychology). During the years of 1964-1974 he held a number of assistantships. From February 1, 1975, until July 1, 1978, he worked at the Institute for Perception TNO in Soesterberg (Utrecht, Netherlands), for the research project "The perception of tones in polyphonic music" (see this dissertation). From September 1, 1977, he has been at the Institute of Musicology of the University of Utrecht, as a teacher of music theory and acoustics. He was the coordinator of the section on "Musical Perception" of the Netherlands Psychonomics Foundation from 1974 to 1978. From 1976 on, he has been a member of the board of the Society for Dutch Musical History. He has published articles on Dutch music history (mainly 17th century), musical psychology (i.a. about Julius Bahle), musical acoustics (i.a. subjective room acoustics) and tone perception (i.a. topics from this dissertation). He has edited modern editions of early music, from the 17th and 18th centuries, such as 't Uitnement Kabinet, Cantiones Natalitiae, violin sonatas by Philippus van Wichel, and cello sonatas by Pieter Hellendaal. He plays the harpsichord and the (baroque) violin.