The Perception of Pure and Tempered Musical Intervals

c

f g

b♭ d

e♭ a

g♯ e

c♯ f♯ b

Joos Vos

The Perception of Pure and Tempered Musical Intervals – Joos Vos – 1987

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Vos, Joos

The perception of pure and tempered musical intervals /
Joos Vos. - [S.1. : s.n.]. - Fig., tab.
Proefschrift Leiden. - Met lit. opg.
ISBN 90-9001470-5
SISO 781.5 UDC 159.932:781.2/.4
Trefw.: muziekpsychologie.

The Perception of Pure and Tempered Musical Intervals

Proefschrift

ter verkrijging van de graad van Doctor aan de Rijksuniversiteit te Leiden, op gezag van de Rector Magnificus Dr. J.J.M. Beenakker, Hoogleraar in de Faculteit der Wiskunde en Natuurwetenschappen, volgens besluit van het college van dekanen te verdedigen op 28 januari 1987 te klokke 14.15 uur

door

Joos Vos

geboren te Vlissingen in 1949

Drukkerij Elinkwijk bv — Utrecht

Promotiecommissie:

Promotores	:	Prof. Dr. G.B. Flores d'Arcais Prof. Dr. Ir. R. Plomp
Referenten	:	Prof. Dr. W.J.M. Levelt Dr. R.A. Rasch
Overige leden	:	Prof. Dr. J.P. van de Geer Dr. Ir. L.P.A.S. van Noorden Prof. Dr. W.A. Wagenaar

Het in dit proefschrift beschreven onderzoek werd uitgevoerd op de afdeling Audiologie van het Instituut voor Zintuigfysiologie TNO te Soesterberg.
Een gedeelte van dit onderzoek werd uitgevoerd met steun van de Nederlandse Organisatie voor Zuiver-Wetenschappelijk Onderzoek ZWO. De door ZWO gehonoreerde subsidieaanvrage ("De waarneming van zuivere en getemperde consonante tweeklanken", nr. 15-29-05) werd via de werkgemeenschap Muziekperceptie van de Nederlandse Stichting voor Psychonomie ingediend door Dr. Rudolf A. Rasch en Prof. Dr. Ir. Reinier Plomp.

Opgedragen aan Rudolf en Reinier

CONTENTS

CHAPTER 7

CHAPTER 8

1

Introduction

1.1 General Background

Pure consonant intervals are characterized by small-integer frequency ratios, for example, 1:2 for the octave, 2:3 for the fifth, 4:5 for the major third, etc. Tempered consonant intervals consisting of simultaneous complex tones are marked by frequency differences between those harmonics that coincide in pure intervals.

Perceptually, the interference of nearly coinciding harmonics in these tempered intervals gives rise to beats or roughness. In addition to beats, changes in the size of these intervals (stretched or compressed relative to the size of the pure intervals) may be detected. In a musical context, both the sensation of beats or roughness and the detection of changes in interval size may affect the degree to which these tempered intervals are experienced to be "impure" or "out-of-tune."

In tuning fixed-pitch keyboard instruments such as the organ, the harpsichord, and the piano, it is inevitable that some of the intervals are tempered. Which intervals are tempered, and to what extent, is described in musical tuning systems. In the history of music, numerous tuning systems have been proposed (Barbour, 1951). In the Pythagorean tuning system, for example, 11 of the 12 fifths are pure, as a result of which eight major thirds are considerably stretched. In the meantone tuning system, eight major thirds are pure, as a result of which 11 fifths are compressed. In equal temperament, all 12 fifths are slightly compressed, and the 12 major thirds are moderately stretched. More variation in the degree of tempering of the intervals can be found in, for example, one of the tuning systems constructed by Werckmeister (1691). This system contains eight pure and four compressed fifths; the major thirds are all stretched, but to four different degrees.

A given tuning system may be cordially recommended by one and disapproved of or totally rejected by another. Very diverging points of view concerning the values and the advantages of the various tuning systems are held by historians and music theorists like, for example, Werckmeister, Neidhardt, Mattheson, Sorge, Kirnberger, and Marpurg. Details on these different evaluations can be found in Dupont (1935), van Biezen (1974), Kelletat (1981), and Rasch (1985).

It is, indeed, not easy to decide which tuning system has to be applied in a musical performance, because this choice depends, in addition to personal preferences, also on the music to be played, the instrument to be used, and the ease with which the temperings of the intervals can be set.

More systematic evaluations of different tuning systems have been proposed by Hall (1973) and by Rasch (1983). In view of the great interest that has existed over the last few decades in performing Baroque and earlier music on authentic instruments and according to the tuning systems that were used in that period, such evaluations may be significant for musical practice.

The theoretical evaluations just mentioned are based, among other things, on the assumption that the tempering of the intervals should be minimized. Because of the lack of experimental data, Hall (1973) assumed a quadratic relationship between tempering (in cents) of the interval and its subjective "purity" or "suitability to fulfil its usual musical role". Rasch (1983) was less definite about the characteristics of the relation between tempering and purity. In the evaluation of, for example, the major and minor triads in various regular 12-tone tuning systems, he used three different measures for mean tempering, all of which were basically intended to represent overall subjective purity.

Naturally, in the discussion and the evaluation of the different tuning systems, it is of great interest to know how important the differences between the pure and tempered intervals are to our perception. To make up for the lack of detailed experimental results reported in the literature, therefore Chapters 2, 3, 4, and 5 of this study are concerned with the investigation of different aspects of the capacity to discriminate between pure and tempered intervals. We expected that the differences between the pure and tempered intervals would be more prominent for intervals with simultaneous tones (harmonic intervals) than for intervals with successive tones (melodic intervals), especially for harmonic intervals with complex tones. The common experience that the effect of tempering depends on the kind of musical interval led us to study discrimination for various musical intervals. In addition to musical interval and degree of tempering, we investigated the effects of tone duration, spectral content and fundamental frequency on discrimination. The experiments on discrimination were supplemented by studies in which pure and tempered intervals were rated: In Chapter 6 we examine the relationship between tempering and subjective purity and in Chapter 7 we study subjective acceptability of tuning systems in musical fragments. An epilogue of the present study is given in Chapter 8.

1.2 Survey of the Chapters

Chapter 2

In Chapter 2, we determined thresholds for discrimination between pure and tempered fifths and major thirds for various degrees of tempering and at tone durations within a range of 0.25 to 1 s. Since both the sensation of beats and roughness and the detection of changes in interval size seemed to be relevant in discrimination, thresholds for beats and thresholds for the identification of the direction of tempering were also determined.

Chapter 3

In Chapter 3, we attempted to answer three questions: (1) To what extent are the thresholds determined in Chapter 2 a result of the interference of nearly coinciding harmonics? (2) Is the beat frequency of the first pair of these harmonics equal to the perceived beat frequency in the tempered intervals? (3) Are differences in discriminability at threshold related to differences in perceived strength of beats in supraliminal conditions?

The answer to the first question was sought by investigating the effect of the spectral content of the tones on the discrimination thresholds for the fifth and the major third. The dominantly perceived beat frequency and the perceived beat strength, which were determined in supraliminal conditions for various spectra and degrees of tempering, provided the basis for answering the second and third questions.

Chapter 4

The results obtained in Chapters 2 and 3 showed that the discrimination thresholds were lower (higher sensitivity to tempering) for fifths than for major thirds. It was the purpose of the experiments reported in Chapter 4 to find a general rule to describe the relation between discriminability and type of interval. Therefore, we determined discrimination thresholds for 13 different musical intervals that varied in size from unison to twelfth. In addition to musical interval, spectral content of the tones was the main independent variable.

Chapter 5

In the experiments reported in Chapters 2, 3, and 4, the discrimination thresholds were only determined for a restricted frequency range. For the fifths and the major thirds, for example, the fundamental frequencies had always been between middle C (C4 = 261.6 Hz) and the octave above middle C (C5).

The aim of the experiment described in Chapter 5 was to determine to what extent previous findings might be generalized to a much wider frequency range. Thresholds for discrimination between pure and tempered intervals were therefore determined over a frequency range of four octaves between C2 and C6.

Chapter 6

In Chapter 6 we investigated the relationship between tempering and subjective purity of fifths and major thirds. Since the experiments discussed in the previous chapters had shown that duration and spectral content of the tones affect discriminability between pure and tempered intervals, these two factors were also included in the first experiment reported in Chapter 6. We reasoned that the purity ratings obtained may have been determined by a combination of the sensation of beats and the detection of changes in interval size. In the second experiment attempts were made to gain more insight into the relevance of these hypothetical aspects by reducing or even deleting the potential interference between the harmonics of the two tones.

The degree of interference was manipulated by presentation of two simultaneous sinusoidal instead of two simultaneous complex tones, and by presenting the tones successively. In addition, the experimental results obtained in this chapter were compared with the predictions from the consonance/dissonance models by Plomp and Levelt (1965) and Kameoka and Kuriyagawa (1969).

Chapter 7

In Chapter 7 we investigated the subjective overall acceptability of various regular 12-tone tuning systems in two-part musical fragments. Again, the relative importance of beats and roughness for the overall evaluation was determined by varying duration and spectral content of the tones. The acceptability ratings could be accurately predicted from a linear combination of the purity ratings of the harmonic fifths and the harmonic major thirds reported in Chapter 6.

Chapter 8

An epilogue of the present study is given in Chapter 8. Suggestions for future research are outlined.

1.3 References

Barbour, J.M. (1951). Tuning and temperament (Michigan State College, East Lansing, reprinted 1972, Da Capo, New York).

Biezen, J.van. (1974). Stemmingen, speciaal bij toetsinstrumenten (Centrum voor de kerkzang, 's-Gravenhage; reprinted 1975, 1978).

Dupont, W. (1935). Geschichte der musikalischen Temperatur (Baerenreiter, Kassel).

Hall, D.E. (1973). "The objective measurement of goodness-of-fit for tunings and temperaments", J. Music Theory 17, 274-290.

Kameoka, A., and Kuriyagawa, M. (1969). "Consonance theory, Part II: Consonance of complex tones and its calculation method," J. Acoust. Soc. Am. 45, 1460-1469.

Kelletat, H. (1981). Zur musikalischen Temperatur, Band I (Baerenreiter, Kassel).

Plomp, R., and Levelt, W.J.M. (1965). "Tonal consonance and critical bandwidth," J. Acoust. Soc. Am. 38, 548-560.

Rasch, R.A. (1983). "Description of regular twelve-tone musical tunings", J. Acoust. Soc. Am. 73, 1023-1035.

Rasch, R.A. (1985). "Does "well-tempered" mean "equal-tempered"? In P. Williams (Ed.), Bach, Handel, Scarlatti tercentenary essays, 293-310 (Cambridge University Press, Cambridge, England).

Werckmeister, A. (1691). Musicalische Temperatur (Quedlinburg).

Perception & Psychophysics
1982, 32 (4), 297-313

2

The perception of pure and mistuned musical fifths and major thirds: Thresholds for discrimination, beats, and identification

JOOS VOS
Institute for Perception TNO, Soesterberg, The Netherlands

In Experiment 1, the discriminability of pure and mistuned musical intervals consisting of simultaneously presented complex tones was investigated. Because of the interference of nearby harmonics, two features of beats were varied independently: (1) beat frequency, and (2) the depth of the level variation. Discrimination thresholds (DTs) were expressed as differences in level (ΔL) between the two tones. DTs were determined for musical fifths and major thirds, at tone durations of 250, 500, and 1,000 msec, and for beat frequencies within a range of .5 to 32 Hz. The results showed that DTs were higher (smaller values of ΔL) for major thirds than for fifths, were highest for the lowest beat frequencies, and decreased with increasing tone duration. Interaction of tone duration and beat frequency showed that DTs were higher for short tones than for sustained tones only when the mistuning was not too large. It was concluded that, at higher beat frequencies, DTs could be based more on the perception of interval width than on the perception of beats or roughness. Experiments 2 and 3 were designed to ascertain to what extent this was true. In Experiment 2, beat thresholds (BTs) for a large number of different beat frequencies were determined. In Experiment 3, DTs, BTs, and thresholds for the identification of the direction of mistuning (ITs) were determined. For mistuned fifths and major thirds, sensitivity to beats was about the same. ITs for fifths and major thirds were not significantly different; deviations from perfect at threshold ranged from about 20 to 30 cents. Comparison of the different thresholds revealed that DTs are mainly determined by sensitivity to beats. Detailed analysis, however, indicated that perception of interval width is a relevant aspect in discrimination, especially for the fifths.

One of the classic problems of music theory is that of how to tune our 12-note chromatic scale. This tuning problem takes a very concrete form in the fixed-pitch keyboard instruments such as the organ, the harpsichord, and the piano. Since perfect fifths and pure major thirds are incompatible within one tuning system, it is inevitable that one introduces temperaments in which the fifths and/or major thirds are adjusted, that is, mistuned to some extent.

For complex tones, mistuned intervals are characterized by small frequency differences between those harmonics which coincide completely in pure intervals. [In this paper, interval represents a *harmonic* interval, in which a low tone (tone 1) and a high tone (tone 2) are presented simultaneously.] The interference of these just-noncoinciding harmonics gives rise to the perception of beats or roughness (see Helmholtz, 1877/1954, chap. 10). When the interfering harmonics have equal amplitude, these beats are most prominent. In line with current models on *tonal* or *sensory* consonance, that is, the perception of consonance for isolated intervals (Kameoka & Kuriyagawa, 1969; Plomp & Levelt, 1965; Terhardt, 1976), we propose that the subjectively experienced out-of-tuneness in music may in some way be related to the perception of beats. Since it is one aspect in the complex process of evaluating different tuning systems, it may be worthwhile to know our sensitivity to mistuning.

It was the aim of the present study to investigate the perception of pure and mistuned fifths and major thirds, particularly with respect to (1) discriminability between pure and mistuned intervals, (2) sensitivity to beats, and (3) identification of the direction of mistuning, that is, whether the interval will have been stretched or compressed. Before we give a detailed description of the experimental method, we shall elaborate the musical, physical, and psychophysical aspects that are relevant to our experimental study.

MUSIC THEORY

In Western music, consonant intervals such as the octave, fifth, and major third play an important role. A consonant interval is a musical interval in which

This research was supported by Grant 15-29-05 from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.). The author is indebted to Rudolf Rasch and to Reinier Plomp, who made critical comments on earlier versions of this paper. Thanks are also due Ben van Vianen, for his participation in running the third Experiment. Requests for reprints should be sent to Joos Vos, Institute for Perception TNO, Kampweg 5, 3769 DE Soesterberg, The Netherlands.

0031-5117/82/100297-17$01.95/0

the ratio of the fundamental frequencies of tones 1 and 2 can be described by means of small integers: 1:2 for the octave, 2:3 for the fifth, and 4:5 for the major third. In music theory, a major third equals the ratio between four fifths and two octaves:

$$\beta = \alpha^4/\omega^2, \tag{1}$$

in which α, β, and ω are the ratios of the fundamental frequencies of the fifth, major third, and octave, respectively. In Equation 1, we have two degrees of freedom, but given that the perfect octave ($\omega = 2/1$) is an axiom in all tuning systems, Equation 1 can be reduced to:

$$\beta = \alpha^4/4. \tag{2}$$

If the fifth is perfect ($\alpha = 3/2$), the major third is a Pythagorean major third (81/64), one syntonic comma (81/80) larger than the pure major third. If the major third is pure ($\beta = 5/4$), the fifth equals $5^{1/4}$, or 1.495349. This illustrates that it is not possible to construct tuning systems in which both the fifths and major thirds are pure. Either the fifths or the major thirds, or both, have to be mistuned. A mistuned major third, β, comprises the pure major third and a very small tempering interval, b ($\beta = 5b/4$). Likewise, a mistuned fifth, α, comprises the pure fifth and a small interval a ($\alpha = 3a/2$). Substituting (5b/4) and (3a/2) in Equation 2 for β and α, respectively, results in:

$$5b/4 = (3a/2)^4/4, \tag{3}$$

from which it follows that

$$b = (81/80)a^4. \tag{4}$$

Both in musical and perceptual contexts, it is considered to be more appropriate to convert frequency ratios of musical intervals into logarithmic quantities (Barbour, 1940; Pikler, 1966; Young, 1939). Logarithmic conversion of Equation 4 results in

$$B = S + 4A, \tag{5}$$

in which S equals the syntonic comma. The relation between the mistuning of the major third and the mistuning of the fifth, as given in Equation 5, is depicted in Figure 1. Mistuning is expressed both with S and with the cent as unity. The cent is defined as the interval between two tones having as a basic frequency ratio the 1,200th root of 2. Figure 1 clearly illustrates that there is only a restricted range in which an increase in mistuning of the major third results in a decrease in mistuning of the fifth, and vice versa. Only this range, which is printed boldly in Figure 1, is relevant for optimizing the purity of both the fifths and the major thirds.

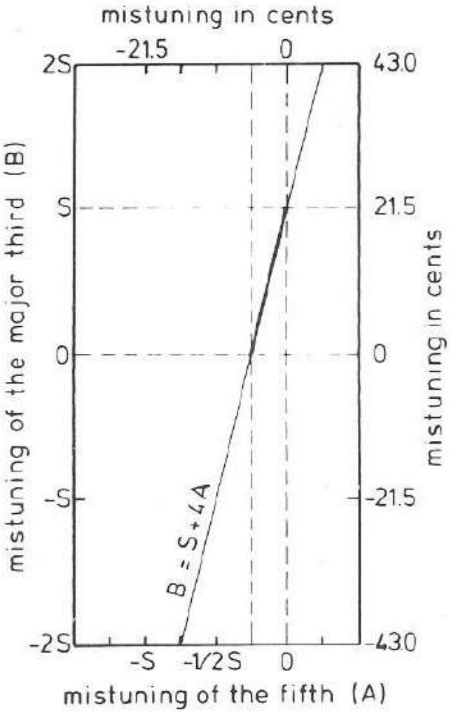

Figure 1. Relation between mistuning of the major third and mistuning of the fifth. Mistuning is expressed both with the syntonic comma (S) and with the cent as unity. The range in which the purity of both the fifths and the major thirds can be optimized is in bold print.

Physical Aspects

Beat Frequency

In a consonant interval, in which the ratio of the fundamental frequency f_1 of tone 1 and the fundamental frequency f_2 of tone 2 is equal to p:q, with p and q ($p < q$) small integers, the nqth harmonic of tone 1 coincides with the npth harmonic of tone 2 ($n = 1, 2, \ldots$).

In mistuned intervals, however, where $f_1 : f_2$ slightly departs from p:q, these harmonics do not coincide exactly in frequency. The superposition of the different pairs of just-noncoinciding harmonics with frequencies of nqf_1 and npf_2 Hz and equal amplitudes is equal to

$$\sin 2\pi(nqf_1)t + \sin 2\pi(npf_2)t$$

$$= 2\cos[2\pi \tfrac{1}{2}(npf_2 - nqf_1)t] \cdot \sin[2\pi \tfrac{1}{2}(npf_2 + nqf_1)t]. \tag{6}$$

For a given value of n, the compound waveform can be interpreted as a tone with frequency $\frac{1}{2}(npf_2 + nqf_1)$ Hz, which is the mean of the frequencies of the two harmonics, and a slowly fluctuating amplitude

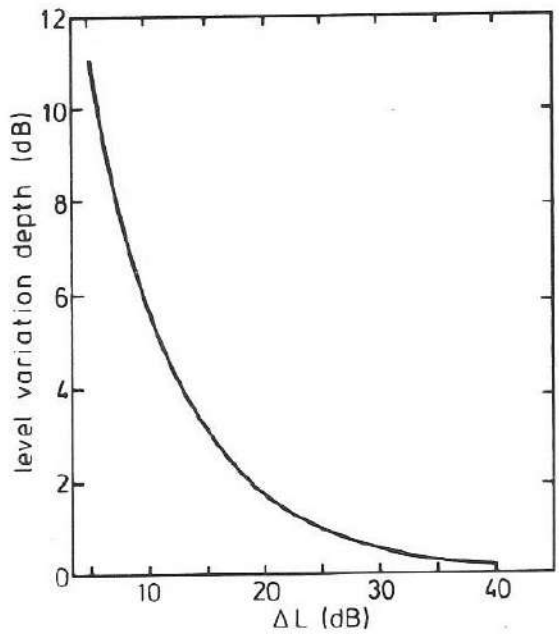

Figure 2. Depth of level variation of two beating tones, given as a function of the level difference ΔL of tones 1 and 2. ΔL equals 0 dB when the spectral envelopes of tones 1 and 2 coincide.

given by $| 2 \cos 2\pi \frac{1}{2}(npf_2 - nqf_1)t |$. Since this amplitude varies $(npf_2 - nqf_1)$ times per second between 0 and 2, the frequency of the beating sensation f_{bn} for the n^{th} pair of interfering harmonics is equal to

$$f_{bn} = | npf_2 - nqf_1 |. \tag{7}$$

If $npf_2 - nqf_1 > 0$, the musical interval is stretched $(f_2/f_1 > q/p)$, if $npf_2 - nqf_1 < 0$, the musical interval is compressed $(f_2/f_1 < q/p)$. In our experiments, the degree of mistuning of the intervals will be expressed as the beat frequency for the lowest pair of just-noncoinciding harmonics $(n = 1)$. The relationship between this beat frequency f_b and deviation T (in cents) of the mistuned interval from a perfect interval is given by

$$f_b = | qf_1(2^{T/1,200} - 1) |,$$

or simply by

$$f_b = | qf_1 T/1,731 |. \tag{8}$$

Equation 8 shows that the beat frequency depends on the amount of mistuning T, the value of q, and the frequency of the fundamentals. From Equation 8, the upper boundary of those beat frequencies which are significant in music perception research can be computed. For example, the beat frequency of a major third with fundamental f_1 of 349 Hz (musical F4), which is tuned 21.5 cents sharp, equals 21.8 Hz.

Level Variation Depth

The amplitude p(t) of the two interfering harmonics with waveforms $a_1 \sin 2\pi qf_1 t$ and $a_2 \sin 2\pi pf_2 t$ equals:

$$p(t) = (a_1^2 + a_2^2 + 2a_1 a_2 \cos 2\pi f_b t)^{1/2}. \tag{9}$$

The maximum and minimum values of p(t) are $(a_1 + a_2)$ and $(a_1 - a_2)$, respectively. The level variation depth in decibels of the beating harmonics is given by

$$D = 20 \log \left|\frac{a_1 + a_2}{a_1 - a_2}\right| = 20 \log \left[\frac{10^{\Delta L/20} + 1}{10^{\Delta L/20} - 1}\right], \tag{10}$$

in which ΔL is the absolute level difference between a_1 and a_2 in decibels $[\Delta L = | 20 \log (a_1/a_2) |]$. D approaches ∞ for $a_2 \rightarrow a_1$. In Figure 2, the relation between the level difference ΔL and level variation depth D, as defined in Equation 10, is given.

Psychophysical Aspects

The Perception of Beats

To our knowledge, no experimental results on thresholds for the perception of beats for intervals consisting of two complex tones have been reported. Only the discrimination of beats, caused by adding two sinusoids of nearly the same frequency, has been studied.

As a function of beat rate, Riesz (1928) found a U-shaped curve for the detection threshold of intensity fluctuations, with a minimum for about 3 Hz. Further experiments at this beat frequency revealed that sensitivity increases with increasing level above absolute hearing threshold (sensation level) up to 60 dB, and remains constant at higher levels. This effect was verified by Harris (1963). In addition, Harris found that the threshold varies in the same way as a function of sensation level both for beats resulting from two interfering sinusoids and for sine-wave amplitude modulation (AM) of one sinusoid.

Since sensitivity for the perception of beats is high, that is, the difference between a_1 and a_2 at threshold is large, the amplitude p(t) as given in Equation 9 can be approximated for threshold conditions by:

$$p(t) = a_1 + a_2 \cos 2\pi f_b t. \qquad (a_1 >> a_2) \tag{11}$$

Since this function is identical to the amplitude variation of an AM tone, we can relate the results from studies on AM tones to those of Harris (1963) and Riesz (1928). As for two tones, the level variation threshold for AM tones decreases with increasing modulation frequencies up to 4 Hz and increases again for modulation frequencies above this value (Ebel, 1952; Terhardt, 1974; Zwicker, 1952).

Maiwald (1967) and Schöne (1979) have shown that the level variation threshold decreases with increasing sound pressure level (SPL). Zwicker (1952) found that this effect does not depend on modulation frequency.

The Perception of Interval Width

Musical interval recognition. Using *simultaneous* complex tones, Plomp, Wagenaar, and Mimpen (1973) showed that trained listeners are able to categorize the 12 musical intervals between C4 and C5, if tone duration is 120 msec, with a mean accuracy of 83%. In a similar study, Killam, Lorton, and Schubert (1975) found that their subjects reached a mean score of 67% correct. The sharpness of transitions between the categories has been successfully determined by means of *successive* sine-wave tones. In these experiments, subjects were presented with ascending intervals, their widths being varied independently in small steps of 12.5 to 20 cents. A number of studies have demonstrated that, also in such conditions, trained listeners are able to use musical-interval categories consistently, and that their transitions between adjacent categories are sharp and well defined (Burns & Ward, 1978; Siegel & Siegel, 1977a, 1977b).

Identification of the direction of mistuning. Identification of the direction of mistuning was investigated extensively by Stumpf and Meyer (1898). To make their results more accessible, we reanalyzed the data by transforming the frequency differences between the tones into deviations (in cents) of the mistuned intervals from the pure intervals, and by transforming the correct response percentages for the stretched and compressed conditions into z scores. From these values, thresholds were determined by means of linear regression functions. For *melodic* major thirds, fifths, and octaves, averaged over sine-wave tones and complex tones, thresholds were 3, 5, and 4 cents, respectively. In the same way, the data from the *harmonic* intervals were reanalyzed. The results are shown in Figure 3 for the major third, fifth, and octave, separately. The amount of mistuning, which resulted in 75% correct compressed and 75% correct stretched responses, is indicated by arrows. In these conditions, thresholds for the major third and octave are 12 and 20 cents, respectively, and thus considerably higher than in the melodic conditions. This had also been observed by Delezenne (1827). It should be noted, however, that, in Stumpf and Meyer (1898), determination of thresholds for melodic and harmonic intervals was confounded with stimulus complexity. In the melodic conditions, the frequency of either the low or the high tone was fixed, whereas in the harmonic conditions, both tones could be varied in frequency. The low threshold for the fifth (=4 cents, see Figure 3b) can also be attributed to stimulus simplicity since in this particular harmonic condition only the high tone was varied. The effect of stimulus complexity has recently been demonstrated by Zatorre and Halpern (1979), who found for their subjects that consistency of labeling data was entirely dependent on the constancy of the lower tone.

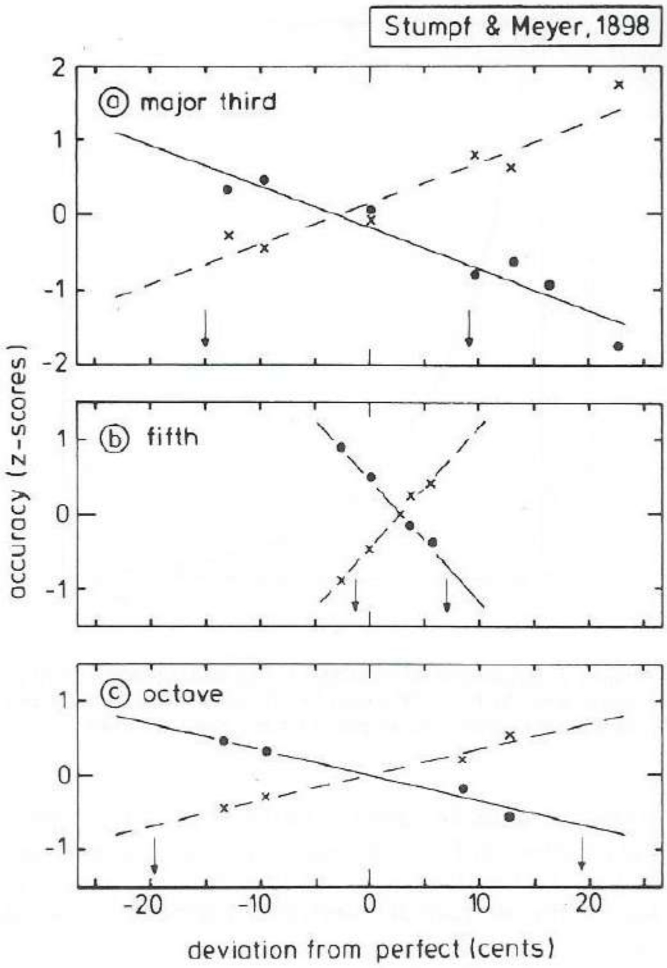

Figure 3. Reanalysis of data from Stumpf and Meyer (1898), who investigated identification of the direction of mistuning for several musical intervals. Frequency differences between the tones are transformed into deviations (in cents) of the mistuned intervals from the perfect intervals. Percentage correct responses for the stretched (×) and compressed (●) conditions are converted to z scores. Thresholds (75% correct), computed by means of linear regression functions, are indicated by arrows. For harmonic intervals, results are presented for the major third, fifth, and octave in panels a, b, and c, respectively.

The idea that thresholds for the perception of interval width are more likely to range between 10 and 20 cents, or even between 10 and 30 cents, than between 4 and 10 cents is in line with adjustment thresholds for musical intervals consisting of simultaneous sine-wave tones (Moran & Pratt, 1926), and with adjustment thresholds for musical intervals consisting of successive sine-wave tones (Rakowski, Note 1).

The Rationale Behind the Experiments

In Experiment 1, discrimination between pure and mistuned fifths and pure and mistuned major thirds was investigated. Two features of the beats were varied: (1) beat frequency, representing the amount of mistuning (see Equations 7 and 8), and (2) level variation depth of the beats (see Equation 10). We shall learn that, for high beat frequencies, thresholds at which subjects correctly discriminated between pure and mistuned complex-tone intervals 75% of the time deviated from the well-known U-shaped curve for the detection threshold of beats. Thresholds for beats in complex-tone intervals, however, do not necessarily have to be similar to those caused by superposition of two sinusoids of nearly the same frequency. In Experiment 2, we therefore determined thresholds for beats in complex-tone intervals for a large number of beat frequencies, including those of the various mistuned intervals from Experiment 1.

The main results from the second experiment were (1) that sensitivity to beats in complex tone intervals does not have a clear maximum such as the one found with superimposed sinusoids, and (2) that there was at least one condition in which the pattern of the beat threshold strongly deviated from the pattern of the discrimination threshold. The latter finding supported our notion that, apart from beats, the perception of differences between the size of the mistuned interval and that of the pure interval may have contributed to the discrimination thresholds.

Experiment 3 was designed to obtain more direct evidence that both the perception of beats and the perception of interval width were relevant aspects in discrimination. In this experiment, not only discrimination and beat thresholds, but also thresholds for identification of the direction of mistuning were determined.

EXPERIMENT 1

It is needless to say that tone duration is a relevant parameter in music. Moreover, while listening to performed music, it can be easily noticed that the perceptibility of beats is enhanced at longer tone durations. Therefore, we decided to determine discrimination thresholds at various tone durations, ranging from 250 to 1,000 msec.

Method

Stimuli

The stimuli were intervals consisting of two complex tones, 1 and 2, with fundamental frequencies of f_1 and f_2 Hz, respectively. The frequency ratio $f_1:f_2$ was equal to or slightly different from p:q. We presented musical fifths (p = 2; q = 3) and musical major thirds (p = 4; q = 5). The complex tones 1 and 2 consisted of 20 harmonics with amplitudes according to:

$$p(t) = \sum_{n=1}^{n=20} \frac{1}{n} \sin(2\pi nft + \phi_n). \qquad (12)$$

This resulted in a stimulus with a spectral envelope slope of −6 dB/octave. The phase (ϕ_n) of the individual harmonics was chosen randomly. Because of this, the waveforms of tones 1 and 2, although having the same spectral envelope, were different. The overall level of tones 1 and 2 together was 85 dB SPL. This level was measured with the help of an artificial ear (Brüel & Kjaer, Type 4152). Rise and decay times of the tone bursts, defined as the time interval between 10% and 90% of the maximum amplitude, were 40 and 20 msec, respectively. Level variation depth of the beating harmonics was obtained by attenuation of tone 1 or tone 2 by ΔL dB. When the spectral envelopes of the two tones coincide, that is, when the amplitudes of the just-noncoinciding harmonics are equal, level variation depth is maximal. We defined ΔL to be 0 dB for this condition. The spectral contents of our intervals in the conditions in which ΔL equals 0 dB are depicted in the left-hand panels of Figure 4. In these panels, only the levels of the first 10 harmonics of tone 1 and the first 7 or 8 harmonics of tone 2 are given. The levels are plotted relative to the level of the first harmonic of tone 1. Coincidence of spectral envelopes occurs when the level of tone 2 is 20 log (q/p) dB lower than the level of tone 1. Increments of ΔL from the base conditions, as illustrated in the left-hand panels of Figure 4, always result in smaller level variation depths. In the middle panels of Figure 4, the spectral contents of our intervals are given for cases in which $\Delta L = 15$ dB and the level of tone 2 (L_2) is lower than the level of tone 1 (L_1). In the right-hand panels, the levels of the harmonics are presented relative to the level of the first harmonic of tone 2. In the latter conditions, when $L_1 < L_2$, ΔL equals 15 dB again.

Apparatus

A block diagram of the apparatus is given in Figure 5. The experiments were run under the control of a PDP-11/10 computer. Tones 1 and 2 were generated in the following way. One period of the waveforms of tones 1 and 2 was stored in 256 discrete samples (with 10-bit accuracy) in external revolving memories. These recirculators could be read out by digital-to-analog converters. Sampling rates were determined by pulse trains, derived from frequency generators 1 and 2. After gating, the tones were filtered by Krohn-Hite filters (Model 3341) with the function switch in the low-pass RC mode and with a cutoff frequency of 8 kHz. The sound-pressure levels of the tones were controlled by programmable attenuators. After appropriate attenuation, the tones were mixed and fed to two headphone amplifiers. The signals were presented diotically (same signal to both ears) by means of Beyer DT 48S headphones. Subjects were seated in a soundproof room.

Subjects

Four musically trained subjects were tested over eight sessions, each session on a different day. Two were students from the Institute of Musicology at Utrecht, who were paid for their participation; the other two were the author and a colleague. Since the experiment aimed at studying a musical-perceptual competence, the choice of musically trained subjects was a deliberate one. Moreover, our subjects were experienced in performing auditory tasks, and instruction was facilitated by the possibility of using musical terminology. We decided to present our stimuli to a rather small number of subjects because, due to the nature of the experimental task, we did not expect to find large discrepancies between the thresholds of the different subjects.

Experimental Design

The independent variables were: (1) tone duration (250, 500, and 1,000 msec); (2) musical interval (fifth and major third); (3) beat frequency of the first pair of just-noncoinciding harmonics (.5, 1, 2, 4, 8, 16, and 32 Hz); (4) attenuation of tone 1 or tone 2; (5) direction of mistuning (stretched vs. compressed interval). Description of the discrimination threshold, being the dependent variable, is given below. For a tone duration of 250 msec and beat frequencies of .5 and 1 Hz, and for a tone duration of 500 msec and a beat frequency of .5 Hz, no discrimination thresholds were determined because these stimuli contained less than half a beating period. This means that, with respect to tone duration, we have an

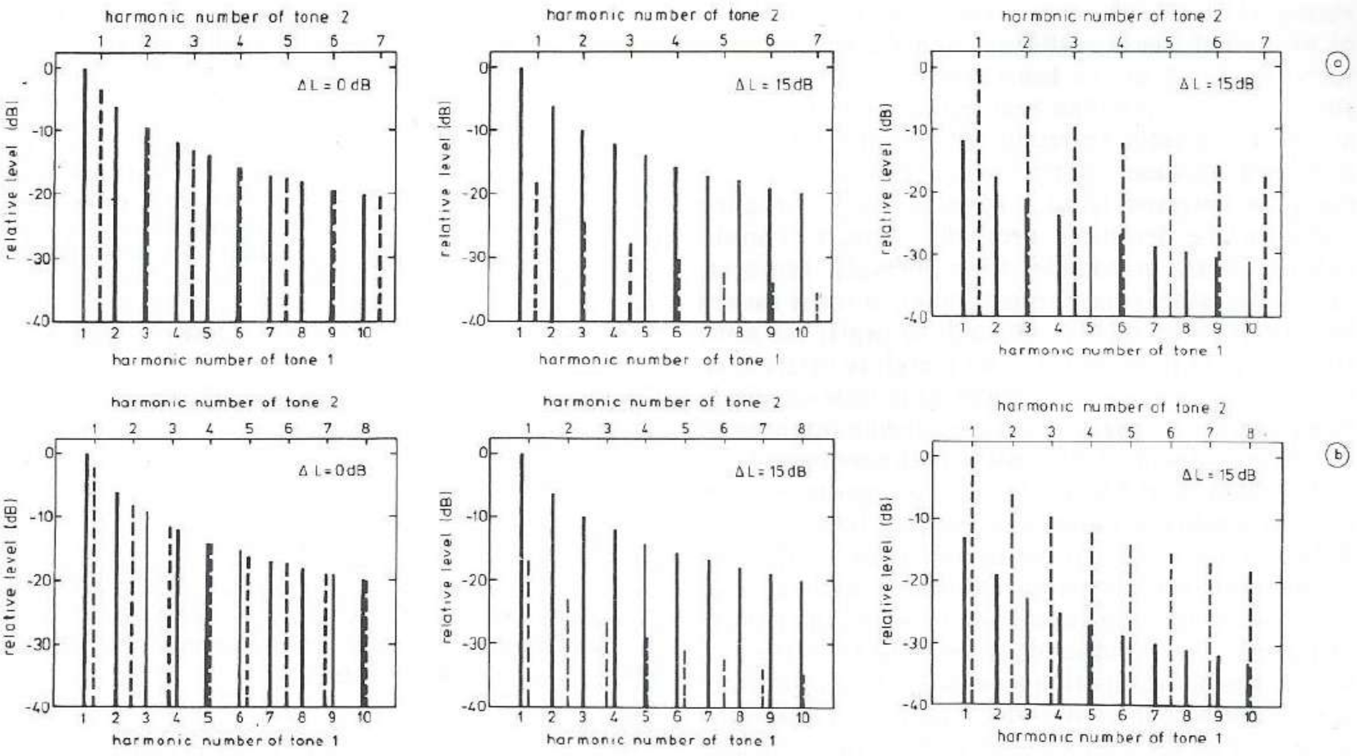

Figure 4. Presentation of the spectral contents of various fifths (panel a) and major thirds (panel b). The levels of the first 10 harmonics of tone 1 (bottom horizontal) and the first 7 or 8 harmonics of tone 2 (upper horizontal) are given on a linear frequency scale. In the left-hand and middle panels, the levels of the harmonics are plotted relative to the level of the first harmonic of tone 1. In the right-hand panels, the levels are given relative to the level of the first harmonic of tone 2.

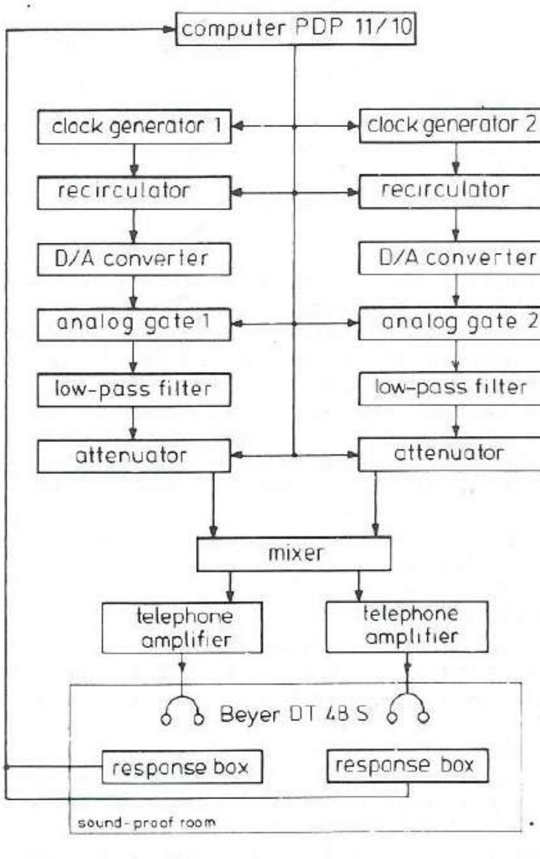

Figure 5. Block diagram of the apparatus used.

incomplete factorial design. Analyses of variance (ANOVAs) on the interactions of tone duration with the other variables could be performed for the five beat rates from 2 to 32 Hz.

Data Acquisition and Data Reduction

The task for the subject was to discriminate between pure and mistuned musical intervals in a 2-AFC paradigm. By means of the method of constant stimuli, all stimulus combinations were presented 10 times for six different values of ΔL in equal steps of 4 or 5 dB. For every stimulus combination, the raw scores for different ΔL values were transformed to percentage correct. Since the response patterns of the stretched and compressed intervals were similar in almost all cases, the percentages of these two conditions were combined. On the assumption that the scores, for a given set of ΔL values, are cumulatively normally distributed (Kling & Riggs, 1972, pp. 11-14), these percentages were converted to z scores. The discrimination threshold, that is, the value of ΔL at which a subject responded correctly in 75% of the cases, was obtained by solving the linear regression function $z = a + b\Delta L$ for ΔL when z equals .67. Computation of coefficients of determination (r^2) for all linear regression functions showed that the explained variability in z by ΔL and the linear rule, was more than .75 in 73% of the cases. Moreover, r^2 was about the same for all conditions and all subjects. Mean r^2 was .80.

Procedure

Every trial consisted of a comparison of a pure interval and a mistuned interval. The mistuned interval was presented first in half of the trials and second in the other half of the trials. Between the intervals was a silent period of 1.5 sec. By means of a response box, the subject indicated whether the first or the second interval was mistuned. Feedback was given by a red light above the correct button. The next trial was presented 1.5 sec after the response to the preceding trial. Dependent upon stimulus duration, which was held constant within the series, a series comprised 240, 288, or 336 trials. The presentation order of the different series was randomized. Within a series, which lasted 15 to 25 min, the stimuli were presented in a random order. At the beginning of a series, new waveforms of tones 1 and 2, with different ϕ_n values, were determined. The fundamental frequencies were varied from trial to trial in such a way that the central frequency f_c [= the geometric mean $(f_1 f_2)^{1/2}$] of the pure interval was equal to 370 Hz ±50, ±100, ±150, or ±200 cents, in a random order. As a result, all fundamental frequencies fell within a range of 261 Hz (musical C4) and 523 Hz (musical C5). The f_1 of the pure interval equaled

$$f_1 = f_c/(q/p)^{1/2}. \quad (13)$$

In the mistuned interval, one of the fundamental frequencies was equal to the corresponding fundamental frequency in the pure interval, whereas the other one was altered. Within a series, the frequencies of tones 1 and 2 were altered equally often, the particular order being determined by sampling without replacement. For the mistuned interval, either a new f_1 was calculated by means of

$$f_1(\text{mistuned}) = (p \cdot f_2 \pm f_b)/q, \quad (14)$$

or a new f_2 was calculated by

$$f_2(\text{mistuned}) = (q \cdot f_1 \pm f_b)/p. \quad (15)$$

Within a series, the particular order of stretched vs. compressed intervals (signs in Equations 14 and 15) was randomized. A session comprised five or six series. During the first and second sessions, in which subjects were trained, the adequate range of ΔL was determined. The choice of ΔL demanded careful planning because the series of six equally spaced values of ΔL had to cover the subjects' transition range between chance performance and 100% correct responses. If the presented range of ΔL values did not result in scores around 75%, a better range was chosen. The range appeared to depend on the particular musical interval and on whether tone 1 or tone 2 was attenuated. In addition, the adequate range differed from subject to subject to a small extent. In most cases, the range of ΔL was from 10 to 40 dB. In the $\Delta L < 10$ dB conditions, L_1 and L_2 were attenuated to a small extent to satisfy our criterion that the overall level of the interval should remain at a constant level of 85 dB SPL. The correction (C) in decibels of L_1 and L_2, being dependent on ΔL and on the musical interval (q/p), is given by

$$C = 10 \log\{1 + 10^{[-\Delta L \pm 20 \log(q/p)]/10}\}. \quad (16)$$

The positive and negative signs apply to the $L_1 < L_2$ and $L_1 > L_2$ conditions, respectively.

Results

Discrimination thresholds (DTs) are presented in terms of level difference, ΔL, rather than of level variation depth, D (see Equation 10). This way of presentation was preferred because the exact contribution of the various pairs of beating harmonics to DT is not known a priori. Moreover, other sources such as the perception of waveform periodicity and combination tones may have influenced DT as well (Plomp, 1976).

DTs were subjected to an ANOVA [4 (subjects) × 3 (tone duration) × 2 (fifth or major third) × 2 (attenuation of tone 1 or tone 2) × 5 (beat frequencies of 2, 4, 8, 16, and 32 Hz), all repeated measures]. Although individual differences occurred [$F(3,24) = 94.0$, $p < .00001$], a number of experimental variables affected our subjects' DTs in similar ways.

In general, mean threshold increases when tone duration becomes shorter [$F(2,6) = 68.4$, $p < .0005$]. Tone duration is most effective at low beat frequencies, whereas the differences become less apparent at the highest beat frequency [$F(8,24) = 12.3$, $p < .00001$]. This interaction effect, being slightly dependent on musical interval [$F(8,24) = 2.33$, $p < .05$], is given separately for the fifth and the major third in Figures 6 and 7, respectively. In these figures, DTs are plotted as a function of beat frequency with tone duration as parameter. The DTs for the fifths are lower than for the major thirds [$F(1,3) =$

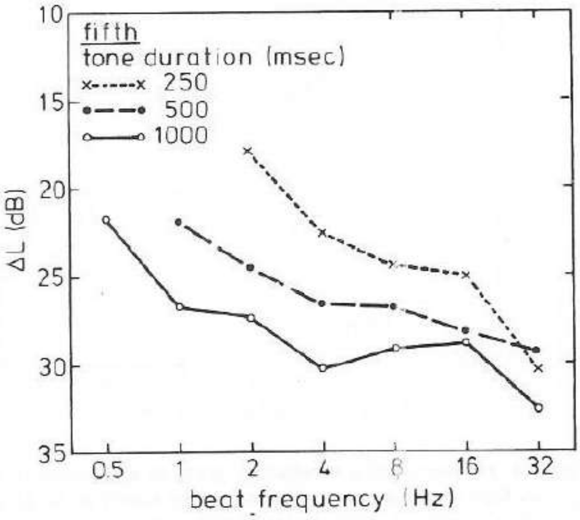

Figure 6. Thresholds for the discrimination between pure and mistuned fifths, plotted as a function of beat frequency, with tone duration as parameter.

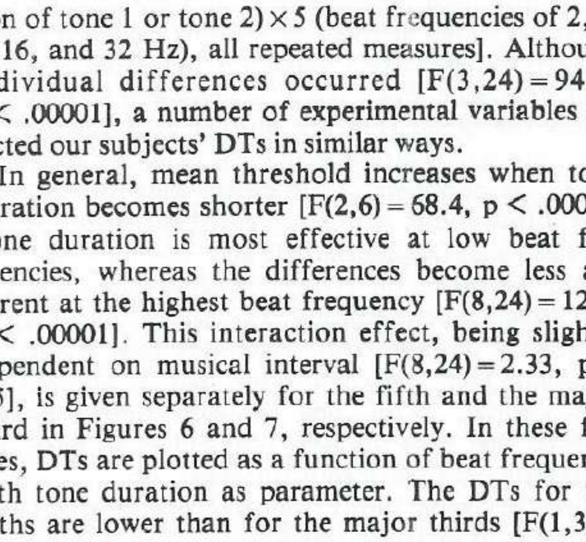

Figure 7. Thresholds for the discrimination between pure and mistuned major thirds, plotted as a function of beat frequency, with tone duration as parameter.

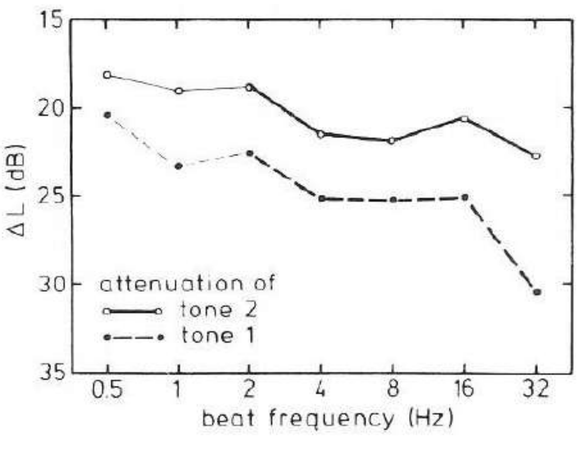

Figure 8. Discrimination thresholds, given as a function of beat frequency with attenuation of either tone 1 or tone 2 as parameter. Mean values, connected by boldly printed lines, are based on thresholds from three different tone durations.

48.0, $p < .005$]. Attenuation of tone 1 results in lower DTs than attenuation of tone 2 [$F(1,3) = 63.2$, $p < .005$]. This effect depended to some degree on beat frequency [$F(4,12) = 3.22$, $p < .05$]. In Figure 8, where the mean DTs are plotted as a function of beat frequency with attenuation of either tone 1 or tone 2 as parameter, it is shown that this attenuation effect is due to the increased difference of the DTs at a beat frequency of 32 Hz.

For the fifths, the effect of beat frequency is similar for all tone durations. The general trend is that these DTs decrease with increasing beat frequency, although three Newman-Keuls paired comparison tests (Winer, 1970), one for each tone duration, showed that DTs for beat frequencies of 4, 8, and 16 Hz are not significantly different.

For the major thirds, however, the effect of beat frequency is different (see Figure 7). At a tone duration of 1,000 msec, tests for trend (Winer, 1970) revealed that most of the variation caused by beat frequency is accounted for by the quadratic trend. The Newman-Keuls paired comparison test showed that the DTs for beat frequencies of .5, 1, and 16 Hz are all significantly different from the DT at 4 Hz ($\alpha = .05$). A similar paired comparison test between the DTs for a tone duration of 500 msec indicated that the mean threshold at 16 Hz is not significantly higher than it is at 8 Hz.

Discussion

The ANOVA on the DTs for beat frequencies from 2 to 32 Hz, as well as three ANOVAs, carried out on the DTs for each tone duration separately, showed that the main effects of musical interval, attenuation of tone 1 or tone 2, and beat frequency were significant in all subgroups of data. In addition, tone duration and its interaction with beat frequency proved to be very effective. These effects will be discussed in detail.

Musical Interval

Mean thresholds for the detection of mistuned fifths were about 7 dB lower than they were for the detection of mistuned major thirds. On the assumption that the perception of various beating pairs of just-noncoinciding harmonics provides one of the most important cues in the discrimination task, we can explain at least part of this difference by refering to the fact that the mean level of interfering harmonics for the fifths was about 5 dB higher than it was for the major thirds. With the fifths, for example, when $L_1 > L_2$, the level of the unattenuated third harmonic of tone 1 is 9.5 dB lower than the level of the fundamental (see middle panel of Figure 4a). When $L_1 < L_2$, the level of the unattenuated second harmonic of tone 2 is 6.0 dB lower than the level of the fundamental (see right-hand panel of Figure 4a). For the major thirds, however, the corresponding values are substantially lower: 14 and 12 dB, respectively. The mean difference of 5.25 dB in the levels of the interfering harmonics holds for all other pairs of just-noncoinciding harmonics. Our hypothesis that the DT depends on the SPL of the interfering harmonics is based on data presented by Maiwald (1967) and by Schöne (1979), who showed that the modulation threshold of AM sinusoids decreases with increasing SPL of the tones. Moreover, Zwicker (1952) found that this threshold decrease does not depend on modulation frequency. The combined results from Maiwald, Schöne, and Zwicker indicate that when the signal level increases from 40 to 80 dB, the level variation depth D (Equation 10) at threshold decreases from about .85 to .35 dB. In terms of ΔL, this threshold decrease is 8 dB. At low sensation levels within a range of 5 to 40 dB, a similar relation between sensation level and threshold was found for beats resulting from the superposition of two sinusoids with a frequency difference of 3 Hz (Riesz, 1928). Since the literature cited above suggests that the effect of sensation level on the DT is rather small, it is unlikely that the difference between the DTs of the fifths and the major thirds can be completely explained by the mean level difference of 5.25 dB between the interfering harmonics.

There is a second factor, which may have also influenced the DTs for the fifths and the major thirds differently. For any beat frequency, f_b, the interval size of the mistuned fifths has to be changed to a greater extent than the interval size of the major thirds (see also Equation 8). From this equation it follows that this difference increases with beat frequency, being 28 cents when f_b equals 32 Hz. The relevance of this second factor is apparent in the trend for the differences between the DTs for the fifths and the major thirds to increase at the higher beat frequencies. The second factor will be discussed in detail in the section on the effect of beat frequency.

Attenuation of Tone 1 or Tone 2

That DTs depended on whether tone 1 or tone 2 was attenuated can be partly explained again by the fact that beats are most easily observed at higher levels of the interfering harmonics, as discussed above: the mean SPLs of the interfering harmonics in the $L_1 < L_2$ conditions are 2.75 dB higher than in the $L_1 > L_2$ conditions. Masking also seems to be a candidate for explaining part of the differences. It is a well-established fact that the masked threshold decreases as a function of the frequency separation of the masker and the maskee (see Plomp, 1976). In the $L_1 < L_2$ condition, the frequency separation of the harmonics of tone 2 is a factor q/p wider than the frequency separation of the harmonics of tone 1 in the $L_1 > L_2$ condition. Thus, less masking of the various pairs of interfering harmonics would be expected in the $L_1 < L_2$ conditions. This is supported by our data, since the DTs were lower in the $L_1 < L_2$ conditions.

Beat Frequency

At tone durations of 500 and 1,000 msec, the DTs for beat frequencies higher than about 4 or 8 Hz tended to decrease when the interval was a fifth and tended to increase when the interval was a major third. We think that these particular patterns can be explained by the combined effect of two different processes, which have opposite effects on DT for beat frequencies from about 4 Hz on. The first mechanism is related to the perception of beats. Recall that both for two superimposed sinusoids and for AM tones, sensitivity to beats increases with increasing beat frequency up to about 4 Hz and decreases again for beat frequencies above this value. It is reasonable to assume a similar sensitivity for the perception of beats for our complex tones.

The second mechanism is based on the perception of interval width and is thus related to musical interval recognition. From Equation 8, it can be inferred that, for a constant beat frequency, the deviation of the mistuned intervals from the pure intervals, $|T|$, decreases with increasing values of f_1. In Figure 9, the DTs for the different beat frequencies, as given in Figures 6 and 7, are replotted as a function of the corresponding values of mean $|T|$, with tone duration as parameter. For beat frequencies of 32 Hz, for example, the corresponding mean $|T|$ of the mistuned fifths and the mistuned major thirds equaled 61 and 34 cents, respectively. We suppose that our subjects, especially in these conditions, identified the mistuned intervals, in at least a number of trials, as not belonging to the pure fifth or pure major third category. Thresholds for identification of the direction of mistuning (Stumpf & Meyer, 1898) and adjustment thresholds for musical intervals (Moran & Pratt, 1926; Rakowski, Note 1) suggest that this proposition is correct.

Tone Duration

There is a downward shift of the DTs with increasing tone duration. This shift is most pronounced for lower beat frequencies. For beat frequencies within a range of .5 to 16 Hz, halving of tone duration of the fifth is equivalent to about quadrupling of beat frequency. For the major third, a simple rule cannot be given because of the strong quadratic trend in the data.

In our experiments, all subjects could discriminate between pure and mistuned intervals, if the mistuned intervals contained at least one half beating period. It

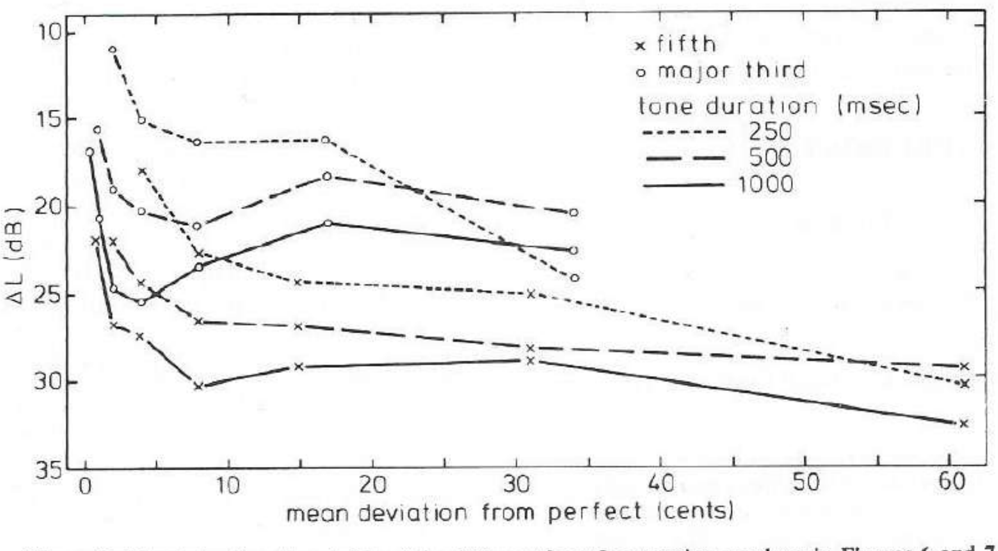

Figure 9. Discrimination thresholds at the different beat frequencies, as given in Figures 6 and 7, replotted as a function of the corresponding values of mean | T |, with musical interval and tone duration as parameters.

would be interesting to test the validity of such a simple rule because, if correct, it would mean that musical intervals, mistuned by only 1/10th of a cent can be discriminated from a pure interval, provided that its duration is long enough.

Presumably, small fluctuations in the pure interval itself and limitations of human memory will determine the lower bounds for discrimination.

The effect of tone duration becomes smaller and less consistent for the highest beat frequencies. The idea that the effect of beat frequency can be explained by the combined effect of two different mechanisms is supported by the interaction effect of beat frequency and tone duration.

In general, perceptibility of beats is enhanced at longer tone durations. Perception of interval size, however, which is based on the ratio of the fundamental frequencies of the two tones, does not seriously deteriorate as a result of tone duration within the range presented in our experiment (Plomp, Wagenaar, & Mimpen, 1973).

In sum, both the perception of beats and the perception of interval size may have influenced the DTs. DTs for fifths at beat frequencies higher than 4 Hz and those for major thirds at beat frequencies higher than 16 Hz deviate from the U-shaped curve for the detection threshold of beats for simple tones. Assuming a similar sensitivity to beats for simple and complex tones, this fact was tentatively explained by the effect of the perception of interval size. In fact, however, no experimental results on thresholds for the perception of beats for intervals consisting of two complex tones have been reported. Therefore, we designed Experiment 2 to determine the sensitivity to beats for both our mistuned fifths and major thirds. This was done for a large number of beat frequencies, including those presented in the first experiment. Comparison of these beat thresholds (BTs) with the DTs should, it was thought, provide an indirect measure of the contribution to the DTs of the perception of differences in interval size.

EXPERIMENT 2

Method

Stimuli

The stimuli were the same as those in Experiment 1.

Apparatus

The apparatus was identical to that of Experiment 1.

Subjects

Four musically trained subjects were tested over three sessions, each session on a different day. The subjects had not participated in Experiment 1 and were naive concerning the purposes of the experiment. Three participants were students from the Institute of Musicology at Utrecht, and the fourth was a student at the University of Amsterdam. All subjects were paid for their services.

Experimental Design

The independent variables were: (1) musical interval (fifth and major third); (2) beat frequency of the first pair of just-noncoinciding harmonics (.5, 1, 2, 4, 6, 8, 12, 16, 24, and 32 Hz); (3) attenuation of tone 1 or tone 2. For every stimulus combination there were four replicas.

Procedure

On each trial, the subjects were presented with a sequence of mistuned intervals. Each interval had a duration of 1,000 msec. Between the intervals, there was a silent period of 600 msec. At the beginning of a trial, ΔL was 0 dB. For every successive interval, ΔL was increased in steps of 2 dB. During such a descending series, the subjects were instructed to press a key as soon as they no longer perceived any beats. After this response, ΔL was decreased in equal steps of 2 dB. In such an ascending series, subjects had to press a second key as soon as they perceived beats again. Within a trial, there were four descending and four ascending series. At the end of a descending series, ΔL was increased by a random amount of 2-10 dB. At the end of an ascending series, ΔL was decreased by a similar variable amount. At the end of every trial, BT was computed by taking the mean of the eight values of ΔL at the times that subjects had pressed the keys. A trial was repeated immediately, if the standard deviation of the eight values of ΔL was greater than 6 dB. Within a run, both the musical interval and the tone with the variable level were held constant. The presentation order of the different beat frequencies was randomized. Presentation orders of the different runs were balanced according to Latin squares. Each session started with some training series. At the beginning of a run, new waveforms of tones 1 and 2 were determined. The fundamental frequency f_1 was held constant at 400 Hz, whereas f_2 depended on interval and beat frequency (see Equation 15). The mistuned intervals were always stretched.

Results

Again, BTs are presented in terms of level difference ΔL. BTs for the fifths are given in Figure 10a as a function of beat frequency, with attenuation of tone 2 ($L_1 > L_2$) and tone 1 ($L_1 < L_2$), respectively. In Figure 10b, BTs for the major thirds are given in a similar way. In both figures, the corresponding DTs from Experiment 1 are given as a reference. BTs were subjected to an ANOVA [4 (subjects) × 2 (fifth and major third) × 2 (attenuation of tone 1 or tone 2) × 10 (beat frequencies), all repeated measures]. The effect of subjects, tested against within-cell variance was significant [$F(3,480) = 37.8$, $p < .00001$]; BTs for fifths and major thirds were not significantly different ($p > .72$); attenuation of tone 1 resulted in lower thresholds than did attenuation of tone 2 [$F(1,3) = 43.5$, $p < .006$]; and BTs were affected by beat frequency [$F(9,27) = 7.19$, $p < .001$]. A Newman-Keuls paired-comparison test revealed that, in general, thresholds for beat frequencies of .5, 1, and 2 Hz are higher than thresholds for 12 and 16 Hz ($\alpha = .05$). In view of the specific purposes of this experiment, however, it is more interesting to relate the BTs to the DTs from Experiment 1.

In a comparison of thresholds from Experiments 1 and 2, it was found that the effect of beat frequency for the major thirds was about the same for both BTs and DTs (see Figure 10b). For the fifths, the DTs

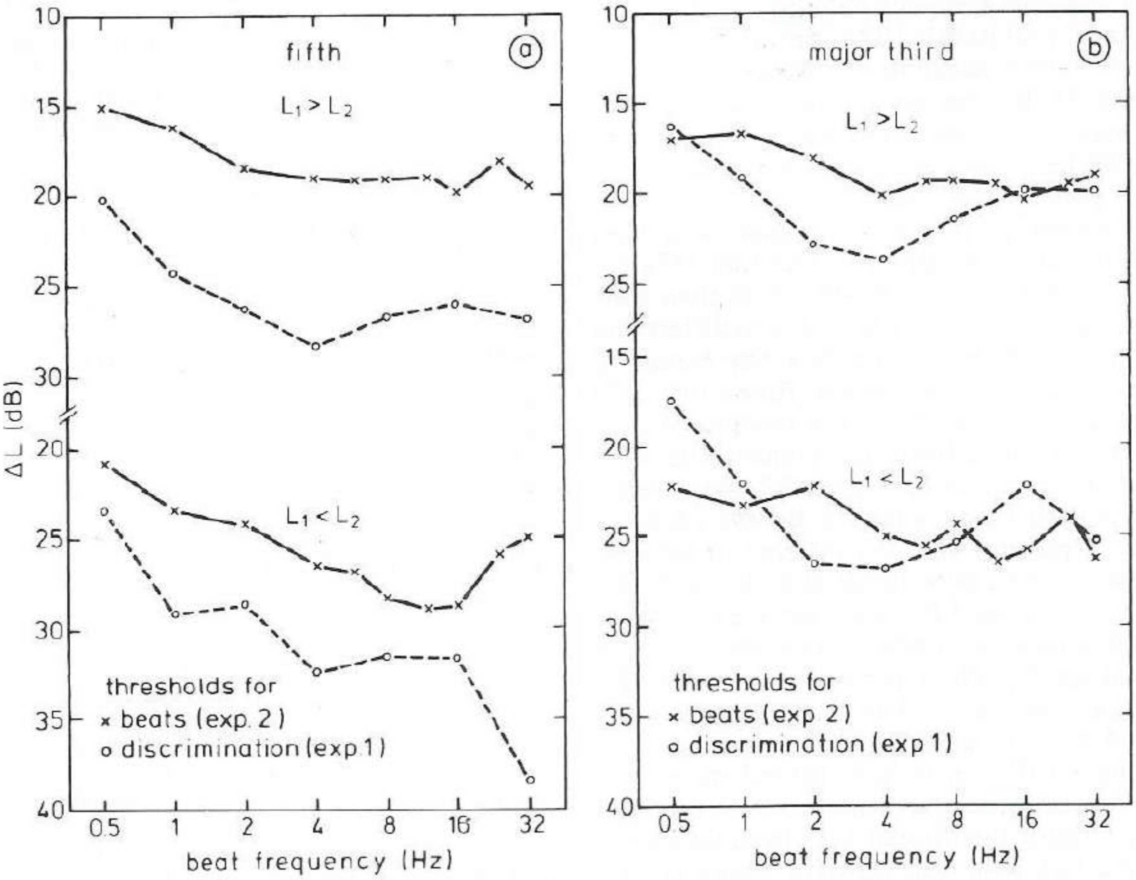

Figure 10. (a) Beat thresholds from Experiment 2 and discrimination thresholds from Experiment 1 for the fifths, plotted as a function of beat frequency with attenuation of tone 2 ($L_1 > L_2$) and attenuation of tone 1 ($L_1 < L_2$), respectively. (b) Beat thresholds from Experiment 2 and discrimination thresholds from Experiment 1 for the major thirds, plotted as a function of beat frequency with attenuation of tone 2 ($L_1 > L_2$) and attenuation of tone 1 ($L_1 < L_2$), respectively.

were lower than the BTs (see Figure 10a). In the $L_1 > L_2$ conditions, the dependence on beat frequency was similar for both types of thresholds. In the $L_1 < L_2$ conditions, however, the patterns deviated for beat frequencies higher than 16 Hz. To test the significance of this deviation, these thresholds were subjected to an ANOVA with threshold type as a between-subjects variable and beat frequency as a within-subjects variable. Only the levels of beat frequency that were presented in both experiments were tested. The interaction between threshold type and beat frequency was significant [$F(6,36) = 3.81$, $p < .005$]

Discussion

The main purpose of Experiment 2 was to ascertain to what extent DTs were due to beats rather than to the perception of differences in interval size. If the perception of differences in interval size contributes to the DTs, DTs and BTs should interact with beat frequency. Deviations between DTs and BTs are expected only for the higher beat frequencies, since the differences in interval width for low beat frequencies are too small to be detected. For the fifth, the interaction pattern in the $L_1 < L_2$ condition strongly suggests that the perception of interval width contributes to DT. For a beat frequency of 32 Hz, DT decreases, whereas BT increases. These interaction effects were not found for either the fifths in the $L_1 > L_2$ conditions or the major thirds in the $L_1 > L_2$ and $L_1 < L_2$ conditions. Moreover, although the BTs in the latter conditions were higher for the lowest beat frequencies, sensitivity to beats in complex tone intervals does not have a clear maximum such as the one found with superimposed sinusoids.

In Experiment 1, a mean deviation of 34 cents from the pure major third is apparently not large enough for the second mechanism to contribute to the DT. On the other hand, one may ask why the DT did not decrease for the fifth in the $L_1 > L_2$ condition. For judgments of interval width to be possible, the levels of both tone 1 and tone 2 should be above masking threshold. It seems plausible to attribute the discrepancy between the $L_1 < L_2$ and $L_1 > L_2$ conditions to the fact that tone 1 is a more effective masker of tone 2 than vice versa.

We had not expected to find that BTs for the fifths would be consistently higher than the DTs, whereas such differences could not be found for the major thirds. We can only speculate about the reason. It is possible that the subjects in Experiment 2 were rather conservative in their decisions whether they perceived beats or not. If this was not the case, the data suggest that apart from beats, mistuned intervals near threshold do bear more information distinguishing them from pure intervals than do beats alone. One additional source of information could have been small differences in timbre. The fact that DTs for major thirds were not consistently lower than their corresponding BTs may have been due to differences between the experimental procedures by means of which thresholds were determined. Recall that DTs were investigated by means of the method of constant stimuli, in which fifths and major thirds were randomized within blocks, whereas BTs were determined by employing an adaptive Békésy up-down method. It is possible that the differences between DTs for fifths and major thirds should not be attributed only to level differences between the just-noncoinciding harmonics and to differences in relative interval size by which pure intervals were mistuned, as was done above. The fact that pure major thirds sound rather rough, compared with the smooth-sounding pure fifths, may have caused more confusion in discrimination between the pure and mistuned major thirds than would have been the case if interval type had been held constant within blocks.

EXPERIMENT 3

Three important questions remain unanswered: (1) Are the differences between BTs and DTs due to differences between the employed experimental procedures? (2) Can these differences be attributed to the fact that DTs and BTs were determined for *different* groups of subjects? (3) What do thresholds for the identification of stretched and compressed intervals look like?

In order to obtain more direct evidence for the relevance of both the perception of beats and the perception of interval width, we designed a new experiment. In Experiment 3, not only DTs and BTs, but also thresholds for identification of the direction of mistuning (ITs) were determined for the *same* group of subjects in similar experimental conditions.

Method

Stimuli

The stimuli were the same as those in Experiment 1. The overall level of tones 1 and 2, however, was 80 dB SPL. Rise and decay times of the tone bursts were 30 and 20 msec, respectively. In contrast to Experiment 1, attenuation of either the low tone or the high tone started from the base condition, as depicted in the left-hand panels of Figures 4a and 4b. During the experimental runs, the levels of both tone 1 and tone 2 were corrected (see Equation 16) to remain at the same overall level of 80 dB SPL.

Apparatus

The apparatus was identical to that in Experiment 1.

Subjects

Six music students from the Conservatory of Utrecht were tested over six sessions, each session on a different day. All of them had high marks for ear training. The subjects had not participated in Experiment 1 or Experiment 2. They were paid for their services.

Experimental Design

Thresholds for the identification of the direction of mistuning. The variables were: (1) musical interval (fifth and major third); (2) level difference ΔL between the tones (0 and 20 dB); (3) deviation of the mistuned intervals from perfect intervals (± 2.5, ± 5, ± 7.5, ± 10, ± 15, ± 20, ± 25, ± 30, ± 35, ± 40, ± 45, ± 50, ± 60, ± 70, and ± 80 cents); and (4) attenuation of tone 1 or tone 2. Variables 1, 2, and 4 resulted in six combinations, because attenuation of tone 1 or tone 2 is relevant only in the $\Delta L = 20$ dB condition. For each of these conditions, the 30 different amounts of mistuning were presented 20 times.

At the end of the experiment it turned out that DTs were lower than 20 dB in most conditions. Since this experiment was designed mainly to relate the three different thresholds to each other, we decided to determine additional ITs in the following conditions: (1) fifth, $L_1 < L_2$, $\Delta L = 33$ dB, (2) fifth, $L_1 > L_2$, $\Delta L = 25$ dB, and (3) major third, $L_1 < L_2$, $\Delta L = 25$ dB. In these conditions, deviations of the mistuned intervals from perfect were ± 10, ± 20, ± 30, ± 40, ± 50, ± 60, ± 70, and ± 80 cents. Again, the 16 different degrees of mistuning were presented 20 times.

Beat thresholds and discrimination thresholds. The independent variables were: (1) musical interval (fifth and major third), (2) beat frequency of the first pair of just-noncoinciding harmonics (1, 5, 15, and 25 Hz), and (3) attenuation of tone 1 or tone 2. For all stimulus combinations, thresholds were determined twice. The dependent variable was ΔL.

Determination of Thresholds for Identification of the Direction of mistuning

The percentage of compressed responses ($n = 20$) for the 30 (or 16) different values of mistuning were converted to z scores. The parameters a and b of the linear regression function $z = b + ax$ were determined, where x equals the deviation of the mistuned interval from perfect in cents. From the linear function, thresholds meeting any desired criterion can be derived (z/a). Moreover, the point of subjective purity, $x = -b/a$, that is, the amount of mistuning that results in 50% stretched and 50% compressed responses, can be computed as well.

Procedure

General. The order in which thresholds for the different aspects were determined was balanced: one (3×3) Latin square for the first determination of BTs and DTs and for presentation of half of the trials in the identification task, another (3×3) Latin square for the second determination and the remaining trials in the identification task. Two subjects were assigned to each different presentation order. The duration of the tone bursts was 1,000 msec.

Identification thresholds for the direction of mistuning. Each trial consisted of a mistuned interval, which was presented twice. The silent interval between these identical stimuli was 1.3 sec. By means of a response box, the subject indicated whether the interval was compressed or stretched, relative to the pure interval. It should be noted (1) that the pure interval was presented only once, at the very beginning of a series of 300 trials, and (2) that no feedback was given. If, at each trial, both the pure and the mistuned interval had been presented, perception of pitch shifts rather than interval width would have been studied. In this experiment, subjects had to imagine a pure interval themselves. Information about the size of this subjective reference interval would have been lost if feedback had been given. Thresholds determined in the conditions described above can be conveniently compared with the DTs, since in the discrimination conditions, the reference was not explicitly given either, but had to be identified by the subjects themselves.

The next trial was presented 1.7 sec after the response to the preceding trial. Within a series, the stimuli were presented in a random order. (In a pilot experiment, a double staircase method had been employed: two intermingled series, one for stretched and one for compressed intervals. Since we felt that in this paradigm undesirable cues could be used, the method of constant stimuli was chosen in our main experiment.) Presentation order of the different series, in which both the level of ΔL and the musical interval were held constant, was randomized. For values of $\Delta L >$ 20 dB, the number of trials within a series was reduced to 160. These series were presented to the same group of subjects in a random order after the main plot, as described in the general procedure, had been finished. Fundamental frequencies were varied from trial to trial in such a way that the central frequency of the mistuned interval was equal to 370 Hz ±25, ±75, ±125, ±175, or ±225 cents, in a random order. The f_1 of the interval equaled

$$f_1 = f_c/2^{(W+T)/2,400}, \quad (17)$$

in which W equals 1,200 $\log_2$ (q/p), and T equals the amount of mistuning relative to W. In the same vein, f_2 was calculated by

$$f_2 = f_c \cdot 2^{(W+T)/2,400}. \quad (18)$$

A session comprised five or six series. The subjects were not trained.

Beat thresholds. The procedure was similar to that in Experiment 2. The silent period between the intervals, however, was 1 sec. The fundamental frequency f_1 was computed by means of Equation 13, with f_c always at 370 Hz. The fundamental frequency f_2 was computed by means of Equation 15. The mistuned intervals were always stretched.

Discrimination thresholds. The task for the subject was to discriminate between pure and mistuned musical intervals in a 2-AFC paradigm. The mistuned interval was presented first in half of the trials and second in the other half of the trials. Between the intervals was a silent period of 1.5 sec. By means of a response box, the subject indicated whether the first or the second interval was mistuned. Feedback was given by a red light above the correct button. One subject was visually handicapped; after each correct response, she received feedback by means of a soft tone with a fundamental frequency of 784 Hz and a duration of 250 msec. The next trial was presented 1.7 sec after the response to the preceding trial. DTs were determined by means of an adaptive procedure, in which ΔL was increased by 2 dB after three correct responses and decreased by the same amount after an incorrect response. Each reversal of the direction of change in ΔL over successive trials was counted as a turnaround. There were six turnarounds. In the first run, however, ΔL, being 0 dB at the very first trial, was increased by 3 dB after each correct response. The effective DT was taken as the mean value of ΔL on the trials that resulted in the last four turnarounds. This value estimates the value of ΔL required to produce 79.4% correct responses in a nonadaptive 2-AFC procedure (Levitt, 1971). A series comprised four blocks, one for each beat frequency. Presentation order of the different blocks was randomized. Presentation order of musical interval and attenuation of tone 1 or tone 2, which were held constant within a series, was balanced. At the beginning of a series, new waveforms of tones 1 and 2, with different ϕ_n values, were determined. Fundamental frequencies of the pure and mistuned intervals were computed in the same way as in Experiment 1. The central frequency, f_c, of the pure interval, however, was equal to 370 Hz ±25, ±75, ±125, ±175, or ±225 cents, in a random order. The direction of mistuning (signs in Equations 14 and 15) was randomized. Subjects received two short training series in the first session.

Results and Discussion

Identification

For each subject, thresholds at which the direction of mistuning was consistently identified in 75% of the cases (.67/a) were determined for nine different conditions. For the same conditions, points of subjective purity were also derived. It turned out that the data from one particular subject were very different from those of all the other five subjects. This subject, having a mean IT of 11 cents and a mean point of subjective purity of −35 cents, was therefore dropped from further analysis.

For the remaining five subjects, mean thresholds and mean points of subjective purity are plotted in panels a and b of Figure 11, for each condition separately. The standard deviation, σ_{n-1}, around these mean values is indicated by bars. Mean ITs for the fifth and the major third are not significantly different and range from about 20 to 30 cents. These findings are more in line with adjustment thresholds for musical intervals than with ITs, as found by Stumpf and Meyer. It is plausible to attribute the discrepancy between Stumpf and Meyer's data and our results to differences in stimulus complexity. For both the fifth and the major third, t tests revealed that a 20-dB attenuation of the higher tone ($L_1 > L_2$) resulted in significantly higher thresholds than a 20-dB attenuation of the lower tone ($L_1 < L_2$; $p < .05$). Moreover, for the fifth, the 25-dB attenuation in the $L_1 > L_2$ condition yielded a higher threshold than the 20- and 33-dB attenuation in the $L_1 < L_2$ condition ($p < .05$). These results may be explained by masking effects.

For both the fifth and the major third, and for all different values of ΔL, the mean points of subjective purity are very close to zero. In other words, mean subjective purity is not significantly different from physical purity.

Discrimination

DTs were subjected to an ANOVA [5 (subjects) × 2 (fifth or major third) × 2 ($L_1 < L_2$ or $L_1 > L_2$) × 4 (beat frequency), all repeated measures]. The effect of subjects, tested against within-cell variance, is significant [$F(4,80) = 4.8$, $p < .005$]. DTs for the fifth are about 6 dB lower than for the major third [$F(1,4) = 137.0$, $p < .001$]. Attenuation of tone 1 results in lower DTs than attenuation of tone 2 [$F(1,4) = 38.2$, $p < .005$]. This effect, however, is most prominent for the fifth [$F(1,4) = 27.5$, $p < .01$]. DTs are highest for the lowest beat frequency [$F(3,12) = 4.9$, $p < .02$]. There is a tendency for the effect of beat frequency to depend both on the kind of musical interval [$F(3,12) = 3.3$, $p < .05$] and on the $L_1 > L_2$ or $L_1 < L_2$ conditions [$F(3,12) = 3.1$, $p < .06$]. The interaction patterns can be inspected in Figure 12, in which DTs for the fifth and major third are plotted as a function of the mean deviation from perfect, with attenuation of tone 2 ($L_1 > L_2$) and attenuation of tone 1 [$L_1 < L_2$), respectively. With respect to the effect of beat frequency for the major third, tests for trend (Winer, 1970) revealed that only the cubic trend is significant [$F(1,12) = 11.5$, $p < .01$].

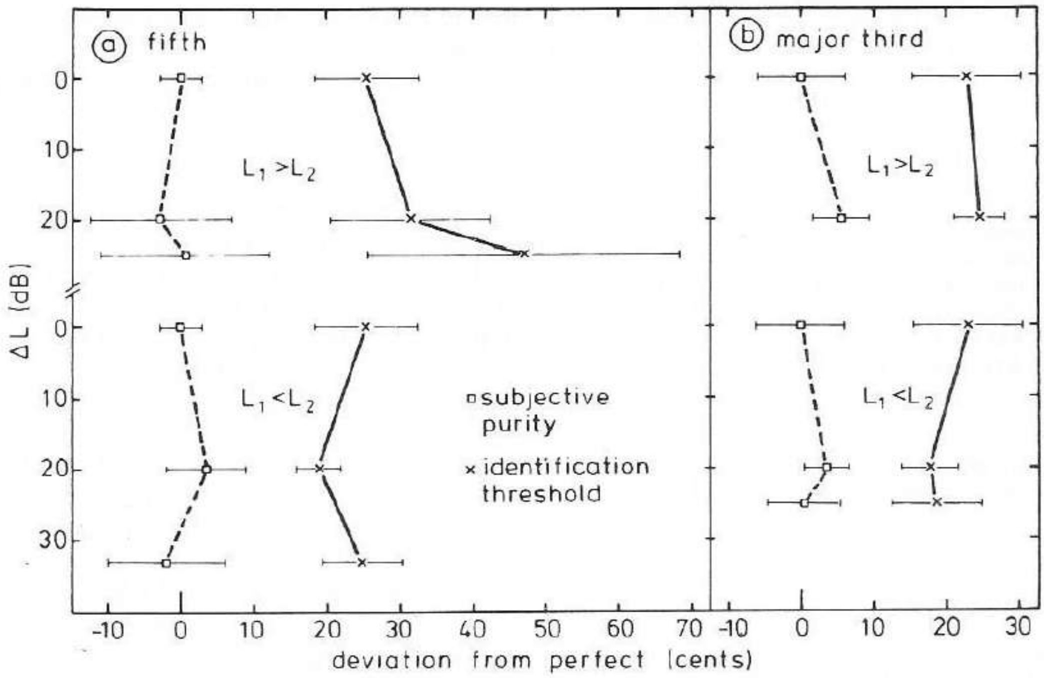

Figure 11. Mean identification thresholds and mean points of subjective purity, plotted with the deviation from a perfect fifth in panel a and a pure major third in panel b as the dependent variable. Data are given for the $L_1 < L_2$ and $L_1 > L_2$ conditions and different values of ΔL, separately. Standard deviations, σ_{n-1}, are indicated by bars. For every subject, the identification threshold, as presented here, is based on two parts: (1) a small interval, added to the pure interval, required for 75% stretched responses, and (2) a small interval, subtracted from the pure interval, required for 75% compressed responses. The identification threshold is equal to the mean of these two tempering intervals. The *sign* of the mean point of subjective purity indicates which interval, the first or the second, was larger.

Beats

BTs are plotted in Figure 12 in the same way as the DTs. An ANOVA with a design similar to that used for the DTs, was carried out. There are substantial differences between subjects [$F(4,80) = 31.3$, $p < .00001$]. There is a weak tendency for the thresholds for the fifth to be lower than those for the major third [$F(1,4) = 5.6$, $p < .08$]; the main effect of beat frequency is far from significant ($p > .22$). Sensitivity to beats is higher in the $L_1 < L_2$ condition than in the $L_1 > L_2$ condition [$F(1,4) = 10.5$, $p < .03$]. Combined for the fifth and major third, sensitivity to beats decreases at the highest beat frequency only in the $L_1 < L_2$ condition [$F(3,12) = 4.7$, $p < .02$].

Comparison of Discrimination and Beat Thresholds

DTs and BTs were subjected to an ANOVA [5 (subjects) × 2 (DT or BT) × 2 (fifth or major third) × 2 ($L_1 < L_2$ or $L_1 > L_2$) × 4 (beat frequency), all repeated measures]. BTs are significantly higher than DTs only for the fifth [$F(1,4) = 17.5$, $p < .01$]. Combined for the fifth and major third, the differences between the DT and BT increase at the highest beat frequency only in the $L_1 < L_2$ condition [$F(3,12) = 4.5$, $p < .02$]. The patterns of DTs and BTs, as presented in Figure 12, can be compared with the corresponding patterns from Figure 10. In fact, the effects for the fifth in the $L_1 > L_2$ and $L_1 < L_2$ conditions, and for the major third in the $L_1 > L_2$ condition, were confirmed. For the major third in the $L_1 < L_2$ condition, however, the significant deviation of DT and BT at the greatest mistuning ($t = 2.33$, $p < .05$) suggests that the perception of interval width, which was claimed to play a part at the high beat frequencies in the $L_1 < L_2$ condition of the fifth (see Figure 10a), has contributed to DT as well.

One may wonder how much weight should be attributed to the fact that the mean BTs are higher than the corresponding DTs. The criterion on which decisions for the presence or absence of beats are based had to be chosen by our subjects themselves. The fact that 34.5% of the variance in BTs from Experiment 3 was explained by the main effect of subjects, whereas this percentage was only 4.7 for the DT shows that this criterion was strongly dependent on the subject. Moreover, correlation coefficients between DTs and BTs per se are low. On the other hand, during the experiment, the criteria for every subject remained relatively stable, since within-cell variance for BTs and DTs were about the same: 36% and 40%, respectively. Therefore, it is more appropriate to com-

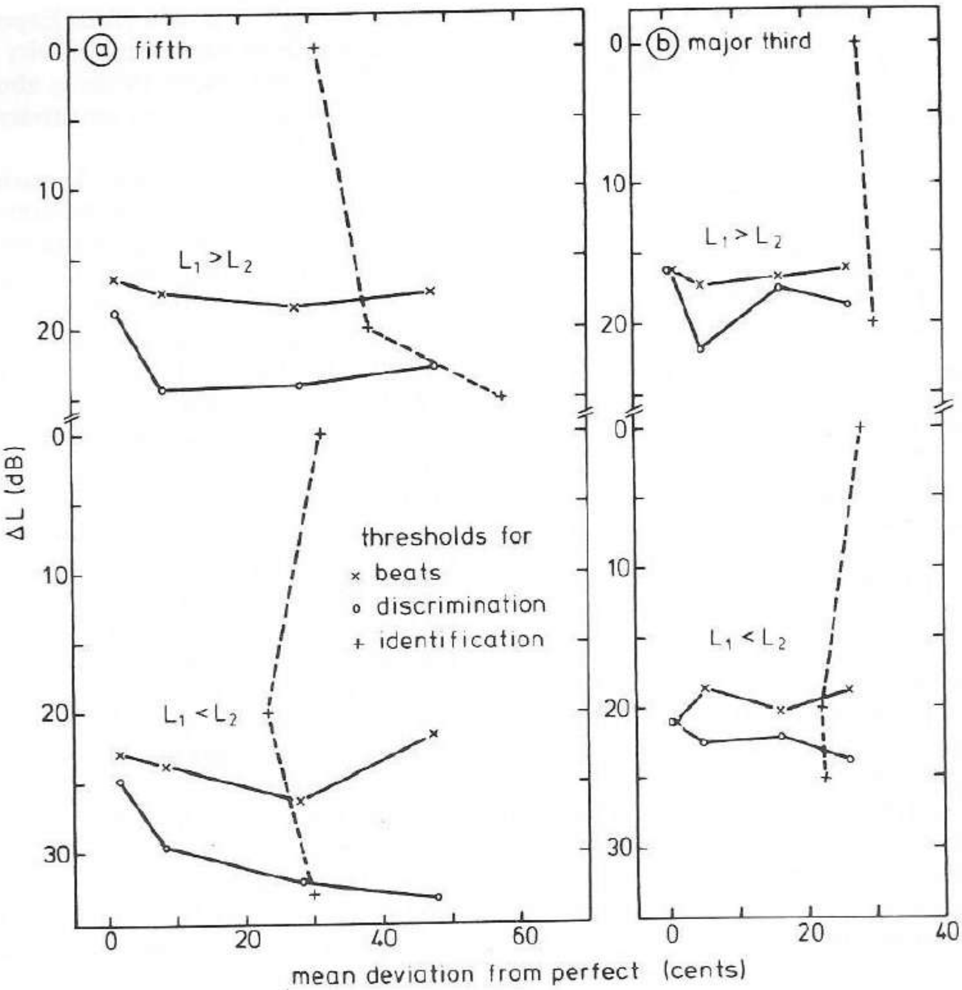

Figure 12. Beat thresholds and discrimination thresholds, plotted as a function of the mean deviation from perfect with attenuation of tone 2 ($L_1 > L_2$) and attenuation of tone 1 ($L_1 < L_2$), respectively. Also presented are identification thresholds from Figure 11, but now for a performance criterion of 79.4%, with the deviation from perfect as the dependent. Thresholds are given for the $L_1 > L_2$ and $L_1 < L_2$ conditions and for the different ΔL values, separately. The different thresholds for the fifth and the major third are given in panels a and b, respectively.

pare the *patterns* of the thresholds and, as we shall see below, define changes in threshold levels relative to thresholds from comparable conditions.

Discrimination, Beat, and Identification Thresholds Together

For the data from Experiment 3, the plausibility of the idea that both sensitivity to beats and the perception of interval width may have contributed to the DT can be supported more directly because ITs are available. Therefore, ITs from Figure 11, but now, like the DTs, for a performance criterion of 79.4%, are replotted in Figure 12. The specific interaction patterns between BTs, ITs, and DTs can be tentatively described by the following rule. The DT depends on BT and IT in a compensatory manner. For example, if the BT at the highest beat frequency is relatively high, but the IT is low, the DT remains low (fifth and major third, $L_1 < L_2$). If the IT is high, but the BT is relatively low, the DT remains low (fifth and major third, $L_1 > L_2$). A consequence of this rule is that if both the BT and the IT are high, the DT has to be high as well. We realize that the patterns shown in Figure 12 are no more than suggestive. A stronger relation between DT and IT can be demonstrated if we correlate the individual decrements of the DT for a beat frequency of 25 Hz, relative to the mean of the corresponding DTs for beat frequencies of 1, 5, and 15 Hz, with the individual ITs for the corresponding $\Delta L = 20$ db conditions. The correlation coefficients are given in the first column of Table 1. Especially for the fifth, where the mean amount of mistuning at a beat frequency of 25 Hz is 22 cents higher than it is for the major third, the r values of $-.83$ and $-.73$ are high. If we correlate the relative decrease of the DT with a similar decrease of the sensitivity to beats, high correlations are found for the $L_1 > L_2$ conditions (see second column of

Table 1
Prediction of the Relative Decrease of the Discrimination Threshold (X_1) From Identification Threshold (X_2) and Relative Decrease of the Beat Threshold (X_3) by Means of the Multiple Linear Regression Equation
$$X_1' = m_1 X_2 + m_2 X_3$$

Condition		r_{12}	r_{13}	$r_{1.23}$	m_1	m_2
Fifth	$L_1 < L_2$	−.83	.25	.93	−.91	.42
	$L_1 > L_2$	−.73	.86	.94	−.42	.67
Major Third	$L_1 < L_2$	−.19	.37	.45	.51	.81
	$L_1 > L_2$	−.56	.85	.85	−.11	.79

Note—All scores were normalized. For both the discrimination and beat thresholds, the relative decrease is defined as the difference between the thresholds at a beat frequency of 25 Hz and the mean of the corresponding thresholds at beat frequencies of 1, 5, and 15 Hz. For all conditions, identification thresholds were taken from the corresponding $\Delta L = 20$ dB conditions. The different correlation coefficients and regression weights are given separately for the fifths and major thirds in both the $L_1 < L_2$ and $L_1 > L_2$ conditions.

Table 1). As shown in the third column in Table 1, the relative decrease of the DT can be predicted even better, if both the IT and the relative decrease of the BT are considered simultaneously. For each condition, the relative contribution of the perception of interval size and sensitivity to beats to the decrease of the DT at the highest beat frequency can be determined. For the fifth in the $L_1 < L_2$ condition, the explained variance increases by 17% if sensitivity to beats is also taken into account. In the $L_1 > L_2$ conditions, however, the explained variance increases by 35% for the fifth and by 41% for the major third. Only for the major third in the $L_1 < L_2$ condition is neither the IT nor the increased sensitivity to beats a very effective predictor of the relative decrease of the DT. For the remaining conditions, however, the compensatory character, in which the DT depends on the IT and BT, is suggested by the values of the regression weights, which are given in columns 4 and 5 of Table 1. For each condition, prediction is optimized by giving either identification (m_1) or beats (m_2) the highest weight.

GENERAL CONCLUSIONS

Combining the results from Experiments 1 and 3, we may conclude that DTs (1) are higher for major thirds than for fifths, (2) are highest for the lowest beat frequencies, and (3) decrease with increasing tone duration. In addition, (4) the effect of tone duration is highest at low beat frequencies.

From Experiment 3, we may conclude that (5) ITs are about the same for fifths and major thirds, deviations from perfect at threshold ranging from 20 to 30 cents, and (6) mean subjective purity is not significantly different from physical purity.

Combining the results from Experiments 2 and 3, we may conclude that (7) sensitivity to beats for mistuned fifths and major thirds is about the same and (8) beat frequency affects sensitivity to beats in only some conditions.

If we are willing to relate the patterns, rather than the absolute values of the different thresholds, we may conclude that (9) discrimination is mainly determined by sensitivity to beats. Detailed analysis, however, indicates that (10) the perception of interval size has been a relevant aspect in discrimination, especially for the fifths in the $L_1 < L_2$ condition.

In the introduction, we proposed that knowledge of our sensitivity to mistuning might provide at least one criterion for the evaluation of different tuning systems. In this context, the first conclusion may be related to the fact that in the tuning system that is applied most frequently, that is, the equal temperament, fifths are mistuned only slightly (2 cents), taking the considerable mistuning of the major thirds (14 cents) into the bargain. Moreover, the decreased sensitivity to mistuning of the major third for beat frequencies around, say, 15 Hz, as found in Experiments 1 and 3, suggests once more how to optimize the purity of the fifth and major third. From conclusions 3 and 4, it can be predicted that the tolerance of mistuned intervals is higher in passages with short tones than with sustained tones, provided the mistuning is not too large. Evidently, tolerance of mistuned intervals, presented within a musical context, will depend on both the sensitivity to mistuning, as has been determined here for isolated intervals, and the musical context itself. The fifth conclusion squares with the general experience that Pythagorean thirds and, especially, the wolf from regular mean tone temperament are easily identified.

REFERENCE NOTE

1. Rakowski, A. *Tuning of isolated musical intervals.* Paper presented at the 91st Meeting of the Acoustical Society of America, Washington, D.C., 1976.

REFERENCES

Barbour, J. M. Musical logarithms. *Scripta Mathematica*, 1940, **7**, 21-31.

Burns, E. M., & Ward, W. D. Categorical perception-phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, 1978, **63**, 456-468.

Delezenne, M. Mémoire sur les valeurs numériques des notes de la gamme. *Recueil des travaux de la Société des Sciences de Lille*, 1826-1827, 1-71.

Ebel, H. Das Hören von Amplitudenmodulationen. *Acustica*, 1952, **2**, 246-250.

Harris, J. D. Loudness discrimination. *Journal of Speech and Hearing Disorders*, 1963, Monograph Supplement II.

Helmholtz, H. L. F. von. [On the sensations of tone as a physiological basis for the theory of music] (A. J. Ellis, trans.). New York: Dover, 1954. (Originally published, 1877.)

KAMEOKA, A., & KURIYAGAWA, M. Consonance theory, Part II: Consonance of complex tones and its calculation method. *Journal of the Acoustical Society of America*, 1969, **45**, 1460-1469.
KILLAM, R. N., LORTON, P. V., & SCHUBERT, E. D. Interval recognition: Identification of harmonic and melodic intervals. *Journal of Music Theory*, 1975, **19**, 212-235.
KLING, J. W., & RIGGS, L. A. *Woodworth and Schlosberg's experimental psychology* (3rd ed.). New York: Holt, Rinehart and Winston, 1972.
LEVITT, H. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 1971, **49**, 467-477.
MAIWALD, D. Ein Funktionsschema des Gehörs zur Beschreibung der Erkennbarkeit kleiner Frequenz- und Amplitudenänderungen. *Acustica*, 1967, **18**, 81-92.
MORAN, H., & PRATT, C. C. Variability of judgments on musical intervals. *Journal of Experimental Psychology*, 1926, **9**, 492-500.
PIKLER, A. G. Logarithmic frequency systems. *Journal of the Acoustical Society of America*, 1966, **39**, 1102-1110.
PLOMP, R. *Aspects of tone sensation. A psychophysical study*. London: Academic Press, 1976.
PLOMP, R., & LEVELT, W. J. M. Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, 1965, **38**, 548-560.
PLOMP, R., WAGENAAR, W. A., & MIMPEN, A. M. Musical interval recognition with simultaneous tones. *Acustica*, 1973, **29**, 101-109.
RIESZ, R. R. Differential intensity sensitivity of the ear for pure tones. *Physical Review*, 1928, **31**, 867-875.
SCHÖNE, P. Messungen zur Schwankungsstärke von amplitudenmodulierten Sinustönen. *Acustica*, 1979, **41**, 252-257.
SIEGEL, J. A., & SIEGEL, W. Absolute identification of notes and intervals by musicians. *Perception & Psychophysics*, 1977, **21**, 143-152. (a)
SIEGEL, J. A., & SIEGEL, W. Categorical perception of tonal intervals: Musicians can't tell sharp from flat. *Perception & Psychophysics*, 1977, **21**, 399-407. (b)
STUMPF, C., & MEYER, M. Maassbestimmungen über die Reinheit consonanter Intervalle. *Beiträge zur Akustik und Musikwissenschaft*, 1898, **2**, 84-167.
TERHARDT, E. On the perception of periodic sound fluctuations (roughness). *Acustica*, 1974, **30**, 201-213.
TERHARDT, E. Ein psychoakustisch begründetes Konzept der musikalischen Konsonanz. *Acustica*, 1976, **36**, 121-137.
WINER, B. J. *Statistical principles in experimental design* (International student ed.). London: McGraw-Hill, 1970.
YOUNG, R. W. Terminology for logarithmic frequency units. *Journal of the Acoustical Society of America*, 1939, **11**, 134-139.
ZATORRE, R. J., & HALPERN, A. R. Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Perception & Psychophysics*, 1979, **26**, 384-395.
ZWICKER, E. Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones. *Acustica*, 1952, **2**, 125-133.

(Manuscript received April 27, 1982;
revision accepted for publication July 12, 1982.)

Perception & Psychophysics
1984, 35 (2), 173-185

3

Spectral effects in the perception of pure and tempered intervals: Discrimination and beats

JOOS VOS
Institute for Perception TNO, Soesterberg, The Netherlands

Thresholds for discrimination between pure and tempered musical intervals consisting of simultaneous complex tones have been investigated previously (Vos, 1982). The aim of the present research was to answer three questions: (1) To what extent are these thresholds determined by the interference of just-noncoinciding harmonics? (2) Is the beat frequency of the first pair of these adjacent harmonics equal to the perceived beat frequency of the tempered intervals? (3) Are differences in discriminability at threshold representative of differences between the perceived strength of beats in supraliminal conditions? In the first three experiments, the answer to the first question was sought by investigating the effect of spectral content of the tones on the discrimination thresholds (DTs) for the fifth (Experiment 1) and the major third (Experiments 2 and 3). The results showed that, for moderately tempered fifths, DTs are determined mainly by the interference of the first pair of adjacent harmonics. For major thirds, however, the data suggest that the interference of other harmonics plays a role as well. In Experiment 4, attempts were made to answer the second and third questions: In supraliminal conditions, both the dominantly perceived beat frequency and the perceived beat strength were determined for various spectral conditions and degrees of tempering. For both the fifth and the major third and for all spectral conditions, the dominantly perceived beat frequency was in most cases equal to the frequency difference of the first pair of adjacent harmonics. Comparison of the DTs and the perceived strength of the beats for corresponding spectral conditions revealed that, especially for the fifth, perceived strength of beats and sensitivity to moderately tempered intervals are highly correlated.

From a classical point of view, a consonant interval is a musical interval in which the ratio of the fundamental frequency f_1 of the lower tone and the fundamental frequency f_2 of the higher tone is equal to p:q, with p and q ($p < q$) being small integers. For two simultaneously sounding complex tones, slightly tempered consonant intervals (f_1:$f_2 \sim$ p:q) are characterized by small frequency differences between those harmonics which coincide completely in pure intervals. The interference of these just-noncoinciding harmonics gives rise to the perception of beats.

The occurrence of beats must have been known for a long time, and attempts to explain this phenomenon were first made in the 17th century. As will be elucidated below, beats play an important role in tuning fixed-pitch keyboard instruments, and are essential in models on tonal or sensory consonance. It has recently been demonstrated (Vos, 1982) that thresholds for discrimination between pure and moderately tempered musical intervals are mainly determined by sensitivity to the beats. This same study showed that when the tempered intervals deviated from purity by more than 20 to 30 cents, perception of interval width could also become a relevant aspect in discrimination.

It was the aim of the present study (1) to investigate which interfering harmonics might have determined the discrimination threshold in the experimental conditions described by Vos (1982), and (2) to determine, for supraliminal conditions, the effect of the spectral content of the tones on both the perceived frequency and the perceived strength of the beats.

This research was supported by Grant 15-29-05 from The Netherlands Organization for the Advancement of Pure Research (ZWO). Part of this paper is an extended written version of a contribution to the Fourth Workshop on Physical and Neuropsychological Foundations of Music, Ossiach/Austria, August 8-12, 1983. The author is indebted to Reinier Plomp and to Rudolf Rasch, who made critical comments on an earlier version of this paper. Thanks are also extended to two referees for their valuable comments. Requests for reprints should be sent to Joos Vos, Institute for Perception TNO, Kampweg 5, 3769 DE Soesterberg, The Netherlands.

Beats and Tuning

Beats have been and still are used very often in the process of tuning musical instruments (see, e.g., Barnes, 1959; Ellerhorst, 1936; Elliston, 1916; Scheibler, 1834, 1838; Smith, 1749/1966). Pure consonant intervals can be obtained by freeing them of beats, for example, the fifths ($p=2$; $q=3$) in the Pythagorean system or the major thirds ($p=4$; $q=5$)

in the mean-tone system for tuning keyboard instruments. Only after the quantitative relations between fundamental frequency, musical intervals, temperaments, and beat frequencies were known was it possible to set the bearings of more complex tuning systems in a reliable way. This was the case in 1749 when Robert Smith published his *Harmonics*, which included tables with the numbers of beats of the fifths in the mean-tone system and in his own system of equal harmony (different from equal temperament). More successful in the propagation of his tuning method, however, was Scheibler (1834, 1838), whose procedure was based on beats from tempered unisons. More specifically, Scheibler's method of tuning instruments consisted of tuning a series of forks on his tonometer 4 beats/sec flatter than the frequencies required. The string, pipe, or reed was then tuned sharper than the fork by 4 beats/sec.

Beats have even been used to determine the fundamental frequencies of two low harmonium reeds, which formed a known and slightly tempered musical interval and sounded simultaneously (Rayleigh, 1878-1879/1966). Frequencies were determined with very high precision by counting for 10 min those beats that resulted from interference of different pairs of higher harmonics.

Beats and Consonance

In models on tonal or sensory consonance (Kameoka & Kuriyagawa, 1969a; Plomp & Levelt, 1965), the total dissonance of musical intervals has been assumed to be equal to the sum of the dissonance of each pair of adjacent harmonics. According to Plomp and Levelt (1965), these subdissonances are roughly about maximum at 25% and minimum at 0% and 100% of the critical bandwidth. Results from Kameoka and Kuriyagawa (1969b) suggest that the subdissonances not only depend on (1) the frequency of the tones and (2) their frequency distance, but also on (3) their sound-pressure levels. The intervals at which the maximum dissonances were found to occur increased with increasing tone level. For example, for two tones at a level of 57 dB SPL, the maximum values of the subdissonances were reached at interval sizes of 300 to 40 cents for frequencies of the lower tone of between 100 and 7000 Hz. According to Kameoka and Kuriyagawa (1969b), these interval sizes range from 460 to 60 cents for two tones at a level of 80 dB SPL.

The essential power of these consonance models is that, from the relationship between beats or roughness and consonance, they can demonstrate that octaves and fifths are much more sensitive to a deviation from their simple frequency ratio than are the major thirds. In widely accepted equal temperament, which permits modulation among keys, the major thirds are tempered considerably (14 cents) in favor of the slightly tempered fifths (2 cents). The greater tolerance for tempered major thirds, as predicted from the consonance models, may explain why equal temperament has not been rejected.

The generality of these models has been demonstrated by Pierce (1966), who showed that consonance can be obtained with intervals comprising tones from an unconventional 8-note scale, having spectra with nonharmonic partials, provided that the distance between the partials remains larger than the critical bandwidth of the ear. Informal listening tests, reported by Slaymaker (1970), illustrate that chords from complex tones with stretched partials and with frequencies derived from correspondingly stretched scales are perceived as consonant, which is consistent with models on sensory consonance.

Beats of Complex Tones

For simple tones constituting a tempered consonant interval, attempts to explain the beats have yielded a vast number of studies (see Plomp, 1967, 1976, for reviews). For complex tones, the explanation of beats in terms of interference between adjacent harmonics with frequencies of pf_2 and qf_1 Hz was so plausible that, apparently, this case was reduced to beats of tempered unisons. It was recognized long ago, however, that beats of the first-order combination tones can be heard very well at the same time as those due to the higher harmonics of complex tones by an ear accustomed to hearing combination tones (Helmholtz, 1877/1954, chap. 11). Especially for very slightly tempered intervals, it has been noted that it is easy to make a mistake in counting the beats, since, in addition to the first pair of just-noncoinciding harmonics, the higher pairs beat as well (Helmholtz, 1877/1954, chap. 8). However, from more formal experiments, with two simultaneous complex tones that deviated slightly from unison ($p=1$; $q=1$), in which moderately well-trained listeners had to count the resulting beats, it was concluded that the beat frequency of the first pair of adjacent harmonics pf_2-qf_1 Hz, was perceptually dominant (Warren, 1978). This was also the case when the fundamental components were absent. The dominantly perceived beat frequency equaled the fundamental beat frequency of the complex beats even when, in contrast to normal situations, the beat frequency of higher pairs of interfering harmonics was lower than that of the first pair of adjacent harmonics (Warren, personal communication, 1981).

Beats and Discrimination

In a previous paper, thresholds for discrimination between pure and tempered fifths and major thirds were presented in terms of level difference between the tones (Vos, 1982). Basically, level differences were introduced to manipulate the depth of the level

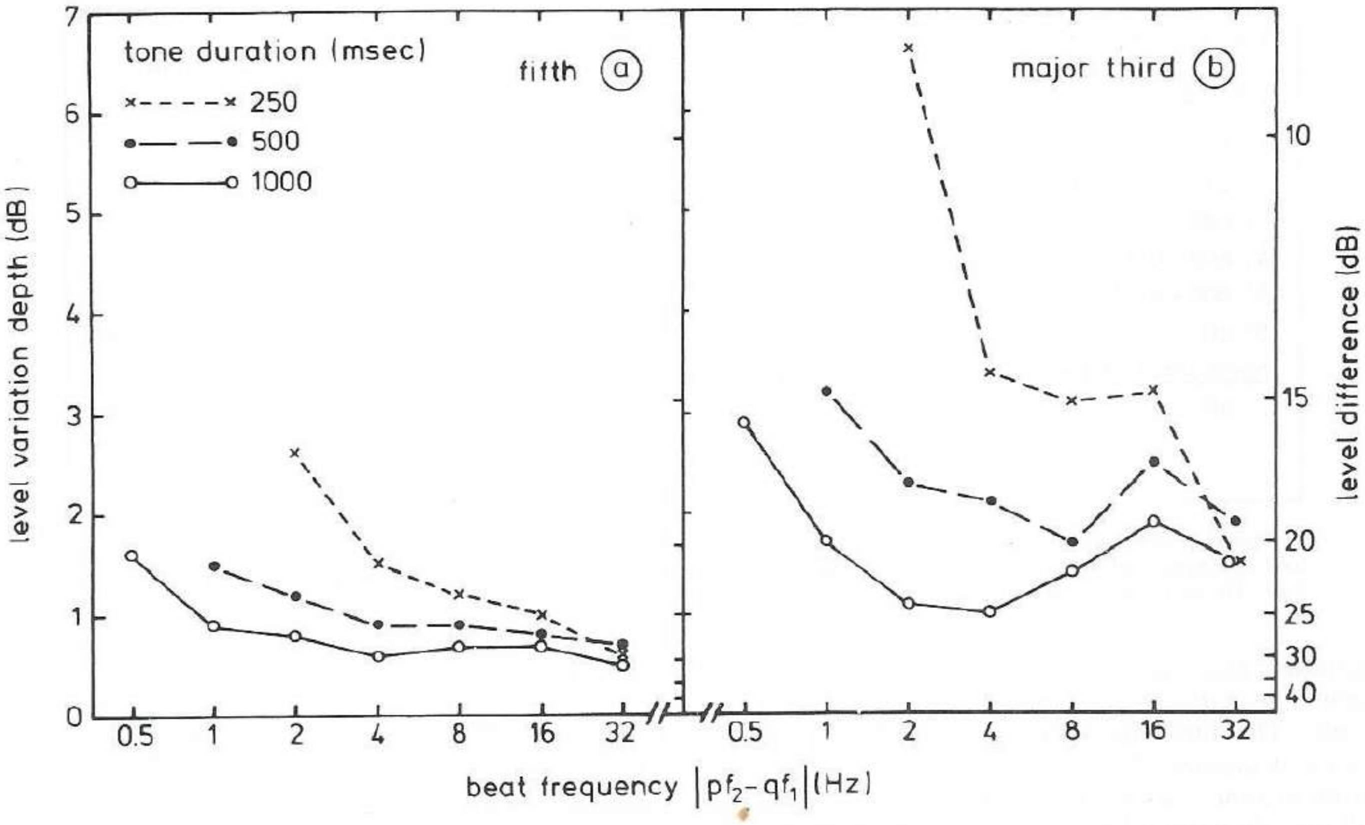

Figure 1. Replotted thresholds for discrimination between pure and tempered intervals (Vos, 1982), given as a function of beat frequency with tone duration as parameter. Thresholds are given separately for the fifth and major third in panels a and b, respectively.

variation of the beating harmonics with frequencies of npf_2 and nqf_1 Hz. In Figure 1, these discrimination thresholds (DTs) are replotted in terms of level variation depth and given as a function of beat frequency, $f_b = | pf_2 - qf_1 |$, with tone duration as a parameter. It was the purpose of the research reported here to answer three questions: (1) Are the DTs determined mainly by the interference of the first pair of just-noncoinciding harmonics with frequencies of pf_2 and qf_1 Hz? (2) Does the beat frequency, $| pf_2 - qf_1 |$, representing at least an objective measure of the amount of tempering also match the *perceived* beat frequency? And (3) are differences between discriminability at threshold representative of differences between the perceived strength of beats in supraliminal conditions? If these questions can be answered positively, the conclusion may be drawn that the DTs in Figure 1, with respect to both the abscissa and the ordinate, are also adequately represented from a perceptual point of view.

Experiments 1, 2, and 3 were designed to answer the first question. In these experiments, the relative importance of various harmonics for discrimination between pure and tempered fifths (Experiment 1) and pure and tempered major thirds (Experiments 2 and 3) was determined by investigating the effect of spectral content on the DTs. In Experiment 4, attempts were made to answer the second and third questions. In supraliminal conditions, both the most dominantly perceived beat frequency and the perceived strength of the beats were determined for various spectral conditions.

EXPERIMENT 1

In a tempered consonant interval, in which the ratio of the fundamental frequency f_1 of tone 1 and the fundamental frequency f_2 of tone 2 departs slightly from p:q, the n^{th} pair of just-noncoinciding harmonics comprises the nq^{th} harmonic of tone 1 and the np^{th} harmonic of tone 2 ($n = 1, 2, \ldots$). Therefore, for the fifth ($p = 2$; $q = 3$) investigated here, the first pair of just-noncoinciding harmonics comprises the third harmonic of tone 1 and the second harmonic of tone 2. However, for the benefit of a unifying terminology, which will be especially useful in Experiment 4, where both fifths and major thirds ($p = 4$; $q = 5$) were investigated, it will be maintained throughout this paper that the first pair of just-noncoinciding harmonics comprises the q^{th} and p^{th} harmonics of tones 1 and 2, respectively.

The various spectral conditions presented in Experiment 1 are depicted in Figure 2. In this figure, the first 12 harmonics of tone 1, represented by

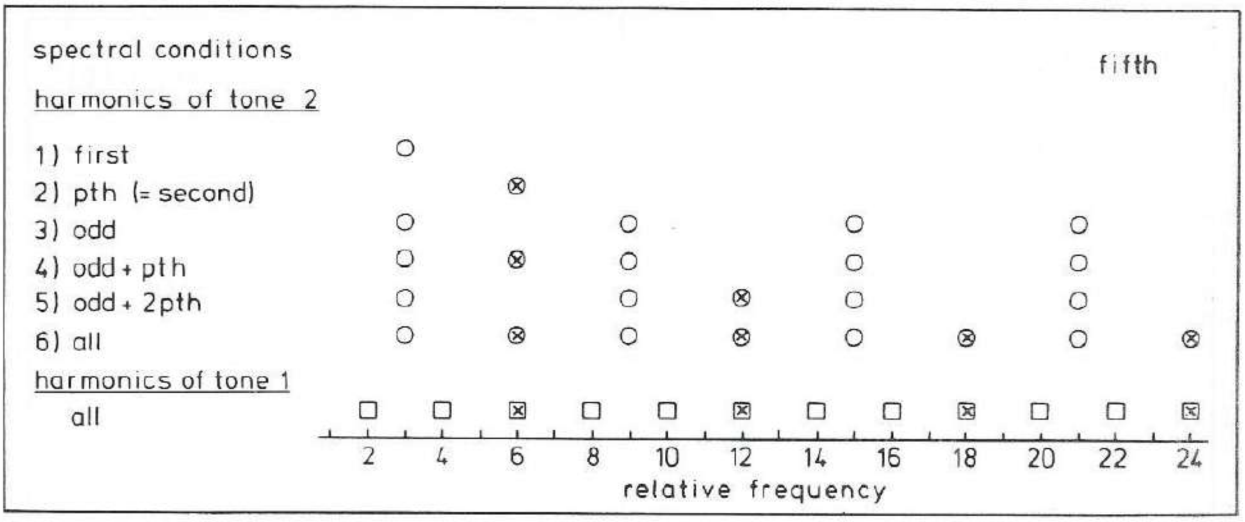

Figure 2. Schematic of the spectral conditions which were investigated in Experiment 1. The first 12 harmonics of tone 1, represented by squares, and various harmonics of tone 2, represented by circles, are given on a linear frequency scale. The just-noncoinciding harmonics are indicated by crosses.

squares, and various harmonics of tone 2, represented by circles, are given on a linear frequency scale. The just-noncoinciding harmonics are indicated by crosses. For five of six conditions, the spectrum of tone 2 always included the first harmonic or all the odd harmonics. As a result, it was guaranteed that in these conditions the perceived interval size equaled that of the fifth. In the third spectral condition, which comprised only the odd harmonics, interference between just-noncoinciding harmonics was not possible because the npth harmonics of tone 2 were deleted. In Conditions 4 and 5, the pth or 2pth harmonics were selectively added to the odd harmonics. The sixth condition, comprising all harmonics, served as a base rate. Comparison of the DTs from Spectral Conditions 2 to 6 should reveal the relative importance of the first, second, and higher pairs of adjacent harmonics for discrimination. To ascertain to what extent the combination of various triads of harmonics, such as harmonics with their relative frequencies in the ratios of 2:3:4, 8:9:10, and 14:15:16, may have contributed to DT, it was decided to include Spectral Condition 1 as well.

Method

Stimuli. For all fifths, tone 1 was a complex tone with amplitudes a_n equal to $1/n$. Therefore, the spectral envelope slope of the stimulus was -6 dB/octave. The waveform is given by:

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

The phase (ϕ_n) of the individual harmonics was chosen randomly. The spectral content of tone 2 was varied. Again, the amplitudes, a_n, of the particular harmonics were equal to $1/n$ and the phases (ϕ_n) of the various harmonics were random.

The overall level of tones 1 and 2 together was 85 dB SPL. This level was measured with the help of an artificial ear (Brüel & Kjaer, Type 4152). The duration of the tone bursts was 500 msec. Rise and decay times of the tones, defined as the time interval between 10% and 90% of the maximum amplitude, were 40 and 20 msec, respectively. Level variation depth of the beating harmonics was manipulated by attenuation of tone 2 by ΔL dB. When the spectral envelopes of the two tones coincide, that is, when the respective amplitudes a_1 and a_2 of the just-noncoinciding harmonics, with frequencies of qf_1 and pf_2 Hz, are equal, then ΔL is defined to be 0 dB. Level variation depth, D, in decibels, being maximal in this condition, is given by:

$$D = 20 \log \left(\frac{a_1 + a_2}{a_1 - a_2} \right) = 20 \log \left(\frac{1 + 10^{\Delta L/20}}{1 - 10^{\Delta L/20}} \right), \qquad (2)$$

in which ΔL is the level difference between a_1 and a_2 in decibels [$\Delta L = 20 \log(a_2/a_1)$]. D approaches ∞ for $a_2 \rightarrow a_1$.

Coincidence of spectral envelopes occurs when the overall level of tone 2 is $20 \log(q/p)$ dB lower than the overall level of tone 1. Increments of ΔL from this base condition always result in smaller level variation depths.

Apparatus. The experiment was run under the control of a PDP-11/10 computer. Tones 1 and 2 were generated in the following way. One period of the waveforms of tones 1 and 2 was stored in 256 discrete samples (with 10-bit accuracy) in external revolving memories. These recirculators could be read out by digital-to-analog converters. Sampling rates were determined by pulse trains, derived from two frequency generators. After gating, the tones were filtered by Krohn-Hite filters (Model 3341) with the function switch in the low-pass RC mode and with a cutoff frequency of 8 kHz. The sound-pressure levels of the tones were controlled by programmable attenuators. After appropriate attenuation, the tones were mixed and fed to two headphone amplifiers. The signals were presented diotically (same signal to both ears) by means of Beyer DT 48S headphones. The subjects were seated in a soundproof room.

Subjects. Four musically trained subjects were tested over seven sessions, each session on a different day. Three of them were conservatory students. They were paid for their services. The author also participated in the experiment.

Experimental design. The independent variables were: (1) spectral content of tone 2 (the spectrum of tone 2 could comprise the following harmonics: first, pth, odd, odd + pth, odd + 2pth, and all); (2) frequency difference $| pf_2 - qf_1 |$ (2, 4, 8, 16, and 32 Hz); and (3) direction of tempering (stretched vs. compressed interval).

Scoring of data. The task for the subject was to discriminate between pure and tempered musical intervals in a 2-AFC paradigm.

By means of the method of constant stimuli, all stimulus combinations were presented 20 times for seven different values of ΔL in equal steps of 3, 4, or 5 dB, depending on stimulus condition. For every stimulus combination, the raw scores for different ΔL values were transformed to percentage correct. Since the response patterns of the stretched and compressed intervals were similar in almost all cases, the percentages of these two conditions were combined. On the assumption that the scores, for a given set of ΔL values, are cumulative normal distributions (Kling & Riggs, 1972), these percentages were converted to z-scores. The discrimination threshold, that is, the value of ΔL at which a subject responded correctly in 75% of the cases, was obtained by solving the linear regression function $z = a + b\Delta L$ for ΔL when z equals .67. Computation of coefficients of determination (r^2) for all linear regression functions showed that the explained variability in z by ΔL and the linear rule was more than .75 in 86% of the cases. Moreover, r^2 did not depend on either the kind of spectrum or beat frequency. Mean r^2 was .87.

Procedure. Every trial consisted of a comparison of a pure interval and a tempered interval. The tempered interval was presented first in half of the trials and second in the other half of the trials. Between the intervals there was a silent period of 1 sec. By means of two response buttons, the subject indicated whether the first or the second interval was tempered. Feedback was given by a red light above the correct button. The next trial was presented 1.5 sec after the response to the preceding trial. Within a series, three different spectra of tone 2 were presented, together with seven values of ΔL, two directions of tempering, and five beat frequencies ($3 \times 7 \times 2 \times 5 = 210$ trials in a series). There were four series types, each with a different combination of three spectra within a series. Thus, the six experimental spectra were assigned to four kinds of series in different combinations. Two subjects were presented first with series types 1 and 2; the other two subjects started the experiment with series types 3 and 4. Within a series, which lasted about 15 min, the stimuli were presented in a random order. At the beginning of a series, new waveforms of tones 1 and 2, with different ϕ_n values, were determined. The fundamental frequencies were varied from trial to trial in such a way that the central frequency f_c [= the geometric mean $(f_1f_2)^{1/2}$] of the pure interval was equal to 370 Hz ± 50, ± 100, ± 150, or ± 200 cents, in a random order. As a result, all fundamental frequencies fell within a range of 261.5 Hz (musical C4) and 523 Hz (musical C5). The f_1 of the pure interval was equal to:

$$f_1 = f_c/(q/p)^{1/2}. \tag{3}$$

In the tempered interval, one of the fundamental frequencies equaled the corresponding fundamental frequency in the pure interval and the other one was altered. For the tempered interval, either a new f_1 was calculated by means of:

$$f_1(\text{tempered}) = (p \cdot f_2 \pm f_b)/q, \tag{4}$$

or a new f_2 was calculated by:

$$f_2(\text{tempered}) = (q \cdot f_1 \pm f_b)/p. \tag{5}$$

Within a series, the frequencies of tones 1 and 2 were altered equally often, the particular order being determined by sampling without replacement. Similarly, the particular order of stretched versus compressed intervals (signs in Equations 4 and 5) was randomized. A session comprised six or seven series. During the first session, the subjects were trained. For most conditions, a rather broad range of ΔL from 7 to 31 dB, in steps of 4 dB, adequately covered the subjects' transition range between chance performance and 100% correct responses. For the spectra comprising the first or the odd harmonics, however, ΔL ranged from 4 to 22 dB, whereas for a beat frequency of 32 Hz, the range of ΔL was from 7 to 37 dB. In the ΔL < 7 dB conditions, L_1 and L_2 were attenuated to a small extent to satisfy our criterion that the overall level of the interval remain at 85 dB SPL.

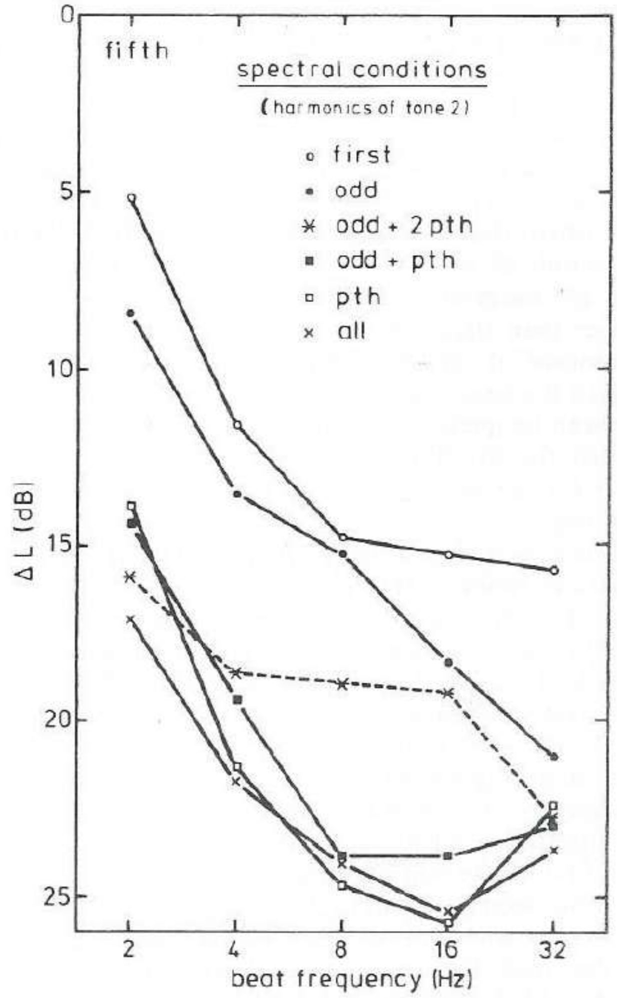

Figure 3. Thresholds for discrimination between pure and tempered fifths, plotted as a function of beat frequency, for various spectral conditions. P^th and 2p^th refer to the second and fourth harmonic, respectively. The spectrum of tone 1 always comprised the first 20 harmonics. For both tone 1 and tone 2, waveforms were determined using Equation 1.

Results

Discrimination thresholds (DTs) are presented in terms of level difference ΔL. DTs were subjected to an ANOVA [4 (subjects) × 6 (spectra of tone 2) × 5 (beat frequencies), all repeated measures]. Although individual differences occurred [$F(3,60) = 83.7$, $p < .000001$], the experimental variables affected our subjects' DTs in similar ways. DTs were lowest for those spectra in which the second harmonic of tone 2 was present [$F(5,15) = 28.2$, $p < .00001$]. In general, DTs decreased with increasing beat frequency [$F(4,12) = 14.5$, $p < .005$], especially from 2 to 8 Hz. This can be seen in Figure 3, where DTs from the various spectral conditions are plotted as a function of beat frequency. For beat frequencies higher than 8 Hz, a further threshold decrement can be seen only for the odd and the odd + $2p^{th}$ harmonics, whereas for the other conditions, the DTs remain the same or even tend to increase again for the beat frequency of 32 Hz [$F(20,60) = 3.45$, $p < .0005$].

Discussion

It was the purpose of Experiment 1 to test the hypothesis that for intervals with complex tones DTs are determined *mainly* by the sensitivity to those beats, which result from the interference of the first pair of just-noncoinciding harmonics with frequencies of pf_2 and qf_1 Hz. The results demonstrate that, for fifths, this is true: in conditions in which the p^{th} harmonic of tone 2 is present and can interfere with the q^{th} harmonic of tone 1, DTs are considerably lower than they are in conditions in which the p^{th} harmonic of tone 2 is absent. For the conditions in which the beat frequency is 32 Hz, corresponding to a mean tempering of about 60 cents, differences between the thresholds are strongly diminished. The DTs for the odd + $2p^{th}$ condition indicate that interference of the second pair of just-noncoinciding harmonics, with frequencies of $2pf_2$ and $2qf_1$ Hz, can also contribute to DT, since these DTs are also lower than the DTs for the spectra that comprise the first, or the odd, harmonics, the effect being more pronounced for the lower beat frequencies. Moreover, for beat frequencies of 2 and 4 Hz, DTs for spectra with odd + p^{th} and odd + $2p^{th}$ harmonics are about the same, suggesting that, within certain limits, beat frequency can compensate for a lower absolute level of interfering harmonics. DTs for spectra with only the first or the odd harmonics show that discrimination between pure and slightly tempered fifths is possible even *without* just-noncoinciding harmonics.

For beat frequencies higher than 8 Hz, further threshold decrements are found for two of the six spectral conditions. These decrements cannot be explained by an increased sensitivity to beats, since it has been shown that beat thresholds for complex tones in comparable $L_1 > L_2$ conditions (Vos, 1982) remain at about the same level for beat frequencies higher than 4 Hz. From experimental data on beats resulting from two interfering sinusoids or from sinewave amplitude modulation of one sinusoid, even a threshold *increase* would be predicted for these beat frequencies (see Vos, 1982, for a review).

It is likely that the perception of interval size is responsible for the further decrement of the DT. For a performance criterion equal to that for the DTs, and for stimulus conditions in which ΔL is equal to 20 or 25 dB, mean thresholds at which the *direction* of tempering was consistently identified were found to be equal to 32 and 47 cents, respectively (Vos, 1982). These temperings correspond to beat frequencies of about 15 and 23 Hz, respectively. Considering these results, it is plausible to explain the decrement of the DTs for beat frequencies higher than or equal to 16 Hz as the result of the contribution of the perception of differences in interval size.

One could object that for the spectral conditions in which the p^{th} harmonic of tone 2 is present, perception of interval size should have caused a further threshold decrement. Especially for $L_1 > L_2$ conditions, however, the higher tone is masked when level differences are greater than about 25 dB.

The condition in which only the first harmonic of the higher tone is present indicates that at least for small amounts of tempering, DTs can be determined by the sensitivity to beats, due to the combination of harmonics that are more than a critical band apart. It has already been observed by ter Kuile (1902) that beats can be heard for chords consisting of three tones with frequencies in the ratio of 2:(3-g):4, g representing a small amount by which the second tone is tempered. These beats are probably due to waveform variations (see Plomp, 1967). We cannot conclude from our experiment that these beats result from the combination of either two harmonics (with frequencies of f_1 and f_2 or f_2 and $2f_1$ Hz) or three harmonics (with frequencies of f_1, f_2, and $2f_1$ Hz).

Equal sensitivity to tempering for conditions that comprised only the p^{th} harmonic or both the odd and p^{th} harmonics suggests that when interference between just-noncoinciding harmonics takes place, added interference due to the combination of other harmonics does not really increase sensitivity to tempering.

EXPERIMENT 2

For the major third investigated here, the five spectral conditions presented are depicted in Figure 4. As in Figure 2, the first 12 harmonics of tone 1 and various harmonics of tone 2 are given on a linear frequency scale. The spectrum of tone 2 included the first harmonic, which ensured for four of the five conditions that the intervals would be perceived as major thirds. In the second spectral condition, comprising only the odd harmonics, interference between just-noncoinciding harmonics was not possible. Comparison of the DTs from the first and third conditions with those from the complete spectral condition should reveal the relative importance of the interference of the first pair of just-noncoinciding harmonics. Condition 4 was included to test whether, in addition to interference of the first pair of adjacent harmonics, sensitivity to tempering could be increased even further by simultaneous interference due to the presence of other even harmonics. Moreover, comparison of the DTs from Conditions 4 and 5 should reveal whether for a more complex spectral condition in which interference between various pairs of just-noncoinciding harmonics takes place, added interference due to the presence of odd harmonics would be able to increase sensitivity to tempering.

Method

Stimuli. The intervals were musical major thirds ($p=4$; $q=5$). Apart from this, the stimuli were the same as those in Experiment 1.

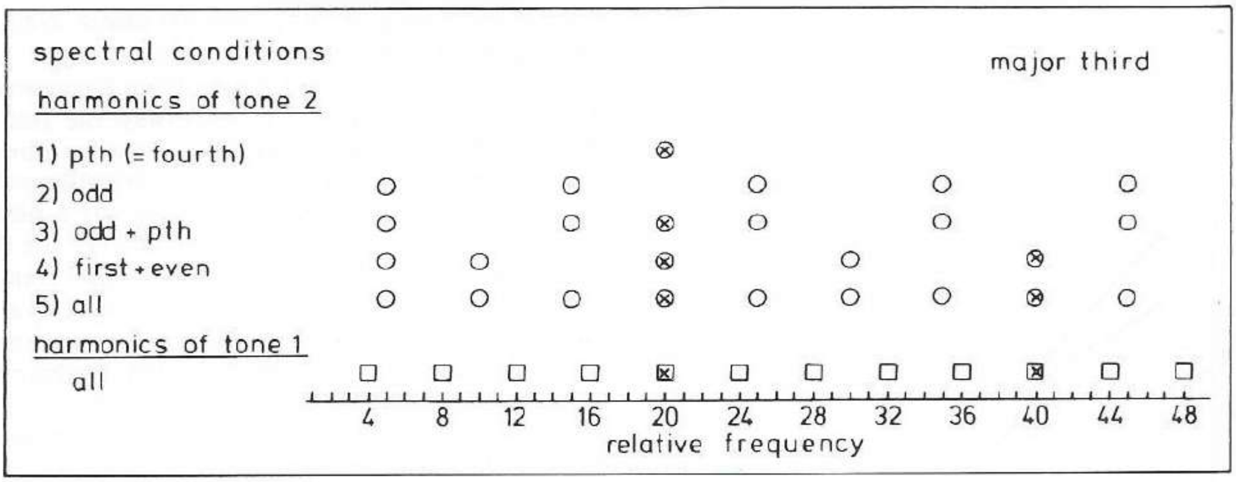

Figure 4. Schematic of the spectral conditions investigated in Experiment 2. The first 12 harmonics of tone 1, represented by squares, and various harmonics of tone 2, represented by circles, are given on a linear frequency scale. The just-noncoinciding harmonics are indicated by crosses.

Apparatus. The apparatus was identical to that of Experiment 1.

Subjects. Five musically trained subjects were tested over three sessions, each session on a different day. The subjects were paid for their services. None of them had participated in Experiment 1.

Experimental design. The independent variables were: (1) spectral content of tone 2 (the spectrum of tone 2 could comprise the following harmonics: p^{th}, odd, odd + p^{th}, first + even, and all); and (2) frequency difference $| pf_2 - qf_1 |$ (2, 4, 8, 16, and 32 Hz). For each of the 25 different conditions, all of which were presented to five subjects, two DTs were determined.

Procedure. As in Experiment 1, the subjects had to discriminate between pure and tempered major thirds in a 2-AFC paradigm. Instead of the method of constant stimuli, however, DTs were determined by an adaptive procedure in which ΔL was increased by 2 dB after three correct responses and decreased by the same amount after an incorrect response. Each reversal of the direction of change in ΔL over successive trials was counted as a turnaround. The effective DT was taken as the mean value of ΔL on the trials that resulted in the last four turnarounds. This value estimates the value of ΔL required to produce 79.4% correct responses in a nonadaptive 2-AFC procedure (Levitt, 1971). The initial value, that is, the value of ΔL on the first "experimental" turnaround was determined as follows. In the first run, ΔL, being 0 dB at the very first trial, was increased by 3 dB after each correct response. After the first incorrect response, ΔL was decreased after each incorrect response and increased after two correct responses, in steps of 2 dB. The initial value equaled ΔL on the trial that resulted in the fourth turnaround. Thus, each DT was based on a block of eight runs (series of trials); the first four runs resulted in an adequate initial value, and the values of ΔL on the peaks and valleys of the last four runs provided an estimate of the DT.

A series consisted of five blocks, one for each frequency difference $| pf_2 - qf_1 |$. Presentation order of the different blocks was randomized. Presentation order of spectrum, which was held constant within a series, was balanced according to two (5×5) Latin squares, one for each replication. At the beginning of a series, new waveforms of tones 1 and 2, with different ϕ_n values, were determined. Fundamental frequencies of the pure and tempered intervals were computed in the same way as in Experiment 1. The central frequency, f_c, of the pure interval, however, was equal to 370 Hz ± 25, ± 75, ± 125, ± 175, or ± 225 cents, in a random order. The direction of tempering (signs in Equations 4 and 5) was randomized. On trials in which ΔL was smaller than 10 dB, L_1 and L_2 were attenuated to a small extent to keep the overall level of the intervals at about 85 dB SPL. Five series (25 blocks of trials) were completed during each session. The subjects were trained in the first session.

Results

In none of the conditions were there any systematic differences between the two replications. Therefore, DTs were subjected to an ANOVA [5 (subjects) $\times$ 5 (spectra of tone 2) $\times$ 5 (frequency difference), all repeated measures], with replication as a within-cell variable. Two subjects had significantly lower thresholds than the other three subjects [$F(4,125) = 9.17$, $p < .00005$].

DTs were lowest for those spectra in which at least the even harmonics were present and highest for the spectrum in which only odd harmonics were present [$F(4,16) = 58.9$, $p < .000001$]. A Newman-Keuls paired comparison test showed that except for the two conditions which rendered the lowest thresholds (largest values of ΔL), the DTs for the other spectra were significantly different at the .05 level or better. In general, DTs decreased with increasing beat frequency [$F(4,16) = 9.65$, $p < .0005$]. However, this effect was most pronounced for those conditions in which tone 2 contained only the odd, odd and p^{th}, or p^{th} harmonics [$F(16,64) = 2.44$, $p < .006$]. This is shown in Figure 5, where DTs for the various spectra are plotted as a function of beat frequency.

Discussion

Relative to the condition in which the spectrum of tone 2 consists of only the odd harmonics, sensitivity to tempering is higher when the p^{th} harmonic of tone 2 is present or added to the odd harmonics. However, our data suggest that since the DTs from conditions in which only the p^{th} harmonic or the odd and p^{th} harmonics are present are not as low as the

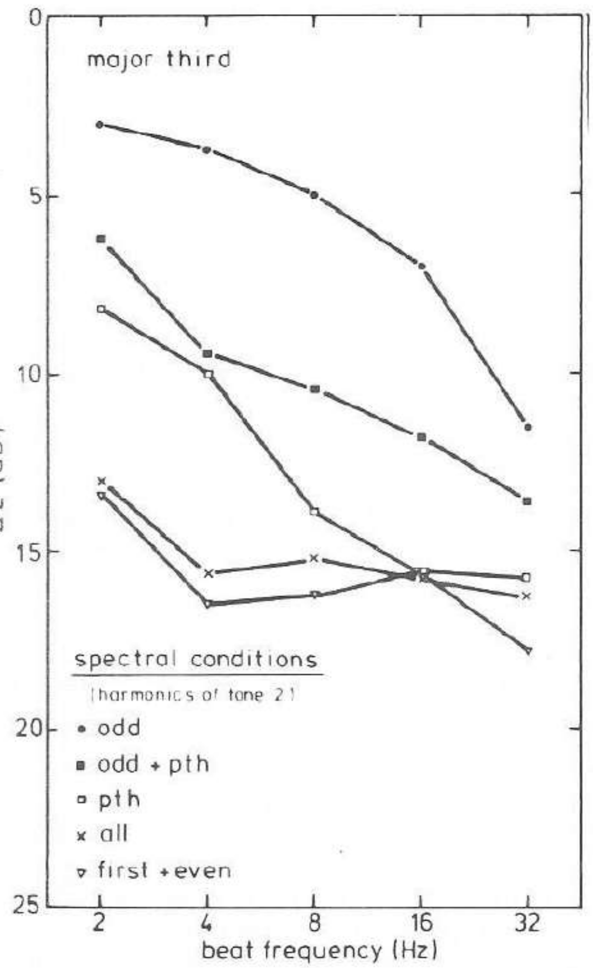

Figure 5. Thresholds for discrimination between pure and tempered major thirds, plotted as a function of beat frequency, for various spectral conditions. Pth refers to the fourth harmonic. The spectrum of tone 1 always comprised the first 20 harmonics. For both tone 1 and tone 2, waveforms were determined using Equation 1.

conditions in which all the harmonics are present, interference between other harmonics must have played a role as well. Given that the conditions with the first and even harmonics and those with all harmonics rendered similar DTs, we must conclude that in addition to the pth harmonic, other even harmonics also contribute significantly to discrimination, especially for the lower beat frequencies. One of these even harmonics responsible for the low DTs might be the 2pth harmonic. Results from Experiment 4, however, which were obtained before those of Experiment 3, suggested that the second and sixth harmonics of the higher tone may also have influenced the DTs. Most likely, the contribution of the second harmonic of the higher tone to the DT is important and is a result of the interference between three harmonics with frequencies of $2f_1$, $2f_2$, and $3f_1$ Hz, respectively. This is supported by observations by ter Kuile (1902), who found that beats for chords consisting of three tones with frequencies in the ratio of 4:(5-g):6 could easily be heard, even when the tuning forks were struck weakly.

In accordance with the results from Experiment 1, the DTs for the spectrum with only the odd harmonics show that discrimination between pure and slightly tempered major thirds is possible *without* just-noncoinciding harmonics. However, a pilot experiment revealed that, in contrast to DTs for fifths, those for major thirds in which the higher tone contained only the first harmonic could, in general, not be determined. This finding agrees with experimental data for sinusoidal tones (Plomp, 1967), which show that beats for slightly tempered major thirds with a frequency of the lower tone higher than or equal to 250 Hz could not be heard by all subjects.

For the various spectral conditions in which the beat frequency was 32 Hz, which corresponds to a mean tempering of about 34 cents, differences between the DTs were diminished. The same mechanism put forward to explain further threshold decrements for the fifth might have contributed to the DTs for the major thirds. For a performance criterion equal to that for the DTs, and for stimulus conditions in which ΔL is equal to 0 or 20 dB, mean thresholds at which the *direction* of tempering was consistently identified were found to be equal to 28 and 30 cents, respectively (Vos, 1982).

EXPERIMENT 3

We have demonstrated that DTs for moderately tempered fifths are determined mainly by the sensitivity to beats that result from the interference of the first pair of just-noncoinciding harmonics. For the major thirds, however, the data suggest that interference of other even harmonics must have played a role as well. Although the second pair of just-noncoinciding harmonics can be considered to be the second best contributor, results from Experiment 4 show that when both the pth and 2pth harmonics are removed, the perceived strength of the beats for tempered major thirds remains relatively high. Therefore, Experiment 3 was designed to determine the relative contributions of various even harmonics, either on their own or in combination, to discrimination. To provide a baseline, all spectra of the higher tone included the odd harmonics. For reference, the spectral condition that comprised all harmonics was investigated as well.

Method

Stimuli. The stimuli were the same as those in Experiment 2. The duration of the tone bursts, however, was 1 sec.

Apparatus. The apparatus was identical to that of Experiment 1.

Subjects. Eight trained listeners were tested over three sessions, each session on a different day. Seven of them had not participated in any of the other experiments reported in this paper. The author also served as a subject.

Experimental design. The independent variables were: (1) spectral content of tone 2 [the spectrum of tone 2 could comprise the following harmonics: odd, odd and second, odd and p^{th}, odd and sixth, odd and $2p^{th}$, odd and second + p^{th}, odd and p^{th} + $2p^{th}$, and all]; and (2) frequency difference $| pf_2 - qf_1 |$ (2 and 4 Hz).

Procedure. In general, the procedure was similar to that of Experiment 2. The adaptive procedure by which DTs were determined was slightly different from that used in Experiment 2. Again, ΔL was increased by 2 dB after three correct responses and decreased by the same amount after an incorrect response. Each reversal of the direction of change in ΔL over successive trials was counted as a turnaround. There were eight turnarounds. In the first run, however, ΔL was increased by 3 dB after each correct response. The effective DT was taken as the mean value of ΔL on the trials that resulted in the last six turnarounds. A series comprised two blocks, one for each frequency difference presented. Presentation order of the different blocks was randomized. Presentation of spectrum, which was held constant within a series, was balanced according to an 8 × 8 Latin square.

Results and Discussion

For all spectral conditions, thresholds are plotted in Figure 6 as a function of beat frequency. As can be seen in this figure, sensitivity to tempering was highly dependent on the spectral content of the higher tone [$F(7,49) = 37.9$, $p < .00001$], and it increased with increasing beat frequency [$F(1,7) = 72.1$, $p < .0005$]. Thresholds from five of the eight spectral conditions were similarly influenced by beat frequency. As discussed below, the divergence of the DTs caused by beat frequency for the three conditions in which, in addition to the odd harmonics, the higher tone comprised the p^{th}, p^{th} and $2p^{th}$, or second and p^{th} harmonics is rather vital to our proposed model. Two Newman-Keuls paired comparison tests, one for each beat frequency, showed that many DTs were significantly different from each other. In fact, the size of the effects was such that both tests revealed that the DTs from the condition in which the higher tone comprised the odd and p^{th} harmonics were significantly different ($\alpha = .01$) from those in which this tone comprised the odd + sixth or all harmonics.

Apart from being more complicated than for the fifth, the results are straightforward and can be summarized as follows: (1) While the contribution of the sixth harmonic to the DT is negligibly small, that of the second harmonic is substantial; however, (2) both the p^{th} and the $2p^{th}$ harmonics contribute more to the DT than does the second harmonic. (3) For the beat frequency of 4 Hz, the DTs from the spectral conditions in which, in addition to the odd harmonics, either the p^{th} and $2p^{th}$ or the second and p^{th} harmonics have been included suggest that, for major thirds, the lowest DTs are determined by the combined effect of different interfering sources. This model of additivity of interference is consistent with the results of Experiment 2, which showed that only when all relevant even harmonics are included are the DTs as low as those for the condition that contained all harmonics.

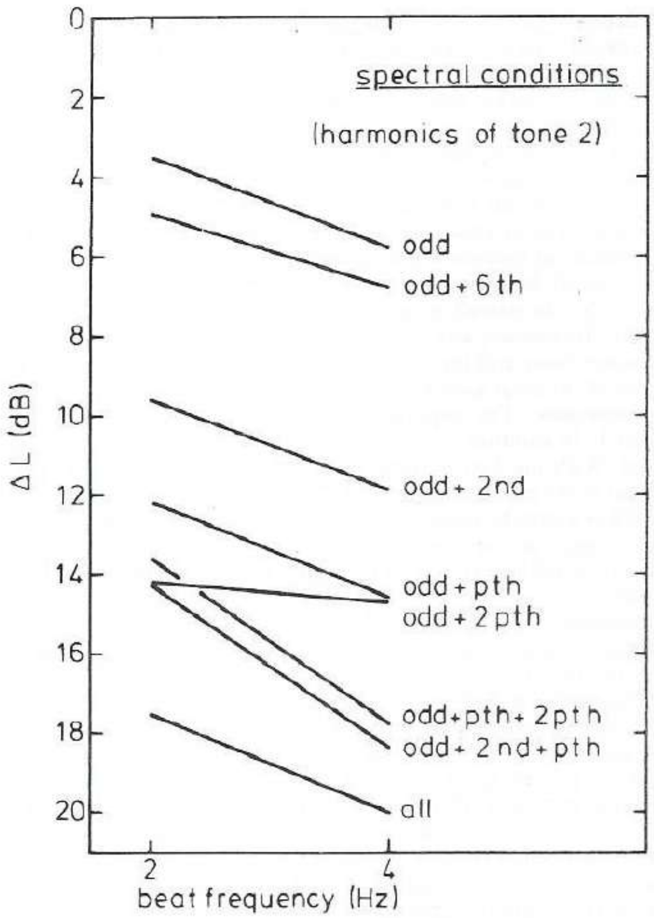

Figure 6. Thresholds for discrimination between pure and tempered major thirds, plotted as a function of beat frequency, for various spectral conditions. The spectrum of tone 1 always comprised the first 20 harmonics. For both tone 1 and tone 2, waveforms were determined using Equation 1.

EXPERIMENT 4

The purpose of this experiment was twofold: (1) to investigate whether for tempered intervals the beat frequency of the first pair of adjacent harmonics is equal to the *perceived* beat frequency, and (2) to determine to what extent differences in discriminability at threshold, as found in Experiments 1, 2, and 3, are representative of differences between the perceived strength of the beats in supraliminal conditions. Therefore, both the most dominantly perceived beat frequency and the perceived strength of these beats were determined for various spectral conditions and various amounts of tempering. With respect to the choice of the spectral contents, the presence or absence of the p^{th} and $2p^{th}$ harmonics was emphasized. Rather than count the beats and estimate the magnitudes of their strengths, it seemed preferable to imitate the rates and the strengths of the beats. This was achieved by adjusting the frequency and the level of a sinusoidal tone relative to the second sinusoidal tone with a fixed frequency of f_2 Hz and a fixed level.

Method

Stimuli. The intervals were major thirds ($p=4$; $q=5$) and fifths ($p=2$; $q=3$). Again, the low tone (tone 1) consisted of the first 20 harmonics, and the spectral content of the high tone (tone 2) was varied. The amplitudes, a_n, of the respective harmonics were equal to $1/n$. Sound-pressure levels of tones 1 and 2 were fixed in such a way that ΔL equaled 0 dB. The overall level of the two tones together was 85 dB SPL. In this experiment, presentation of the complex-tone intervals was alternated with presentation of two simultaneous sinusoidal tones as a comparison sound. To accomplish equal loudness for the complex-tone and sinusoidal-tone intervals, the overall level of the two sinusoids, which had different frequencies, was set at 89 dB SPL. The duration of both the complex tones and the sinusoidal tones was 1 sec. Rise and decay times of the tones were 30 and 20 msec, respectively.

Apparatus. The apparatus was identical to that of Experiment 1. In addition, two potentiometers with blinded knobs were used. With the first potentiometer, the level of one of the sinusoidal tones could be controlled. The frequency of the same tone could be varied by means of the second potentiometer. The voltage across each potentiometer was read out by an analog-to-digital converter and transformed into the appropriate frequency or level measure.

Subjects. Five subjects were tested over four sessions, each session on a different day. Three of the subjects had participated in Experiment 1.

Experimental design. The independent variables were: (1) interval (fifth and major third); (2) spectral content of tone 2 (the spectra could comprise the following harmonics: odd, odd and $2p^{th}$, odd and p^{th}, all, and all *minus* p^{th} and $2p^{th}$; and (3) frequency difference ($pf_2 - qf_1$) (2, 3, 4, 5, and 6 Hz). The stimulus combinations were presented four times to the five subjects. There were two dependent variables: (1) the adjusted beat frequency of the two interfering sinusoidal tones, and (2) the level variation depth of the two interfering sinusoids.

Procedure. The subjects were instructed to listen for beats in the intervals. If beats were perceived, they were to focus on two aspects of these beats: (1) the most dominant beat frequency, and (2) the subjectively experienced strength of the beats. The first part of their task was to imitate the rate of the beats in the musical intervals by adjusting the frequency of one of the sinusoidal tones. The other sinusoidal tone had a constant frequency of f_1 Hz (in a pilot investigation it had been found to be most natural to fix the tone at this frequency). The frequency of the variable tone was equal to f_1 Hz only at the beginning of a trial. During the adjustment, the complex-tone intervals and the sinusoidal-tone intervals were presented alternately in a long sequence. The time interval between the successive tone bursts was 700 msec. The second part of the task was to imitate the strength of the beats by adjusting the level of one of the sinusoidal tones. The level of the variable sinusoid with amplitude a_2 could only be decreased, relative to the level of the fixed tone with amplitude a_1 ($a_2 \leqslant a_1$). The level variation depth D, in decibels, of the beating sinusoids is given by Equation 2.

When the subject was satisfied that the beats of the two interfering sinusoidal tones reflected the beats of the two interfering complex tones in both respects, he pressed a ready-button. If the subject felt that perception of the beats of the complex-tone interval was hindered by the sinusoidal tones, he had the option of skipping the presentation of the sinusoidal tones. Presentation of the next trial was postponed until both knobs had been returned to their starting positions, that is, to the state that resulted in a frequency difference of 0 Hz and a level difference of 0 dB. The sensitivity of the potentiometers (ΔL or Δf per degree of rotation), however, was varied from trial to trial. Therefore, the maximum frequency difference and the maximum value of ΔL ranged from 20.5 to 32.0 Hz and from 25 to 36 dB, respectively. During the experiment, these extreme values were never used.

A series comprised 25 experimental trials (5 spectra × 5 beat frequencies) and 5 catch trials, in which pure intervals, one for each spectrum, were presented. Since all pure intervals were immediately recognized by our subjects, these data will be dropped from further analysis. Within a series, which lasted about ½ h, stimuli were presented in a random order. The musical interval was varied between series. The different intervals were presented alternately; three subjects started with the musical fifth, and two started with the major third. At the beginning of a series, new waveforms of tones 1 and 2, with different ϕ_n values, were determined. The fundamental frequency f_1 of the tempered musical interval was computed by means of Equation 3, with f_c always at 370 Hz. The fundamental frequency f_2 was computed using Equation 5. The tempered intervals were always stretched. During the first session, subjects received two short training series, one for each interval. If the level difference between the two sinusoids was below 10 dB, the levels of both tones were attenuated to a small extent in order to keep the overall level of the tones at 89 dB SPL.

Results

Beat frequency. For all conditions, the adjusted frequency differences between the two sinusoidal tones, representing the frequencies of those beats that were most dominantly perceived, are divided by their corresponding "theoretical" beat frequencies of ($pf_2 - qf_1$) Hz. The resulting fractions were subjected to an ANOVA [5 (subjects) × 2 (tempered fifth and major third) × 5 (spectra of tone 2) × 5 (frequency difference), all repeated measures]. Although individual differences occurred [$F(4,750) = 5.0$, $p < .001$], a number of experimental variables affected our subjects' adjusted beat frequencies in similar ways. The spectrum with only the odd harmonics rendered lower mean beat frequencies than did the other spectra [$F(4,16) = 6.0$, $p < .005$]. This effect, however, was more prominent for the major third than for the fifth [$F(4,16) = 3.6$, $p < .03$]. Mean adjusted beat frequencies were relatively higher in the conditions in which ($pf_2 - qf_1$) equals 2 Hz [$F(4,16) = 19.4$, $p < .00005$]. For both the fifth and the major third, this effect was most prominent for the two spectral conditions in which only the odd harmonics, or the odd and $2p^{th}$ harmonics were present [$F(16,64) = 2.6$, $p < .005$]. The interaction effect is depicted in Figure 7, where the adjusted beat frequency for the different spectra is plotted as a function of the amount of tempering expressed in terms of the theoretical beat frequency of ($pf_2 - qf_1$) Hz.

It is interesting to look at the distributions of the relative adjusted beat frequencies, because, depending on spectral condition, peaks are expected at .0, 1.0, and 2.0 times the beat frequency of ($pf_2 - qf_1$) Hz: (1) spectral conditions in which beats can be only faintly perceived should render a high proportion of trials in which no beats at all are heard, (2) spectral conditions in which the p^{th} harmonic is present might render fractions with values around 1 most of the time, and (3) spectral conditions in which, in addition to the odd harmonics, the $2p^{th}$ rather than the p^{th} harmonic is given, might render more fractions with values around 2 than the other conditions. For the fifth and the major third, and for all spectra sep-

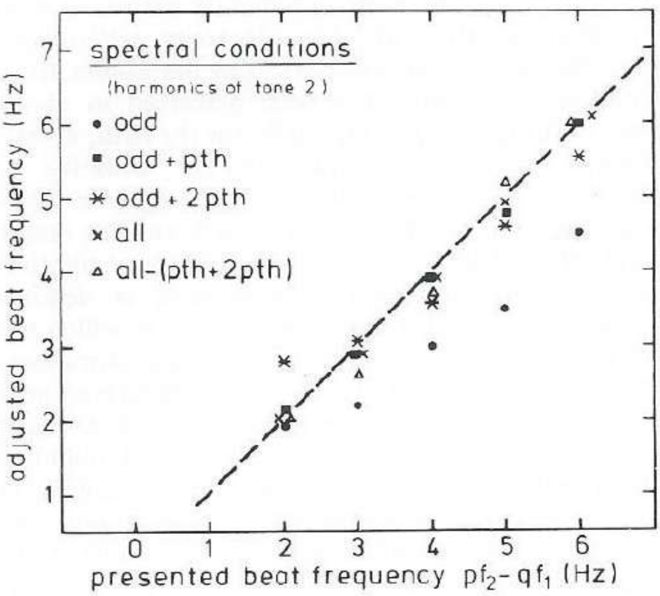

Figure 7. Adjusted beat frequency, plotted as a function of the amount of tempering of musical intervals with fundamental frequencies f_1 and f_2 (f_1:f_2 ~p:q). Data are given for the five spectra of the higher tone separately, and are averaged across the fifth (p = 2; q = 3) and the major third (p = 4; q = 5).

arately, grouped frequency distributions are given in Figure 8. For all histograms (the class-interval size is equal to .25), midpoints of the class intervals are indicated. The shapes of the distributions are clearly unimodal for spectral Conditions C and D, which contained the p^{th} harmonic. To some extent, the distributions are positively skewed for spectrum B, where the $2p^{th}$ harmonic may have determined the most dominantly perceived beat rate in some of the trials. The shapes of the distributions tend to be bimodal for the fifth in Conditions A and E. Spectrally, these conditions are similar up to the sixth harmonic of the higher tone. For the major third, relative beat frequencies are clearly bimodally distributed in Condition A and to a smaller extent in Conditions B and E. In Condition A, the number of times that no beats were perceived at all exceeded those of Condition E by more than 20%.

Level variation depth. The adjusted level variation depth, D, representing the perceived strength of the beats, was subjected to an ANOVA with a design similar to that used for the adjusted beat frequency. The effect of subjects, tested against within-cell variance, was significant [$F(4,750) = 49.8$, $p < .00001$]. Perceived strength of the beats for the fifths was 1 dB higher than for the major thirds [$F(1,4) = 12.2$, $p < .02$]. The spectral content of the higher tone affected the perceived strength of the beats considerably [$F(4,16) = 21.5$, $p < .00003$]. Mean perceived strength of the beats is presented in Figure 9, for each spectrum and for each musical interval separately. Perceived strength of the beats for fifths and major thirds was the same for only the spectrum in which the p^{th} and $2p^{th}$ harmonic had been removed [$F(4,16) = 3.9$, $p < .02$]. For four of the five spectra, perceived beat strength did not depend on beat frequency. For the spectrum that comprised the odd and p^{th} harmonics, there was a tendency to judge the beats as stronger for beat frequencies of 3 and 4 Hz [$F(16,64) = 1.8$, $p < .05$]. For tempered fifths, a Newman-Keuls paired comparison test showed that perceived strength of beats was significantly higher

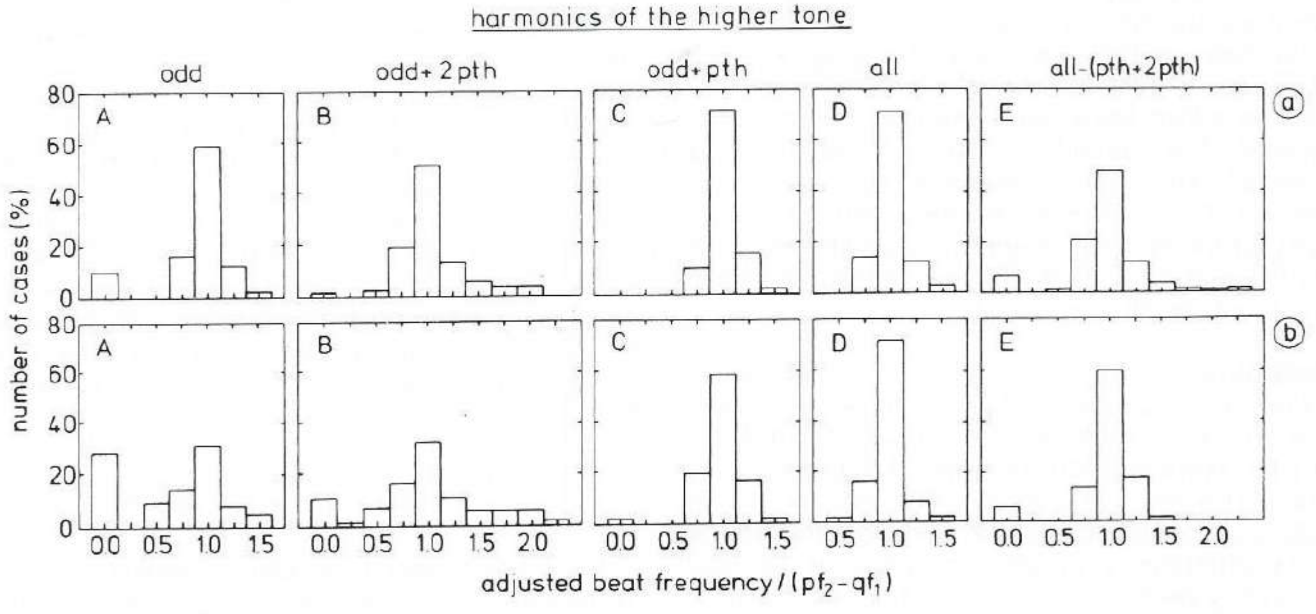

Figure 8. Grouped frequency distributions of the relative adjusted beat frequency. This frequency corresponds to that of the most dominantly perceived beats of tempered fifths (upper panels) and tempered major thirds (lower panels). Distributions are given for five different spectral conditions in panels A, B, C, D, and E, respectively. For each label, the corresponding spectral content of the higher tone is indicated. For the fifth and the major third, the values of p are 2 and 4, respectively.

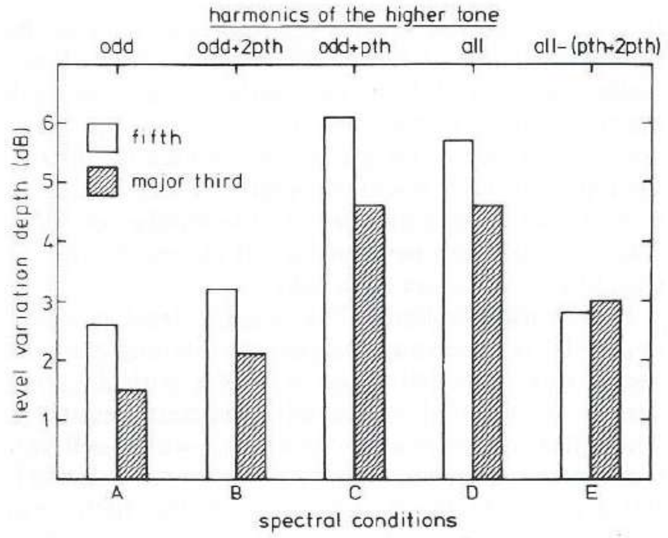

Figure 9. Adjusted level variation depth of two beating sinusoidal tones, which corresponds to the perceived strengths of the beats of tempered fifths (p = 2) and major thirds (p = 4). Mean values are given for five different spectral conditions, labeled A, B, C, D, and E, separately. Spectral conditions are specified at the top of the figure.

for the two spectra that contained the p^{th} harmonic than for the other spectra ($\alpha = .01$). For the major thirds, a similar test revealed that the perceived strength of beats was even higher for the spectrum from which the p^{th} and $2p^{th}$ harmonic had been removed than for the spectra that comprised the odd or the odd and $2p^{th}$ harmonics ($\alpha = .05$).

One could object that differences between the mean adjusted depth of the beating sinusoidal tones are due to the fact that the trials in which no beats are perceived at all are unevenly distributed over the different spectral conditions, as summarized in Figure 8. It has been verified that essentially the same pattern of results is obtained when the data for only those trials in which beats *were* perceived are taken into account. For example, if we combine all values obtained with relative adjusted beat frequencies between .75 and 1.25 Hz, the mean shifts in the patterns for the fifth and major third, as depicted in Figure 9, equal .20 and .26 dB, and standard deviations are .1 and .24 dB, respectively.

Discussion

One of the purposes of Experiment 4 was to ascertain to what extent the beat frequency of $(pf_2 - qf_1)$ Hz represents the *perceived* frequency of the beats. This was a relevant question because acoustically it can be shown (see Vos, 1982) that for complex tones interference of different pairs of just-noncoinciding harmonics results in a harmonic series of beat frequencies.

From the histograms given in Figure 8, it can be concluded that, for complete spectra of the tones (Condition D), the most dominantly perceived beat frequency is $(pf_2 - qf_1)$ Hz. Moreover, ***without*** interfering pairs of just-noncoinciding harmonics, as in Condition A, beats have been perceived in about 80% of the cases with, especially for the fifth, a well-defined frequency of $(pf_2 - qf_1)$ Hz. Relative to Condition D, however, the strength of these beats was much smaller. For both the fifth and the major third, the significance of the interference of the first pair of just-noncoinciding harmonics is demonstrated by the data from Condition C, in which the p^{th} harmonic had been added to the odd harmonics. The results show that (1) the relative perceived beat frequency is 1 Hz or close to 1 Hz, and (2) that the perceived strength of the beats equals that found in the complete spectral conditions. In Condition B, in which the higher tone contained the odd plus $2p^{th}$ harmonics, dominant beats with a frequency of $2(pf_2 - qf_1)$ Hz were observed in only a few cases. In this condition, the mean perceived strength of the beats is only slightly higher than in Condition A. When the p^{th} and $2p^{th}$ harmonics are removed (Condition E), the perceived beat strength for tempered fifths is equal to that in Condition A. This suggests that, apart from the second pair, interference of higher pairs of just-noncoinciding harmonics does not contribute to the perceived strength of the beats. Most likely, this also holds for the major third. It can be assumed that, relative to Conditions A and B, the significantly higher strength of the beats in Condition E is the result of the interference of the *second* harmonic of the higher tone with the second and third harmonics of the lower tone. For discrimination, the relevance of the same mechanism is demonstrated in Experiment 3.

In four of the five conditions, the perceived strength of beats is 1 dB higher for tempered fifths than for tempered major thirds. From Figure 1, where the plotted DTs are also expressed in terms of level variation depth, it can be seen that major thirds are less sensitive to tempering than are fifths. Moreover, these differences increase from .8 to 2.3 dB for tone durations decreasing from 1,000 to 250 msec. For a similar tone duration of 1 sec and for beat frequencies within a range of 1 to 8 Hz, this difference is .5 dB. These corresponding effects of musical interval once more suggest that DTs are mainly determined by sensitivity to beats.

GENERAL CONCLUSIONS

From Experiment 1, we may conclude (1) that for fifths consisting of two complex tones with spectral envelope slopes of -6 dB/octave sensitivity to tempering is determined mainly by the interference of the *first* pair of just-noncoinciding harmonics. The results from Experiment 4 indicate (2) that the dominantly perceived beat frequency of these tempered fifths is $(pf_2 - qf_1)$ Hz. These two facts lead us to conclude (3) that the thresholds for the discrimination between pure and tempered fifths, as given in Figure 1, are also adequately represented from a perceptual point of view.

The data from Experiment 4 show (4) that for the major thirds the dominantly perceived beat frequency is also equal to $(pf_2 - qf_1)$ Hz. From Experiments 2 and 3, we may conclude (5) that, for major thirds, discrimination is *partly* determined by the interference of the first pair of just-noncoinciding harmonics. The results from Experiment 3 suggest (6) that, for major thirds, DTs are determined by the combined effect of different interfering sources. Moreover, thresholds from Experiment 3 revealed (7) that, in addition to interference from just-noncoinciding harmonics, interference due to the combination of the second harmonic of the higher tone with harmonics of the lower tone plays a role as well. With respect to the ordinate, it is therefore (8) not justified to apply the third conclusion to this interval.

Restricting ourselves to corresponding conditions and the same range of beat frequencies, the spectral content of fifths had, qualitatively speaking, the same effect on sensitivity to tempering as on perceived strength of the beats. This also held for the major third in the two extreme conditions in which the higher tone comprised the odd or all harmonics, and in the condition in which both the p^{th} and the $2p^{th}$ harmonics were deleted. However, the large difference between the beat strengths found for the conditions in which either the $2p^{th}$ or the p^{th} harmonics were added to the odd harmonics was not predicted by the discrimination thresholds.

For all spectral conditions, the beats perceived as dominant usually had a frequency equal to that of the first pair of just-noncoinciding harmonics. Perceptual dominance of beats resulting from a specific source of interference is in agreement with the fact that beats are a reliable tool for tuning various musical instruments. This finding does not mean that, under certain conditions and when deliberately scrutinized, beats resulting from other sources of interference cannot be perceived.

The relatively important contribution of the second harmonic of the higher tone of the major third suggests that, for other musical intervals, harmonics different from just-noncoinciding harmonics may also play a role in discrimination and perceived strength of beats. This may be true for the second harmonic of the higher tone of the fourth ($p = 3$; $q = 4$) and the major sixth ($p = 3$; $q = 5$) and for the second, third, and fourth harmonics of the higher tone of the minor third ($p = 5$; $q = 6$) and the minor sixth ($p = 5$;$q = 8$). This hypothesis should be tested in future research. It is also interesting to determine to what extent our conclusions are independent of the spectral envelope slopes of the tones.

REFERENCES

BARNES, W. H. (1959). *The contemporary American organ* (pp. 95-101). Glen Rock, NJ: Fisher.

ELLERHORST, W. (1936). *Handbuch der Orgelkunde*. Einsiedeln, Switzerland: Benziger.

ELLISTON, T. (1916). *Organs and tuning*. London: Weekes.

HELMHOLTZ, H. L. F. VON. (1877). *Die Lehre von den Tonempfindungen als physiologische Grundlage fur die Theorie der Musik* (4th ed.). Braunschweig: Vieweg.[*On the sensations of tone as a physiological basis for the theory of music* (A. J. Ellis, Trans.). New York: Dover, 1954.]

KAMEOKOA, A., & KURIYAGAWA, M. (1969a). Consonance theory, Part II: Consonance of complex tones and its calculation method. *Journal of the Acoustical Society of America*, **45**, 1460-1469.

KAMEOKA, A., & KURIYAGAWA, M. (1969b). Consonance theory, Part I: Consonance of dyads. *Journal of the Acoustical Society of America*, **45**, 1451-1459.

KLING, J. W., & RIGGS, L. A. (1972). *Woodworth and Schlosberg's experimental psychology* (3rd ed.). New York: Holt, Rinehart and Winston.

KUILE, TH. E. TER (1902). Einfluss der Phasen auf die Klangfarbe. *Archiv für die gesammte Physiologie*, **89**, 332-426.

LEVITT, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, **49**, 467-477.

PIERCE, J. R. (1966). Attaining consonance in arbitrary scales. *Journal of the Acoustical Society of America*, **40**, 249.

PLOMP, R. (1967). Beats of mistuned consonances. *Journal of the Acoustical Society of America*, **42**, 462-474.

PLOMP, R. (1976). *Aspects of tone sensation*. London: Academic Press.

PLOMP, R., & LEVELT, W. J. M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, **38**, 548-560.

RAYLEIGH, LORD (1966). On the determination of absolute pitch by the common harmonium. In *Proceedings of the Musical Association* (Vol. 5; pp. 15-26). Vaduz, Liechtenstein: Kraus Reprint. (Original work published 1878-1879).

SCHEIBLER, H. (1834). *Der physikalische und musikalische Tonmesser*. Essen: Bädeker.

SCHEIBLER, H. (1838). Über mathematische Stimmung, Temperaturen und Orgelstimmung nach Vibrations-Differenzen oder Stössen. In *Schriften über musikalische und physikalische Tonmessungen und deren Anwendung auf Pianoforte- und Orgelstimmung*. Krefeld.

SLAYMAKER, F. H. (1970). Chords from tones having stretched partials. *Journal of the Acoustical Society of America*, **47**, 1569-1571.

SMITH, R. (1966). *Harmonics or the philosophy of musical sounds* (1st ed.). New York: Da Capo Press, 1966. (Original work published 1749)

VOS, J. (1982). The perception of pure and mistuned musical fifths and major thirds: Thresholds for discrimination, beats, and identification. *Perception & Psychophysics*, **32**, 297-313.

WARREN, R. M. (1978). Complex beats. *Journal of the Acoustical Society of America*, **64**, Suppl. 1, S38.

(Manuscript received June 20, 1983;
revision accepted for publication October 18, 1983.)

J. Acoust. Soc. Am. 77 (1), January 1985

4

Thresholds for discrimination between pure and tempered intervals: The relevance of nearly coinciding harmonics[a)]

Joos Vos and Ben G. van Vianen
Institute for Perception TNO, Kampweg 5, Soesterberg, The Netherlands

(Received 14 May 1984; accepted for publication 10 August 1984)

Thresholds for discrimination between pure and tempered musical intervals consisting of simultaneous complex tones (fundamental frequencies f_1 and f_2) were investigated. For these tones the main clue for the discrimination of pure intervals (f_1:$f_2 = p$:q; p and q small integers) from moderately tempered intervals (f_1:$f_2 \sim p$:q) is absence versus presence of beats. The strength of the beats (level difference between envelope maximum and minimum or level-variation depth D) was manipulated by introduction of differences in level (ΔL) between the two tones. In each of three experiments the discrimination thresholds (DTs) were determined for 13 intervals with different values for p and/or q. Experiment 1 showed that there is a simple relation between frequency-ratio complexity and discriminability: DTs gradually increased (smaller values of ΔL) with increasing $p + q$. Experiment 2, in which tones with harmonics of equal amplitude were used, indicated that level of the interfering harmonics was not responsible for the relation between DT and $p + q$. Yet, Experiment 3, in which the spectral content of the tones was varied, clearly showed that for all intervals DT had been determined by the interference between nearly coinciding harmonics. Detailed analysis of the results revealed that the relation between DT and ratio complexity might have been the result of masking.

PACS numbers: 43.66.Fe, 43.66.Lj, 43.75.Cd

INTRODUCTION

Presentation of two tones results in the perception of a musical interval. The character of such a musical interval is determined by the ratio between the fundamental frequencies of the tones. Traditionally, our Western tonal system is based on a set of musical intervals that are characterized by small-integer frequency ratios, e.g., 1:2 for the octave, 2:3 for the fifth, 3:4 for the fourth, 4:5 for the major third, etc. These intervals are associated with the concept of consonance. Throughout this paper, they will be denoted as pure consonant intervals.

It can easily be demonstrated that, for our 12-tone scale, pure fifths and pure major thirds are incompatible within one tuning system (see, e.g., Rasch, 1983; Vos, 1982). In tuning fixed-pitch keyboard instruments such as the organ, the harpsichord, and the piano, it is therefore necessary to temper at least part of the intervals. In tempered consonant intervals the frequency ratio is slightly different from that of the corresponding pure consonant intervals.

In a previous study (Vos, 1982), thresholds for the discrimination between pure and tempered fifths and major thirds have been determined. The results showed that the discrimination thresholds (DTs) were lower (higher sensitivity to tempering) for 2:3 than for 4:5 and it was suggested that this difference might be partly explained by differences in the absolute levels of the interfering harmonics.

It was the purpose of the present study to find and comprehend a general rule to describe a possible relation between discriminability and type of interval. Therefore, DTs were determined for a large number of intervals varying in size between and including unison and twelfth (see Table I).

Before we give a detailed description of the experimental method, we shall elaborate the physical and perceptual aspects that are relevant to our experimental study.

I. PHYSICAL AND PERCEPTUAL ASPECTS

A. Interference of nearly coinciding harmonics

1. Amplitude variation

In a pure consonant interval the ratio of the fundamental frequency f_1 of the lower tone (tone 1) and the fundamental frequency f_2 of the higher tone (tone 2), is equal to p:q, with p and q ($p < q$) being small integers. For such an interval the harmonics with frequencies of npf_2 and nqf_1 Hz ($n = 1,2,...$) coincide. This is illustrated in Fig. 1 for the octave, fifth, and major third. In this figure, the frequencies of a number of harmonics of the lower and the higher tones are represented. For 1:2 (octave), for example, the first (pth) harmonic of the higher tone coincides with the second (qth) harmonic of the lower tone. In fact, for 1:2, all harmonics of the higher tone coincide with harmonics of the lower tone. It can be seen in Fig. 1, that for 2:3 (fifth) only 50% of the harmonics of the higher tone coincide with harmonics of the lower tone. For 4:5 (major third) this percentage falls to 25%.

Physically, for two simultaneous complex tones, a moderately *tempered* interval is characterized by small frequency differences between those harmonics which coincide in pure intervals. The interference of these nearly coinciding harmonics results in both amplitude variations and phase or frequency variations (see Rasch, 1984).

[a)] Preliminary results of part of the present research have been presented by Rudolf A. Rasch at the 11th International Congress on Acoustics, Paris, France, 19–27 July 1983, and are included in the *Proceedings*, Vol. 4, 419–422.

TABLE I. Summary table of the 13 musical intervals which were presented in our experiments. The intervals are pure when the ratio of the fundamental frequencies of the lower tone and the higher tone, f_1:f_2, is equal to p:q. The interval size, $1200 \log_2(q/p)$, is given in the fourth column. Thresholds for the discrimination between pure and tempered intervals, expressed in terms of the absolute difference (ΔL) between the level of tone 1 (L_1) and the level of tone 2 (L_2) in dB, are given for experiment 1 and experiment 2, separately. In experiment 1, the complex tones 1 and 2 consisted of 20 harmonics with amplitudes, a_n, proportional to $1/n$, whereas in experiment 2, all harmonics had the same amplitude. Beat frequency, defined as $|pf_2 - qf_1|$, is given in Hz.

				Discrimination threshold in level difference (ΔL) between the tones in dB											
				Experiment 1						Experiment 2					
		Frequency ratio	Interval size in	$L_1 < L_2$ beat frequency			$L_1 > L_2$ beat frequency			$L_1 < L_2$ beat frequency			$L_1 > L_2$ beat frequency		
Stim. No.	Name of musical interval	p:q	cents	2	4	8	2	4	8	2	4	8	2	4	8
1	unison	1:1	0	29	32	29	28	29	28	27	29	30	28	29	28
2	octave	1:2	1200	30	33	34	23	25	26	27	29	32	23	25	26
3	twelfth	1:3	1902	28	31	32	21	23	25	25	27	29	19	22	24
4	fifth	2:3	702	22	26	27	18	22	26	20	25	27	28	21	24
5	major tenth	2:5	1586	21	25	26	16	16	23	22	26	27	13	17	19
6	fourth	3:4	498	18	22	25	17	21	22	18	21	25	15	18	22
7	major sixth	3:5	884	20	22	26	16	17	22	18	20	21	13	17	19
8	major third	4:5	386	18	21	22	17	19	21	19	20	23	14	17	18
9	subminor seventh	4:7	969	16	20	23	11	13	15	15	19	19	8	14	15
10	minor third	5:6	316	16	18	21	13	16	21	15	19	18	13	17	15
11	subminor fifth	5:7	583	12	18	21	10	15	15	14	17	14	11	17	14
12	minor sixth	5:8	814	16	20	20	10	14	15	14	18	18	10	13	14
13	subminor third	6:7	267	13	17	17	12	15	19	12	15	16	11	15	15

In short, the depth (in dB) of the variation in level (= envelope) of two interfering harmonics with amplitudes a_1 and a_2 is given by

$$D = 20 \log \left| \frac{a_1 + a_2}{a_1 - a_2} \right| = 20 \log \left(\frac{10^{\Delta L/20} + 1}{10^{\Delta L/20} - 1} \right), \qquad (1)$$

in which ΔL is the absolute level difference between a_1 and a_2 in dB [$\Delta L = |20 \log(a_1/a_2)|$]. D approaches ∞ for $a_2 \to a_1$. For the nth pair of interfering harmonics, the frequency of the level variation is equal to

$$f_{bn} = |npf_2 - nqf_1|. \qquad (2)$$

Perceptually, the interference of the nearly coinciding harmonics gives rise to beats or roughness; the strength of the beats increases with increasing D. For various spectral conditions, Vos (1984) has shown that the dominantly perceived beat frequency is equal to that of the first pair of nearly coinciding harmonics.

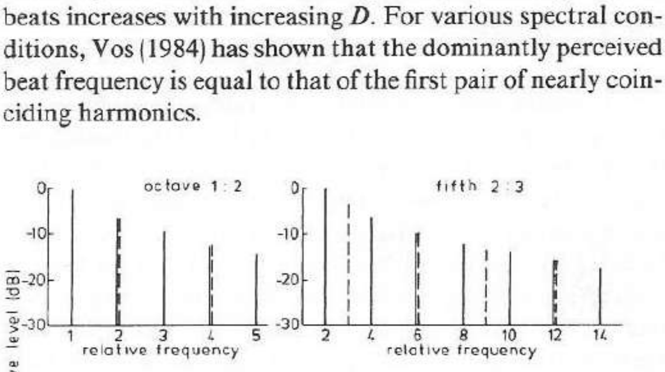

FIG. 1. Presentation of part of the spectral contents of the octave, fifth, and major third. The levels of a number of harmonics of the lower tone (continuous lines) and the higher tone (broken lines) are given on a linear frequency scale. The levels of the harmonics are plotted relative to the level of the first harmonic of the lower tone. In this figure the spectral-envelope slope is −6 dB/oct.

In our experiments, level-variation depth of the beats was manipulated by changing the difference in level (ΔL) between the two tones. DTs will be expressed either in terms of ΔL or in terms of D. Throughout this paper beat frequency (f_b), representing the degree of tempering, is defined as the absolute frequency difference between the harmonics in the first pair of nearly coinciding harmonics.

2. Frequency variation

When nearly coinciding harmonics have equal amplitudes, the pitch is equal to the mean of the frequencies of the two harmonics. When they are different in both frequency and amplitude, there is a shift in the instantaneous frequency during the moment when the envelope goes through a minimum (Helmholtz, 1877/1954, App. XIV). Moreover, the degree of frequency variation strongly decreases with increasing difference between the levels of the harmonics. Because of these two factors, the perception of the frequency variation is, in general, less conspicuous than that of the amplitude variation (Wever, 1929; Jeffress, 1968).

For small level differences (0 dB $< \Delta L \leqslant 3$ dB) between two sinusoids, however, it has been demonstrated that frequency fluctuations may be relevant in perception (Feth, 1974; Feth and O'Malley, 1977; Feth *et al.*, 1982).

In the experiments reported in the present paper, discrimination between pure and tempered intervals was studied in threshold conditions. It can be seen in Table I that, in most cases, the thresholds obtained were rather low (10 dB $< \Delta L <$ 30 dB). In this paper, frequency variation will therefore be left out of consideration. This is supported by the finding that for modulation frequencies lower than, say 25 Hz, thresholds for frequency-modulated sinusoids are, in terms of modulation depth, much higher than thresholds for amplitude-modulated sinusoids (Zwicker, 1952; Hartmann and Hnath, 1982).

B. Interference of other harmonics

It has been shown that, for moderately tempered fifths, discriminability between pure and tempered intervals is mainly determined by the interference of nearly coinciding harmonics (Vos, 1984). For major thirds, however, Vos concluded that the interference of other harmonics plays a role as well.

Physically, the small-integer frequency ratios for various triads of adjacent harmonics are also changed by tempering. As depicted in Fig. 1, for the fifth this is the case for harmonics with their relative frequencies in the ratios of 2:3:4 and 8:9:10. For the major third, the triad, comprising the second and third harmonics of the lower tone and the second harmonic of the higher tone, with their frequencies in the ratio of 4:5:6, provides another example.

Beats can also be perceived in a number of these slightly tempered triads. By manipulating the spectral content of the higher tones, Vos (1984) has shown that for the fifth, beats are heard in the triad with frequencies in the ratio of $2:(3-g):4$, g representing a small amount of tempering. Similarly, he showed that for the major third, beats result from interference of the harmonics with frequencies in the ratio of $4:(5-g):6$. These findings are in agreement with early observations by ter Kuile (1902). The extent to which interference of harmonics that are different from nearly coinciding harmonics contributes to DT will be investigated in experiment 3.

II. EXPERIMENT 1

A. Method

1. Stimuli

The stimuli consisted of two complex tones 1 and 2 with fundamental frequencies of f_1 and f_2 Hz, respectively. The frequency ratio $f_1:f_2$ was equal to or slightly different from $p:q$. Thirteen musical intervals were presented. They are given in Table I, together with the corresponding values of p and q. In the fourth column of this table, the size of the pure intervals is given in cents.

The complex tones 1 and 2 consisted of 20 harmonics with amplitude a_n proportional to $1/n$. For both tones, the spectral-envelope slope was therefore -6 dB/oct. This spectrum closely resembles that of a bowed string instrument in its lower frequency register. The phase of the individual harmonics was chosen randomly. Because of this the waveforms of tones 1 and 2, although having the same spectral envelope, were different. The overall level of tones 1 and 2 together was 75 dB(A). This level was measured by means of an Artificial Ear (Brüel & Kjaer, type 4152). Rise and decay times of the tone bursts, defined as the time interval between 10% and 90% of the maximum amplitude, were 30 and 20 ms, respectively. Level-variation depth of the beating harmonics was manipulated by attenuation of tone 1 or tone 2 by ΔL dB. When the spectral envelopes of the two tones coincide, that is, when the amplitudes of the nearly coinciding harmonics are equal, ΔL is defined to be 0 dB. Level-variation depth is maximal in this condition (see Eq. 1), which occurs when the level of tone 2 is $20 \log(q/p)$ dB lower than the level of tone 1 (see Fig. 1).

2. Apparatus

The experiment was run under the control of a PDP-11/10 computer. Tones 1 and 2 were generated in the following way. One period of the waveforms of tones 1 and 2 was stored in 256 discrete samples (with 10 bit accuracy) in external revolving memories. These recirculators could be read out by digital-to-analog converters. Sampling rates were determined by pulse trains, derived from two frequency generators. After gating, each tone passed a Krohn-Hite filter (model 3341) with the function switch in the low-pass RC mode and with a cutoff frequency of 8 kHz. The sound-pressure levels of the tones were controlled by programmable attenuators. After appropriate attenuation, the tones were mixed and fed to a headphone amplifier. The signals were presented diotically (same signal to both ears) by means of Beyer DT 48S headphones. Subjects were seated in a soundproof room.

3. Subjects

Six musically trained subjects were tested over five half-day sessions. They were paid for their participation.

4. Experimental variables

The independent variables were (1) musical interval (see Table I), (2) attenuation of tone 1 versus attenuation of tone 2, and (3) beat frequency of the first pair of nearly coinciding harmonics (2,4, and 8 Hz). For each of the 78 different conditions, all of which were presented to the six subjects, a DT was determined.

5. Procedure

The task for the subject was to discriminate between pure and tempered musical intervals in a two-Alternative Forced Choice (2 AFC) paradigm. The tempered interval was presented first in half of the trials and second in the other half of the trials. The duration of the tone bursts was 500 ms. Between the intervals there was a silent period of 800 ms. By means of two response buttons, the subject indicated whether the first or the second interval was tempered. Feedback was given by a red light above the correct button.

The next trial was presented 1.7 s after the response to the preceding trial. Thresholds were determined by means of an adaptive procedure, in which ΔL was increased by 2 dB after three correct responses and decreased by the same amount after a single incorrect response. Each reversal of the direction of change in ΔL over successive trials was counted as a turnaround. There were eight turnarounds. In the first run (a number of successive trials), however, ΔL, being 0 dB at the very first trial, was increased by 3 dB after each correct response. The DT was defined as the mean value of ΔL on the trials that resulted in the last six turnarounds. This value estimates the value of ΔL required to produce 79.4% correct responses in a nonadaptive 2 AFC procedure (Levitt, 1971). A block comprised three series, each resulting in the DT for one beat frequency. Presentation order of the different beat frequencies was randomized. There were 26 different blocks: 13 musical intervals in which the level of tone 1 (L_1) was lower than the level of tone 2 (L_2), and 13 musical intervals in which $L_1 > L_2$. Presentation order of the various blocks was randomized. At the beginning of a block, new waveforms of tones 1 and 2, with different randomly chosen phases of the individual harmonics, were computed. In order to avoid absolute pitch recognition as a cue, fundamental frequencies were varied from trial to trial in such a way that the central frequency f_c [= the geometric mean $(f_1 f_2)^{1/2}$] of the pure interval could be equal to 370 Hz ± 25, ± 75, ± 125, ± 175, or ± 225 cents, with equal probability of occurrence. The f_1 of the pure interval equaled

$$f_1 = f_c/(q/p)^{1/2}. \quad (3)$$

In the tempered interval, one of the fundamental frequencies equaled the corresponding fundamental frequency in the pure interval, whereas the other one was altered. For the tempered interval, either a new f_1 was calculated by means of

$$f_1(\text{tempered}) = (p f_2 \pm f_b)/q, \quad (4)$$

or a new f_2 was calculated by

$$f_2(\text{tempered}) = (q f_1 \pm f_b)/p. \quad (5)$$

For each trial, the probability that the frequency of either tone 1 or tone 2 was altered, and that the tempered interval was either stretched or compressed [signs in Eqs. (4) and (5)] was 50%. Throughout this paper, data have been combined for stretched ($pf_2 - qf_1 > 0$) and compressed ($pf_2 - qf_1 < 0$) intervals, because previous research (Vos, 1982, 1984) indicated that DTs do not depend on the direction of tempering.

A session comprised five or six blocks. In the first session the subjects received two training series; at subsequent sessions one training series was given before the experimental series started. In the $\Delta L < 10$ dB conditions, L_1 and L_2 were attenuated to a small extent to satisfy our criterion that the overall level of the interval should remain at 75 dB(A). This correction (C) in dB of L_1 and L_2, being dependent on ΔL and on the musical interval (q/p), is given by

$$C = 10 \log(1 + 10^{[-\Delta L \pm 20 \log(q/p)]/10}). \quad (6)$$

B. Results

Discrimination thresholds (DTs), presented in terms of level difference ΔL, were subjected to an ANOVA [6 (subjects) $\times$ 13 (musical intervals) $\times$ 2 (attenuation of tone 1 or tone 2) $\times$ 3 (beat frequency), all repeated measures]. Although individual differences occurred [$F(5,120) = 68.4$, $p < 0.00001$], the variance explained by the three experimental variables was about six times as high as that explained by differences between the subjects. Averaged across subjects, the DTs are given in Table I for all conditions. DTs, being highly dependent on interval [$F(12,60) = 69.6, p < 0.00001$], are plotted in Fig. 2(a) as a function of the interval size of the various pure intervals. The interval size, given in cents, is defined as 1200 $\log_2(q/p)$. A Newman–Keuls paired-comparison test (Winer, 1970) revealed that the differences between the DTs of the various intervals could be nicely

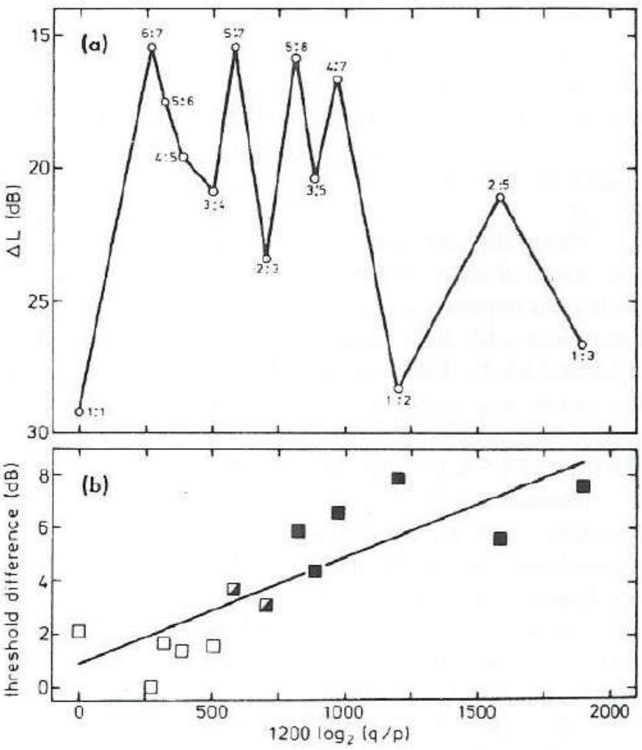

FIG. 2. (a) Thresholds for discrimination between pure and tempered intervals, as a function of the interval size of the various pure intervals in cents. The thresholds are averaged across beat frequency and the $L_1 < L_2$ and $L_1 > L_2$ conditions. (b) Differences between the discrimination thresholds for the $L_1 < L_2$ and $L_1 > L_2$ conditions, as a function of the interval size of the various pure intervals in cents. Significance levels, determined by means of t tests, are indicated: □ (not significant), ◪ ($p < 0.05$), ■ ($p < 0.01$). The line fitted through the mean values has been determined by linear regression; the linear equation is $y = 0.85 + 0.004x$; the coefficient of determination r^2 equals 0.71.

summarized by means of four groups of intervals, the groups being significantly different from each other at the 0.01 level. Ranked from high to low sensitivity to tempering, these groups comprise (1) 1:1, 1:2, and 1:3, (2) 2:3, (3) 4:5, 3:4, 3:5, and 2:5, and (4) 6:7, 5:6, 5:7, 5:8, and 4:7. Attenuation of tone 1 results in lower DTs than attenuation of tone 2 [$F(1,5) = 179.0$, $p < 0.0005$]. More importantly, however, the differences between the DTs for the $L_1 < L_2$ and $L_1 > L_2$ conditions increase with increasing size of the intervals [$F(12,60) = 5.6$, $p < 0.00005$]. The interaction pattern is shown in Fig. 2(b). The inserted linear regression line, fitted through the mean threshold differences for the various intervals, accounts for 70% of the variance in the threshold differences. Sensitivity to tempering increases with increasing beat frequency [$F(2,10) = 23.2$, $p < 0.0005$]; mean DTs for beat frequencies of 2, 4, and 8 Hz are 18, 21, and 23 dB, respectively. The effect of beat frequency was to some extent dependent on the kind of musical interval [$F(24,120) = 2.0$, $p < 0.01$]. Discrepancies, however, were too unsystematic to be interpretable. The interaction pattern can be inspected in the upper part of Fig. 3, where the DTs are plotted, as will be explained in the discussion, as a function of the sum of p and q, with beat frequency as parameter. The inserted linear regression lines, fitted through the mean thresholds of the various intervals for beat frequencies of 2, 4, and 8 Hz, separately, account for 94%, 94%, and 96% of the variance in ΔL, respectively.

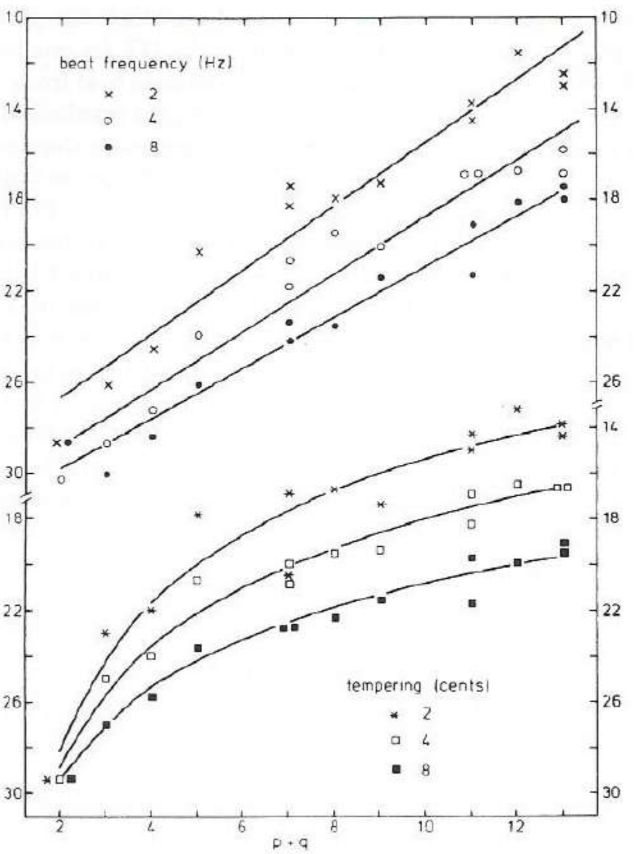

FIG. 3. Thresholds for discrimination between pure and tempered intervals, as a function of frequency-ratio complexity, $p+q$. Beat frequency of the first pair of nearly coinciding harmonics (upper part of the figure) or the degree of tempering in cents (lower part of the figure) are given as parameters. The inserted regression functions were determined by the method of least squares.

C. Discussion

1. Frequency-ratio complexity

Sensitivity to tempering was highly dependent on the kind of musical interval. It is evident from Fig. 2(a) that there is no relation between discriminability and the size of the intervals. However, a simple relation can be given if musical interval is defined in terms of the parameters p and q, representing the frequency ratio of the tones: DTs increased with increasing complexity of the frequency ratio. Very high correlation coefficients r were found between the DTs and measures of frequency-ratio complexity such as $\log_{10}(pq)$, $p+q$, and $(pq)^{1/2}$, corresponding values of r being -0.99, -0.98, and -0.98, respectively. Complexity measures such as $\log_{10}(pf_2)$, q, p, and pq correlated highly with DT as well. Evidently, the various complexity measures, as given above, are highly intercorrelated. Therefore, as a rule of thumb, it can be concluded that for musical intervals with f_1:f_2 slightly different from p:q, DT decreases linearly as a function of the sum of p and q. This relationship is shown in the upper part of Fig. 3 for the three presented beat frequencies separately.

One could object that the strong relationship between the DTs and, for example, $p+q$, as found in our experiment, is a trivial one, since frequency-ratio complexity and the levels of the interfering harmonics with frequencies of pf_2 and qf_1 Hz had changed together. In fact, linear regression reveals that about 90% of the variance in the mean DTs can be accounted for by the level of the qth harmonic with a frequency of qf_1 Hz, relative to the level of the first harmonic. [Recall that the level of the qth harmonic is 20 log(q) lower than the level of the first harmonic.] However, it is unlikely that the differences between the DTs can be completely explained by the level differences of the interfering harmonics. For our set of intervals, the highest and lowest levels of corresponding pairs of interfering harmonics are found for 1:1 and 5:8, respectively, the difference being 18 dB. The combined results from Maiwald (1967), Schöne (1979), and Zwicker (1952) indicate that when the sound-pressure level of an amplitude-modulated sinusoid decreases from 80 to 60 dB, the level-variation depth at threshold increases from about 0.3 to 0.5 dB. In terms of ΔL, this threshold increase is only 5 dB. Since our stimulus conditions are much more complex than those in which the amplitude of only one sinusoid is modulated, it is still worthwhile to test to what extent the relation between discriminability and frequency-ratio complexity depends on these level differences rather than on some other mechanism. Therefore, we designed experiment 2, which was in fact a replication of experiment 1, but now for tones with harmonics of equal amplitude.

2. Measures of the amount of tempering

A second point that can be raised concerns the way in which the amount of tempering was expressed. In our experiment, tempering was represented by beat frequency f_b, but it can be argued that it is more appropriate to express tempering as the difference between the logarithmically defined interval size of the tempered interval and that of the pure interval (see, e.g., Barbour, 1940). As suggested by Rasch (1984), the relationship between this difference, expressed in cents, and f_b can simply be given by

$$T=(1731\,f_b)/(qf_1). \tag{7}$$

Since f_b was defined as the absolute difference between pf_2 and qf_1, T in Eq. (7) also represents tempering in absolute terms. Equation (7) shows that T does not only depend on f_b, but also on q and the frequency of the fundamentals. In short, since the actually presented magnitude of T depended on musical interval, the strength of the relationship between the DTs and ratio complexity $p+q$, as found in our experiment, may be exaggerated if T rather than f_b is the preferred representation of the degree of tempering.

Therefore, for each musical interval separately, but combined for the $L_1<L_2$ and $L_1>L_2$ conditions, the relation between DT and T was determined. For example, at 5:6, beat frequencies of 2, 4, and 8 Hz correspond to mean temperings of 1.7, 3.4, and 6.9 cents, respectively. A regression function was based on the three mean DTs for these conditions and the three values of T. From this function, new DTs were computed for temperings of 2, 4, and 8 cents. For 12 of the 13 intervals, computation of regression functions was significant (mean $r^2=0.92$; standard deviation $=0.07$). As can be seen in Fig. 3, there was no similar effect of tempering on the DTs for 1:1. In the lower part of Fig. 3, DTs are plotted as a function of $p+q$, with tempering in cents as a parameter. The inserted power functions $[\mathrm{DT}'=a(p+q)^b]$

account for about 94% of the variance in the derived DTs.

Comparing the upper and lower parts of Fig. 3, it can be seen that taking the whole set of intervals into account, the strength of the effect of frequency-ratio complexity on the DTs is about the same for both measures of tempering. If 1:1 is left out of consideration, the threshold differences attributed to ratio complexity, found at the various beat frequencies, are reduced by about 30% when the thresholds are given for various degrees of tempering in cents. From this we can conclude that the significance of the effect of ratio complexity on the DTs does not strongly depend on a specific measure of the degree of tempering.

III. EXPERIMENT 2

The main result of experiment 1 was that we found a very high correlation between DT and frequency-ratio complexity as defined by, e.g., the sum or the product of p and q, the two integers in the ratio. It has been explained before that, at least for a number of the intervals, discrimination performance is mainly determined by the interference of nearly coinciding harmonics. Due to the spectral-envelope slope of our tones, frequency-ratio complexity and the levels of these nearly coinciding harmonics were interrelated. To test to what extent the relation between discriminability and ratio complexity depends on these level differences instead of on some other mechanism, experiment 1 was repeated for tones with harmonics of equal amplitude.

A. Method

1. Stimuli and apparatus

The stimuli were the same as those in experiment 1. However, the spectral-envelope slope of complex tones 1 and 2 was flat, that is, all harmonics had the same amplitude. Again, the overall level of tones 1 and 2 together was 75 dB(A). Coincidence of spectral envelopes occurs when the levels of tones 1 and 2 are equal. The apparatus was the same as in experiment 1.

2. Subjects

Six musically trained subjects were tested over five half-day sessions. Two of them had also participated in experiment 1; their mean thresholds were in between those of the other four subjects in both experiments 1 and 2. All subjects were paid for their participation.

3. Experimental design

For the main experiment, in which the tones had a flat spectral-envelope slope, the design was identical to that of experiment 1. In addition, thresholds were determined for three "representative" musical intervals, using tones with a spectral-envelope slope of -6 dB/oct. These intervals were 1:1, 2:3, and 5:8. For each of these intervals in this reference experiment, thresholds were determined for the $L_1 < L_2$ and $L_1 > L_2$ conditions and for beat frequencies of 2, 4, and 8 Hz.

4. Procedure

In general, the procedure was similar to that in experiment 1. Presentation order of the various conditions, however, was different to some extent. A subblock comprised three or four series, each resulting in the DT for one beat frequency. Presentation order of the different beat frequencies was randomized. There were 32 different subblocks: 16 musical intervals (13 with a flat spectral-envelope slope and three with a spectral-envelope slope of -6 dB/oct) in which $L_1 < L_2$, and the same 16 intervals, in which $L_1 > L_2$. For the $L_1 < L_2$ and $L_1 > L_2$ conditions, separately, the 16 intervals were assigned to three blocks. One block comprised 1:1, 1:2, 1:3, and 2:3 from the main experiment and 1:1 from the reference experiment. A second block comprised 5:6, 4:5, 3:4, 3:5, and 2:5 from the main experiment and 2:3 from the reference experiment. A third block comprised 6:7, 5:7, 5:8, and 4:7 from the main experiment and 5:8 from the reference experiment. Thus, combined for the $L_1 < L_2$ and $L_1 > L_2$ conditions, there were six different blocks. Presentation order of these blocks was balanced by means of a (6×6) Latin square.

B. Results

For the conditions in which all harmonics had equal amplitudes, the DTs were subjected to an ANOVA having a design similar to that used in experiment 1. Individual differences $[F(5,120) = 274.0, p < 0.00001]$ were higher than those found in experiment 1. However, the variance in the DTs explained by the experimental variables was more than twice as high as that explained by the main effect of subjects. DTs were highly dependent on musical interval $[F(12,60) = 70.7, p < 0.00001]$. Attenuation of tone 1 results in lower DTs than attenuation of tone 2 $[F(1,5) = 63.3, p < 0.001]$. Similar to the results from experiment 1, the differences between the DTs for the $L_1 < L_2$ and $L_1 > L_2$ conditions increase with increasing size of the musical intervals $[F(12,60) = 3.9, p < 0.0005]$. Again, sensitivity to tempering increases with increasing beat frequency $[F(2,10) = 15.7, p < 0.001]$. The effect of beat frequency is to some extent dependent on the kind of interval $[F(24,120) = 2.6, p < 0.0005]$. Discrepancies, however (see Table I), could not be interpreted, not even on the basis of a comparison with the interaction pattern between beat frequency and interval, found in experiment 1.

The results of experiment 2 were essentially the same as those of experiment 1 (see Table I): in general, all the main and interaction effects discussed above were similar. For a small set of intervals, comprising 1:1, 2:3, and 5:8, the same conclusion was reached for thresholds obtained in conditions in which spectral-envelope slope was *not* confounded with a particular group of subjects. Collapsed over beat frequency, neither the main effect of spectral content $(p > 0.18)$ nor any interaction effect with spectral content was significant.

IV. EXPERIMENT 3

On the one hand, it is nice to conclude that the simple relation between discriminability and ratio complexity, as found in experiment 1 for tones with a spectral-envelope slope of -6 dB/oct, is not affected by a large change in this slope as in experiment 2. Thus the results of experiment 2 suggest that the relation between DT and $p+q$ cannot be explained by the level of the interfering harmonics. The results also indicate that the rule can be applied to a large number of spectrally different tones. On the other hand, however, we are not confident of the true nature of the mechanism responsible for the relationship we have established. The finding (Vos, 1984) that for the major third the interference of the second harmonic of the higher tone with the adjacent harmonics of the lower tone (see Fig. 1) contributes significantly to DT suggests, that with other musical intervals, harmonics different from nearly coinciding harmonics may also play a role in discrimination. This may be true for all intervals with sums of p and q larger than 5. Moreover, it might even be hypothesized that with increasing sum of p and q, DTs are increasingly determined by the combined effect of these different interfering sources, and that the contribution of the interfering nearly coinciding harmonics becomes negligibly small. This was tested in experiment 3 by selectively presenting or deleting the nearly coinciding harmonics.

A. Method

1. Stimuli and apparatus

The stimuli were the same as those in experiment 1. The spectral-envelope slope of the complex tones was also -6 dB/oct. The spectral content of tones 1 and 2, however, could be varied. The apparatus was identical to that in experiment 1.

2. Subjects

Six subjects, all having prior experience with this kind of discrimination task, were tested over two half-day sessions. Four of them, including two subjects who had also participated in experiments 1 and 2, were paid for their services. The remaining two subjects were the authors.

3. Experimental design

The independent variables were (1) musical interval (the set of intervals of Table I), (2) attenuation of tone 1 ($L_1 < L_2$) or tone 2 ($L_1 > L_2$), and (3) spectral content of the tones. In the $L_1 < L_2$ conditions, the spectrum of tone 1 could comprise the following harmonics: (a) all (i.e., the first 20 harmonics), (b) all—nqth, and (c) nqth, whereas tone 2 always comprised all harmonics. In the $L_1 < L_2$ conditions, the spectrum of tone 2 could comprise the following harmonics: (a) all, (b) all—npth, and (c) npth, whereas tone 1 always comprised all harmonics.

4. Procedure

In general, the procedure was similar to that in experiment 1. For the tempered intervals, the beat frequency was always 4 Hz. In both the $L_1 < L_2$ and $L_1 > L_2$ conditions, there were 33 series: 13 intervals with tones 1 and 2 comprising all harmonics, ten intervals for which either the nqth harmonics of tone 1, or the npth harmonics of tone 2 were deleted (the spectral variations are not significant for 1:1, 1:2, and 1:3), and ten intervals in which tones 1 or 2 only comprised the nqth or npth harmonics. For each subject separately and for both the $L_1 < L_2$ and $L_1 > L_2$ conditions, the 33 series were randomly assigned to six blocks, each comprising five or six series. Blocks for the $L_1 < L_2$ and $L_1 > L_2$ conditions were presented alternately. Three subjects started with an $L_1 < L_2$ block, the other subjects started with an $L_1 > L_2$ block.

B. Results

For all conditions in which the frequency-ratio complexity, $p + q$, was larger than 4, the DTs were subjected to an ANOVA [6 (subjects) $\times$ 10 (intervals) $\times$ 2 (attenuation of tone 1 or tone 2) $\times$ 3 (spectral conditions), all repeated measures]. Although individual differences occurred [$F(5,90) = 96.8, p < 0.00001$], the variance explained by the three experimental variables was about six times as high as that explained by differences between the subjects.

For the three spectral conditions, mean thresholds are plotted in Fig. 4 as a function of frequency-ratio complexity, $p + q$. From this figure it can be seen that, for all spectral conditions, sensitivity to tempering decreased with increasing ratio complexity. The ANOVA showed that this effect was very significant [$F(9,45) = 34.7$, $p < 0.00001$]. In the conditions in which the npth or nqth harmonics were deleted the DTs were much higher than in the conditions in which these harmonics were present [$F(2,10) = 157.0$, $p < 0.00001$].

By and large, the effect of musical interval on DT was similar for those conditions in which nearly coinciding harmonics were present; the linear regression line, fitted through the mean thresholds for the various intervals accounts for 92% of the variance in ΔL. In the condition in which one of each pair of nearly coinciding harmonics was absent, the inserted linear regression line explains 89% of the variance in the mean thresholds.

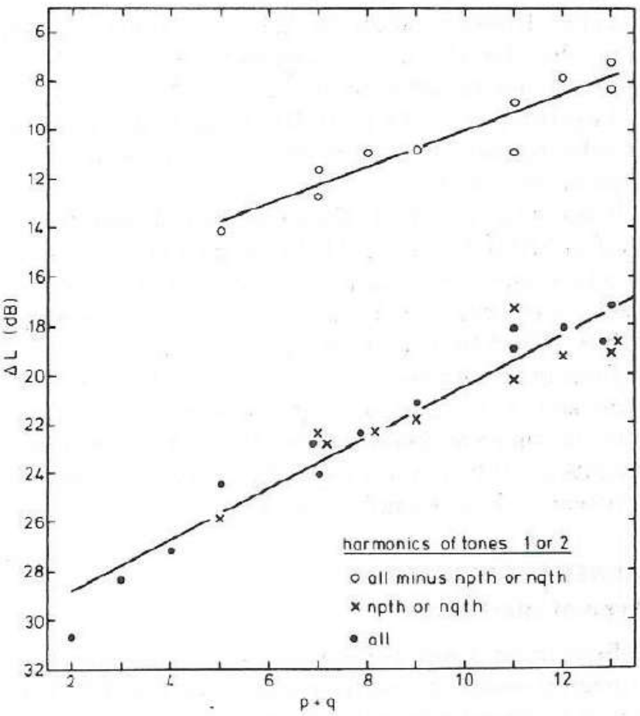

FIG. 4. Thresholds for discrimination between pure and tempered intervals, as a function of frequency-ratio complexity, for three spectral conditions. Thresholds are averaged across the $L_1 < L_2$ and $L_1 > L_2$ conditions.

TABLE II. Thresholds for discrimination between pure and tempered intervals, determined in experiment 3, expressed in terms of the absolute difference (ΔL) between the level of tone 1(L_1) and the level of tone 2(L_2). The various spectral conditions are indicated. The spectrum of tone 1 was varied in the $L_1 < L_2$ conditions; the spectrum of tone 2 was varied in the $L_1 > L_2$ conditions. Spectral variations are not significant for 1:1, 1:2, and 1:3. In the last two columns, thresholds, averaged across experiments 1, 2, and 3, are given for the conditions in which all harmonics were present and the beat frequency was 4 Hz. These thresholds are expressed in terms of the level-variation depth of nearly coinciding harmonics.

		Experiment 3: DT in level difference ΔL (dB)						Experiments 1, 2, and 3: DT in level-variation depth (dB)	
		$L_1 < L_2$ harmonics of tones 1 and 2			$L_1 > L_2$ harmonics of tones 1 and 2				
Stim. No.	Frequency ratio p:q	tone1: all tone 2: all	nqth all	all-nqth all	all all	all npth	all all-npth	$L_1 < L_2$	$L_1 > L_2$
1	1:1	31	...	...	30	...	...	0.5	0.6
2	1:2	31	...	...	26	...	...	0.6	1.0
3	1:3	31	...	...	24	...	...	0.6	1.3
4	2:3	26	28	17	23	24	11	0.9	1.5
5	2:5	27	26	18	21	19	8	0.9	2.4
6	3:4	24	24	14	21	22	9	1.4	1.8
7	3:5	24	25	15	21	20	7	1.5	2.7
8	4:5	22	24	12	20	20	10	1.6	2.2
9	4:7	21	20	14	15	15	8	2.1	3.9
10	5:6	20	22	10	18	19	8	2.2	2.7
11	5:7	21	21	10	15	17	6	2.4	3.4
12	5:8	19	21	10	15	17	4	2.1	3.9
13	6:7	20	20	10	17	19	7	2.6	3.4

In addition to these robust main effects, the ANOVA showed a significant interaction effect of interval and spectral content [$F(18,90) = 2.55, p < 0.002$]. In fact (see Table II) only the small deviation from the general trends, found for 4:7, is worth mentioning. In this interval, variation of spectral content had been less effective.

In Fig. 4, the plotted DTs were combined for the $L_1 < L_2$ and $L_1 > L_2$ conditions because neither the effect of spectral content ($p > 0.23$) nor the interaction effect of spectrum and interval ($p > 0.10$) depended on which tone was attenuated. However, again the DTs were lower for the $L_1 < L_2$ than for the $L_1 > L_2$ conditions [$F(1,5) = 154.0$, $p < 0.0003$], and this difference increased with increasing interval size [$F(9,45) = 3.88, p < 0.001$]. Inspection of Table II shows that the latter two effects were similar to those found in experiments 1 and 2.

The similarity of these effects was confirmed by the results of an ANOVA with the DTs from experiments 1, 2, and 3 as a between-subjects variable and musical interval and attenuation of the lower or the higher tone as within-subjects variables. [The ANOVA was only performed on the thresholds from the conditions in which (1) all harmonics were present and (2) the beat frequency was equal to 4 Hz.] Neither the differences between the DTs from the three experiments ($p > 0.39$) nor any interaction effect with kind of experiment were significantly different.

V. GENERAL DISCUSSION

A. Type of interference

Experiment 3 was designed to test the significance of two different sources of interference for discrimination. The first source originates from the interference of various pairs of nearly coinciding harmonics. The second source, in fact, represents all other kinds of interference which can be thought to be the result of the combination of various pairs or triads of adjacent harmonics.

In contrast to a previous study (Vos, 1984), the present study tested the relevance of these two different sources in a general sense. Our data were not decisive on the relative contributions of the various pairs of nearly coinciding harmonics. Similarly, with respect to the second source, it was our aim neither to investigate which pair or triad of neighboring harmonics produced most interference nor to find out whether these harmonics gave rise to the perception of combination tones or waveform variations (see Plomp, 1976).

These restrictions, however, do not prevent us from concluding that for all intervals comprising tones with a

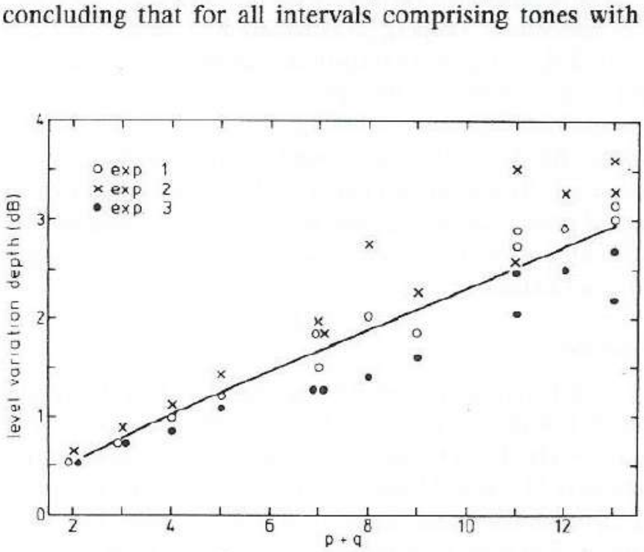

FIG. 5. Thresholds for discrimination between pure and tempered intervals, presented in terms of level-variation depth and plotted as a function of frequency-ratio complexity, are given for experiments 1, 2, and 3, separately. For all experiments, thresholds are given only for conditions in which all harmonics were present and the beat frequency was 4 Hz. Thresholds are averaged across the $L_1 < L_2$ and $L_1 > L_2$ conditions.

complete series of harmonics, discriminability between pure and tempered intervals is determined by the interference of nearly coinciding harmonics. That is, when the harmonics with frequencies of npf_2 and nqf_1 Hz are present, the interference between other harmonics plays a minor part in discrimination. Having localized the most significant source of interference, we prefer to express the DTs in terms of level-variation depth (D). For this purpose, the individual ΔL values at threshold were converted into corresponding values of D for the conditions in which all harmonics were present and the beat frequency was 4 Hz. The mean thresholds are plotted in Fig. 5 as a function of frequency-ratio complexity for experiments 1, 2, and 3, separately.

The regression line fitted through the mean thresholds, shown in Fig. 5, explains 91% of the variance. A straight line, hardly discriminable from this power function, would explain 85% of the variance.

B. Discriminability and frequency-ratio complexity

One of the aims of the present study was to find and comprehend a possible relation between type of musical interval and the performance level at which, for each of these intervals, their pure versions could be discriminated from their tempered versions.

We have shown that, indeed, there is a simple relation between frequency-ratio complexity and discriminability. Moreover, we have demonstrated that the decreasing discriminability with increasing ratio complexity was not simply a result of (1) the way in which the degree of tempering had been defined (see Fig. 3), and (2) the choice of a spectrum in which, from the fundamental upwards, the levels of the harmonics dropped off by 6 dB/oct (see Fig. 5). We have also shown (3) that, although discrimination performance was much higher when interference between nearly coinciding harmonics was possible, the nature of the relationship remained the same, even when the main source of interference was eliminated (see Fig. 4).

More research, probably with further refinements in spectral variations, is needed before a satisfactory explanation can be given. It may be important to know the relative contributions of the various interfering pairs of nearly coinciding harmonics to discrimination. Therefore, it may be preferred to start future experiments with simpler stimuli. However, as a working hypothesis for future research, we would like to suggest at least one explanation. Our attention will be focused on harmonics within a restricted frequency range around the first pair of nearly coinciding harmonics. We assume that ignoring a possible interval-dependent contribution of higher pairs of interfering harmonics to discrimination does not affect the essential content of our explanation.

Recall that for each interval the estimation of our subjects' overall discrimination performance was based on thresholds for the $L_1 < L_2$ and $L_1 > L_2$ conditions. Since differences between the DTs for the $L_1 < L_2$ and the $L_1 > L_2$ conditions depended on the kind of interval in a reliable way, it seemed worthwhile to analyze these thresholds in more detail. Because of the similarity of the results the thresholds, given in the last two columns of Table II, are collapsed over experiments 1, 2, and 3.

In the $L_1 < L_2$ conditions, the differences between the DTs for intervals with equal values of p are very small and can therefore be grouped. With respect to the value of q, the same holds for the DTs in the $L_1 > L_2$ conditions. At threshold, the levels of the harmonics of tone 2 in the $L_1 < L_2$ conditions are much higher than those of tone 1 (see upper part of Fig. 6). The interference between the low level qth harmonic of tone 1 and the high level pth harmonic of tone 2 can be thought to be masked by the high level (pth $- 1$) and (pth $+ 1$) harmonics of tone 2. Similarly, it seems plausible to assume that in the $L_1 > L_2$ conditions (see lower part of Fig. 6), the interference between the low level pth harmonic of tone 2 and the high level qth harmonic of tone 1 is masked by the high level (qth $- 1$) and the (qth $+ 1$) harmonics of tone 1.

It is a well-established fact that, for sinusoidal tones, the masked threshold decreases as a function of the frequency separation between the masker and the masked tone. In addition, it has been confirmed repeatedly (see Plomp, 1976) that the low-frequency slope of the masking pattern is considerably steeper than the high-frequency slope. On the assumption that the masking pattern for the detection of amplitude variations resembles that for the detection of

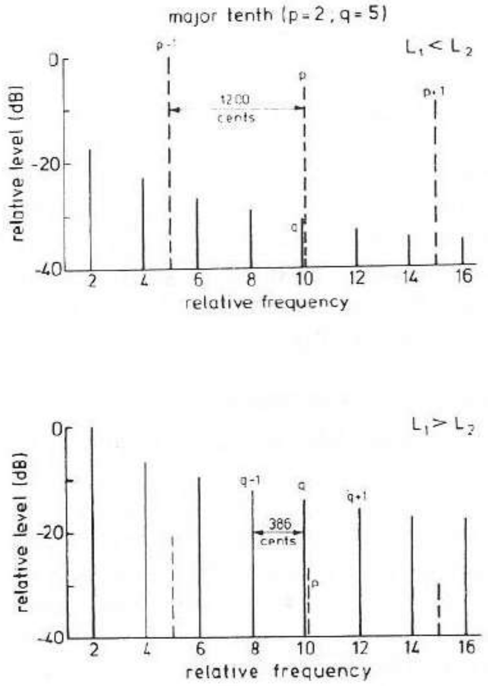

FIG. 6. Presentation of part of the spectral content of a major tenth. The levels of a number of harmonics of the lower tone (continuous lines) and the higher tone (broken lines) are given on a linear frequency scale. Interference of the harmonics p and q results in the sensation of level variations. In this paper we suggest that this sensation is masked partially, especially by high-level harmonics with frequencies lower than those of p and q. In the $L_1 < L_2$ condition (upper part) partial masking may be expected to be due mainly to the first harmonic ($p - 1$) of the higher tone. In the $L_1 > L_2$ condition (lower part) this may be the case for the fourth harmonic ($q - 1$) of the lower tone. This figure illustrates that the interval size between the masker and the partially masked sensation is much larger for the $L_1 < L_2$ condition than for the $L_1 > L_2$ condition.

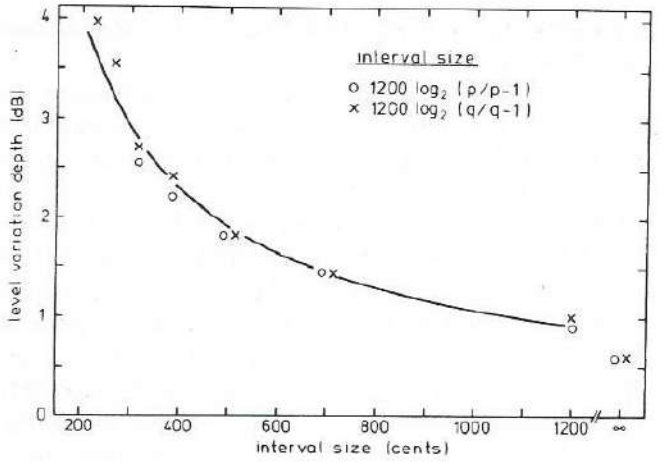

FIG. 7. Thresholds for discrimination between pure and tempered intervals, combined for experiments 1, 2, and 3, and given for the $L_1 < L_2$ and $L_1 > L_2$ conditions, separately. For these conditions, thresholds are plotted as a function of the interval size between the $(p\text{th} - 1)$ and pth harmonics (○) and the $(q\text{th} - 1)$ and qth harmonics (×), respectively. For the conditions in which p and q are equal to 1, the interval size is indicated by ∞. Thresholds are given only for conditions in which all harmonics were present and the beat frequency was 4 Hz.

sinusoids, especially with increasing interval size between the $(p\text{th} - 1)$ and pth, or $(q\text{th} - 1)$ and the qth harmonics, a decrease of the DTs is expected. As a function of this interval size, the DTs are plotted in Fig. 7 for the $L_1 < L_2$ and $L_1 > L_2$ conditions, separately. For the $L_1 < L_2$ conditions, DTs are collapsed over intervals with equal values of p; for the $L_1 > L_2$ conditions, DTs are combined for intervals with equal values of q. For the conditions in which p or q was equal to 1, beats could not be masked by an adjacent low-frequency harmonic. The DTs for these conditions are plotted on the extreme right of the abscissa. In Fig. 7 it can be seen that all DTs can be fitted perfectly by a declining function. (The inserted power function explains 98% of the variance in the mean thresholds.) This finding strongly suggests that the relation between type of interval and sensitivity to tempering is the result of masking.

C. Tonal consonance and discriminability

The degree to which the difference between a pure and a tempered interval can be perceived may be relevant for its function in music. The rank order of intervals in terms of their DTs can, indeed, be compared to the rank order found in the musicological literature throughout the ages. In Western music theory, the intervals with frequency ratios equal to 1:1, 1:2, 2:3, and 3:4, are labeled as perfectly consonant intervals, and those with frequency ratios of 3:5, 4:5, 5:6, and 5:8 are classified as imperfectly consonant intervals. In this paper, tonal or sensory consonance rather than musical consonance is considered (see Cazden, 1945). Tonal consonance refers to the perception of consonance for isolated intervals without any musical context.

Our DTs provide an additional quantitative measure for tonal consonance. Other objective criteria for consonance ranks have been given by models that are based on the overall perceived beats and roughness which result from interfering harmonics. From the computed relative degree of tonal consonance for intervals with complex tones, Plomp and Levelt (1965), Kameoka and Kuriyagawa (1969), and Hutchinson and Knopoff (1978) all found that tonal consonance decreased with increasing sum of p and q. Small differences in rank order obtained in the various studies can be explained by, for example, added refinements within the models, spectral differences between the tones, and whether tonal consonance/dissonance had been computed for pure intervals or for intervals tuned according to equal temperament.

In addition to these psychoacoustical explanations for tonal consonance, a number of psychological studies have tried (1) to describe the perceptual qualities of consonance/dissonance, and (2) to obtain rank orderings for various intervals. With respect to the ways in which musical intervals can be judged, especially the first of two main dimensions mentioned here is relevant.

First, the subjective characteristic "smoothness" seemed to be an independent dimension for judging musical intervals. Results from Guernsey (1928) and Malmberg (1917–1918) showed that "smoothness" was highly dependent on $p + q$: it decreased almost linearly with increasing $p + q$. Although less prominently than in the other two studies, a similar conclusion could be drawn from the results of van de Geer *et al.* (1962).

Second, when intervals were judged to be "consonant," they were also considered "pleasant" (Guthrie and Morrill, 1928), or "beautiful" and "euphonious" (van de Geer *et al.*, 1962). Data from Guernsey (1928) and van de Geer *et al.* (1962) showed that for $p + q \leq 13$, "pleasantness" and "beautifulness" were about the same. However, ratings of these perceptual qualities decreased rapidly for $p + q > 13$. This conclusion holds for both laymen and musicians, trained to various degrees, if the low ratings for the octave and fifth given by the advanced musical performers in Guernsey's study are left out of consideration.

Surprisingly, the two general effects that were found in the studies discussed above did not depend on harmonic spectrum in a systematic way. This may be partly explained by the fact that, more than half a century ago, the experimental conditions were less well controlled. For a number of intervals ranging from the major third up to the octave, however, Kaestner (1909) showed that with complex tones, but not with simple tones, "pleasantness" was rated much higher for simple frequency ratios. Using the nonverbal method of triadic comparisons, the relevance of frequency-ratio complexity in the perception of complex-tone intervals has also been demonstrated by Levelt *et al.* (1966).

D. Tuning procedures and discriminability

Pure consonant intervals can be obtained by freeing them of beats, for example, the fifths in the Pythagorean system or the major thirds in the meantone system for tuning keyboard instruments. In the majority of tuning systems proposed, however, all or at least some of these intervals are tempered (see Barbour, 1951, and Rasch, 1983). In general, for a certain pitch range such as, for example, the octave above middle C, the temperaments are set by using the frequencies of the beats in the fifths and fourths, with the major thirds as checks. With respect to audibility of beats, which, with the degree of tempering used in this paper, can be assumed to have been the main clue for discrimination, these intervals belong to the middle group ($5 \leqslant p + q \leqslant 9$).

In addition, the DTs (see Tables I and II) suggest that to set the temperaments in a reliable way, it is preferable to use twelfths and major tenths, rather than fifths and major thirds, respectively. This finding squares with observations already made by Praetorius and Romieu. According to Praetorius (1619, p. 154) an absolutely pure major third can be obtained much more easily by tuning its octave enlargement, the major tenth. For tuning the organ or the harpsichord, Romieu (1758, p. 501) preferred using the double-octave enlargement of the major third, 1:5, to the major third itself, and, similarly, he preferred the twelfth to the fifth.

When, for a certain pitch range, the temperaments have been set, the remaining pipes or strings are tuned with the help of unisons and octaves. Sensitivity to tempering is highest for these intervals. The beats for tempered unisons at a rate of 4 beats/s are prominent to such an extent (see also Riesz, 1928; Zwicker, 1952) that a successful tuning procedure was based exclusively on this kind of beat: Scheibler's method (1834) of tuning instruments consisted of tuning a series of forks on his tonometer 4 beats/s flatter than the frequencies required. The string or pipe was then tuned sharper than the fork by 4 beats/s.

Intervals with $p + q > 9$ are never utilized in tuning procedures. This can be fully understood from the high discrimination thresholds for these intervals.

VI. GENERAL CONCLUSIONS

(1) From experiment 3, we may conclude that for all intervals investigated, which comprise tones with a complete series of harmonics, discrimination threshold DT is determined by the interference between nearly coinciding harmonics. Because of this, these DTs should be expressed in terms of level-variation depth D.

(2) Combining the results from experiments 1 and 2, we may conclude that the DTs are highly correlated with frequency-ratio complexity: DTs gradually increase (higher value of D) with increasing frequency-ratio complexity, $p + q$.

(3) The extent to which DT depends on $p + q$ decreases with increasing degree of tempering.

(4) The only difference between experiment 1 and 2 was that, in experiment 1, the level of the harmonics decreased at a rate of 6 dB/oct, starting from the fundamental, whereas in experiment 2 all harmonics had the same level. Since the results of experiment 2 were essentially the same as those of experiment 1, we may conclude that the relation between DT and $p + q$ was not simply the result of a specifically chosen spectrum of the tones, but that this relation also holds for tones with strongly deviating spectral-envelope slopes.

(5) Apparently, the simple relation between DT and $p + q$ cannot be explained by absolute level differences between the harmonics within the various first pairs of nearly coinciding harmonics. Detailed analysis of the results, taking into account differences between the conditions in which either the lower or the higher tone was attenuated led us to hypothesize that the established relationship between DT and $p + q$ is the result of masking.

(6) The rank order of sensitivity to tempering, as obtained with the various intervals investigated in our study, resembles that predicted by various models on tonal consonance, based on overall perceived beats and roughness resulting from interfering harmonics. Moreover, the obtained rank order is similar to that obtained when consonant intervals are judged with respect to their degree of "smoothness." These similarities support our notion that DT provides an additional quantitative measure of tonal consonance.

(7) The observation that our results are consistent with the role which various intervals play in tuning procedures is worthy of notice.

ACKNOWLEDGMENTS

Part of this research was supported by Grant 15-29-05 from the Netherlands Organization for the Advancement of Pure Research (ZWO). The authors are grateful to Rudolf Rasch for his stimulating discussions which led to the initiation of this series of experiments. In addition, the authors are indebted to Reinier Plomp, to Ino Flores d'Arcais, and to Rudolf Rasch for their comments on an earlier draft of this paper.

Barbour, J. M. (1940). "Musical logarithms," Scr. Math. 7, 21–31.

Barbour, J. M. (1951). *Tuning and Temperament* (Michigan State College, East Lansing, reprinted 1972, Da Capo, New York).

Cazden, N. (1945). "Musical consonance and dissonance: A cultural criterion," J. Aesthe. Art Crit. 4, 3–11.

Feth, L. L. (1974). "Frequency discrimination of complex periodic tones," Percept. Psychophys. 15, 375–378.

Feth, L. L., and O'Malley, H. (1977). "Two-tone auditory spectral resolution," J. Acoust. Soc. Am. 62, 940–947.

Feth, L. L., O'Malley, H., and Ramsey, J. W. (1982). "Pitch of unresolved, two-component complex tones," J. Acoust. Soc. Am. 72, 1403–1412.

Geer, J. P. van de, Levelt, W. J. M., and Plomp, R. (1962). "The connotation of musical consonance," Acta Psychol. 20, 308–319.

Guernsey, M. (1928). "The role of consonance and dissonance in music," Am. J. Psychol. 40, 173–204.

Guthrie, E. R., and Morrill, H. (1928). "The fusion of non-musical intervals," Am. J. Psychol. 40, 624–625.

Hartmann, W. M., and Hnath, G. M. (1982). "Detection of mixed modulation," Acustica 50 (5), 297–312.

Helmholtz, H. L. F. von (1877). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (F. Vieweg & Sohn, Braunschweig, Germany), 4th ed., translated by A. J. Ellis, *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (Dover, New York, 1954).

Hutchinson, W., and Knopoff, L. (1978). "The acoustic component of Western consonance," Interface 7, 1–29.

Jeffress, L. A. (1968). "Beating sinusoids and pitch changes," J. Acoust. Soc. Am. 43, 1464.

Kameoka, A., and Kuriyagawa, M. (1969). "Consonance theory, Part II: Consonance of complex tones and its calculation method," J. Acoust. Soc. Am. 45, 1460–1469.

Kaestner, G. (1909). "Untersuchungen über den Gefühlseindruck unanalysierter Zweikiänge," Psychol. Stud. 4, 473–504.

Kuile, Th. E. ter (1902). "Einfluss der Phasen auf die Klangfarbe," Arch. Ges. Physiol. 89, 332–426.

Levelt, W. J. M., Geer, J. P. van de, and Plomp, R. (1966). "Triadic comparisons of musical intervals," Brit. J. Math. Stat. Psychol. 19, 163–179.

Levitt, H. (1971). "Transformed up–down methods in psychoacoustics," J. Acoust. Soc. Am. 49, 467–477.

Maiwald, D. (1967). "Ein Funktionsschema des Gehörs zur Beschreibung der Erkennbarkeit kleiner Frequenz- und Amplitudenänderungen," Acustica **18**, 81–92.

Malmberg, C. F. (**1917–1918**). "The perception of consonance and dissonance," Psychol. Mono. **25**, 93–133.

Plomp, R. (**1976**). *Aspects of Tone Sensation* (Academic, London).

Plomp, R., and Levelt, W. J. M. (**1965**). "Tonal consonance and critical bandwidth," J. Acoust. Soc. Am. **38**, 548–560.

Praetorius, M. (**1619**). *Syntagma Musicum, Tomus Secundus: De Organographia* (Holwein: Wolfenbüttel, Germany).

Rasch, R. A. (**1983**). "Description of regular twelve-tone musical tunings," J. Acoust. Soc. Am. **73**, 1023–1035.

Rasch, R. A. (**1984**). "Theory of Helmholtz-beat frequencies," Music Percept. **1** (3), 308–322.

Riesz, R. R. (**1928**). "Differential intensity sensitivity of the ear for pure tones," Phys. Rev. **31**, 867–875.

Romieu, J. B. (**1758**). "Mémoire théorique et pratique sur les systèmes tempérés de musique," Mem. Acad. R. Sci. Annee **1758**, 483–519.

Scheibler, H. (**1834**). *Der physikalische und musikalische Tonmesser* (Bädeker, Essen, Germany).

Schöne, P. (**1979**). "Messungen zur Schwankungsstärke von amplitudenmodulierten Sinustönen," Acustica **41**, 252–257.

Vos, J. (**1982**). "The perception of pure and mistuned musical fifths and major thirds: Thresholds for discrimination, beats, and identification," Percept. Psychophys. **32** (4), 297–313.

Vos, J. (**1984**). "Spectral effects in the perception of pure and tempered intervals: Discrimination and beats," Percept. Psychophys. **35** (2), 173–185.

Wever, E. G. (**1929**). "Beats and related phenomena resulting from the simultaneous sounding of two tones—I," Psychol. Rev. **36**, 402–418.

Winer, B. J. (**1970**). *Statistical Principles in Experimental Design* (McGraw-Hill, London), International student ed.

Zwicker, E. (**1952**). "Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones," Acustica **2**, 125–133.

Perception & Psychophysics
1985, 37, 507-514

5

The effect of fundamental frequency on the discriminability between pure and tempered fifths and major thirds

JOOS VOS and BEN G. van VIANEN
Institute for Perception TNO, Soesterberg, The Netherlands

Thresholds were determined for discrimination between pure and tempered musical intervals (DTs) consisting of simultaneous complex tones. Tempered intervals are characterized by small frequency differences between those harmonics that coincide in pure intervals. Interference of these nearly coinciding harmonics gives rise to the perception of beats. Two parameters of beats were varied independently: (1) beat frequency and (2) depth of level variation (temporal envelope) as a measure of DT. DTs were determined for musical fifths and major thirds at a tone duration of 0.5 sec. The geometric mean of the fundamental frequency of the tones was varied in octave steps from 92.5 to 740 Hz. The beat frequency of the intervals was varied within a range of 2 to 64 Hz. The major result of our experiment was that, for a frequency range comprising 3 to 4 octaves, DTs do not depend on the frequency of the fundamentals, provided that the DTs are ordered according to beat frequency. When tempering (T) is expressed in cents and only those conditions are considered in which the degree of T corresponds to that found in the relevant tuning systems, it can be concluded that perceptual differences between tempered and pure fifths are increasingly irrelevant for tones lower than about middle C (262 Hz). Tempered major thirds, even those in a compromise tuning system such as that proposed by Silbermann (T=6 cents), can be discriminated from pure major thirds for all fundamental frequencies that are relevant in musical composition.

Discrimination between pure and tempered musical intervals consisting of simultaneous complex tones has been investigated in previous studies. Tempered intervals are characterized by small frequency differences between harmonics that coincide in pure intervals. Interference of these nearly coinciding harmonics gives rise to the perception of beats. Thresholds for discrimination (DTs) are expressed in terms of the depth of the level variation (temporal envelope) of these beating harmonics.

The main results obtained in one of the first studies (Vos, 1982) were that the DTs are higher (lower sensitivity) for major thirds than for fifths, are highest for low degrees of tempering, and decrease with increasing tone duration. In a second paper, Vos (1984) showed that for moderately tempered fifths, the value of DT depends mainly on the degree of interference of the first pair of nearly coinciding harmonics, whereas for major thirds, interference of other harmonics may play a role as well. For a large set of intervals varying in size between and including unison and twelfth, Vos and van Vianen (1985) showed that DTs gradually increase with the sum of the integers in the frequency ratio of the pure interval. In addition, they concluded that for all intervals investigated, DT is determined by the interference between nearly coinciding harmonics. DT, providing an objective measure for sensitivity to tempering, was related to tonal consonance and tuning procedures (Vos & van Vianen, 1985).

This research was supported by Grant 15-29-05 from the Netherlands Organization for the Advancement of Pure Research (ZWO). The authors are indebted to Reinier Plomp, to Ino Flores d'Arcais, and to Rudolf Rasch for their comments on an earlier draft of this paper.

Requests for reprints may be sent to Joos Vos, Institute for Perception TNO, Kampweg 5, 3769 DE Soesterberg, The Netherlands.

In these experiments, DTs were determined for only a restricted frequency range: for fifths and major thirds, the fundamental frequencies were between 261.6 and 523.2 Hz, that is, using staveless notation (Young, 1939), between C4 and C5. It was the aim of the present research to determine to what extent previous findings might be generalized to a much wider frequency range. Since most music comprises several octaves, knowledge of these results may be significant for students in music perception.

Physical and Perceptual Aspects

For a pure interval in which, by definition, the ratio of the fundamental frequency, f_1, of the lower tone (tone 1) and the fundamental frequency, f_2, of the higher tone (tone 2) is equal to p:q, with p and q ($p<q$) being small integers, the harmonics with frequencies of npf_2 and nqf_1 Hz ($n=1, 2, \ldots$) coincide. For two simultaneous complex tones, a moderately tempered interval is characterized by small frequency differences between the harmonics that coincide in pure intervals. Perceptually, the most relevant aspect of the interference of these nearly coinciding harmonics is variation in amplitude. The depth (in decibels) of the level variation (temporal envelope) of two interfering harmonics with amplitudes a_1 and a_2 is given by

$$D = 20 \log \left| \frac{a_1+a_2}{a_1-a_2} \right| = 20 \log \left(\frac{10^{\Delta L/20}+1}{10^{\Delta L/20}-1} \right), \quad (1)$$

in which ΔL is the absolute level difference between a_1 and a_2 in decibels [$\Delta L = |20 \log(a_1/a_2)|$]. D approaches ∞ for $a_2 \rightarrow a_1$. Because discriminability between pure and tempered intervals is determined by the interference of nearly coinciding harmonics (Vos & van Vianen, 1985), DTs will be expressed in terms of D.

For the nth pair of interfering harmonics, the frequency of the level variation is equal to

$$f_{bn} = |npf_2 - nqf_1|. \quad (2)$$

For tones with many relatively strong harmonics, Vos (1984) showed that the dominantly perceived beat frequency was equal to that of the first pair of nearly coinciding harmonics (f_b). In our experiment, the amount of tempering will therefore be expressed in terms of f_b.

Tuning systems, however, are described differently (see, e.g., Rasch, 1983): tempering (T) of the various intervals is expressed as the difference between the logarithmically defined interval size of the tempered interval and that of the pure interval. The relationship between f_b and T (in cents) is given by

$$f_b = |qf_1(2^{T/1200}-1)|,$$

and, since $(2^{T/1200}-1) \sim \ln 2^{T/1200} = T/1731$, it follows that

$$T = (1731 \cdot f_b)/(qf_1). \quad (3)$$

In Figure 1, T is plotted for the fifth and the major third, separately, as a function of f_1 with f_b as parameter. For the octave range between C2 and C3, even relatively slow beats correspond to large temperings, especially for the fifth. For the equally tempered major third (T = 13.7 cents), high values of f_b are obtained for the octave range between C5 and C6.

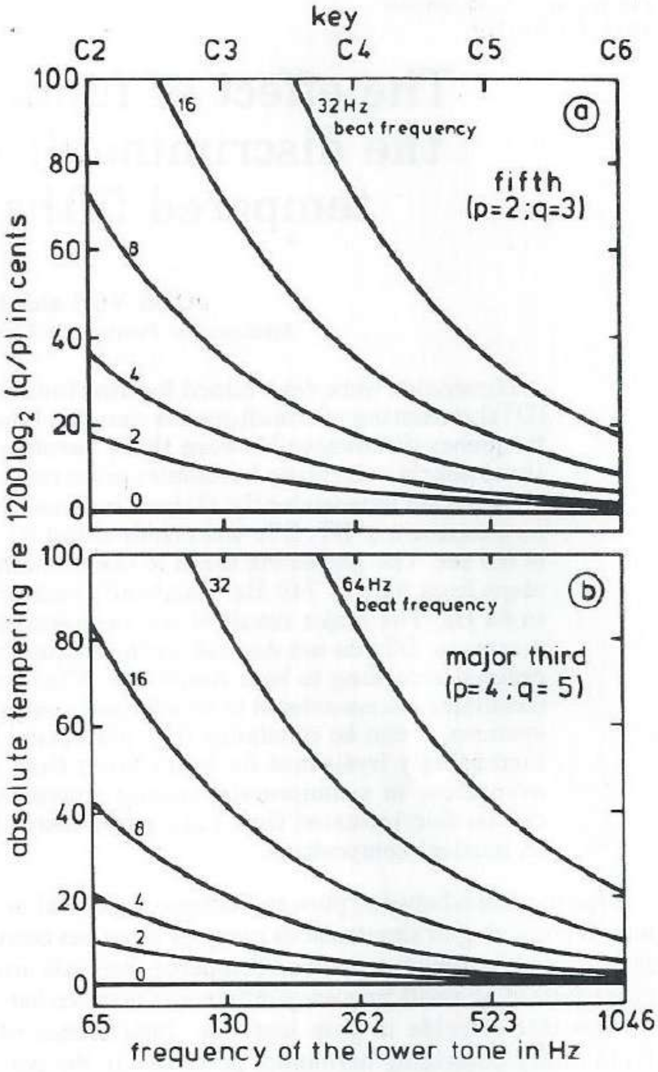

Figure 1. Absolute tempering according to Equation 3 as a function of the fundamental frequency of the lower tone, with beat frequency $|pf_2 - qf_1|$ as a parameter, given for the fifth and the major third in panels a and b, respectively.

Critical Bandwidth

In order to constitute a relatively smooth-sounding musical interval, the frequencies of two simultaneous tones have to be separated by a certain minimum value. This minimal frequency distance is required because the resolving power of the ear is limited. This resolving power may be expressed in the critical bandwidth of the ear (see Scharf, 1970, for a survey). Thus, when the frequencies of two simultaneous tones are within a single critical band, the result is an overall roughness sensation. This sensation becomes more and more manifest with decreasing frequency distance of the tones, up to a maximum at about 25% to 40% of the bandwidth; for smaller frequency differences, the beat sensation decreases to zero when the tones coincide. Comparison of estimates of critical bandwidth that are based on various experimentally investigated phenomena, such as loudness summation, masking, and frequency analysis, indicates that the values may differ by a factor of 2 (see Plomp, 1976).

With respect to discrimination between pure and tempered intervals, estimation of the critical bandwidth from experiments in which presence versus absence of beats or roughness was investigated seems to be most relevant here. Such experiments have been carried out by Cross and Goodwin (1893), by Mayer (1894), and by Plomp and Steeneken (1968). The results of these three studies are very similar; with the help of curvilinear regression (least squares solution), we found that critical bandwidth (CBW), in terms of the frequency distance between its lower and higher bounds—being dependent on the mean frequency ($\bar{f}$) of these two bounds—can be described adequately ($r^2 = 0.98$) by

$$CBW = 1.22\ (\bar{f})^{0.69}. \quad (4)$$

Comparison of the width of the critical band with that of musical intervals is facilitated by the expression of CBW in cents (CBW′). CBW′, derived from the data reported by Cross and Goodwin (1893), Mayer (1894), and Plomp and Steeneken (1968) is plotted in Figure 2 as a function of $\bar{f}$ for the three studies separately. Based on all data points in Figure 2, CBW′ can be represented excellently ($r^2 = 0.92$) by

$$CBW' = 2371\ [\log_{10}(\bar{f})]^{-2.06}. \quad (5)$$

For the octave range between C1 and C2, CBW′ may be expected to be larger than a musical fifth. Extending the frequency range investigated in our experiment to frequencies lower than C2 would not be interesting for the perception of music. For the octave between C2 and C3, CBW′ is about 270 to 150 cents larger than a pure major third. Because of this, discrimination thresholds will not be determined for major thirds with fundamental frequencies below 130 Hz. Interval distributions computed from the chords of a number of musical compositions reveal that major thirds and fifths are avoided for lower octaves and therefore strongly suggest that critical bandwidth plays an important role in music (Plomp & Levelt, 1965).

Sensitivity to Beats as a Function of Frequency

It has been shown (Vos, 1982) that, for moderately tempered intervals ($T < 20$ cents), thresholds for discrimination are similar to thresholds for the detection of beats. This same study demonstrated that when the tempered intervals deviated from purity by more than 20 to 30 cents, perception of the difference between the size of the tempered interval and that of the pure interval could also become a relevant cue for discrimination. Since, in the present experiment, absolute tempering was smaller than 20 cents in more than 75% of the conditions, investigation of whether DT depends on fundamental frequency or, say, mean frequency of the nearly coinciding harmonics seems to be of primary concern here.

To our knowledge, no experimental results on the relation between fundamental frequency and threshold for the perception of beats that can be heard in tempered intervals consisting of two complex tones have been reported. Only the level-variation thresholds for beats caused by the addition of two sinusoids of nearly the same frequency (Riesz, 1928) and for beats resulting from amplitude modulation (AM) of one sinusoid (Zwicker, 1952a) have been studied systematically for various (fundamental) frequencies.

For a beat frequency of 3 Hz, Riesz (1928) varied the mean frequency of the interfering sinusoids from 35 Hz up to 10 kHz at various sensation levels of the tones. An effect of mean frequency on the threshold for level variation was found at very low sensation levels. For the fre-

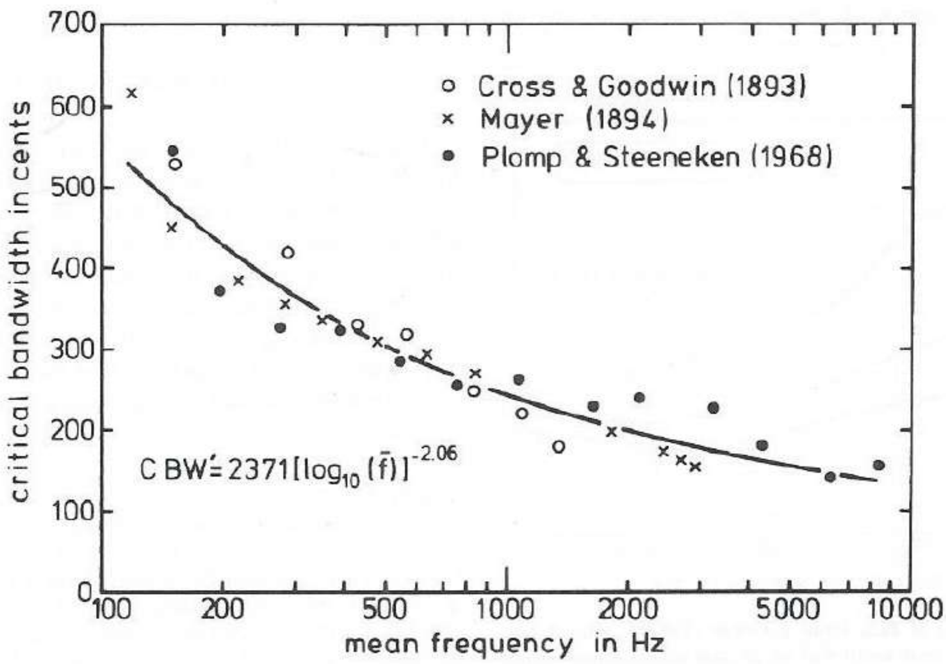

Figure 2. Estimates of critical bandwidth as a function of mean frequency of two simultaneous tones that are just separated enough to be free of beats and roughness, adopted from three studies. Computation of the inserted regression function was based on all data points ($r^2 = 0.92$).

quency range relevant to the present investigation and at sensation levels exceeding 20 dB, however, no effect could be found.

Zwicker (1952a) varied carrier frequency of AM and FM sinusoids in octave steps from 65 to 8000 Hz. Since he was primarily interested in relative differences between thresholds for the detection of amplitude and frequency modulation, Zwicker never reported an effect of carrier frequency on these absolute thresholds. Our detailed analysis of his data (Zwicker, 1952b), however, revealed that there is a moderate, but consistent, effect of carrier frequency on the threshold for the detection of amplitude modulation. Averaged across modulation frequencies of 2, 4, and 8 Hz, the thresholds are plotted in Figure 3 for three loudness levels of the tones, separately. As can be seen in Figure 3, the threshold decrease with increasing frequency of the carrier is most prominent at a low loudness level. The threshold increase for carrier frequencies higher than 4 kHz was confirmed in an additional experiment (Zwicker, 1952b) in which frequencies higher than 8 kHz were also included.

Carrier frequency of AM sinusoids was also varied by Terhardt (1968). At a modulation frequency of 5 Hz, he found that the perceived strengths of beats at carrier frequencies of 125, 500, and 2000 Hz were about equal. For higher modulation frequencies, Terhardt's data are not conclusive on the effect of carrier frequency on perceived strength of the beats because, in his pertinent experiment, carrier frequency and modulation frequency were confounded.

In conclusion, the results discussed above are mutually inconsistent with respect to the effect that mean frequency of the interfering sinusoids may have on sensitivity to beats.

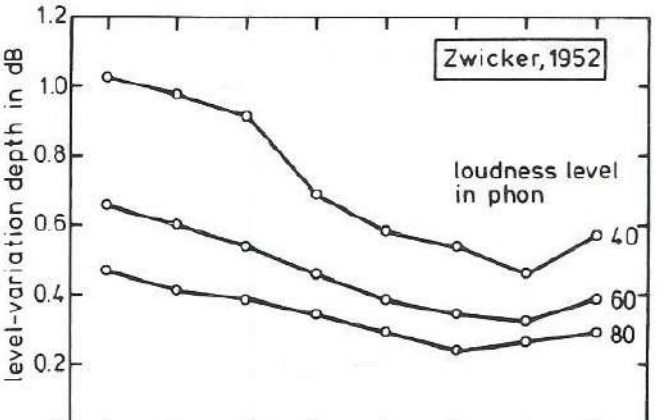

Figure 3. Reanalysis of data from Zwicker (1952b), who determined thresholds for the detection of amplitude modulation. These thresholds were converted into the depth of the level variation (Equation 1). Averaged across modulation frequencies of 2, 4, and 8 Hz, they are plotted separately as a function of carrier frequency for three loudness levels of the tones.

Masking and Overall Roughness in Complex Tones

For complex tones, perception of level variations may be obscured by the masking effect of adjacent components (see Vos & van Vianen, 1985). Zwicker (1954) has shown that, for noise bands masked by two sinusoidal tones at frequencies below and above the noise band, the masked threshold remains constant up to a certain frequency separation, but decreases progressively beyond that value. The interval size in cents between the two sinusoids up to which the masked threshold remains at maximum decreases as a function of the mean frequency of the maskers. On the assumption that the masking pattern for the detection of level variations resembles that for the detection of a band of noise, fundamental frequency, which will be varied in our experiment, may have an effect on discrimination between pure and tempered intervals.

A second factor that may be relevant is the overall roughness of the pure intervals: harmonics that fall within one single critical band are responsible for a roughness sensation. It may be expected that relatively rough pure intervals are less well differentiated from, again relatively rough, tempered intervals than are smooth intervals. For example, the intervals between the harmonics of the lower and the higher tones of the fifth are, up to the fourth harmonic of the lower tone, larger than or equal to about 500 cents, whereas for the major third this value is about 100 cents. Because estimates of the critical bandwidth, based on the presence versus absense of beats or roughness (see Figure 2) strongly depend on the frequency of the sinusoids, an additional effect of frequency on the discrimination thresholds is not precluded a priori, especially for the major thirds.

THE EXPERIMENT

The aim of the present experiment was to determine to what extent previous findings on discriminability between pure and tempered intervals might be applied to a much wider range of fundamental frequencies. The frequency range in the experiment corresponds to that given in Figure 1. It has been shown above that, due to the limited resolving power of the ear, extending the range to frequencies lower than C3 (for the major thirds) or C2 (for the fifths) would not be relevant for the perception of music.

Method

Stimuli and Apparatus. The stimuli were intervals consisting of two simultaneous complex tones, 1 and 2. The frequency ratio $f_1:f_2$ was equal to or slightly different from p:q. We presented fifths (p=2, q=3) and major thirds (p=4, q=5). The overall sound level of tones 1 and 2 together was about 85 dB SPL. The duration of the tone bursts was 0.5 sec. The complex tones consisted of 20 harmonics with amplitude a_n proportional to 1/n. A detailed description of the waveforms and the temporal envelopes of the tones, as well as the way in which the level-variation depth of the beating harmonics was manipulated (Equation 1), was given in a previous paper (Vos, 1982, Experiment 1). Also, details on the apparatus used are given in full in that report.

Subjects. Six subjects, students of either the Institute of Musicology or the Conservatory at Utrecht, were tested over four or five half-day sessions. Their hearing thresholds were determined with pure tones between 125 and 8000 Hz. None of them had a hearing loss larger than 15-dB hearing level (ISO 38 g)in any part of the audiogram. The subjects were paid for their participation.

Experimental design. The independent variables were: (1) interval (fifth and major third); (2) fundamental frequency [mean central frequency, $\bar{f}_c$ (= the geometrical mean $(\bar{f}_1\bar{f}_2)^{1/2}$), of the pure intervals could be 92.5, 185.0, 370.0, and 740.0 Hz]; (3) attenuation of tone 1 versus attenuation of tone 2; and (4) degree of tempering (beat frequency, f_b, of the first pair of nearly coinciding harmonics could be 2, 4, 8, 16, 32, and 64 Hz). For the lowest frequency range ($\bar{f}_c$ = 92.5 Hz), the fundamental frequencies of the major third were clearly within the same critical band; therefore, no thresholds were determined for these conditions. Beat frequencies lower than 2 Hz were not presented because, for a tone duration of 0.5 sec, these beat frequencies would result in stimuli that contained less than one beating period. In addition, temperings larger than ±35 cents were not investigated. This means that, with respect to both fundamental frequency and beat frequency, we have an incomplete factorial design. For the fifth, thresholds were determined for the following 14 combinations of $\bar{f}_c$ and f_b: $\bar{f}_c$ = 92.5 Hz, f_b = 2 and 4 Hz; $\bar{f}_c$ = 185 Hz, f_b = 2, 4, and 8 Hz; $\bar{f}_c$ = 370 Hz, f_b = 2, 4, 8, and 16 Hz; and $\bar{f}_c$ = 740 Hz, f_b = 2, 4, 8, 16, and 32 Hz. For the major third, there were 15 combinations: $\bar{f}_c$ = 185 Hz, f_b = 2, 4, 8, and 16 Hz; $\bar{f}_c$ = 370 Hz, f_b = 2, 4, 8, 16, and 32 Hz; and $\bar{f}_c$ = 740 Hz, f_b = 2, 4, 8, 16, 32, and 64 Hz. For each of these 29 conditions, the level of tone 1 (L_1) could be decreased relative to that of tone 2 (L_2), and vice versa. For the resulting 58 conditions, all of which were presented to the six subjects, two thresholds were determined.

Procedure. The task for the subject was to discriminate between pure and tempered musical intervals in a 2-AFC paradigm. The tempered interval was presented at random in the first or the second interval of a trial. Between the intervals, there was a silent period of 1.5 sec. By means of two response buttons, the subject indicated whether the first or the second interval was tempered. It was clear to the subjects that, in addition to beats, differences between the sizes of the tempered and the pure intervals (stretched or compressed) could also provide a valid cue for discrimination. Feedback was given by a light above the correct button. The next trial was presented 1.7 sec after the response to the preceding trial. Discrimination thresholds (DTs) were determined by means of an adaptive procedure, in which the absolute difference between L_1 and L_2, ΔL, was increased by 2 dB after three correct responses and decreased by the same amount after a single incorrect response. Each reversal of the direction of change in ΔL over successive trials was counted as a turnaround. There were eight turnarounds. In the first run (a number of successive trials), however, ΔL, which was 0 dB at the very first trial, was increased by 3 dB after each correct response. The DT was defined as the mean value of ΔL on the trials that resulted in the last six turnarounds, corresponding with 79.4% correct responses in a nonadaptive 2-AFC procedure (Levitt, 1971). This value was converted into a D value (Equation 1). The standard deviation of the values of ΔL for the last six turnarounds was considered as a measure of the reliability of DT. Standard deviations greater than 5 dB were exceptional. If they occurred, the series was repeated until results with a standard deviation less than 5 dB were obtained. Three subjects were presented first with the block that comprised the 28 series of the fifth; the remaining three subjects were presented first with the block that comprised the 30 series of the major third. When thresholds had been determined for all 58 series, the blocks were presented for the second time. Presentation order of the series was randomized within each block. To avoid the use of absolute pitch as a cue, fundamental frequencies were varied from trial to trial in such a way that f_c of the pure interval could be equal to 92.5, 185, 370, or 740 Hz ±25, ±75, ±125, ±175, or ±225 cents, with equal probability of occurrence. Details about the computation of f_1 and f_2 of the pure and tempered intervals are given in Vos (1982, 1984). For each trial, the probability that either f_1 or f_2 had been altered and that the tempered interval had been stretched ($pf_2 - qf_1 > 0$) or compressed ($pf_2 - qf_1 < 0$) was 50%. The data were combined for stretched and compressed intervals. A session comprised about 30 series. The subjects received a few training series before a new experimental block (fifth or major third) started. In the $\Delta L < 10$ dB conditions, L_1 and L_2 were attenuated to a small extent to satisfy our criterion that the overall level of the interval should remain at 85 dB SPL.

Results

The DTs, averaged across subjects, replications, and the $L_1 < L_2$ and $L_1 > L_2$ conditions, are presented in Figure 4 as a function of beat frequency. In this figure, DTs are given separately for the fifth and major third in panels a and b, respectively, with mean central frequency as a parameter. DTs determined the second time were, on the average, 1 dB lower than those determined the first time, the difference being statistically significant. However, because replication did not interact significantly with the other variables, it will be treated as a within-cell variance.

Differences between the DTs which were due to whether D was manipulated in the $L_1 < L_2$ or the $L_1 > L_2$ conditions did not depend on mean central frequency. Because large differences between L_1 and L_2 are only rarely found in performed music, and differences between the DTs for the $L_1 < L_2$ and $L_1 > L_2$ conditions have been described at length in Vos (1982) and in Vos and van Vianen (1985), this variable will not be considered any further here.

As can be seen in Figure 4a, for the fifth, the DTs from the various frequency ranges, when plotted as a function of f_b, are very similar. Although there is more scatter in the data and the DTs for the mean central frequency of 740 Hz tend to be higher than those for the other frequencies, the DTs for the major thirds were also not significantly affected by the frequency of the fundamentals. For both intervals, the DTs decreased when beat frequencies increased from 2 to 8 Hz. It can also be seen in Figure 4 that, in general, the DTs for the fifths were considerably lower than those for the major thirds. The significance of these effects was tested by subjecting various sets of data to ANOVAs. For example, two sets of threshold data consisted of a 6 (subjects) × n (mean central frequency) × m (beat frequency) × 2 (interval) × 2 ($L_1 < L_2$ vs. $L_1 > L_2$) factorial design, all repeated measures and with replication as within-cell variance. Set I comprised three different mean central frequencies and the three lowest beat frequencies, whereas Set II comprised mean central frequencies of 370 and 740 Hz and beat frequencies of 2, 4, 8, and 16 Hz. The two ANOVAs showed (see Table 1 for F values and significance levels) that DTs did not depend on mean central frequency, and that the effects of beat frequency and musical interval were significant. Furthermore, neither first-order nor second-order interaction effects with mean central frequency were sta-

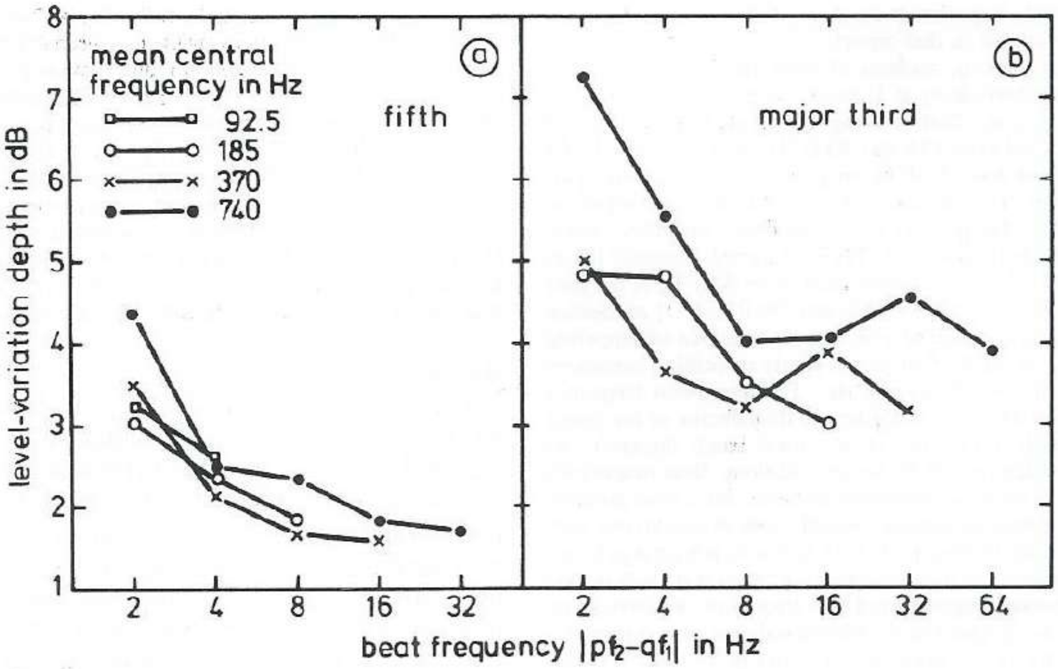

Figure 4. Discrimination thresholds as a function of beat frequency, with mean central frequency as a parameter, given for the fifth and the major third in panels a and b, respectively.

tistically significant. The highly significant differences between subjects tested against within-cell variance are due to the DTs for one subject, whose thresholds were considerably higher than those of the other five subjects. The results, summarized in Table 1, were confirmed by two ANOVAs that were performed on additional sets of DTs for either the fifth or the major third.

DISCUSSION

Measures of the Amount of Tempering

The major result of our experiment was that, over a frequency range of 4 octaves, discriminability between pure and tempered intervals did not depend on the frequency of the fundamentals. It should be emphasized, however, that the DTs shown in Figure 4 were ordered according to beat frequency. As discussed above, both in musical and perceptual contexts, it may be more appropriate to present the data as a function of tempering, T, in cents (Equation 3). Therefore, the DTs from Figure 4 were replotted as a function of T in Figure 5.

Table 1
Summary of the F-Values and Corresponding Probabilities of Two ANOVAs, One for Each Data Set (see Text)

	Data Set	
	I	II
A	$F(2,10) = 2.90$ n.s.	$F(1,5) = 2.89$ n.s.
B	$F(2,10) = 7.24$ $p < .01$	$(F(3,15) = 4.39$ $p < .02$
C	$F(1,5) = 18.2$ $p < .01$	$F(1,5) = 19.7$ $p < .007$
A × B	n.s.	n.s.
A × C	n.s.	n.s.
B × C	n.s.	n.s.
A × B × C	n.s.	n.s.
S	$F(5,216) = 26.8$ $p < .00001$	$F(5,192) = 21.5$ $p < .00001$

Note—A = mean central frequency, B = beat frequency, C = musical interval, and S = subjects.

Fifth

Figure 5a, in which the DTs for the fifths are given, suggests that tempered fifths in the meantone tuning system ($T = -5.4$ cents) and in equal temperament ($T = -2.0$ cents) can be discriminated from pure fifths, as long as the tones are not lower than about middle C. Thus, within the range of relevant tuning systems, and for intervals with a tone duration that has a high probability of occurrence in many musical compositions, perceptual differences between tempered and pure intervals are expected to be increasingly irrelevant for tones lower than about C4.

The DTs in Figure 5a suggest that the inferred effect of frequency for small temperings is maintained for temperings larger than 15 cents. This was tested by subjecting all DTs with mean absolute temperings of 15.5 and 31 cents to an ANOVA. The effect of fundamental frequency was, indeed, significant [$F(3,15) = 8.1$, $p < .002$]. A Newman-Keuls paired-comparison test (Winer, 1970) showed that only the DTs for the fifths between C2 and C3 were significantly higher than those with higher fundamental frequencies ($\alpha \geq .05$).

Major Third

Comparison of the DTs for the fifth with those for the major third reveals, in accordance with Equation 3, that, for corresponding mean central frequencies, the smallest temperings presented for the major third are nearly twice

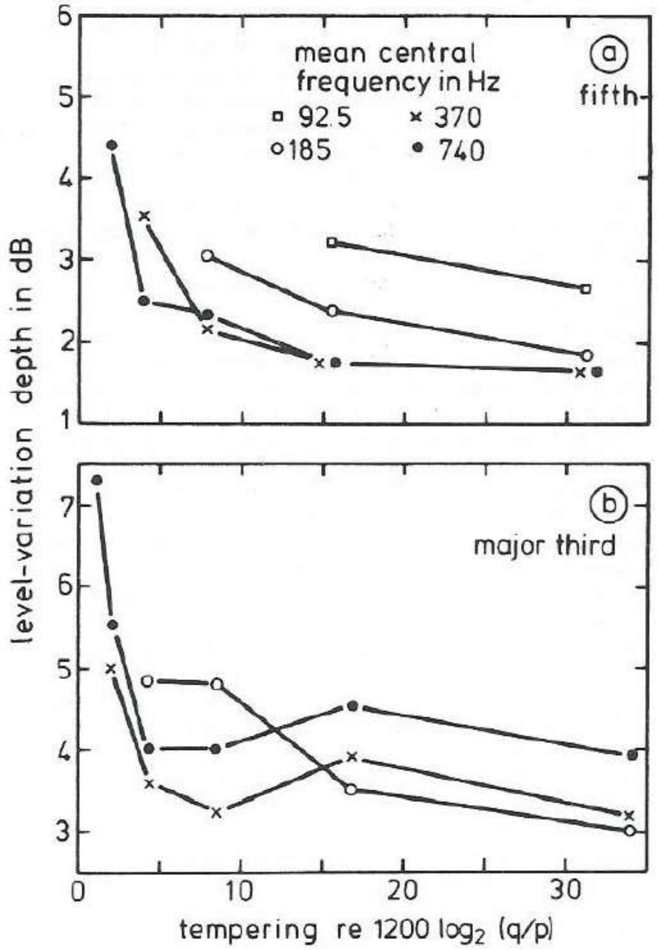

Figure 5. Thresholds for discrimination between pure and tempered intervals at the different beat frequencies, as given in Figure 4, replotted as a function of the corresponding mean absolute values of T in cents. Thresholds are given for the fifth and the major third in panels a and b, respectively.

as small as those for the fifths. In addition, Figure 5b indicates that tempered major thirds in the Pythagorean tuning system (T = 21.5 cents) and even those in compromise tuning systems such as, for example, that proposed by Silbermann (T = 5.9 cents) and equal temperament (T = 13.7 cents), can be discriminated from pure major thirds for all fundamental frequencies that are relevant in musical composition.

It can be seen in Figure 5b, however, that discriminability is very low for small temperings. For temperings of between 3 and about 13 cents, the data suggest that DTs at a mean central frequency of 185 Hz are higher than those at the other frequencies, whereas the reverse seems to be the case for temperings higher than 13 cents. This interaction effect, however, was not significant ($p > .07$).

Implications for Intervals Played on Keyboard Instruments

If we assume that the amount of tempering does not depend on the central frequency of the tones constituting the interval, our experimental results may be related to the perception of intervals played on fixed-pitch keyboard instruments. For musical instruments that produce tones with strictly harmonic partials, such as the pipe organ, the assumption of equality of tempering for the whole keyboard seems to be justified.

For a restricted frequency range of 2 octaves between C4 and C6, fundamental frequencies for two organ stops were reported by Grützmacher and Lottermoser (1935). From their data it can be concluded that although the temperaments set did not correspond exactly to equal temperament, the standard deviations (SDs) of the differences were, in general, not greater than about 2 cents. More importantly, a trend towards either stretched or compressed tuning could not be found.

From more extended frequency measurements, performed on four different pipe organs for the octaves between C2 and C6 (Vos, 1981), it was also concluded that convincing evidence for stretched or compressed tuning could be found for neither 4-ft nor 8- or 16-ft stops. For a set of 15 stops, the slopes of the linear functions that fitted the deviation of the successive tones from equal temperament were very flat ($M = -0.3$ cents/octave; $SD = 1.1$ cents/octave) and the variance (r^2) explained by the linear functions was low ($M = 0.10$; $SD = 0.11$).

Similar frequency measurements performed on pianos and grand pianos indicate that our experimental results can be applied only to a limited part of the keyboards of pianos. For a number of grand pianos, Grützmacher and Lottermoser (1935) found that for the keys between C4 and C6, a trend towards stretched or compressed tuning could not be found. By investigating the whole frequency range of pianos and grand pianos, Railsback (1938; see Schuck & Young, 1943, for more details) and Meinel (1954, 1957) found that, indeed, for tones between about C4 and C6, the frequencies hardly deviated from equal temperament in a systematic way. For tones higher than about C6 and lower than C3, however, there was in most cases a tendency towards stretched tuning, extending to about 4-10 cents/octave. This phenomenon has been related to the inharmonicity of piano strings (Schuck & Young, 1943; Young, 1954; see also Rasch & Heetvelt, in press).

In conclusion, the experimental results described above can be applied to intervals played on the pipe organ. The same probably holds for the intervals played on the harpsichord, because the inharmonicity of its long and thin strings is very low. For the frequency range between C3 and about C6, the results can also be applied to the intervals played on the piano. Stretched tuning, found for the extreme upper and lower frequency ranges of the piano, may result in fifths and major thirds that are, relative to a tuning system with pure octaves, less tempered. For octaves that are stretched by 3.3 cents, for example (see Kolinski, 1959), pure fifths in a regular tuning system result in major thirds that are only 1.2 cents larger than those in equal temperament. Similarly, fifths that are tempered by 2 cents result in major thirds that deviate 6.8

cents from the pure major third. For intervals consisting of tones with inharmonic partials, the relation between tempering and beat frequency is different from that given for tones with strictly harmonic partials (Equation 3). It has recently been suggested (Rasch & Heetvelt, in press), however, that at least for the middle and descant registers of the piano, both the increase of inharmonicity with increasing frequency and the stretched tuning of the octaves lead to beat frequencies that are not much different from those expected for intervals with harmonic tones.

GENERAL CONCLUSIONS

Over a frequency range of 4 octaves between C2 and C6, thresholds for discrimination between pure and tempered fifths do not depend on the frequency of the fundamentals, provided that the thresholds are ordered according to beat frequency.

For a frequency range between C3 and C6, the same holds true for major thirds, although there is more scatter in the data.

Because frequency failed to show an effect on discrimination, it seems justified to generalize previous findings with respect to, for example, the effect of tone duration (Vos, 1982), and the effect of spectral content of the tones (Vos, 1984; Vos & van Vianen, 1985), to a wider frequency range.

When tempering is expressed in cents, and only those conditions are considered in which the degree of tempering (T) corresponds to that found in the relevant tuning systems ($-5.4 \leq T < 0$), it can be concluded that perceptual differences between tempered and pure fifths are expected to be increasingly irrelevant for tones lower than about C4.

Within the same musical constraints, tempered major thirds ($0 < T \leq 21.5$), even those in a compromise tuning system such as that proposed by Silbermann ($T = 5.9$ cents), can be discriminated from pure major thirds for most fundamental frequencies that are relevant in musical composition.

Differences in discriminability between fifths and major thirds, as well as the effect of beat frequency, reported in previous studies (Vos, 1982, 1984), could be confirmed.

Analysis of temperaments, determined for various organ stops and pianos, reveals that our experimental results may be related to the perception of intervals played on the pipe organ, and, for the frequency range between C3 and about C6, they may also be related to the perception of intervals that are played on the piano.

REFERENCES

Cross, C. R., & Goodwin, H. M. (1893). Some considerations regarding Helmholtz's theory of consonance. *Proceedings of the American Academy of Arts and Sciences*, **27**, 1-12.

Grützmacher, M., & Lottermoser, W. (1935). Über die Stimmung von Flügeln. *Physikalische Zeitschrift*, **36**, 903-911.

Kolinski, M. (1959). A new equidistant 12-tone temperament. *Journal of the American Musicological Society*, **12**, 210-214.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, **49**, 467-477.

Mayer, A. M. (1894). Researches in acoustics, No. IX. *Philosophical Magazine Journal of Science, Series 5*, **37**, 259-288.

Meinel, H. (1954). Zur Stimmung der Musikinstrumente. *Acustica*, **4**, 233-236.

Meinel, H. (1957). Musikinstrumentenstimmungen und Tonsysteme. *Acustica*, **7**, 185-190.

Plomp, R. (1976). *Aspects of tone sensation*. London: Academic Press.

Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, **38**, 548-560.

Plomp, R., & Steeneken, H. J. M. (1968). Interference between two simple tones. *Journal of the Acoustical Society of America*, **43**, 883-884.

Railsback, O. L. (1938). A study of the tuning of pianos. *Journal of the Acoustical Society of America*, **10**, 86.

Rasch, R. A. (1983). Description of regular twelve-tone musical tunings. *Journal of the Acoustical Society of America*, **73**, 1023-1035.

Rasch, R. A., & Heetvelt, V. (in press). String inharmonicity and piano tuning. *Music Perception*.

Riesz, R. R. (1928). Differential intensity sensitivity of the ear for pure tones. *Physical Review*, **31**, 867-875.

Scharf, B. (1970). Critical bands. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 1, Chap. 5, pp. 157-202). New York: Academic Press.

Schuck, O. H., & Young, R. W. (1943). Observations on the vibrations of piano strings. *Journal of the Acoustical Society of America*, **15**, 1-11.

Terhardt, E. (1968). Über akustische Rauhigkeit und Schwankungsstärke. *Acustica*, **20**, 215-224.

Vos, J. (1981). [*Tuning of pipe organs, as found in practice.*] Unpublished manuscript (in Dutch), Institute for Perception TNO, Soesterberg, The Netherlands.

Vos, J. (1982). The perception of pure and mistuned musical fifths and major thirds: Thresholds for discrimination, beats, and identification. *Perception & Psychophysics*, **32**, 297-313.

Vos, J. (1984). Spectral effects in the perception of pure and tempered intervals: Discrimination and beats. *Perception & Psychophysics*, **35**, 173-185.

Vos, J., & van Vianen, B. G. (1985). Thresholds for discrimination between pure and tempered intervals: The relevance of nearly coinciding harmonics. *Journal of the Acoustical Society of America*, **77**, 176-187.

Winer, B. J. (1970). *Statistical principles in experimental design* (International Student ed.). London: McGraw-Hill.

Young, R. W. (1939). Terminology for logarithmic frequency units. *Journal of the Acoustical Society of America*, **11**, 134-139.

Young, R. W. (1954). Inharmonicity of piano strings. *Acustica*, **4**, 259-262.

Zwicker, E. (1952a). Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones. *Acustica*, **2**, 125-133.

Zwicker, E. (1952b). *Die Grenzen der Hörbarkeit der Amplituden- und Frequenzmodulation von Tönen und ihre Berücksichtigung in der Übertragungstechnik und der Hörphysiologie*. Doctoral thesis, Technical University Stuttgart.

Zwicker, E. (1954). Die Verdeckung von Schmalbandgeräuschen durch Sinustöne. *Acustica*, **4**, 415-420.

(Manuscript received November 26, 1984;
revision accepted for publication May 23, 1985.)

Music Perception
Spring 1986, Vol. 3, No. 3, 221–258

6

Purity Ratings of Tempered Fifths and Major Thirds[1]

JOOS VOS
*Institute for Perception TNO,
The Netherlands*

In this study, the relationship between the degree of tempering of musical intervals and the subjective purity of these intervals was investigated. Subjective purity was determined for fifths and major thirds. The intervals were presented in isolation, that is, they were not given in a musical context. For two simultaneous complex tones the relationship between subjective purity and tempering could be described by exponential functions. These functions were obtained both for ratings on a 10-point equal-interval scale and for subjective distances derived from preference data collected by means of the method of paired comparisons. To verify to what extent subjective purity had been determined by interference of nearby harmonics, the spectral content of the tones was varied. For both the fifths and the major thirds, interference of the various pairs of nearly coinciding harmonics was canceled by deletion of the even harmonics of the higher tone. This deletion resulted in higher purity ratings, the effect being most prominent for the major third. A further reduction of the potential interference between harmonics was still more effective: for simultaneous sinusoidal tones, subjective differences between pure and tempered intervals were much smaller than for complex tones. Purity ratings for simultaneous sinusoids presented at a low sound level were about equal to the ratings for successive tones. The purity ratings were compared with dissonance patterns predicted by models for tonal consonance/dissonance. Only in a few conditions do the patterns predicted by the model of Plomp and Levelt resemble the rating patterns obtained, and the dissonance patterns predicted by the model of Kameoka and Kuriyagawa are at variance with the purity ratings in all conditions. Suggestions for revision of the models are given.

Introduction

Following tradition in Western music theory, we define a pure consonant interval as an interval in which the ratio of the fundamental frequency, f_1, of the lower tone (Tone 1) and the fundamental frequency, f_2, of the higher tone (Tone 2) is equal to $p{:}q$, with p and q ($p < q$) being small integers. For

Requests for reprints may be sent to Joos Vos, Institute for Perception TNO, Kampweg 5, 3769 DE Soesterberg, The Netherlands.

1. Preliminary results of part of the present research were presented at the 11th International Congress on Acoustics, Paris, France, July 19–27, 1983, and are included in the Proceedings, Vol. 4, 423–426.

two simultaneous complex tones, slightly tempered consonant intervals ($f_1:f_2 \sim p:q$) are characterized by small frequency differences between those harmonics that coincide in pure intervals. Perceptually, the interference of the nearly coinciding harmonics in tempered intervals gives rise to beats or roughness.

Discriminability between such pure and tempered intervals has been investigated previously (Vos, 1982, 1984; Vos & van Vianen, 1985a, b). Briefly, the experimental results showed (a) that discrimination performance decreases with increasing sum of p and q, (b) that discriminability is poorer for low degrees of tempering than for high degrees of tempering, and (c) that this difference is more prominent at short tone durations. Further, the results showed (d) that for moderately tempered intervals discrimination is mainly determined by the interference of nearly coinciding harmonics, and (e) for higher degrees of tempering discriminability may be enhanced by differences in the size of the tempered and the pure interval.

The purpose of this study is to investigate the relationship between tempering of fifths ($p = 2; q = 3$) and major thirds ($p = 4; q = 5$) and the subjective *purity* of these intervals. Subjective purity will be determined both for complex and sinusoidal tones. In the discussion of the results obtained, it will be argued that both the evaluation of purity and the evaluation of tonal or sensory consonance are highly affected by beats or roughness. Because of this correspondence our experimental results will be used to test the validity of two consonance models (Plomp & Levelt, 1965; Kameoka & Kuriyagawa, 1969a). A number of suggestions for revision of these models will be given.

Evaluation of Tuning Systems

Systematic evaluations of different tuning systems, as proposed by Hall (1973) and by Rasch (1983), are based on the idea that tempering should be minimized as much as possible. These theoretical evaluations assume that the appreciation of a tuning system, or its suitability for the performance of a particular piece of music, is related to the number of cents by which a tempered interval deviates from the corresponding pure interval.

Hall (1973) assumed a quadratic relationship between tempering of the interval and its subjective "undesirability" or the "unsuitability to fill its usual musical role." This presumed relationship is shown in Figure 1. Thus, in a two-part piece of music that consists of, for example, two fifths, three minor thirds, and four major thirds, the root of the mean of the squared temperings, as found in equal temperament, is equal to $\{[2(-2.0)^2 + 3(-15.6)^2 + 4(13.7)^2]/9\}^{1/2} = 12.9$ cents. In the meantone tuning system, quadratic mean tempering is only $\{[2(-5.4)^2 + 3(-5.4)^2 + 4(0.0)^2]/9\}^{1/2} = 4.0$ cents, and as a result this evaluation suggests that for the perform-

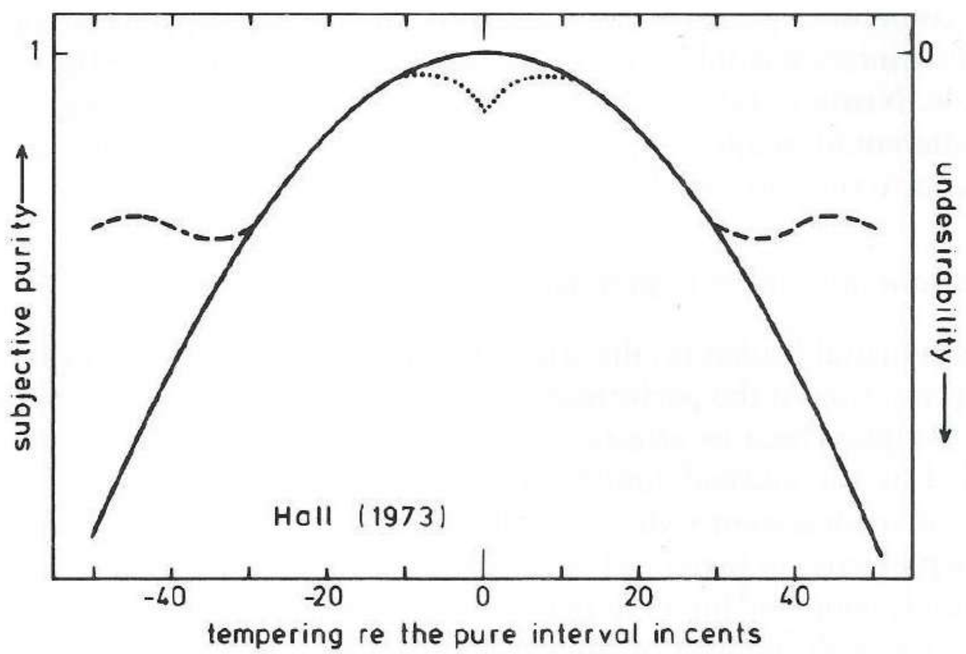

Fig. 1. Quadratic relationship between tempering and subjective "purity" or "desirability," which formed the basis for Hall's evaluation of various tuning systems. The dotted curves indicate that slightly tempered intervals may be preferred to the pure interval; the dashed curves indicate that "undesirability" of considerably tempered intervals may remain rather constant.

ance of this particular piece of music the meantone tuning system is preferred to equal temperament.

Regular 12-tone musical tuning systems have been described by Rasch (1983). In evaluating these tuning systems, he used three measures for mean tempering, which were basically meant to represent mean subjective purity. The three measures were quadratic mean tempering, linear mean tempering, and maximum tempering. The measure of linear mean tempering suggests that even small amounts of tempering will affect the degree to which music is experienced to be "out-of-tune." The use of maximum tempering is based on the idea that the worst interval in a piece of music is indicative of the quality of its performance.

Although Hall (1973) felt it rather plausible to assume that an increase of the degree of tempering by a factor of two corresponded to "four times as undesirable," he mentioned two disadvantages of adoption of the quadratic relationship. First, some people may rate pure intervals to be "insipid" and therefore prefer intervals slightly tempered as expressing greater "warmth." This hypothetical effect is indicated by the dotted curve in Figure 1. Roberts and Mathews (1984) found that, indeed, this effect may be important for at least some listeners. Second, for intervals tempered considerably, which have beats too rapid to be clearly audible one by one, "undesirability" may remain rather constant at higher degrees of tempering. This is indicated by the broken curve in Figure 1.

The common experience that the effect of tempering depends on the kind of musical interval would suggest the use of unequal weights for the various intervals. Neither Hall (1973) nor Rasch (1983) attempted to introduce such differential weights, because results from, for example, perceptual experiments to support these weights were lacking.

The Rationale of the Experiments

Experimental studies on the effect that different tuning systems have on the appreciation of the performance of a particular piece of music may be rather complex, because effects of roughness and intonation are likely to be modified by the musical context. For example, in compositions that are written in a consonant style, especially for sustained tones, pure intervals may be preferred to tempered intervals, whereas in a more dissonant style moderately tempered intervals may be as acceptable as pure intervals. Some support for such an effect of context is given by results obtained by Shackford (1961, 1962a, b), who analyzed the size of the intervals performed by various string trios in a set of four pieces in which a number of musical functions of the intervals had been varied. Tonal versus atonal music, and leading tones and the specific ways in which diminished fifths and augmented fourths are resolved, are other examples of context effects that seem to have influenced the intonation behavior of Shackford's string players. Understanding of such results may be facilitated by detailed knowledge of the relationship between tempering and purity rating for isolated intervals.

We assume that purity ratings of various musical intervals can be different only when these intervals can be discriminated, although the purity of discriminable intervals does not necessarily have to be rated differently. As thresholds for the discrimination between pure and tempered intervals depend on tone duration (Vos, 1982) and on the spectral content of the tones (Vos, 1984; Vos & van Vianen, 1985a), tone duration and spectral content are also included in the present experiment. Further, we hypothesized that the perceived purity of the isolated intervals depends both on the sensation of beats and roughness and on the identification of the size of the intervals. More specifically, for temperings larger than about 25 cents, at which also the direction of tempering (stretched or compressed relative to the pure interval) can be identified (Vos, 1982), purity rating should be low, irrespective of the spectral content of the tones.

In this study, subjective purity is measured in two ways. First, purity is rated by means of a 10-point equal-interval scale. Because the ratings depend both on the perceived similarities/dissimilarities of the various stimuli and on the way in which the rating scale is used, these ratings yield measurements at the ordinal level. This means that, in fact, any monotonic (order-preserving) transformation on these data is allowed. To be able to test the significance of the sharp peaks with relatively low dissonances, as predicted by the consonance/dissonance models of Plomp and Levelt (1965) and of Kameoka and Kuriyagawa (1969a), the ratings were supplemented with paired comparison judgments in which the subjects had to select the purer interval from each pair. For the analysis of these preferences the method of Thurstone's case V model is applied. This model (see Torgerson, 1958, Ch. 9) processes pairwise probabilities and is presumed to yield values for subjective purity at the level of an interval scale.

Experiment 1

Method

Stimuli and Apparatus

For all intervals, Tone 1 was a complex tone that consisted of the first 20 harmonics with amplitude a_n proportional to $1/n$. Consequently, the spectral-envelope slope of the tone was − 6 dB/octave. This spectrum resembles that of a bowed string instrument in its lower frequency register. The phase of the individual harmonics was chosen randomly. The spectrum of Tone 2 comprised the first 20 harmonics or the first 10 odd harmonics (1, 3, 5, .., 19). Again, the amplitudes of the particular harmonics were proportional to $1/n$ and the phases of the various harmonics were random. Rise and decay times of the tones, defined as the time interval between 10 and 90% of the maximum amplitude, were 30 and 20 msec, respectively. The overall level of Tone 2 was 20 log(q/p) dB lower than that of Tone 1. As a result, the spectral envelopes of the two tones coincided, that is, the amplitudes of the nearly coinciding harmonics were equal. For both the tempered fifths and the tempered major thirds, level-variation depth of the beating harmonics was therefore maximal. The overall level of Tone 1 was about 74dB(A). This level was measured by means of an Artificial Ear (Brüel & Kjaer, Type 4152).

The experiment was run under the control of a PDP-11/10 computer. Tones 1 and 2 were generated in the following way. One period of the waveforms of Tones 1 and 2 was stored in 256 discrete samples (with 10-bit accuracy) in external revolving memories. These recirculators could be read out by digital-to-analog converters. Sampling rates were determined by pulse trains, derived from two frequency generators. After gating, each tone passed a Krohn-Hite filter (Model 3341) with the function switch in the low-pass RC mode and with a cutoff frequency of 8 kHz. The sound-pressure levels of the tones were controlled by programmable attenuators. After attenuation, the tones were mixed and fed to an amplifier set, to which four Beyer DT 48S headphones were connected. By means of these headphones the musical intervals were presented diotically (same signal to both ears). Subjects were seated behind a terminal in a soundproof room. They responded by pressing the keys of a keyboard. Parts of the instructions were presented visually on a display.

Subjects

Twenty-four musically trained subjects, 8 females and 16 males, participated in the experiment. Most of them were students, either of the Institute of Musicology or of the Conservatory at Utrecht. The subjects were paid for their participation.

Experimental Design and Procedure

Ratings. Four factors were varied independently: (1) musical interval (fifth and major third), (2) the amount of tempering (0, ±2, ±4, ±6, ±8, ±10, ±15, ±20, ±25, ±35, and ±50 cents), (3) spectral content of the higher tone (the first 20 harmonics vs. the first 10 odd harmonics), and (4) tone duration (0.25 and 0.5 sec). The 84 stimulus combinations of the first three variables were presented in a random order within a block. There were two blocks, one for each tone duration; each block was presented twice. The subjects were requested to rate each interval in terms of its "purity," by using the whole range of scale values from "very impure" (1) to "very pure" (10). At the beginning of the first experimental block with a certain tone duration, the subjects were presented with a familiarization block, comprising the trials for only one type of spectral content, but covering the whole range of tempering to be given in the experimental blocks. Subjects were told that the purpose of this block was to give them a frame of reference within which they had to rate the purity of the intervals.

Paired Comparison Judgments. Four factors were varied: (1) the amount of tempering (0, ±2, ±4, ±6, ±8, ±15, ±25, and ±35 cents), (2) musical interval (fifth and major third), (3) spectral content of the higher tone (the first 20 harmonics vs. the first 10 odd harmonics), and (4) tone duration (for the fifth: 0.25 and 0.5 sec; for the major third 0.5 sec only). From the 15 temperings included, we composed 105 $[n(n-1)/2]$ different pairs of intervals within a block. From each pair the subjects had to select the purer interval by pressing one of two possible keys of the keyboard. Musical interval, spectral content of the higher tone, and tone duration were held constant within a block. Since the major thirds were only presented at a tone duration of 0.5 sec, there were 6 blocks of 105 pairs in all.

General. Fundamental frequencies of Tones 1 and 2 together were varied from trial to trial in such a way that the central frequency, $(f_1 f_2)^{1/2}$, of an interval was equal to 370 Hz ± 25, ±75, ±125, ±175, or ±225 cents, with equal probability (10%) of occurrence. For paired-comparison judgments the probability that the central frequency of the second interval was equal to that of the first interval was 10% as well.

The experimental session lasted about 4 hr. The subjects were tested in six groups of four subjects. Three groups were presented first with the intervals of 0.25 sec, whereas the other three groups started with the intervals of 0.5 sec. Likewise, the order in which the ratings or the paired-comparison judgments had to be given was balanced with respect to tone duration. The various successive blocks (with intervals at a fixed tone duration), in which paired comparison judgments had to be given, were presented to the groups in different orders.

Results

Ratings

Reliability. Making direct quantitative judgments of the degree of subjective purity associated with each musical interval is, relative to the indirect method of paired-comparison judgments, a difficult task. To exclude the data of those subjects who had not been able to give their ratings in a reasonably consistent way, we computed the correlation coefficient (r) between the ratings of the first and the second block of 84 trials for each tone duration separately. If, for a given subject, r was less than 0.50 for both the tone durations of 0.25 and 0.5 sec, the data were dropped from further analysis. This was the case for four subjects. In addition, the ratings from two subjects could not be used because they had not completed the experimental program. For the remaining 18 subjects the mean values of r were 0.66 and 0.73, with standard deviations of r equal to 0.12 and 0.09, for tone durations of 0.25 and 0.5 sec, respectively.

Mean Rating as a Measure of Central Tendency. For all 336 conditions [84 (conditions) × 2 (replications) × 2 (tone durations)], differences between the median and the arithmetic mean were determined. The differences did not depend systematically on specific conditions. Because of the more advanced statistical analyses available, we preferred to present the results in terms of mean values. Since, on the average, the difference between the median and the arithmetic mean was 0.04 and, more importantly, the standard deviation of the differences was only 0.4, we considered this to be justified.

Effect of Tempering. The ratings for all conditions, averaged across the 18 subjects and the two replications, are given in Figure 2 as a function of tempering in cents. For both the tone durations of 0.5 and 0.25 sec, and the two spectrally different conditions, subjective purity was highly affected by the degree of tempering (see Table 1, in which for each tone duration separately, the results of an analysis of variance performed on the ratings are summarized). In general, the relationship between mean purity rating (y) and the amount of tempering (x) in cents can be adequately described by means of the exponential function $y = a \exp(bx)$. For all conditions pre-

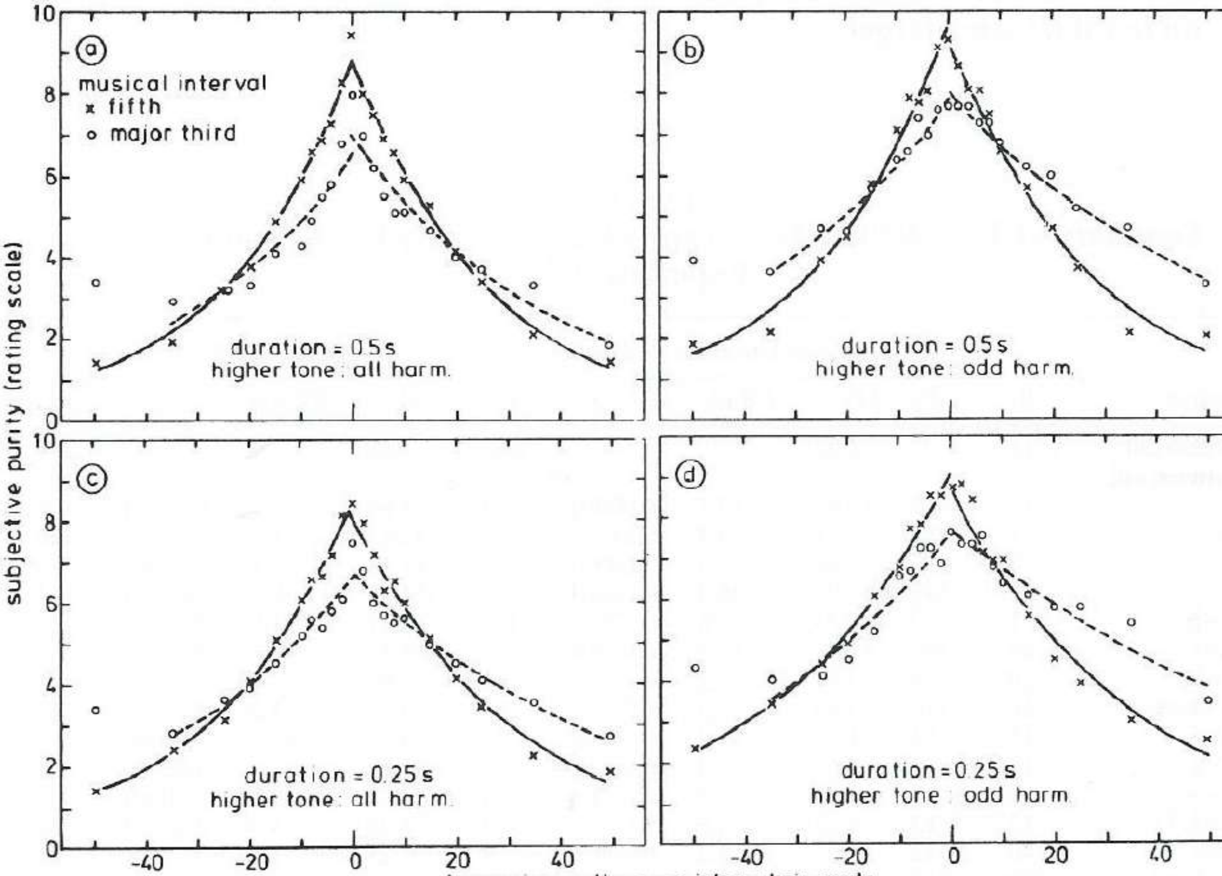

Fig. 2. Mean subjective purity ratings, plotted as a function of tempering for the fifth and the major third, separately. The curves represent the best exponential fit. In panels a and b tone duration is 0.5 sec, whereas in panels c and d, the duration is 0.25 sec. In panels a and c, the spectrum of the lower tone and that of the higher tone comprised the first 20 harmonics (all). In panels b and d, the spectrum of the higher tone comprised the first 10 odd harmonics only.

sented in Figure 2, the values of a and b, together with the coefficients of determination (r^2), are given in Table 2. The mean values of r^2 are 0.98 and 0.95 for the fifths and the major thirds, respectively. In terms of r^2, the exponential fit is about 8% better than the linear fit $y = a + bx$ for the spectral conditions in which also the higher tone contained all harmonics. For the conditions in which the spectrum of the higher tone comprised only the odd harmonics, the quality of the exponential fit exceeds that of the linear fit by about 3%.

The exponential fits of the mean ratings emphasize that especially at temperings close to physical purity, the differences between the ratings are prominent. This was confirmed by the results of Newman–Keuls paired comparison tests (Winer, 1970) performed on the 16 different sets of data, each set being represented by one exponential function. For the spectral conditions in which both tones contained all harmonics, mean ratings at temperings of ±2 or, atmost, ±4 cents, were significantly lower ($\alpha = 0.01$) than the mean ratings at physical purity (see Table 3). Especially for the major third in the conditions with the reduced spectrum of the higher tone, these values were larger.

TABLE 1
Summary of Two ANOVAs Performed on the Purity Ratings Obtained in Experiment 1[a]

			Tone Duration = 0.25 sec				Tone Duration = 0.5 sec			
Source	df_1	df_2	MS	F Ratio	p	$\hat{w}^2$	MS	F Ratio	p	$\hat{w}^2$
Within cell	1512		2.27			35.8	1.87			26.4
Between cell										
S	17	1512	30.62	13.5	0.00001	2.7	46.64	24.9	0.00001	3.8
A	1	17	27.43	0.8	0.61	0.0	42.62	1.0	0.66	0.0
B	1	17	706.80	93.3	0.00001	3.6	781.2	53.5	0.00002	3.6
C	20	340	425.90	98.4	0.00001	43.9	551.0	136.0	0.00001	51.1
A×B	1	17	15.43	3.6	0.07	0.1	98.22	11.9	0.003	0.4
A×C	20	340	42.98	15.5	0.00001	4.2	44.41	19.8	0.00001	3.9
B×C	20	340	3.61	1.2	0.24	0.1	6.49	2.6	0.0004	0.4
A×B×C	20	340	2.63	1.2	0.27	0.0	2.15	1.2	0.28	0.0
A×S	17	1512	34.39	15.1	0.00001	3.0	44.52	23.8	0.00001	3.6
B×S	17	1512	7.57	3.3	0.00003	0.5	14.61	7.8	0.00001	1.1
C×S	340	1512	4.33	1.9	0.00001	3.9	4.05	2.2	0.00001	3.7
A×B×S	17	1512	4.33	1.9	0.01	0.2	8.24	4.4	0.00001	0.5
A×C×S	340	1512	2.77	1.2	0.008	0.9	2.25	1.2	0.01	0.6
B×C×S	340	1512	2.98	1.3	0.0006	1.3	2.49	1.3	0.0004	1.0
A×B×C×S	340	1512	2.23	1.0	0.58	0.0	1.85	1.0	0.55	0.0

[a] For both tone durations we used an 18 [subjects (S)] × 2 [musical intervals (A)] × 2 [spectra (B)] × 21 [temperings (C)] factorial design, all repeated measures and with replication as a within-cell variable. MS = mean of squares; df_1 and df_2 represent the number of degrees of freedom for the numerator and the denominator, respectively; p = probability of F ratio; $\hat{w}^2$ = percentage of estimated component of variance.

TABLE 2
Values of *a*, *b*, and r^2 for All Conditions in Experiment 1[a]

Tone Duration (sec)	Musical Interval	Tempering (cents)	Harmonics of Tone 2: All *a*	All *b*	All r^2	Odd *a*	Odd *b*	Odd r^2
Ratings								
0.5	Fifth	≤0	8.8	0.040	0.99	9.7	0.037	0.97
		≥0	8.9	−0.038	0.99	9.4	−0.035	0.96
	Major third	≤0	6.6	0.028	0.88	7.9	0.023	0.97
		≥0	7.0	−0.026	0.96	8.1	−0.017	0.98
0.25	Fifth	≤0	8.6	0.037	0.99	9.2	0.028	0.99
		≥0	8.2	−0.033	0.98	8.8	−0.028	0.96
	Major third	≤0	6.7	0.025	0.97	7.7	0.022	0.91
		≥0	6.7	−0.019	0.97	7.8	−0.014	0.93
Paired comparison judgments[b]								
0.5	Fifth	≤0	3.1	0.041	0.98	2.9	0.034	0.97
		≥0	3.5	−0.058	0.99	3.3	−0.051	0.98
	Major third	≤0	2.7	0.027	0.90	2.4	0.018	0.88
		≥0	2.8	−0.028	0.97	2.6	−0.021	0.99
0.25	Fifth	≤0	2.9	0.034	0.99	2.7	0.028	0.99
		≥0	3.1	−0.045	0.99	3.0	−0.039	0.98

[a]The relationship between subjective purity (y) and the amount of tempering (x) in cents can be adequately described by means of the exponential function $y = a \exp(bx)$. Subjective purity was expressed either as ratings between 1 (very impure) and 10 (very pure) or as scale values that were derived from preference data collected by means of the method of paired comparisons.

[b]Predicted z scores, as plotted in Figure 3, can be obtained when the values that result from the exponential functions are reduced by 2.

TABLE 3
Temperings in Cents at which the Mean Subjective Purity Ratings Were Significantly Lower Than the Mean Ratings at the Tempering of 0 Cents[a]

Tone Duration (sec)	Musical Interval	Harmonics of Tone 2: All Compressed	All Stretched	Odd Compressed	Odd Stretched
0.5	Fifth	−2	2	−4	4
	Major third	−2	2*	−10*	15
0.25	Fifth	−4	4	−10	6
	Major third	−2	4	−15	10*

[a]Differences between the mean values were tested by means of Newman–Keuls paired comparison tests (Winer, 1970) for the 16 conditions separately. For 13 conditions the mean values were significantly different at the 0.01 level; in the remaining conditions (marked with asterisks) this level was 0.05.

The conditions in which the major third was tempered by −50 cents were dropped from the regression analysis, because it can be argued that this interval is a stretched minor third (tempered by 21 cents) rather than a compressed major third.

Other Effects. The effect of tempering depends on the kind of interval. Fifths, compressed by up to about 15 to 25 cents or stretched by up to about 10 to 20 cents are perceptually purer than the corresponding major thirds. For larger temperings, the reverse tends to be the case. The exact temperings at which the fifths and major thirds are rated equally depend on (1) the spectral content of the higher tone and (2) whether the pure interval is stretched or compressed. First, it can be seen in Figure 2 that deletion of the even harmonics of the higher tone increased the subjective purity ratings of the major thirds more than those of the fifths. At a tone duration of 0.5 sec, this effect is statistically significant. Second, the tempering at which the rating curves of the fifth and major third intersect depends on the direction of tempering. For both the tone duration of 0.5 and 0.25 sec, stretched major thirds are rated significantly purer than the compressed intervals, whereas for the fifth this effect could not be found.

Paired-Comparison Judgments

For each block of 105 pairs separately, the preference scores were analyzed under the assumption of Thurstone's case V (see Torgerson, 1958). In this model, subjective scale values of the stimuli are estimated from these observed preferences. Again, the relationship between subjective purity and the amount of tempering can be described slightly better by means of exponential functions than by linear functions. For all conditions investigated, the values of a, b, and r^2 are given in Table 2. The mean values of r^2 are 0.98 and 0.94 for the fifth and the major third, respectively. For the conditions in which the fifths and the major thirds were presented at a tone duration of 0.5 sec, the subjective scale values are given in Figure 3. In Figure 3a, both the lower and the higher tones comprised all harmonics, whereas in Figure 3b, the spectrum of the higher tone comprised the 10 odd harmonics only. In Figure 3 it can be seen that the subjective distances between the physically pure and tempered fifths are greater than those between the pure and tempered major thirds. It can also be seen in Figure 3 that, especially for the major thirds, the subjective distances between the pure and tempered intervals tend to decrease with a reduction in the spectral content of the higher tone. Both the difference between the fifths and the major thirds, described above, and the effect of spectral content are consistent with the effects that were found for the ratings on the 10-point equal-interval scale [cf. Figures 2a,b].

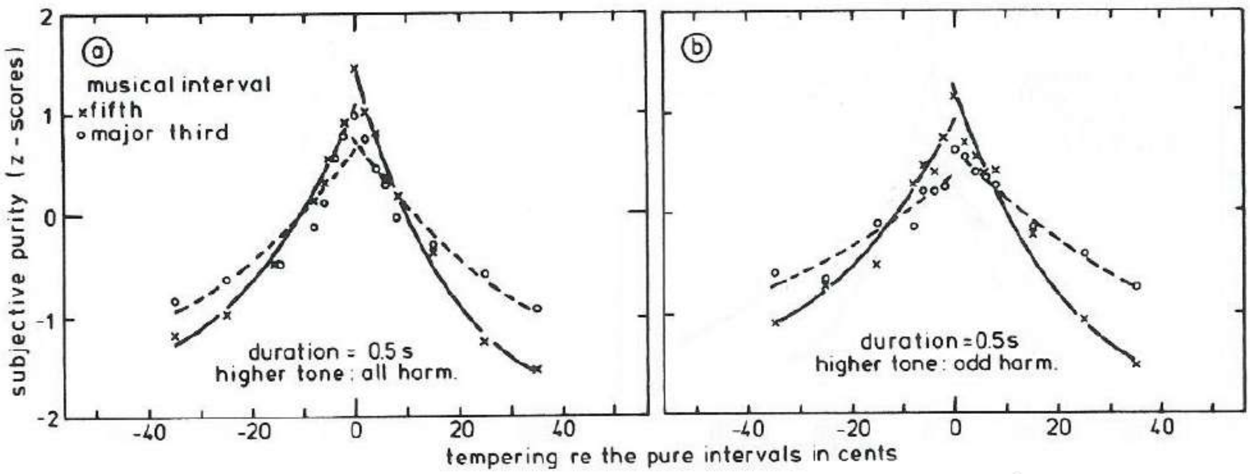

Fig. 3. Scale values for subjective purity, derived from preference data collected by means of the method of paired comparisons and plotted as a function of tempering for the fifth and the major third, separately. The curves represent the best exponential fit. In panel a the spectrum of the higher tone comprised the first 20 harmonics (all); in panel b the higher tone comprised the first 10 odd harmonics only. In both panels the lower tone comprised all harmonics and tone duration was 0.5 sec.

In the present experiment, fifths and major thirds were never given within the same pairs. Because, for the sets of fifths and major thirds separately, the sum of the z scores is, by definition, equal to zero, absolute differences between the subjective distances of the two intervals are not meaningful. This is not the case for the ratings given in Figure 2.

Discussion

Differences between Fifth and Major Third

It was shown in Figures 2 and 3 that the relationships between tempering and subjective purity for the tempered fifths were different from those for the major thirds in all conditions. In these figures, subjective purity is plotted as a function of cents, but, since the ratings seem to have been at least partly influenced by the sensation of beats, it may be worthwhile to plot the ratings as a function of beat frequency. For a number of different spectral conditions, including those adopted in this experiment, it has been shown that the beat frequency of (pf_2-qf_1) Hz is perceived more clearly than multiples of this beat frequency (Vos, 1984). The ratings of the upper panels of Figure 2 were therefore replotted in Figure 4 as a function of (pf_2-qf_1). Comparing Figures 4a and 4b with Figures 2a and 2b, it can be concluded that, especially for the larger temperings, plotting the ratings as a function of beat frequency increases rather than decreases the differences between the fifth and the major third. Part of this difference, however, may be explained by the perceived strength of the beats. Independently of whether the

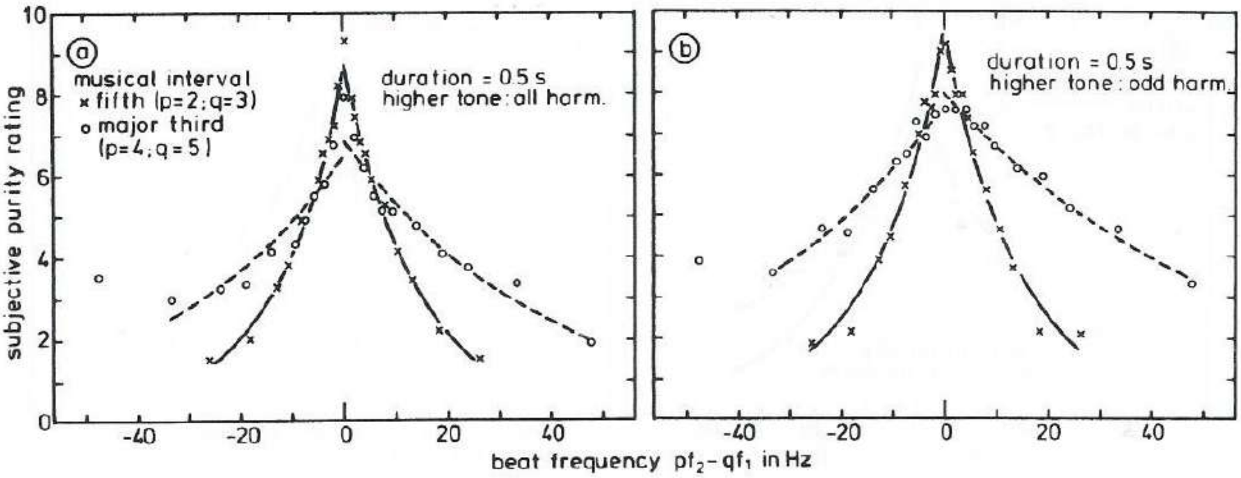

Fig. 4. Subjective purity ratings from Figures 2a and 2b, replotted as a function of the beat frequency of $pf_2 - qf_1$ Hz. The regression functions represent the best exponential fit.

spectrum of the higher tone comprised either all (Figures 2a, 4a) or only the odd harmonics (Figures 2b, 4b), the perceived strengths of the beats of tempered fifths were about 1 dB higher than those of tempered major thirds (Vos, 1984); perceived strength of the beats had been expressed in the level-variation depth of two beating sinusoidal tones that was matched with the tempered fifths or major thirds for equal beat strength.

Except for the physically pure intervals, subjective purity was higher when the spectrum of the higher tone comprised only odd harmonics (see, e.g., Figure 4b) than when that tone comprised all harmonics (Figure 4a). In Figure 5, where the regression functions from Figures 4a and 4b are combined in one plot, it can be seen more easily (1) that the increase in subjective purity resulting from a reduction in spectral content of the higher tone is about equal for all degrees of tempering and (2) that this increase is larger for the major third than for the fifth. Vos (1984) has shown that when the higher tone comprises only odd harmonics, the perceived strength of the beats is about 3 dB lower than when all harmonics are included. Our result that subjective purity increased more for the major third than for the fifth was not to be expected on the basis of that study, since the difference of 3 dB was found for both the fifth and the major third. It is possible, however, that for the major third the reduced spectral content led to more stimuli that received higher ratings because no beats were perceived at all. Differences in tone duration may be relevant here because in the present experiment tone duration was 0.5 sec or less, whereas it was 1 sec in Vos (1984). That tone duration might be responsible for the larger increase in ratings of the major third is supported by the finding that a reduction in tone duration diminishes discriminability between pure and tempered intervals more for major thirds than for fifths (Vos, 1982, 1984).

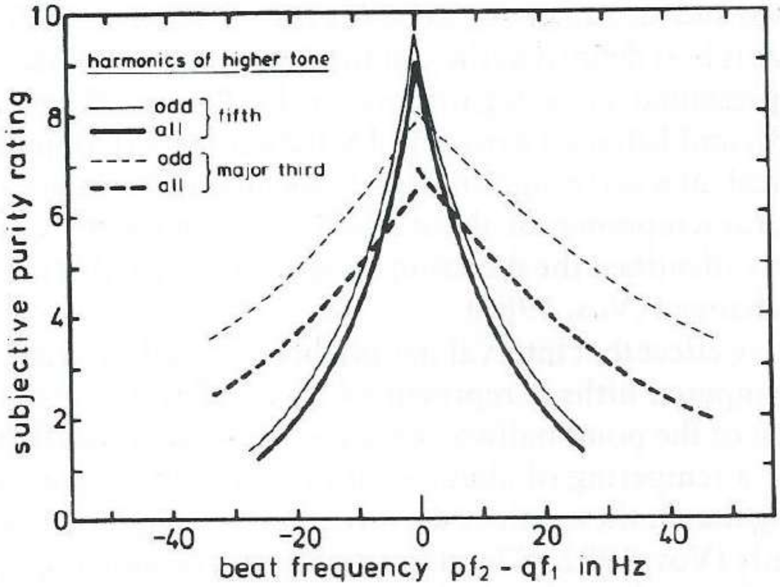

Fig. 5. Regression functions from Figures 4a and 4b, combined in one plot.

Stretched major thirds were rated to be slightly purer than compressed ones, an effect that was not found for fifths. The greater tolerance for stretched than for compressed major thirds may be related to stretched intonation, observed in both solo and ensemble music performed on string and woodwind instruments (Ward, 1970). In these actual performances the major thirds were, on the average, stretched by 22 cents, whereas the fifths were stretched by only 4 cents, which is consistent with our experimental results. Considerably stretched major thirds but relatively pure fifths have also been found for simultaneous intervals sung by oratorio choirs (Lottermoser & Meyer, 1960).

Two-Component Models for Subjective Purity

The extent to which subjective purity was rated higher in the spectrally reduced conditions was independent of the degree of tempering. This does not necessarily mean that perceived strength of beats was the sole criterion that determined the relationship between physical and subjective purity.

It can be hypothesized that in the perception of tempered intervals both (1) the sensation of beats or roughness and (2) the detection of differences between the size of the tempered interval and that of the pure interval are relevant. In the recent papers by Hall (1983) and Hall and Hess (1984), the significance of these two aspects is also emphasized. Interval size is a relevant aspect of perceptual experience. In melodic intervals (successive tones), interval size is even the sole criterion on which purity ratings can be based. This second component should not be confused with Terhardt's (1984) concept of harmony, representing virtual-pitch perception. For both musically trained and untrained subjects, Attneave and Olson (1971) found that interval size is best defined as the distance between two tones with their frequencies represented on a logarithmic scale. Plomp, Wagenaar, and Mimpen (1973), and Killam, Lorton, and Schubert (1975) found that confusions in musical interval recognition are based mainly on interval size. As a last example, for temperings of about 20–30 cents, musically trained subjects consistently identified the direction in which the size of fifths and major thirds was changed (Vos, 1982).

The qualitative effect that interval size might have on the purity rating of, for example, tempered fifths is represented by the dashed curve in Figure 6a. The position of the point halfway between the maximum and the minimum ratings at a tempering of about −20 to −30 cents is suggested by a previous finding that at these values the direction of tempering can be identified consistently (Vos, 1982). The effect that beat sensation might have on purity rating in the condition in which the spectrum of the higher tone comprises all harmonics is represented by the heavily dashed/dotted curve in Figure 6a. The ratings strongly decrease with increasing tempering while there is a rather broad minimum between about −10 and −30 cents. For higher degrees of tempering the curve again suggests a small increase in the

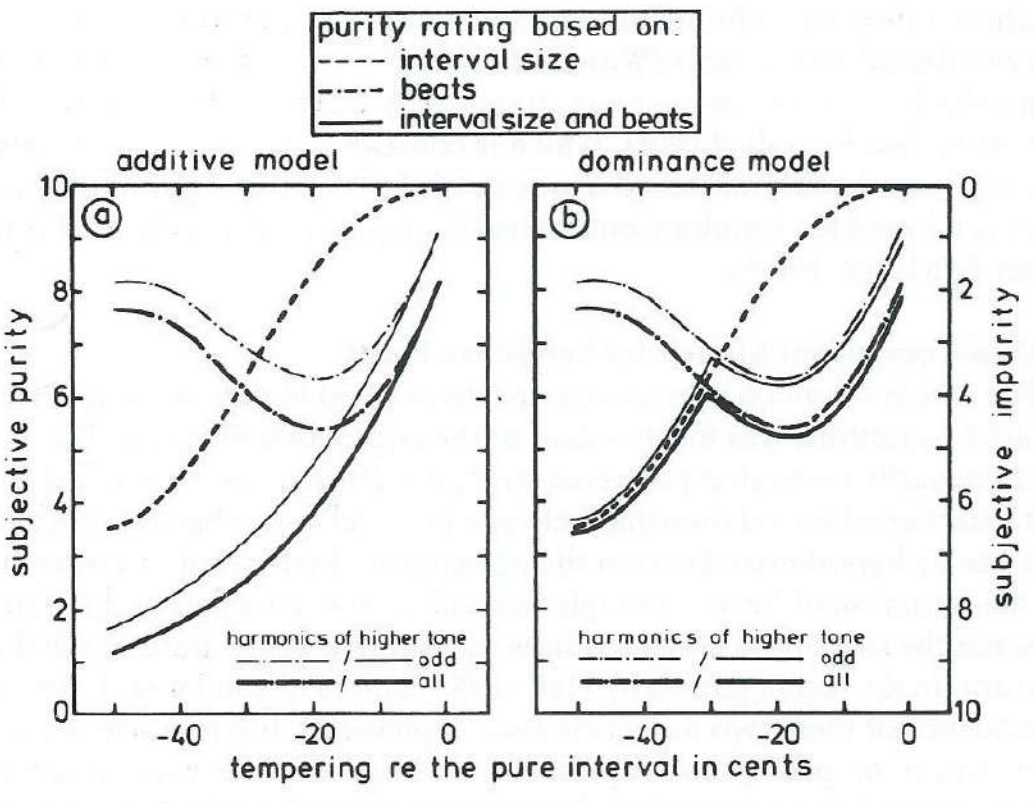

Fig. 6. Illustration of two models that both assume that subjective purity is determined by the combined effect of the perception of interval size and the sensation of beats or roughness. Predictions are separately given for the conditions in which the higher tone comprised either all or only odd harmonics, while in both conditions the lower tone comprised all harmonics.

ratings. The global characteristics of this plotted function are prompted by beat thresholds for tempered fifths in conditions that resemble those in the present experiment with respect to the spectral density of the stimuli (see Vos, 1982). Application of a simple model in which the ratings that are based on the two hypothetical aspects are *added,* results in the final exponential relationship represented by the bold solid curve. The upward shift of the purity ratings that resulted from a reduction in the spectral content of the higher tone was about equal for all degrees of tempering (see Figure 5). The present model is proposed to demonstrate that this experimental effect can be caused by a change in beat sensation, while the contribution of interval size remains unaffected. This is also summarized in Figure 6a, where the reduction of spectral content of the higher tone (thin curves) has diminished the effect of beat sensation (the dashed/dotted curve) only.

The rather uniform upward shift in the overall purity ratings can be explained both by the *additive* model and by an *average model.* This effect, however, can in no way be predicted from a model in which the overall purity rating is set equal to the lowest rating of the two aspects. This is illustrated in Figure 6b, where the hypothetical relationships from Figure 6a are adopted again. For high degrees of tempering, this *dominance model* predicts equal ratings for the two spectrally different conditions.

Experiment 2

Discussing the results of Experiment 1, we have suggested that the purity ratings may have been determined by the combined effect of two different aspects in the perception of tempered intervals. In Experiment 2, attempts were made to gain more insight into the relevance of these two hypothetical aspects of beats and interval size. First, it was reasoned that if the sensation of beats or roughness is relevant in purity rating, a strong reduction in the potential interference between harmonics will affect the ratings. This reduction was carried out by presentation of two simultaneous sinusoidal rather than complex tones. Two sinusoids that are more than a critical band apart may also result in the sensation of beats, provided that the tones are presented at relatively high sound-pressure levels (Plomp, 1967). The latter observation may be related to the finding of Kameoka and Kuriyagawa (1969b) that the evaluated sound quality of isolated intervals at a high sound level is different from that at a lower sound level. Therefore, the sinusoids, constituting tempered fifths or tempered major thirds, are presented both at a moderate and at a very low sound level.

The most powerful method to preclude beats or roughness is by presenting the tones successively. It was reasoned that the degree of tempering of successive tones can only be evaluated by comparing the interval size between the tones with the size of an imagined pure interval. For the simultaneous sinusoids that are presented at a very low level, it can also be hypothesized that the ratings are based on the output of such a comparison procedure. To make a comparison between the ratings for simultaneous and successive sinusoids possible, the latter intervals were included as well.

Method

Stimuli and Apparatus

For all intervals, Tones 1 and 2 were sinusoidal tones. The tones were 0.5 sec in duration. Their temporal envelopes were identical to those in Experiment 1. As in Experiment 1, the two tones had a level difference of $20 \log(q/p)$ dB. The level of Tone 1 was about 83 dB(A) in the high and 58 dB(A) in the low level conditions. The apparatus was the same as in Experiment 1.

Subjects

Twenty-four musically trained subjects participated in the experiment. As in Experiment 1, most of them were students either of the Institute of Musicology or the Conservatory at Utrecht. Nine of them had also participated in Experiment 1, about 2 years earlier. The subjects were paid for their services.

Experimental Design and Procedure

Ratings. Four factors were varied: (1) musical interval (fifth and major third); (2) the amount of tempering (0, ±2, ±4, ±6, ±8, ±10, ±15, ±20, ±25, ±35, and ±50 cents); (3) mode of presentation (simultaneous vs. successive tones); and (4) sound level [high (attenuation of 0 dB) vs. low (attenuation of 25 dB)]. The 42 stimulus combinations of the first two variables were presented twice in a random order within a block of 84 trials. There were three such blocks, two in which the tones were presented simultaneously at either the high or the low sound level, and one block in which the tones were presented successively at the high sound level. Prior to the first experimental block, the subjects were presented with a familiarization block. They were told again that the purpose of this block was to give them a frame of reference within which they had to rate the purity of the intervals, using the whole range of scale values from very impure (1) to very pure (10).

Paired-Comparison Judgments. The design was similar to that of the ratings. In addition, for the fifth, the successive tones were presented at the low sound level too. Musical interval, mode of presentation, and sound level were held constant within a block. For the blocks in which the tones were presented at the high sound level, the amount of tempering was 0, ±2, ±4, ±8, ±15, ±25, ±35, and ±50 cents. For the blocks with the low sound levels, temperings of ±2 and ±35 cents (−50 and +35 cents for the major third) were not included.

General. In the instructions it was emphasized that the subjects had to rate the intervals relative to either "pure fifths" or "pure major thirds." Therefore, tempered intervals that resembled minor thirds, fourths, or minor sixths had to be interpreted as very impure.

The way in which the frequencies were varied was identical to that in Experiment 1. Similarly, the subjects were tested in six groups of four subjects. To restrict the number of changes in response modality during the experimental session, three groups were presented first with all the blocks in which they had to give their ratings, while the other three groups first completed all blocks in which they had to give the paired-comparison judgments. The order in which the subjects were presented with the successive and the simultaneous tones was balanced for each response modality. Presentation order of the various blocks within the four subsets, each comprising a different combination of response mode and presentation mode, was randomized.

Results

Ratings. For sinusoidal tones, subjective differences between tempered intervals were less pronounced than they were for complex tones. Because of this, the rating task in this experiment was much more difficult than that in the previous experiment. This can also be concluded from the high number of subjects that failed to give consistent ratings.

For each subject and for various sets of ratings the rating reliability was determined by computing the correlation coefficient (r) between the ratings given after the first and after the second presentation. In fact, r was computed for six sets of ratings separately (major third and fifth, each for the three different combinations of mode of presentation and sound level). Only 7 of the 24 subjects had r values ≥ 0.50 for more than one of the six sets described above. Even at the low criterion of $r \geq 0.35$ for at least two of the six sets, the data of 13 of the 24 subjects would have been dropped from further analysis. As the experimental effects for the group of subjects with the relatively high r values were about the same as those for the group of subjects with the relatively low r values, we decided to present the results for the whole group of 24 subjects together. The similarity of the results was confirmed by the results of three ANOVAs (one for each combination of mode of presentation and sound level) with relatively high versus relatively low r values as a between-subjects variable and musical interval and tempering as within-subjects variables. All three ANOVAs showed that neither the overall differences between the two groups of subjects, nor any interaction effect with level of rating reliability, were significantly different.

Similar to the procedure applied in Experiment 1, differences between the median and the arithmetic mean were determined. The differences were small (mean difference = 0.14; standard deviation = 0.53) and there was only a weak tendency for this difference to be negative for low values and positive for high values of the median and the mean. We concluded that the ratings may be presented in terms of their mean values.

For the conditions in which the sinusoidal tones were presented simultaneously, mean ratings are given in Figures 7a and 7b for the high and the low sound levels, respectively. In this figure the ratings, averaged across the 24 subjects and the two replications, are plotted as a function of tempering in cents. For the successive tones presented at the high sound level, these ratings are given in Figure 8.

Again the relationship between purity ratings and the amount of tempering in cents can be described by exponential functions. Although the exponential fit is slightly better than the linear fit in only a few conditions, trends in the data will be summarized by exponential functions for reasons of comparability with the results from Experiment 1. For all conditions presented in Figures 7 and 8, the values of the parameters of the function, together with r^2, are given in Table 4.

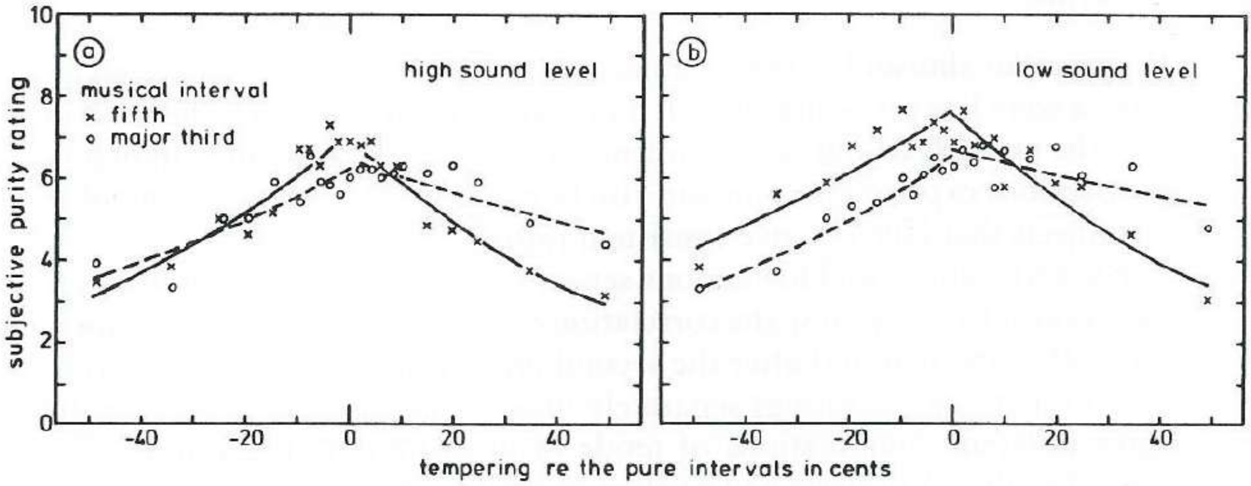

Fig. 7. Mean subjective purity ratings for intervals that comprised two simultaneous sinusoidal tones. The ratings are given for the fifth and the major third, separately, and are plotted as a function of tempering. In panel a, the sound pressure levels of the tones were 25 dB higher than the levels of the tones in panel b. The curves represent the best exponential fit.

TABLE 4
Values of *a*, *b*, and r^2 for All Conditions in Experiment 2a[a]

Mode of Presentation	Sound Level	Musical Interval	Tempering (cents) ≤0 *a*	≤0 *b*	≤0 r^2	≥0 *a*	≥0 *b*	≥0 r^2
Ratings								
Simultaneous	High	Fifth	7.1	0.016	0.92	6.9	−0.017	0.95
		Major third	6.2	0.011	0.75	6.4	−0.007	0.74
	Low	Fifth	7.7	0.011	0.81	7.6	−0.016	0.88
		Major third	6.6	0.014	0.94	6.7	−0.005	0.48
Successive	High	Fifth	7.4	0.009	0.75	7.2	−0.007	0.82
		Major third	6.4	0.017	0.93	6.3	−0.001	0.03
Paired comparison judgments[b]								
Simultaneous	High	Fifth	2.5	0.012	0.96	2.6	−0.017	0.96
		Major third	2.2	0.006	0.88	2.3	−0.008	0.97
	Low	Fifth	2.4	0.010	0.92	2.5	−0.015	0.83
		Major third	2.2	0.006	0.77	2.3	−0.008	0.84
Successive	High	Fifth	2.2	0.008	0.83	2.4	−0.008	0.91
		Major third	2.2	0.011	0.99	2.3	−0.004	0.80
	Low	Fifth	2.3	0.012	0.95	2.4	−0.007	0.68

[a]The relationship between subjective purity (y) and the amount of tempering (x) in cents can be described by means of the exponential function $y = a\exp(bx)$. Subjective purity was expressed either as ratings between 1 (very impure) and 10 (very pure) or as scale values that were derived from preference data collected by means of the method of paired comparisons.

[b]Predicted z scores can be obtained when the values that result from the exponential functions are reduced by 2.

Although for most sets the exponential fits still explain considerable amounts of the variances in the mean ratings, it is clear that (especially for small degrees of tempering) the trends in the data are less stable than those found in Experiment 1. This is quantitatively expressed in Table 5 where, exactly as in Table 3, the results of Newman–Keuls paired comparison tests are summarized.

In Figure 7a, it can be seen that, for simultaneous tones at the high sound level, especially the subjective purity of fifths was highly affected by tempering (see Table 6, in which for each set of ratings presented in Figures 7a, 7b, and 8, the results of an analysis of variance performed on these ratings are summarized). The effect of tempering depended on musical interval, the most prominent effect being that, for temperings greater than about 10 cents, stretched major thirds were rated significantly purer than the corresponding fifths. The interaction pattern resembles that depicted in, for example, Figure 2d. In Figure 7b it can be seen that, for intervals presented at a low sound level, compressed fifths were rated higher than the corresponding major thirds, whereas for stretched intervals the rating patterns are similar to those in the high level condition. Put differently, the 25-dB decrease in sound level increased all purity ratings, except those of the compressed major thirds. This effect, tested in a separate analysis of variance, is significant ($p < .001$).

Differences between the ratings for compressed fifths and major thirds were still larger for successive tones (Figure 8). The ratings of stretched fifths decreased with degree of tempering. The extent to which major thirds were stretched, however, had hardly any effect on the ratings.

It is interesting to compare the ratings of the tempered intervals in the successive mode of presentation with the ratings of the simultaneous sinusoids at the low sound level because it can be expected that for both condi-

TABLE 5

Temperings in Cents at which the Mean Subjective Purity Ratings Were Significantly Lower Than the Mean Ratings at the Tempering of 0 Cents[a]

		Musical Interval			
		Fifth		Major Third	
Mode of Presentation	Sound Level	Compressed	Stretched	Compressed	Stretched
Simultaneous	High	−15	15	−35	50
	Low	−35*	35	−35	50
Successive	High	−50	50	−35	—

[a]Differences between the mean values were tested by means of Newman–Keuls tests (see Table 3) for the 12 sets of means separately. For 10 conditions the mean values were significantly different at the 0.01 level; in the condition marked with an asterisk, this level was 0.05.

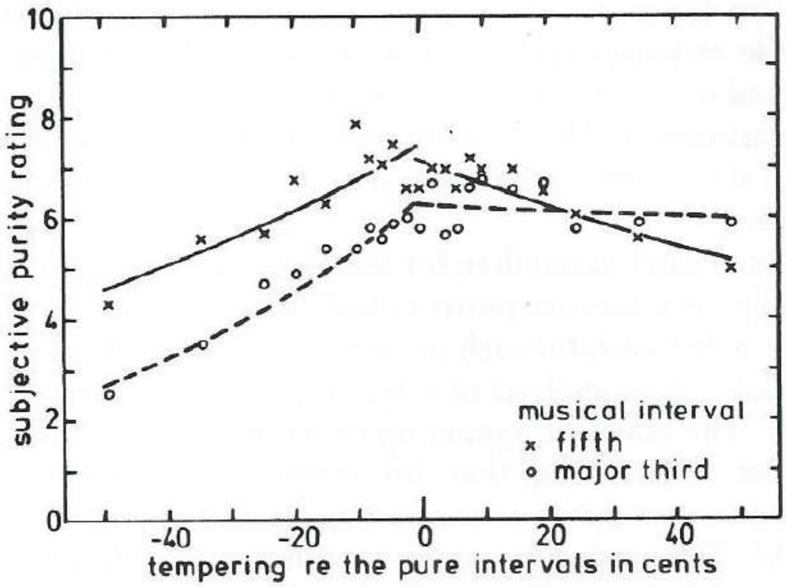

Fig. 8. Mean subjective purity ratings for intervals that comprised two successive sinusoidal tones presented at the high sound level (cf. Figure 7a). The ratings are given for the fifth and the major third, separately, and are plotted as a function of tempering. The curves represent the best exponential fit.

tions, differentiation in purity ratings will be based on differences in perceived interval size rather than on differences in the perception of beats. Therefore, the results from Figures 7b and 8 are replotted in Figures 9a and 9b for the fifth and the major third, respectively. In these figures it can be seen that, both for compressed fifths and major thirds, the ratings for the simultaneous and the successive tones were about equal. This was confirmed by the results of analyses of variance, performed separately on the ratings for the fifth and the major third. For moderately tempered intervals, the same holds for the stretched intervals. When the intervals are stretched by more than about 25 cents (for the fifths, $p < .005$) or by about 50 cents (for the major thirds, $p < .03$), however, successive tones are rated purer than simultaneous tones.

Paired-Comparison Judgments. In general, the results are similar to those obtained for the ratings. Again, the exponential fit (see Table 4 for values of a, b, and r^2) is slightly better than the linear fit in only a few conditions. It can be concluded from the data in Figure 10, that for the conditions in which the simultaneous tones were presented at a high sound level, the exponential fit is better for the fifth, but that for the major third the linear fit would be about equally adequate.

For successive tones, subjective distances between tempered fifths presented at a low sound level were of the same magnitude as those for fifths presented at a high sound level, as was expected.

The results that were obtained in the successive and the simultaneous modes of presentation are given in Figure 11. These results closely resemble those obtained with the ratings on the 10-point scale (cf. Figure 9).

TABLE 6

Summary of Three ANOVAs Performed on the Purity Ratings from Experiment 2[a]

			Simultaneous Tones								Successive Tones			
			High Sound Level				Low Sound Level							
Source	df_1	df_2	MS	*F* ratio	*p*	$\hat{\omega}^2$	MS	*F* Ratio	*p*	$\hat{\omega}^2$	MS	*F* Ratio	*p*	$\hat{\omega}^2$
Within cell	1008		4.14			61.4	4.28			67.8	4.90			76.7
Between cell														
S	23	1008	36.42	8.79	0.00001	5.7	28.65	6.7	0.00001	4.6	24.01	4.90	0.00001	3.6
A	1	23	1.56	0.03	0.86	0.0	112.40	4.0	0.054	0.7	402.7	28.30	0.00009	3.0
B	20	460	97.95	19.10	0.00001	13.6	88.88	17.1	0.00001	13.1	70.88	12.20	0.00001	10.1
A×B	20	460	22.31	4.03	0.00001	2.5	26.56	5.4	0.00001	3.4	17.60	3.70	0.00001	2.0
A×S	23	1008	54.30	13.10	0.00001	8.8	27.93	6.5	0.00001	4.5	14.24	2.91	0.00003	1.7
B×S	460	1008	5.13	1.24	0.003	3.5	5.20	1.2	0.007	3.5	5.81	1.18	0.02	3.4
A×B×S	460	1008	5.54	1.34	0.0002	4.9	4.93	1.2	0.04	2.5	4.76	0.97	0.64	0.0

[a]For all three data sets we used a 24 [subjects (*S*)] × 2 [musical intervals (*A*)] × 21 [temperings (*B*)] factorial design, all repeated measures and with replication as a within-cell variable. MS, df_1, df_2, p, and $\hat{\omega}^2$ are defined in Table 1.

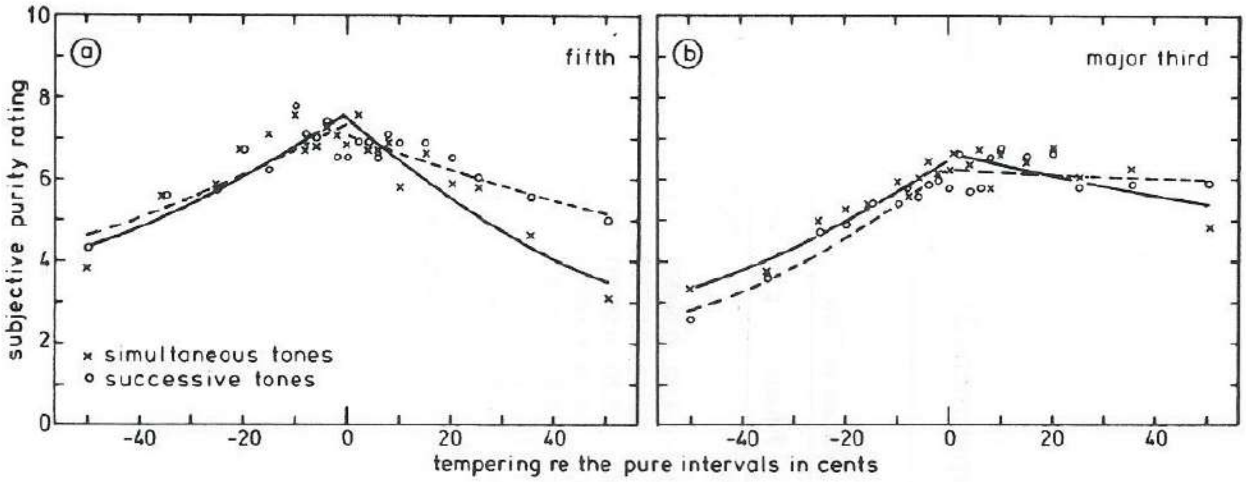

Fig. 9. Mean subjective purity ratings for intervals that consisted of two sinusoidal tones presented at the low sound level in the simultaneous condition and at the high sound level in the successive condition. The ratings are plotted as a function of tempering for the fifth and the major third in panels a and b, respectively. The curves represent the best exponential fit.

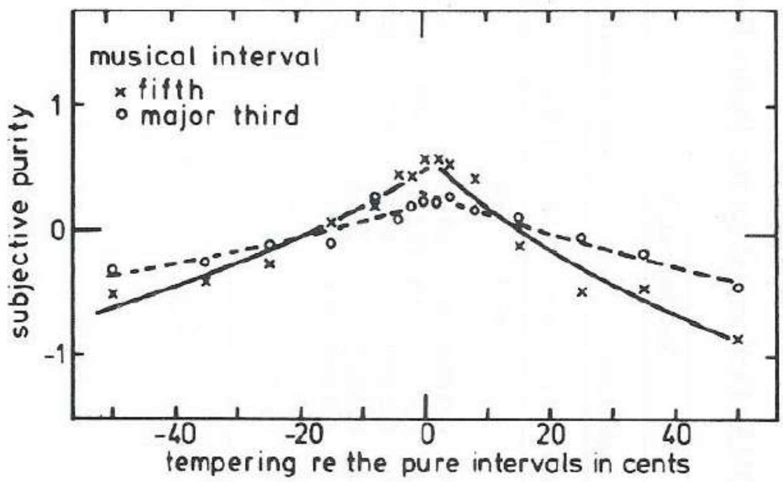

Fig. 10. Scale values for subjective purity for intervals that comprised two simultaneous sinusoidal tones, presented at the high sound level. The scale values, derived from preference data, are plotted as a function of tempering for the fifth and the major third, separately. The curves represent the best exponential fit.

General Discussion

Subjective purity ratings of tempered musical intervals were mainly determined by the interference of the higher harmonics of the lower tone with those of the higher tone. This may be concluded from our finding that the subjective differences between pure and tempered intervals strongly decreased when the spectral content of the tones was reduced from very rich to minimum. For simultaneous tones presented in various conditions, the size of the effect is summarized in Figure 12 for the fifth and the major third,

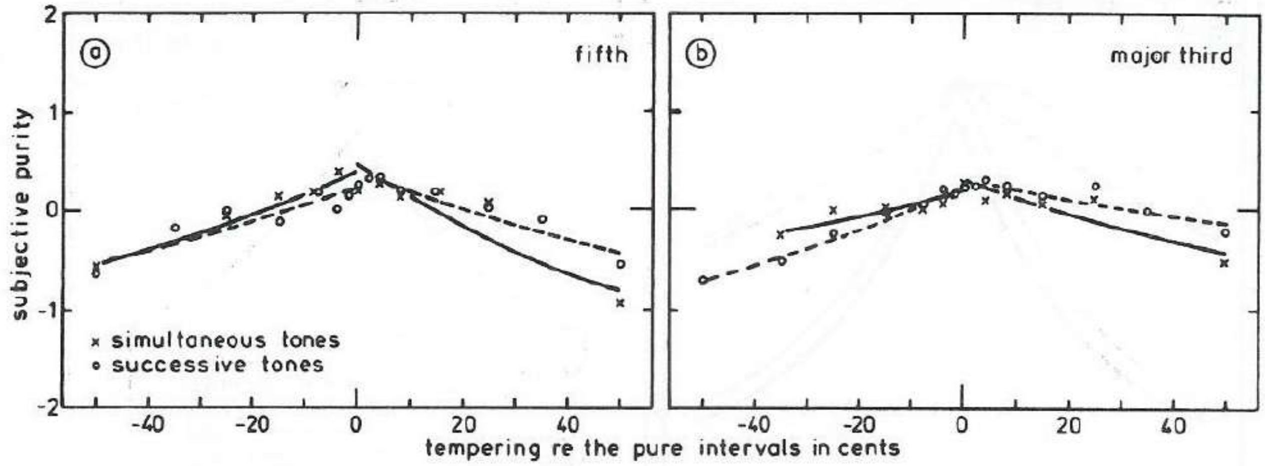

Fig. 11. Scale values for subjective purity derived from preference data, plotted as a function of tempering for the fifth and the major third in panels a and b, respectively. The intervals consisted of two sinusoids presented at the low sound level in the simultaneous condition and at the high sound level in the successive condition.

separately. In this figure, subjective ratings are given relative to the ratings at physical purity. To emphasize that for three of the four conditions in which sinusoidal tones were presented the ratings for temperings between ±35 cents were not significantly different (see Table 5), we summarized those ratings by smoothed curves, fitted by eye.

For all conditions presented in Figure 12, perception of differences in interval size may have influenced the ratings. In the conditions in which two simultaneous sinusoids were presented at the low sound level we tried to isolate the contribution of this aspect from the aspect of beats and roughness. Because the ratings in this simultaneous condition were about equal to the ratings for successive sinusoids (Figures 9 and 11), isolation of the two different aspects seems to have been successful. Therefore, we may conclude that the strong differentiation in rating the pure and tempered intervals, as found for complex tones, can be mainly ascribed to the presence of beats or roughness.

On the basis of the present conditions, it cannot be determined to what extent the effect of beats has been intensified by evaluation of the interval size on its own, as predicted by, for example, the additive model (see Figure 6). A comparison of the ratings obtained in the present experiments with those obtained for amplitude modulated pure intervals at modulation frequencies equal to $(pf_2 - qf_1)$ Hz would provide a better basis for answering this question.

On the other hand, it is conceivable that for complex tones the rating curves are determined by the sensation of beats or roughness only. For a number of conditions this is suggested by the resemblance of our subjective purity functions to the consonance functions that are predicted by models of Plomp and Levelt (1965) and Kameoka and Kuriyagawa (1969a).

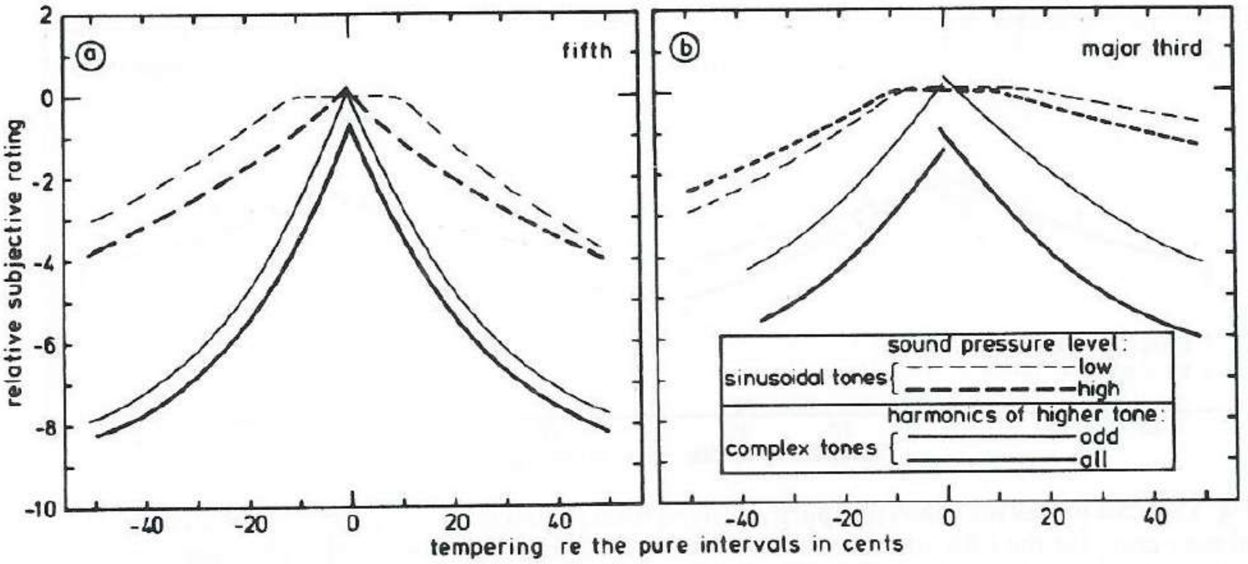

Fig. 12. Mean subjective purity ratings expressed relative to the mean rating for the physically pure interval, presented as a function of tempering for the fifth and the major third in panels a and b, respectively. Spectral content and sound-pressure level of the tones are given as parameters.

Tonal Consonance/Dissonance

Subjective Purity and Tonal Consonance

For isolated intervals, consonance/dissonance, as the terms were used by Helmholtz (1877/1954), was equivalent to "pleasantness/unpleasantness," and according to him this experience, being sensory in nature, was directly dependent on absence or presence of beats. Consonance of simultaneous sinusoids has been studied in, among others, the rating experiments of Plomp and Levelt (1965). When their subjects asked for the meaning of "consonant," this term was defined as "beautiful" and "euphonious." Results obtained by means of Osgood's method of the semantic differential, in which a large set of musical intervals had to be rated on various verbally labeled scales, had shown that "consonant," "beautiful," and "euphonious" are highly correlated (van de Geer, Levelt, & Plomp, 1962). In addition, analysis of the scores reported by van de Geer et al. (1962) reveals that there was also a high negative correlation (−0.87) between "rough" and "consonant." Although Kameoka and Kuriyagawa (1969b) defined consonance as the sensation of "clearness" and dissonance as the sensation of "turbidity," they felt that their concept of consonance was basically equivalent to the tonal consonance defined by Plomp and Levelt (1965).

Terhardt (1977) required his subjects to adjust the modulation depth of sinusoidally amplitude-modulated sinusoids with fixed carriers and modulation frequencies in such a way that the resulting "roughness" was equal to the "roughness" experienced from various pairs of simultaneous sinusoids. Since the obtained roughness curves resembled inverted consonance functions, Terhardt's data suggest that, for isolated simultaneous sinusoids, evaluation of tonal consonance is equivalent to evaluation of roughness.

Guthrie and Morrill (1928) defined consonance as a combination of "purity," "smoothness," and "good-blending." Their experimental results indicated that "consonance" as described and "pleasantness" led to similar ratings. In a recent experiment in which various triads with complex tones were evaluated, Roberts and Mathews (1984) found that for tempered triads "in-tuneness," "smoothness," and "pleasantness" judgments were all very similar, which is consistent with the results of Guthrie and Morrill (1928). In the evaluations of complex-tone intervals, reported by Pierce (1966) and by Slaymaker (1970), consonance was defined in terms of "concordance" and "smoothness" or "in-tuneness," respectively.

The definitions of consonance, summarized so far, all seem to emphasize that the degree of consonance corresponds to the degree to which roughness is absent, while in our experiments, too, subjective purity is considerably affected by beats or roughness. On the assumption that consonance and purity are basically equivalent concepts, it is worthwhile to compare our results more closely with the predictions from the models on tonal consonance.

Previous Tests of Predictions from Consonance Models

The models on tonal consonance/dissonance, which are basically an extension of Helmholtz's consonance theory, seem to be of significance to explain a number of aspects in music perception and are therefore of interest to the psychology of music (see, e.g., Deutsch, 1982). In view of this interest, it is surprising that for complex tones the predictions from Plomp and Levelt's (1965) model have never been tested, and the predictions from Kameoka and Kuriyagawa's model (1969a) have only partially been tested.

Because Kameoka and Kuriyagawa did not present pure and tempered major thirds in their verification experiments, predictions from their model could not be tested for these intervals. Although the set of intervals to be judged by their subjects included a pure fifth, the nearest two intervals tested were 136 cents smaller and 90 cents larger than a pure fifth, respectively, and can therefore not be considered as tempered fifths. Also, for tones with either four even or four odd harmonics, the two intervals closest to the pure fifth deviated by −52 and 98 cents from that fifth, respectively.

Our ratings for the pure and tempered major thirds and fifths provide the first occasion for a careful test of the validity of the present consonance models. Because in our Experiment 1 both musical interval type and spectral content of the tones were varied within the same block of intervals, the relative effects of these factors can be assessed as well.

The Model of Plomp and Levelt: Description and Predictions

In models on tonal consonance, total dissonance of musical intervals is given by somehow adding the dissonance of each pair of interfering harmonics. As a first approximation, Plomp and Levelt (1965) computed total dissonance of the musical intervals by simply adding the various subdissonances. According to Plomp and Levelt, these subdissonances are roughly about maximum at 25% and minimum at 0% and 110% of the critical bandwidth. Since the width of the critical band depends on the mean frequency of its lower and higher bounds, predicted total dissonance of a musical interval therefore depends on (1) the fundamental frequency of the tones and (2) the frequency distance between the various interfering harmonics. For the various conditions investigated in Experiment 1, we computed total dissonance with the help of a computer program. In our computations, we used an estimation of the critical bandwidth that is based on experiments in which presence versus absence of beats or roughness, rather than loudness summation, was investigated. Such experiments have been carried out by Cross and Goodwin (1893), by Mayer (1894), and by Plomp and Steeneken (1968). According to Vos and van Vianen (1985b), this critical bandwidth (CBW, in cents), being dependent on the mean frequency ($\bar{f}$) of its lower and higher bounds, can be represented by

$$\mathrm{CBW} = 2371\,[\log_{10}(\bar{f})]^{-2.06} \qquad (1)$$

For a number of spectral conditions, predicted dissonance is given in Figures 13 and 14 for a lower tone having a fixed frequency of 250 Hz and a higher tone that varies in frequency from 250 to 500 Hz.

In Figures 13a and 13b, dissonance predicted by the model of Plomp and Levelt is given for two tones both consisting of the first n harmonics, n varying from 1 to 20. In Figure 13a it can be seen how the number of peaks for intervals with simple fundamental frequency ratios increases with the number of harmonics. For tempered fifths between, say, 6.5 and 7.5 semitones (1 semitone = 100 cents), for example, the increased dissonance for $n = 3$ is caused by the interference of the third harmonic of the lower tone and the second harmonic of the higher tone. The point with relatively low dissonance (or high consonance) for the pure fifth (2:3, or 7.02 semitones) results from coincidence of these two harmonics.

In Figures 14a and 14b the number of harmonics of the lower tone is in the same way varied between 3 and 20. For the higher tone, however, the even harmonics have been deleted. As a result, there will be no peaks for pure intervals with fundamental frequency ratios in which the first integer is an even number. In Figure 14a it is shown that for $n = 5$ the area with relatively low dissonance for tempered fifths is caused by the dissonance of stretched fourths (3:4) and compressed major sixths (3:5).

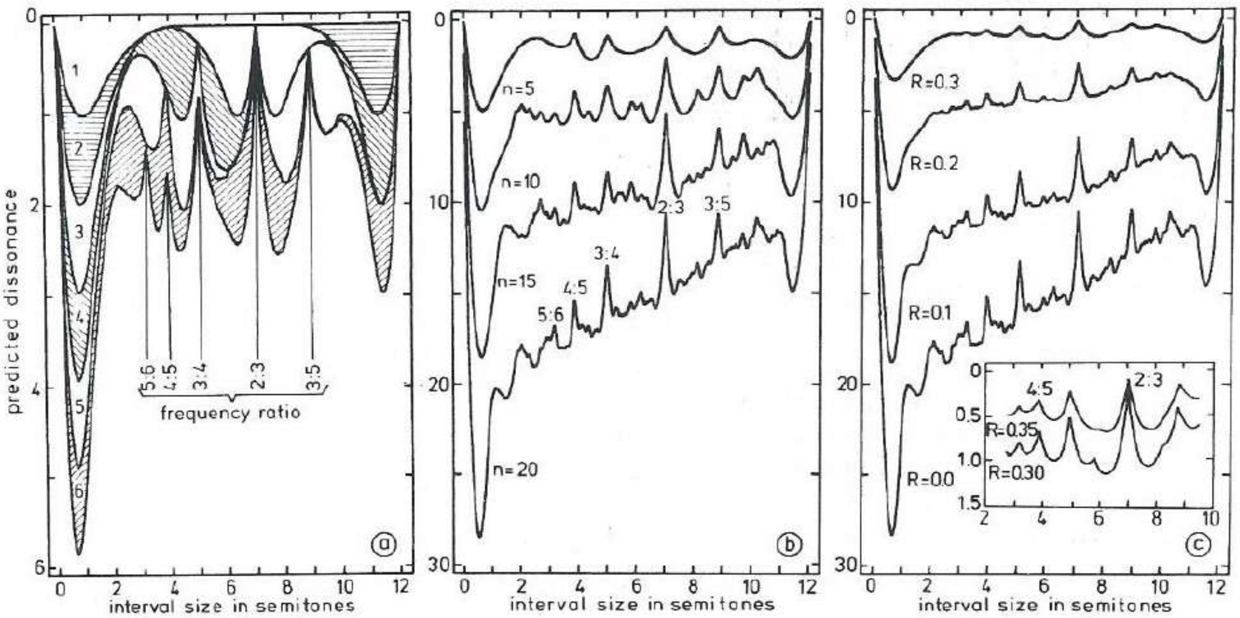

Fig. 13. Dissonance, predicted from the model of Plomp and Levelt (1965) for a lower tone having a fixed fundamental frequency of 250 Hz and a higher tone that varies in fundamental frequency from 250 to 500 Hz. In panels a and b the number of harmonics (n) of these two tones is given as a parameter. For $n = 20$ the reduction (R) of the subdissonances, expressed in a fraction per octave, is given as a parameter in panel c.

Comparison of Our Purity Ratings with These Predictions

For simultaneous sinusoids ($n = 1$), the dissonance predicted by the model of Plomp and Levelt (1965) is given in Figure 13a. For fifths, predicted dissonance is zero and for major thirds, it is close to zero. For both intervals that were presented at the low sound level in Experiment 2, the small discrepancy between ratings and predictions may be explained by the effect of perceived differences in interval size. For the fifths at the high sound level, however, differentiation of the ratings is consistently more pronounced than for the ratings at the low sound level and for those given in the successive conditions. This suggests that, at least for fifths at the high sound level, the purity pattern is determined by beats and roughness, which is consistent with previous findings (see Plomp, 1967). Plomp and Levelt's model does not account for this effect.

To compare the purity ratings of the conditions in which both the lower and the higher tone comprised 20 harmonics, we had to extend the computations. The results, as well as a number of intermediate stages with $n = 5$, 10, and 15 harmonics, are given in Figure 13b. It can be seen in this figure that for the fifth the predicted pattern resembles the rating pattern obtained. Also, for the major third, pure and slightly tempered intervals are predicted to be less dissonant than the considerably tempered intervals.

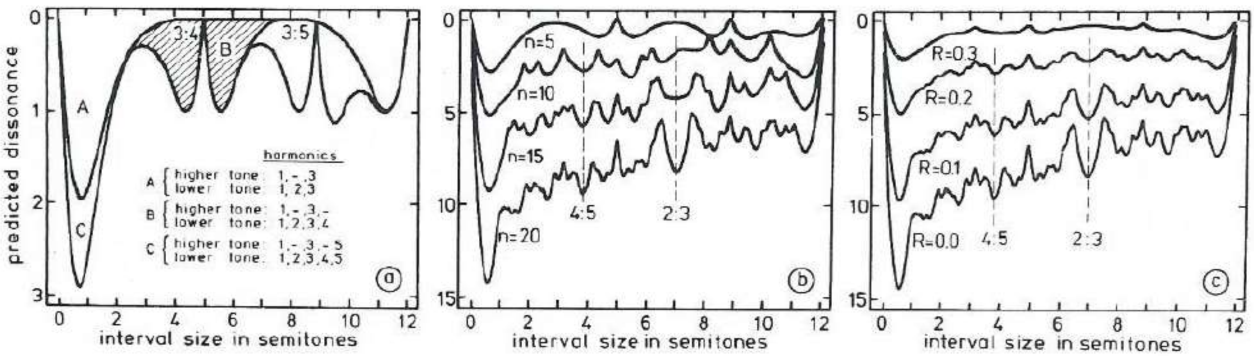

Fig. 14. Dissonance, predicted from the model of Plomp and Levelt (1965), again, as in Figure 13, for a lower tone with a fixed and for a higher tone with a variable fundamental frequency. As indicated in panel a, however, the spectrum of the higher tone comprises only odd harmonics; *n* refers to the number of harmonics of the lower tone.

Quantitatively, however, the predicted difference between the peaks for the major third and the fifth is much larger than the difference obtained in our experiments (see Figures 2a and 2c).

Because in Experiment 1 the levels of the harmonics were not equal but, from the fundamental upwards, dropped off by 6 dB/octave, the subdissonances resulting from interference of higher harmonics may be given some weighting. Plomp and Levelt's model is not explicit about application of a weighting procedure. It seems plausible to apply a weighting (W) of the subdissonances that reduces the subdissonances by a certain fraction (R) per octave [= 12 semitones (S)]:

$$W = 1 - (S/12)R. \qquad (2)$$

For our stimuli, Kameoka and Kuriyagawa's model would prescribe a weighting procedure in which the subdissonances are reduced by about 10% per octave. In addition to $R = 0.1$, we extended computation of total dissonance for $R = 0.2$–0.35. It can be seen in Figure 13c that the difference between the peaks of the major third and the fifth decreases with our measure of weighting the subdissonances. In the insert of Figure 13c it is shown that, to obtain a reasonable correspondence between predictions and experimental results, a reduction of 0.35 is needed. The high reduction of 0.35 per octave implies that subdissonances that are the result of interference with the eighth and higher harmonics of the lower tone are all set to zero.

Purity ratings from the conditions in which the higher tone comprised only odd harmonics can be compared with the predictions given in Figure 14b. First, it can be seen that for this spectral condition predicted dissonance is about a factor of two lower than for the spectral condition in which both tones comprised all first 20 harmonics. Neither for the fifth, nor for the major third (see, e.g., Figures 5 and 12) were the ratings affected to such an extent. Second, for the fifth and the major third, consonance valleys rather than peaks are predicted. It is shown in Figure 14c that reduction of subdissonances resulting from higher harmonics only predicts that, in the end, tempering of these intervals has hardly any effect on dissonance.

The Model of Kameoka and Kuriyagawa: Description and Predictions

Basically, the model of Kameoka and Kuriyagawa (1969a) is an extension of Plomp and Levelt's model. In their model, the subdissonances not only depend on (1) the frequency of the tones and (2) their frequency distance, but also on (3) their sound-pressure levels. In addition (4) the way in which Kameoka and Kuriyagawa add these subdissonances is much more complex than the procedure followed by Plomp and Levelt.

We computed total absolute dissonance with the help of the equations given by Kameoka and Kuriyagawa (1969a), and also the exponents and constants were set at the values proposed. However, one adjustment had to be made: When the frequency difference between two interfering harmonics is equal to 1% of the frequency of the lower harmonic (that is, for a tempering of 17 cents), Kameoka and Kuriyagawa set relative dissonance at zero (see Kameoka & Kuriyagawa, 1969a, Eq. 7, p. 1462). Moreover, in their model relative dissonance is undefined when this frequency difference is less than 1% of the frequency of the lower harmonic. To make it possible for our computations to be performed in steps of 1 cent, the relative dissonance for temperings of between 0 cents (at which relative dissonance is always 0) and 36 cents was obtained by interpolation.

From a comparison of our predicted absolute dissonance, computed in steps of 1 cent, with the predictions that, for identical acoustic stimuli, were obtained by Kameoka and Kuriyagawa, computed in steps of 25 cents, we concluded that the peaks for intervals with simple fundamental frequency ratios are not as sharp as Kameoka and Kuriyagawa led us to believe (see Appendix).

Again, we computed predicted absolute dissonance for a lower tone, having a fixed frequency of 250 Hz and a higher tone that varied in frequency from 250 to 500 Hz. The spectral contents of these tones were identical to those presented in our Experiment 1. The spectral envelopes of these tones coincided for both the major third and the fifth. This means that for those harmonics that result in relative dissonance values larger than 0, the sound-pressure levels were not very different. Because of this, the effect of sound level on absolute dissonance could be accounted for by using Equation 13 from the model of Kameoka and Kuriyagawa (1969a).

In Figure 15a, predicted absolute dissonance is plotted for the two tones ($n = 20$) used in our Experiment 1, with the sound-pressure level of the first harmonic of the lower tone equal to 75 dB. In addition a number of intermediate stages with $n = 5$, $n = 10$, and $n = 15$ harmonics are given. In

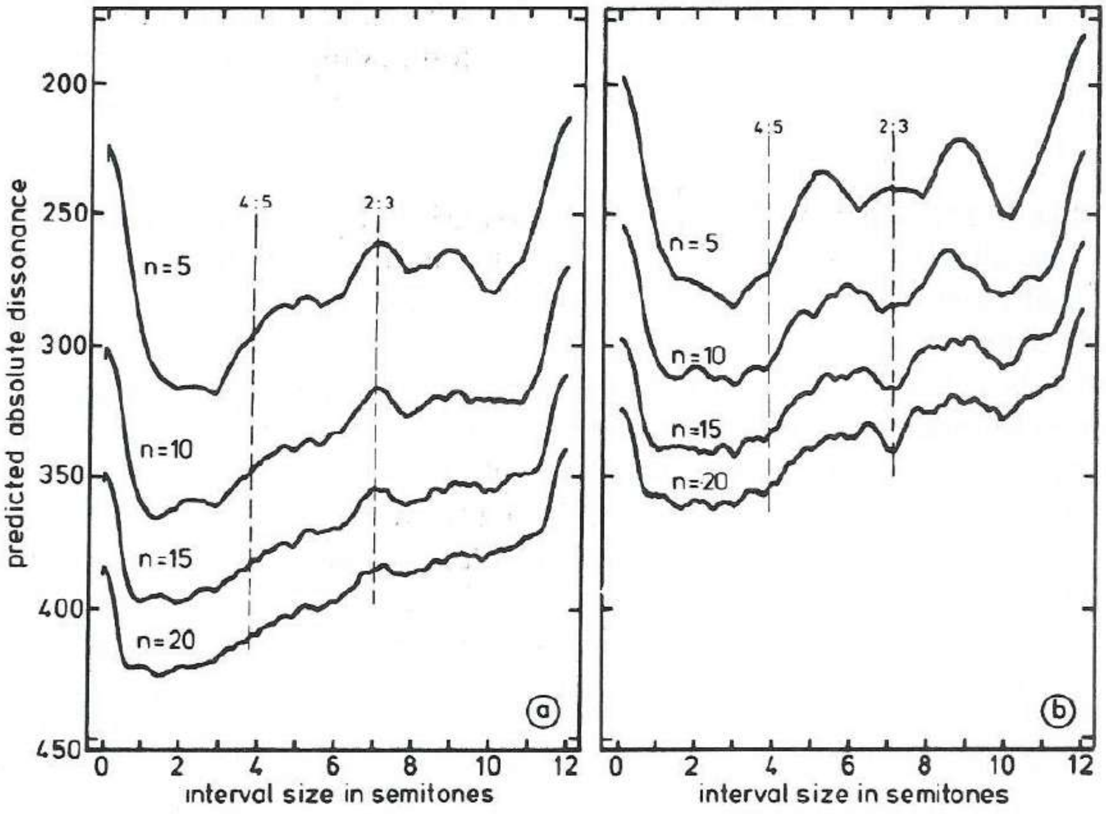

Fig. 15. Absolute dissonance, predicted from the model of Kameoka and Kuriyagawa (1969a) for a lower tone having a fixed fundamental frequency of 250 Hz and a higher tone that varies in fundamental frequency from 250 to 500 Hz. The number of harmonics (n) of the lower tone is given as a parameter. In panel a the spectrum of the higher tone was identical to that of the lower tone, whereas in panel b, the higher tone comprised only odd harmonics. For all conditions sound-pressure levels of the various harmonics were equal to those presented in our Experiment 1.

Figure 15b, predictions are depicted for similar conditions, except that the higher tone comprised only odd harmonics.

Comparison of Purity Ratings with Predictions of Absolute Dissonance

It can be seen in Figure 15a that for the conditions in which both the lower and the higher tone comprised 20 harmonics, predicted absolute dissonance is completely different from our experimental results. Dissonance minima for pure and slightly tempered intervals could not be found. This holds for both the fifths and the major thirds.

In the condition in which the higher tone comprised only odd harmonics, predictions (see Figure 15b) are also at variance with the purity ratings obtained (see, e.g., Figures 5 and 12). For a pure fifth, total absolute dissonance is predicted to be even higher than that for tempered fifths. The same prediction resulted from the model of Plomp and Levelt.

Possible Weaknesses of the Present Models for Tonal Consonance/ Dissonance

In the dissonance model of Kameoka and Kuriyagawa (1969a), total dissonance of two simultaneous complex tones was computed by combining all subdissonances resulting from the various interfering harmonics according to a power function (exponent = 0.25). Kameoka and Kuriyagawa failed to recognize that many of these subdissonances are the result of interference within different critical bands. Therefore, their model is at variance with Zwicker's established model of loudness summation (see, e.g., Zwicker & Scharf, 1965). In the Zwicker model, the power function (again with the exponent of 0.25) was used only to determine the specific loudness for every critical band, whereas the overall loudness of the total sound was proportional to the sum of the specific loudnesses.

In the model of Plomp and Levelt (1965), the subdissonances were added arithmetically. As long as these subdissonances have their origin in different critical bands, this procedure is comparable to integration of specific loudnesses as it occurs in Zwicker's model. For the fifth in Figure 13, for example, there are no pairs of interfering harmonics that fall within the same critical band (Eq. 1), provided that the spectra of the tones do not comprise more than about eight or nine successive harmonics. For smaller intervals, such as the major third, various pairs of interfering harmonics would have maximum excitation at places of the basilar membrane that fall within the same critical band. For these cases it may be preferred to combine the subdissonances according to a power function.

In the model of Kameoka and Kuriyagawa the concept of absolute dissonance was developed to account for the additional effect that, for example, the background noise in an auditorium may have on perceived dissonance. Kameoka and Kuriyagawa considered this noise a component that had to be added to the various (relative) subdissonances. This procedure is completely different from what one would expect to be adequate on the basis of the effect of partial masking: the loudness of a sound is decreased rather than increased when another sound with components in the same frequency range is added (see Zwicker & Scharf, 1965; Plomp, 1976). It is unlikely, however, that the predictive power of their model would benefit from a revision with respect to the way in which the effect of noise on perceived dissonance is accounted for (see Appendix).

The suggestions for extension or revision of these two models of tonal consonance/dissonance raised here are based on the assumption that the perception of beats and roughness is related to the perception of overall loudness. Further research will be required to test this hypothesis and, especially with respect to the effect of the spectral variations investigated in Experiment 1, to develop a model that is able to account more precisely for the purity ratings we have obtained.

Prediction of Subjective Purity in Two-Part Music Pieces

In the introduction, we argued that knowledge of the relationship between tempering and purity rating of isolated intervals would be of help in understanding how context affects overall acceptability of a particular piece of music. It seems that the results of studies such as the present one are badly needed if context effects are to be studied systematically: Even in contextually neutral conditions, it is not a priori clear how the purity ratings for isolated intervals, as obtained in our Experiment 1, are integrated when an overall rating of a piece of music is required.

In a two-part piece of music, which in our example only consists of fifths and major thirds (with complex tones), the overall rating may be determined by the rating of the worst interval. For regular 12-tone tuning systems, in which the temperings of the intervals are all determined by the tempering of the fifth (Rasch, 1983), this hypothetical result is represented by the curve for the isolated major third in Figure 16. This curve, adopted from Figure 2a and replotted as a function of the tempering of the fifth in the corresponding tuning system, predicts that only the meantone tuning system (pure major thirds) will be acceptable. If overall rating is determined by the mean of the ratings of the fifths and the major thirds (average model), or by the sum of the impurity ratings (additive model), all systems between meantone and Pythagorean tuning (pure fifths) are about equally acceptable, whereas other regular systems are unacceptable (see Figure 16). The implications of the latter hypotheses are in agreement with the observation that the majority of the 12-tone regular tuning systems proposed in the literature have fifths that are indeed tempered by between 0 and −5.4 cents, whereas tuning systems with fifths tempered beyond this range, such as, for example, systems proposed by Zarlino (−6.1 cents) and Salinas (−7.2 cents) have largely remained theoretical.

General Conclusions

For two simultaneous complex tones that constitute either tempered major thirds or tempered fifths, the relationships between subjective purity and the amount of tempering can be adequately described by exponential functions. These exponential functions are obtained both for ratings on a 10-point equal-interval scale and for subjective distances derived from preference data collected by means of the method of paired comparisons.

Fifths compressed by up to about 15–25 cents or stretched by up to about 10–20 cents were rated to be purer than the corresponding major thirds. Especially for stretched intervals, the reverse tends to be the case with larger temperings.

Deletion of the even harmonics of the higher tone, as a result of which interference of various pairs of nearly coinciding harmonics was canceled,

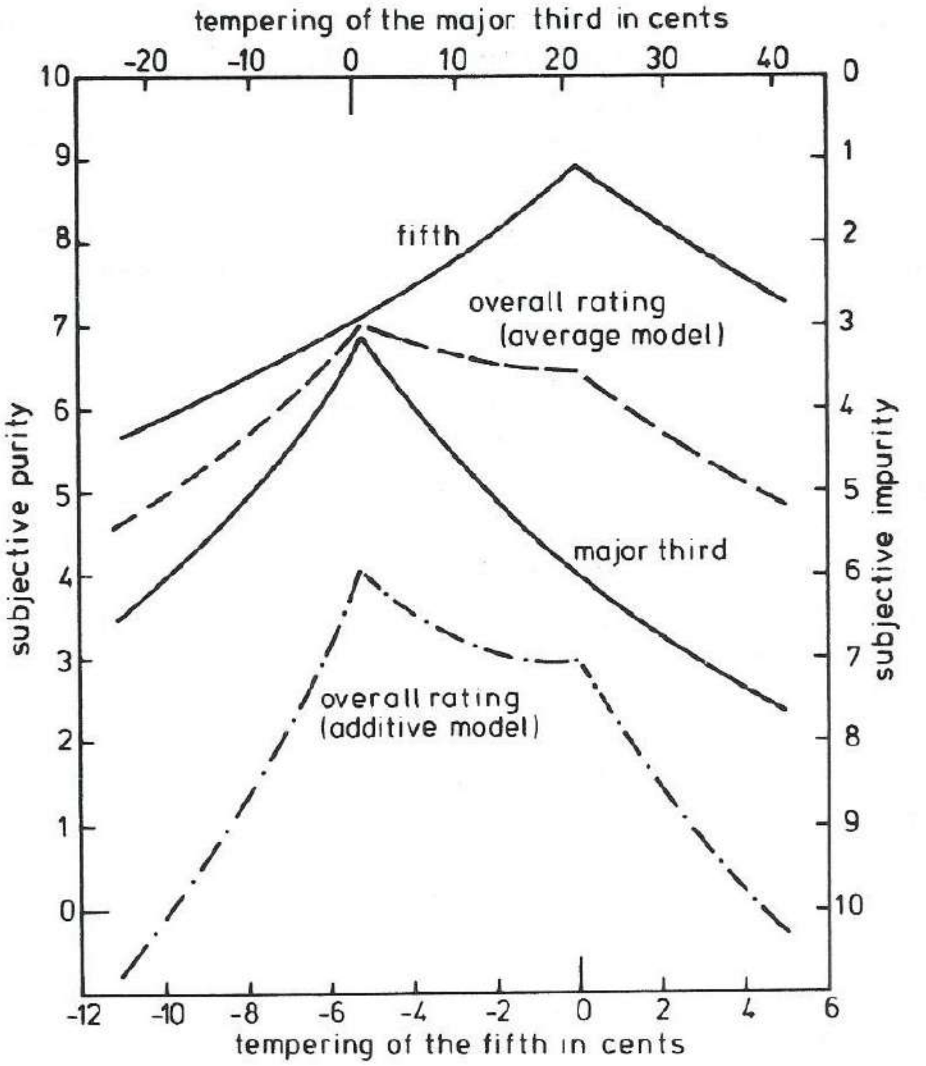

Fig. 16. Purity ratings for the fifth and the major third (from Fig. 2a) plotted as a function of the tempering of the fifth in regular 12-tone tuning systems. The dashed curves represent hypothetical conditions in which the overall purity rating of a piece of music comprising fifths and major thirds, is determined by either the mean of the purity ratings or the sum of the impurity ratings of these intervals. Both overall rating curves suggest that in contextually neutral conditions, all tuning systems between 0 cents (Pythagorean tuning) and −5.4 cents (meantone tuning) would be about equally acceptable.

resulted in higher purity ratings. This effect was most prominent for the major third.

Presentation of the purity ratings as a function of beat frequency (pf_2-qf_1) Hz increases rather than decreases the differences between the fifth and the major third. Part of this difference, however, may be explained by differences in the perceived strength of the beats.

For two simultaneous sinusoidal tones subjective differences between pure and tempered intervals are much smaller than for complex tones. When the sinusoidal tones are presented at a low sound level, purity ratings for simultaneous tones are about equal to the ratings for successive tones.

The strong differentiation in rating between the pure and tempered intervals, as found for complex tones, may be ascribed mainly to the presence of beats or roughness. Therefore, the purity ratings obtained in the present study may be compared with dissonance patterns predicted by two models for tonal consonance/dissonance.

For tones comprising a complete series of successive harmonics the patterns predicted by the model of Plomp and Levelt (1965) resemble the rating patterns obtained. Quantitatively, however, the difference between the peaks for the major third and the fifth, as obtained in our experiments, is predicted when interference due to only the first 7 instead of the first 20 harmonics is taken into account. For the same spectral conditions, the dissonance patterns predicted by the model of Kameoka and Kuriyagawa (1969a) are completely different from our experimental results.

For the conditions in which the higher tone comprised only odd harmonics, the predictions from both the model of Plomp and Levelt and that of Kameoka and Kuriyagawa are at variance with the purity ratings obtained.

For conditions in which the tones comprise many relatively strong harmonics, the predictive power of the two consonance/dissonance models may be enhanced by adoption of the principles underlying the loudness summation model of Zwicker.[2]

Appendix

Comments on Computations in the Consonance/Dissonance Model of Kameoka and Kuriyagawa

For a lower tone with a fixed frequency of 440 Hz and a higher tone that varies in frequency from 440 to 880 Hz, Kameoka and Kuriyagawa (1969a) computed absolute and relative dissonance from their consonance/dissonance model. In their Figure 8, for example, predicted dissonance was given for these tones, both of which consisted of the first n harmonics, n varying from 1 to 6. In our efforts to simulate their complex model we found, first, that we could not obtain peaks that were as sharp as those given by Kameoka and Kuriyagawa. If they had computed their dissonance curves in steps of 1 cent, as we did, instead of steps of 25 cents, they would have obtained the curves shown in Figure 17a, rather than the smoothed ones. For tempered fifths, fourths, and major thirds, predicted absolute dissonance curves are closer to the quadratic relationship given in our Figure 1 than to exponen-

2. Part of this research was supported by Grant 15-29-05 from the Netherlands Organization for the Advancement of Pure Research (ZWO). The author is grateful to Ino Flores d'Arcais, to Rudolf Rasch, and to Reinier Plomp for their comments on an earlier draft of this paper.

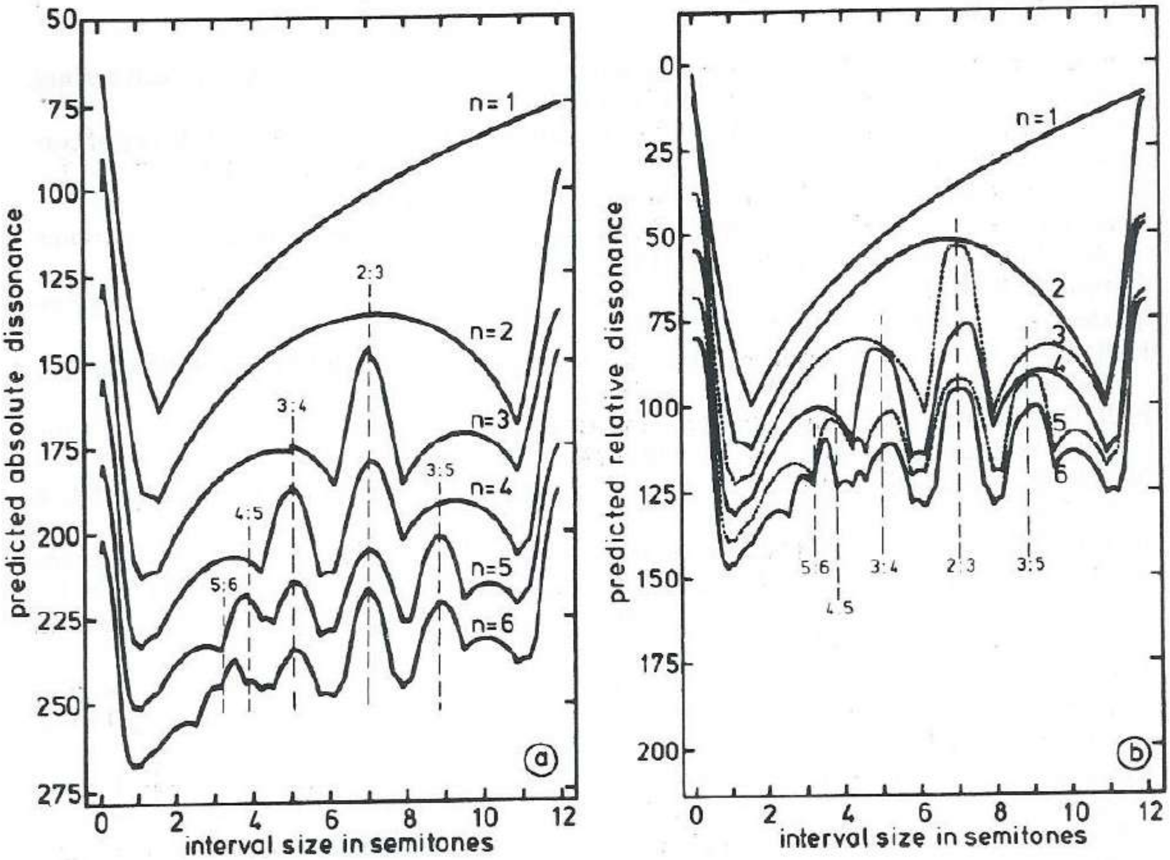

Fig. 17. Absolute dissonance (panel a) and relative dissonance (panel b), predicted from the model of Kameoka and Kuriyagawa (1969a) for a lower tone having a fixed fundamental frequency of 440 Hz and a higher tone that varies in fundamental frequency from 440 to 880 Hz. The number of harmonics (n) is given as a parameter. All harmonics have an equal sound-pressure level of 57 dB.

tial relationships. Second, it can be seen in Figure 17a that for $n = 6$ the dissonance model fails to predict dissonance minima for the intervals with small integer ratios of 4:5 and 5:6. Third, for $n > 1$, small discontinuities in the curves are obtained, which are the result of the difference between the dissonance for an interval size of 0 and 12 semitones. For $n = 3$ at an interval size of 498 cents, for example, the distance between the second harmonic of the lower tone and the third harmonic of the higher tone is equal to 1200 cents and, according to Kameoka and Kuriyagawa, produces a small amount of dissonance. At an interval size of 499 cents, however, the distance exceeds the octave, and the subdissonance is therefore set at 0.

Fourth, in Figure 17b, we demonstrate that, contrary to what Kameoka and Kuriyagawa suggest, the difference between relative and absolute dissonance implies more than a mere shift in the amount of dissonance. If the dissonance caused by, for example, the background noise in an auditorium, is not taken into account, even for fifth, fourth, and major sixth, temperings of between −20 and 20 cents are predicted to have hardly any effect on perceived relative dissonance at all.

References

Attneave, F., & Olson, R. K. Pitch as a medium: a new approach to psychophysical scaling. *American Journal of Psychology,* 1971, *84,* 147–166.

Cross, C. R., & Goodwin, H. M. Some considerations regarding Helmholtz's theory of consonance. *Proceedings of the American Academy of Arts and Sciences,* 1893, *27,* 1–12.

Deutsch, D. (Ed.). *The psychology of music.* New York: Academic Press, 1982.

Geer, J. P. van de, Levelt, W. J. M., & Plomp, R. The connotation of musical consonance. *Acta Psychologica,* 1962, *20,* 308–319.

Guthrie, E. R., & Morrill, H. The fusion of non-musical intervals. *American Journal of Psychology,* 1928, *40,* 624–625.

Hall, D. E. *The objective measurement of goodness-of-fit for tunings and temperaments. Journal of Music Theory,* 1973, *17,* 274–290.

Hall, D. E. Perception of musical interval mistuning. *Proceedings 11th International Congress on Acoustics,* Paris, 1983, *4,* 411–414.

Hall, D. E., & Hess, J. T. Perception of musical interval tuning. *Music Perception,* 1984, *2,* 166–195.

Helmholtz, H. L. F. von. *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (4th ed.). Braunschweig: Vieweg, 1877. [*On the sensations of tone as a physiological basis for the theory of music* (A. J. Ellis, Trans.). New York: Dover, 1954.]

Kameoka, A., & Kuriyagawa, M. Consonance theory, Part II: Consonance of complex tones and its calculation method. *Journal of the Acoustical Society of America,* 1969a, *45,* 1460–1469.

Kameoka, A., & Kuriyagawa, M. Consonance theory, Part I: Consonance of dyads. *Journal of the Acoustical Society of America,* 1969b, *45,* 1451–1459.

Killam, R. N., Lorton, P. V., & Schubert, E. D. Interval recognition: identification of harmonic and melodic intervals. *Journal of Music Theory,* 1975, *19,* 212–235.

Lottermoser, W., & Meyer, J. Frequenzmessungen an gesungenen Akkorden. *Acustica,* 1960, *10,* 181–184.

Mayer, A. M. Researches in Acoustics, No. IX. *Philosophical Magazine Journal of Science, Series 5,* 1894, *37,* 259–288.

Pierce, J. R. Attaining consonance in arbitrary scales. *Journal of the Acoustical Society of America,* 1966, *40,* 249.

Plomp, R. Beats of mistuned consonances. *Journal of the Acoustical Society of America,* 1967, *42,* 462–474.

Plomp, R. *Aspects of tone sensation.* London: Academic Press, 1976.

Plomp, R., & Levelt, W. J. M. Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America,* 1965, *38,* 548–560.

Plomp, R., & Steeneken, H. J. M. Interference between two simple tones. *Journal of the Acoustical Society of America,* 1968, *43,* 883–884.

Plomp, R., Wagenaar, W. A., & Mimpen, A. M. Musical interval recognition with simultaneous tones. *Acustica,* 1973, *29,* 101–109.

Rasch, R. A. Description of regular twelve-tone musical tunings. *Journal of the Acoustical Society of America,* 1983, *73*(3), 1023–1035.

Roberts, L. A., & Mathews, M. V. Intonation sensitivity for traditional and nontraditional chords. *Journal of the Acoustical Society of America,* 1984, *75,* 952–959.

Shackford, C. Some aspects of perception, Part I: Sizes of harmonic intervals in performance. *Journal of Music Theory,* 1961, *5,* 162–202.

Shackford, C. Some aspects of perception, Part II: Interval sizes and tonal dynamics in performance. *Journal of Music Theory,* 1962a, *6,* 66–90.

Shackford, C. Some aspects of perception, Part III: Addenda. *Journal of Music Theory,* 1962b, *6,* 295–303.

Slaymaker, F. H. Chords from tones having stretched partials. *Journal of the Acoustical Society of America,* 1970, *47,* 1569–1571.

Terhardt, E. The two-component theory of musical consonance. In E. F. Evans and J. P. Wilson (Eds.), *Psychophysics and physiology of hearing*. London: Academic Press, 1977, 381–390.

Terhardt, E. The concept of musical consonance: a link between music and psychoacoustics. *Music Perception,* 1984, *1*(3), 276–295.

Torgerson, W. S. *Theory and methods of scaling*. New York: Wiley & Sons, 1958.

Vos, J. The perception of pure and mistuned musical fifths and major thirds: Thresholds for discrimination, beats, and identification. *Perception & Psychophysics,* 1982, *32*(4), 297–313.

Vos, J. Spectral effects in the perception of pure and tempered intervals: discrimination and beats. *Perception & Psychophysics,* 1984, *35*(2), 173–185.

Vos, J., & Vianen, B. G. van. Thresholds for discrimination between pure and tempered intervals: the relevance of nearly coinciding harmonics. *Journal of the Acoustical Society of America,* 1985a, *77*(1), 176–187.

Vos, J., & Vianen, B. G. van. The effect of fundamental frequency on the discriminability between pure and tempered fifths and major thirds. *Perception & Psychophysics,* 1985b, *37,* 507–514.

Ward, W. D. Musical perception. In J. Tobias (Ed.,), *Foundations of modern auditory theory* (Vol. 1). New York: Academic Press, 1970, 407–447.

Winer, B. J. *Statistical principles in experimental design* (International student ed.). London: McGraw-Hill, 1970.

Zwicker, E., & Scharf, B. A model of loudness summation. *Psychological Review,* 1965, *72,* 3–26.

7

Subjective acceptability of various regular twelve-tone tuning systems in two-part musical fragments[1]

Musically trained subjects rated the overall acceptability of the performance of two-part musical fragments. With the help of a computer these fragments were performed according to various regular 12-tone tuning systems: Pythagorean tuning (tempering, A, of the main fifths equal to 0.0 cents), equal temperament (A=-2.0 cents), Silbermann (A=-3.9 cents), meantone (A=-5.4 cents), and Salinas tuning (A=-7.2 cents). In experiment 1 two systems in which A=2.0 or A=-10.0 cents were also included. The tones in both the lower and the higher parts were complex tones with a spectral-envelope slope of -6 dB/oct. Mean ratings were about the same for $-5.4 \leq A \leq 0.0$ cents, whereas for $A > 0.0$ and $A < -5.4$ cents the ratings strongly decreased. This effect of tuning system was also found when acceptability was determined by means of the method of paired comparisons. The specific effect of tuning system was not affected by the tempo (in the fast tempo tone duration was 75% shorter than in the slow tempo) at which the fragments were played. The perception of beats in the harmonic intervals was manipulated in experiment 2 by varying the spectral content of the tones. The condition in which the interference of the nearly coinciding harmonics was canceled resulted in higher acceptability. The effect of tuning system was the same as in experiment 1. In both experiments acceptability could be accurately predicted from a linear combination of the purity ratings of the harmonic fifths and major thirds. It is not precluded, however, that the subjects based their acceptability ratings partly on the subjective purity of the melodic intervals. This has to be tested in future research.

INTRODUCTION

Pure consonant intervals are characterized by small-integer frequency ratios, e.g., 1:2 for the octave, 2:3 for the fifth, 4:5 for the major third, etc. For simultaneous complex tones, slightly tempered consonant intervals are characterized by small frequency differences between those harmonics which coincide in pure intervals.

Perceptually, the interference of nearly coinciding harmonics in these tempered intervals gives rise to beats or roughness. In tuning fixed-pitch

---

[1]Parts of this research were presented at the Fifth Workshop on Physical and Neuropsychological Foundations of Music, Ossiach, Austria, August 6-10, 1985, and at the 12th International Congress on Acoustics, Toronto, Canada, July 24-31, 1986. A preliminary paper is included in the ICA Proceedings, Vol. 3, K5-8.

keyboard instruments such as the organ, the harpsichord, and the piano, it is inevitable that some of the intervals are tempered. Which intervals are tempered, and to what extent, is described in musical tuning systems. In the history of music, numerous tuning systems have been proposed (Barbour, 1951).

In this study we explored the differences in subjective evaluation between a number of these tuning systems. In the experiments musically trained subjects were requested to rate the overall musical acceptability of two-part fragments performed according to the various tuning systems.

Before describing the experiments, we shall elaborate on some music-theoretical and perceptual issues that are relevant to our experimental study.

I. RELEVANT ASPECTS

A. Regular 12-tone tuning systems

1. Tempering of the intervals

The majority of tuning systems applied in music are based on 12 different tones within the octave. Tuning systems are most appropriately characterized by the temperings of the fifths. In irregular 12-tone tuning systems, such as those constructed by, for example, Werckmeister and Neidhardt (Barbour, 1951; Dupont, 1935), at least two or more groups of fifths are tempered by different amounts. In regular 12-tone tuning systems, however, 11 fifths are tempered by the same amount. [In our experiments, enharmonically equivalent intervals -such as the wolf fifth G#-Eb, and the wolf major thirds C#-F, G#-C, B-Eb, and F#-Bb in the meantone tuning system - have never been used. In this study these intervals will therefore be left out of consideration. See Rasch (1983) for more details.] As a consequence, the temperings of the other musical intervals, such as, e.g., the major third, are fixed as well. As will be shown below, the temperings of the other intervals can be derived by adding the appropriate number of fifths and octaves.

In music theory, a major third equals the ratio between four fifths and two octaves:

$$\beta = \alpha^4/\omega^2, \qquad (1)$$

in which α, β, and ω are the ratios of the fundamental frequencies of the fifth, major third, and octave, respectively. Since the octave is pure (ω=2/1) in all tuning systems (see Kolinski, 1959, for an exception), Eq. (1) can be reduced to:

$$\beta = \alpha^4/4. \qquad (2)$$

A tempered major third, β, comprises the pure major third (5/4) and a very small tempering interval, b (β=5b/4). Likewise, a tempered fifth, α, comprises the pure fifth (3/2) and a small interval a (α=3a/2). Substituting (5b/4) and (3a/2) in Eq. (2) for β and α, respectively, results in:

$$5b/4 = (3a/2)^4/4, \qquad (3)$$

from which it follows that

$$b = (81/80)\ a^4. \qquad (4)$$

Both in musical and perceptual contexts, it is considered to be more appropriate to convert frequency ratios of musical intervals into logarithmic quantities (Barbour, 1940; Pikler, 1966; Young, 1939). Commonly, the logarithmic interval is expressed in cents; for a frequency ratio i, this interval is given by $1200\log_2(i)$. Logarithmic conversion of Eq. (4) results in:

$$B = S + 4A, \qquad (5)$$

in which S equals the syntonic comma (= 21.5 cents). The minor third equals the fifth minus the major third. The tempering (in cents) of the minor third, C, equals A-B. Substituting (S+4A) in C=A-B for B [see Eq. (5)] results in:

$$C = -(S + 3A). \qquad (6)$$

If the fifths are pure (A=0 cents), which is the case in the Pythagorean tuning system, the major thirds are stretched by the syntonic comma (B=21.5 cents), while the minor thirds are compressed by the same tempering interval (C=-21.5 cents). If B=0 cents, which is the case in the meantone tuning system, both A and C are equal to -5.4 cents. If C=0 cents, which is the case in the Salinas tuning system, both A and B are equal to -7.2 cents. This shows that it is not possible to construct regular tuning systems in which both the fifths and the major or minor thirds are pure. For the tuning systems mentioned so far, the temperings of the fifth, the major and minor thirds, as well as the temperings of a few smaller intervals, are summarized in Table I.

The temperings of the fifths in equal temperament and the Silbermann tuning system are based on the ditonic comma, D, which equals the difference between 12 pure fifths and 7 pure octaves (D=23.5 cents). In equal temperament A=-D/12, whereas in the Silbermann tuning system, A=-D/6. The temperings for a number of musical intervals in these tuning systems are also given in Table I.

2. Short historical description

In the Middle Ages polyphony was based on the pure consonant unison, octave, fifth, and fourth. Therefore, the Pythagorean tuning system was very

Table I Temperings of the main fifths, major thirds and minor thirds in the regular tuning systems that were used in the experiment. In addition the temperings of the major second (whole tone) and the minor second (diatonic semitone) are given. The latter two intervals were used only as melodic intervals. Temperings are given in cents relative to the pure interval. The fundamental frequency ratio, p:q, and the size in cents of these pure intervals are given between brackets. Temperings of complementary intervals are different only with respect to the sign of the tempering.

	Tempering of the intervals in cents				
Name of tuning system	fifth (2:3; 702.0)	major third (4:5; 386.3)	minor third (5:6; 315.6)	major second (8:9; 203.9)	minor second (15:16; 111.7)
-	2.0	29.5	-27.5	4.0	-31.5
Pythagorean tuning	0.0	21.5	-21.5	0.0	-21.5
Equal temperament	- 2.0	13.7	-15.6	- 3.9	-11.7
Silbermann tuning	- 3.9	5.9	- 9.8	- 7.8	- 2.0
Meantone tuning	- 5.4	0.0	- 5.4	-10.8	5.4
Salinas tuning	- 7.2	- 7.2	0.0	-14.3	14.3
-	-10.0	-18.5	8.5	-20.0	28.5

suitable for this kind of music. Thirds and sixths, all tempered by $\pm$ 21.5 cents, were considered as dissonances, only used at unaccented places.

In the Renaissance the thirds and sixths were also considered as consonant intervals and were employed in accented places and in the final chord of a piece of music. As a result, theorists tried to construct tuning systems in which also the major thirds were pure. It has been explained above that pure fifths and pure major thirds are incompatible in one tuning system and that, in spite of all efforts, the just tuning systems render just intervals only for a restricted set of intervals within the octave.

Pure major thirds were successfully combined with tempered fifths in the meantone tuning system, first described by Pietro Aron in 1523. A later description was given by Michael Praetorius in 1619. In the 16th and 17th centuries meantone tuning was the prevailing system and, on a smaller scale, it continued to be applied until the middle of the 19th century. In this system, however, the difference between the enharmonic equivalent tones is so large that without split keys or other ingenious mechanisms modulation and transposition are restricted.

At the end of the 17th century the new Baroque style required a tuning system which would allow playing in all modes. This is feasible in equal temperament, which is a tuning system in which the purity of the intervals is given up in favor of free modulation and transposition, and of a keyboard that is easy to handle. One of the first descriptions of equal temperament is given as early as 1533 by Giovanni Lanfranco. From the 16th century on, this tempering was used for fretted instruments such as the lute and the viol. Although equal temperament was used in keyboard instruments from the 18th century on, it took one and a half century before it was generally applied.

The Silbermann tuning system is a compromise between meantone tuning and equal temperament. This system, connected with the famous organ builder Gottfried Silbermann, represents the more conservative practice during the time of Bach and Handel. The tempering of its fifths is less than that in meantone tuning, the tempering of its major thirds is less than that in equal temperament. The enharmonic intervals are better than in the meantone tuning and therefore, the Silbermann tuning system is, to some degree, more suitable for playing in different keys.

In 1577 Francisco Salinas proposed a tuning system in which the minor thirds are pure. This system has remained largely theoretical.

For more historical details, the reader is referred to, e.g., Barbour (1951), Dupont (1935), di Vèroli (1978), and Cohen (1984).

B. Physical and perceptual aspects

1. Interference of nearly coinciding harmonics

In a pure consonant interval the ratio between the fundamental frequency f_1 of the lower tone and the fundamental frequency f_2 of the higher tone is equal to p:q, with p and q ($p < q$) being small integers. For such an interval, comprising two simultaneous complex tones, the harmonics with frequencies of npf_2 and nqf_1 Hz (n=1,2,...) coincide [see Vos (1984) and Vos and van Vianen (1985a) for illustrations]. When moderately tempered ($f_1:f_2 \sim p:q$), this interval is characterized by small frequency differences between the harmonics which coincide in pure intervals. The interference of these nearly coinciding harmonics results in amplitude variations. Since, in our experiments, the amplitudes of the nearly coinciding harmonics were equal, the depth of the variation in level (= envelope) was maximal (see Vos and van Vianen, 1985a). For the nth pair of interfering harmonics, the frequency of the level variation is equal to $|npf_2 - nqf_1|$.

Perceptually, the interference of the nearly coinciding harmonics gives rise to beats or roughness. The strength of the beats is related to the depth of level variation. For a number of different spectral conditions, including those adopted in the present experiments, it has been shown that the beat frequency, f_b, of $|pf_2-qf_1|$ Hz is perceived more clearly than multiples of this beat frequency (Vos, 1984). The relationship between f_b and tempering, T, in cents, is given by

$$f_b = |qf_1\ (2^{T/1200} - 1)|\ ,$$

or, as demonstrated earlier (Rasch, 1984; Vos and van Vianen, 1985b), simply by

$$f_b = |qf_1 T/1731|\ . \tag{7}$$

In addition to beats, changes in the size of the interval (stretched or compressed relative to the size of the pure interval) may be detected. An experiment with musically trained subjects has shown that for fifths and major thirds, identification of the direction in which the interval size has been changed becomes relevant at, say, $|T| \geq 25$ cents (Vos, 1982).

2. Subjective rating of pure and tempered intervals

In section A1 it was shown that in terms of their temperings of the main intervals, the differences between the most relevant regular tuning systems are rather small. For the fifths $-7.2 < A \leq 0.0$ cents, while for the major and minor thirds maximum (absolute) tempering does not exceed the syntonic comma of 21.5 cents. To our knowledge only two studies on the relation between tempering and subjective evaluation have included a sufficient number of these small degrees of tempering (Hall and Hess, 1984; Vos, 1986).

In these studies the pure and tempered intervals consisted of two simultaneous complex tones. The intervals were presented in isolation, that is, they were not given in a musical context. Although the differences in individual responses in the experiments of Hall and Hess (1984) were larger than in the experiments of Vos (1986), the effects of tempering, as well as the effect of interval type, were about the same in the two studies.

Part of the results obtained in our own experiment (Vos, 1986) is shown in Fig. 1. In this figure, mean subjective purity ratings are plotted as a function of tempering in cents for the fifth and the major third, separately. Both the lower and the higher tones were complex tones, with spectra being identical to those used in experiment 1 of the present paper. The duration of the tones was 0.25 s. It can be seen that in this condition the fifths are rated to be purer than the major thirds between temperings of about -20 and 15 cents, whereas for larger deviations from the pure intervals, the reverse is the case. In addition, it can be seen that the relations between subjective

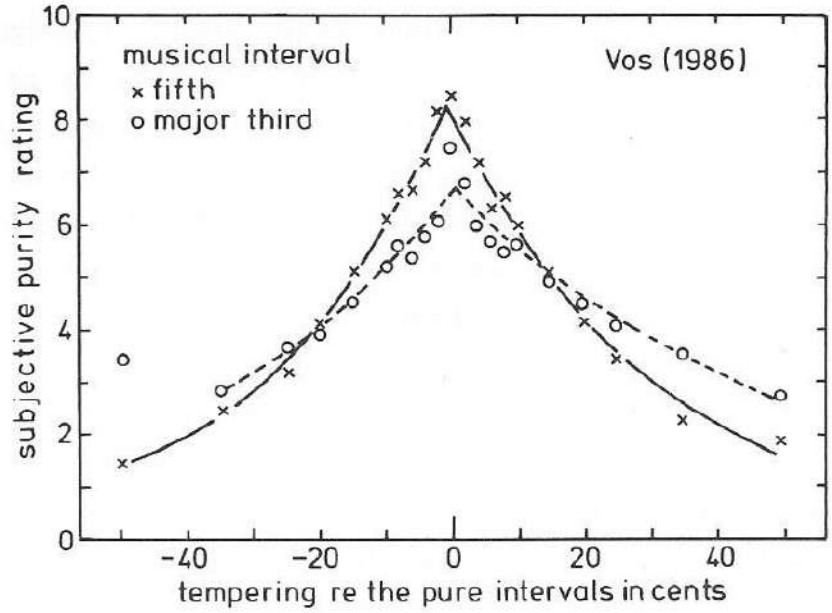

Figure 1 Mean subjective purity ratings (Vos, 1986) as a function of tempering for the fifth and the major third. The curves represent the best exponential fit. Both the lower and the higher tones were complex tones, 0.25 s in duration.

purity and the amount of tempering can be described by exponential functions.

3. Evaluation of different tuning systems

Theoretical evaluations of different tuning systems, as proposed by Hall (1973) and by Rasch (1983), are based on the idea that tempering should be minimized as much as possible. These evaluations assume that the appreciation of a tuning system, or its suitability for the performance of a particular piece of music, is related to the number of cents by which tempered intervals deviate from their corresponding pure intervals. The common experience that the effect of tempering also depends on the kind of musical interval would suggest the use of unequal weights for the various intervals. Neither Hall (1973) nor Rasch (1983) attempted to introduce such differential weights, because at that time results from, for example, perceptual experiments to support these weights were lacking.

Indeed, these weights could not be derived from the few exploratory studies on the perceptual relevance of different tuning systems that have been carried out in the past. Yet, these studies are significant for the present paper.

Van Esbroeck and Monfort (1946) presented various melodic and harmonic musical fragments to about 1000 listeners (high-school pupils, students and teachers from a conservatory of music). These fragments were performed according to four different tuning systems: equal temperament, Pythagorean tuning, just tuning, and "de Meerens tuning", the latter, however, being incompletely defined. The fragments were played pairwise on a pipe organ with a 53-tone keyboard. With the four different tuning systems just mentioned, there were 6 different pairs of performances for each fragment. The listeners were asked whether they heard a difference between the fragments within a pair, and if they did which version they preferred. Overall, a difference could be heard in 74% of the cases, whereas in about 57% of all the cases one version was preferred to the other one. In addition to the overall results, van Esbroeck and Monfort (1946) reported the preference scores for two different subsets: the data in one subset were based on the combined scores for two melodic fragments ("Frere Jacques" and the major diatonic scale), the other subset represented the combined scores for two harmonic fragments (dominant seventh chords, resolved into either the major or the minor triad). The results of these two subsets were further analyzed by us. Both for the melodic and the harmonic fragments, we transformed the preference scores into paired comparison scale values (Thurstone's case V, see Torgerson, 1958). The results are summarized in Fig. 2. It can be seen that for the melodic fragments, equal temperament and, to a lesser extent, the Pythagorean tuning system tended to be preferred to the just tuning system. For the harmonic fragments, however, the magnitudes of the differences force us to conclude that the various tuning systems were evaluated as more or less equivalent.

Loman (1929) asked a large group of musicians to listen to two different performances of the beginning of Wagner's prelude to "Die Meistersinger von Nürnberg". The music was played first on a piano tuned according to just intonation, and then on a piano tuned according to the Pythagorean tuning system. According to Loman, all listeners preferred the Pythagorean system, mainly because in just intonation the expressive character of the major scale was lost.

Kok (1954, 1955) built an experimental electronic organ on which all 12-tone tuning systems could be realised. The organ even allowed changes of the tuning during a performance. Therefore, tuning systems with substantial differences between the enharmonics - such as is the case in the meantone tuning system - could also be applied without getting into trouble while modulating.

From the results of listening tests, Kok (1954, 1955) concluded that musically untrained listeners do not hear any difference between music played

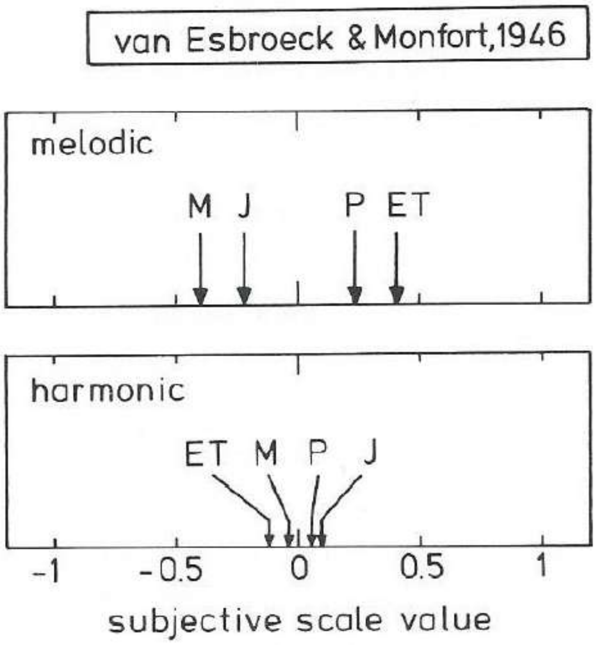

Figure 2 Scale values of four tuning systems for two melodic and two harmonic fragments. The data were paired-comparison preferences based on preferences expressed by the subjects in van Esbroeck and Monfort's (1946) experiment. P = Pythagorean tuning, ET = equal temperament, J = just intonation, M = "de Meerens tuning".

in equal temperament or in the meantone tuning system. Musicians, however, all heard differences between these systems. Especially for the major and minor seconds (see Table I), the size of the melodic intervals provided one source of the differences heard, although these differences did not lead to a higher evaluation of one of the two systems. However, the musicians preferred the meantone tuning to equal temperament in sustained and final chords, and in chorale-like music. None of the organ-players in the group felt that the differences were large enough to justify the additional task of handling the pitch adjusters.

Ward and Martin (1961) designed a number of experiments in which just tuning and equal temperament were compared in ascending diatonic scales. The tones were produced on an electronic organ. They concluded that most people, even most musicians, cannot distinguish short melodic sequences in just intonation from those in equal temperament. In addition, it was concluded that the very few musicians who are able to discriminate seem to prefer equal temperament.

Kolinski (1959) and Martin and Ward (1961) conducted listening tests in which fragments were played on pianos that were tuned either to strict equal temperament or to stretched versions of this temperament. For piano tones higher than about C6 [two octaves higher than middle C (=C4)] and lower than C3, stretched tuning is indeed relevant, and has been related to the inharmonicity of piano strings (Meinel, 1954, 1957; Railsback, 1938; Schuck & Young, 1943; Young, 1954).

Using a stretched scale which was representative of what has been found in practice, Martin and Ward (1961) found that in various (quasi-)musical tonal sequences and chords, this stretched temperament was preferred to strict equal temperament in about 60 to 75% of the cases, a percentage significantly above chance.

Kolinski (1959) applied a more theoretically based stretched scale in which all octaves were stretched by 3.3 cents, resulting in fifths that are physically pure and in major thirds that are tempered by 14.8 cents, which is only 1.2 cents more than the tempering in equal temperament. From listening tests with professional musicians, Kolinski concluded that in pieces of music as well as in isolated intervals and chords, this stretched scale was superior to strict equal temperament.

Terhardt and Zick (1975) investigated perceptual differences between equal temperament and a stretched and even a compressed version of equal temperament. For the tones lower than about E3, however, their stretched version was increasingly more stretched than the version used by Martin and Ward (1961). Especially because Terhardt and Zick (1975) used an electronic organ with harmonic tones,this choice was unexpected. It is therefore not surprising that in chords with normal to close spacing of the tones, the tones comprising at least a number of relatively strong harmonics, they found that their musically trained subjects preferred strictly equal temperament. Only for harmonic intervals with 4 octaves between the descant and the bass parts, comprising tones with a few low-level harmonics, did they find a preference for the stretched scale.

II. RATIONALE BEHIND THE EXPERIMENTS

In our experiments we presented two-part musical fragments that were performed according to all tuning systems (experiment 1) or some of the tuning systems (experiment 2) given in Table I. We asked the subjects to rate the overall acceptability of the way in which these fragments were performed.

A. Relative importance of beats and roughness

Discriminability between pure and tempered harmonic fifths and major thirds consisting of complex tones is poorer for low degrees of tempering than for high degrees of tempering, and this difference is more prominent at short tone durations (Vos, 1982, 1984). This suggests that at least for slightly to moderately tempered intervals discrimination is determined by sensitivity to beats.

The strong differentiation in subjective purity ratings for these intervals also seems to be mainly the result of the presence of beats or roughness (Vos, 1986). If the presence of beats and roughness is also relevant for intervals in a more musical context, the differences in overall acceptability should be smaller at short than at long tone durations for at least some of the tuning systems. In experiment 1 we therefore played the musical fragments in a slow and in a fast tempo. In the fast tempo condition tone duration was about 75% shorter than in the slow tempo condition.

The results of experiment 1 suggested that overall acceptability of the fragments was based on the combined purity ratings of the harmonic fifth and major third. It has been shown earlier that the perceived strength of the beats in these isolated intervals is determined mainly by interference of the nearly coinciding harmonics (Vos, 1984; Vos and van Vianen, 1985a) and that cancellation of this specific kind of interference also resulted in higher subjective purity ratings, the effect being more prominent for the major third than for the fifth (Vos, 1986). If it is true that overall acceptability of our fragments is based on the combined purity ratings of the harmonic fifth and major third, then deletion of the main source of interference should have a positive effect on the acceptability ratings. This hypothesis was tested in experiment 2.

B. Integration of subjective purity of separate harmonic intervals into subjective overall acceptability of tuning systems

In acoustically comparable conditions, the relations between tempering and purity of (isolated) harmonic fifths and major thirds are known (Vos, 1986). To a certain extent, the results of the present experiments can therefore be used to find the way in which the purity ratings of the separate intervals within a fragment are integrated into one overall rating.

For example, the overall rating may be determined by the rating of the

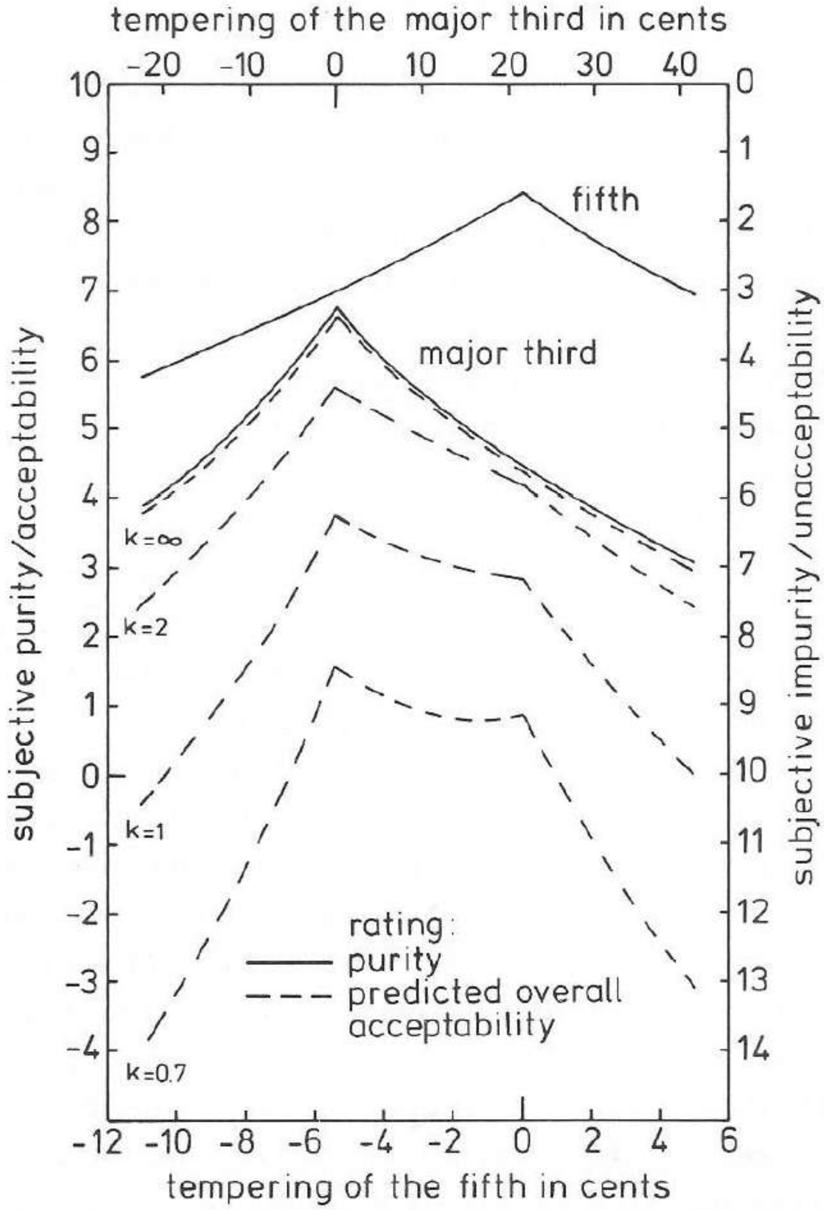

Figure 3 Purity ratings for the fifth and the major third (solid curves) from Fig. 1, replotted as a function of the tempering of the main fifths in regular 12-tone tuning systems. The dashed curves represent subjective acceptability of the two-part piece of music predicted at various values of the exponent k in the family of Minkowski distance functions.

worst interval (dominance model), or, by the linear sum of the (im)purity ratings (additivity model). Both the dominance and the additivity models can be considered to be special cases of the more general family of Minkowski distance functions (see, e.g., Meerling, 1981), in which the subjective impurity ratings for the various intervals X_i ($i=2, 3, \ldots, j$) can be integrated into the overall acceptability rating, X_1:

$$X_1 = \left(\sum_{i=2}^{j} X_i^k \right)^{1/k} . \qquad (8)$$

The specific versions of this family of functions are defined by the exponent k; with k increasing from 0 to ∞, X_1 is increasingly determined by the highest impurity rating.

For a two-part piece of music that only consists of fifths and major thirds, the predictions for a number of these versions are illustrated in Fig. 3. The top two curves have been adapted from Fig. 1, now replotted for various 12-tone regular tuning systems with the temperings of the fifth (lower scale) and the major third (upper scale) as references. The dominance model (k=∞), which is represented by the dashed line just below the curve for the major third, predicts that only the meantone tuning system will be acceptable. The additivity model (k=1), which is represented by the next dashed line but one, predicts that all systems between meantone and Pythagorean tuning are about equally acceptable, whereas other regular systems will receive much lower ratings.

The predictions of the additivity model, as well as of other versions in which $k < 1$ (see Fig. 3), are in agreement with the findings of van Esbroeck and Monfort (1946) and of Kok (1954, 1955) in conditions in which harmonic musical fragments have been presented by means of an organ.

III. EXPERIMENT 1

A. Method

1. Musical fragments

The stimuli were short two-part musical fragments that were based on the first 6-10 tones of the discant and bass parts of four-part chorale settings from the Musae Sioniae, Part VI by Michael Praetorius (1609). From the 200 settingsincluded in Part VI, we selected 24 fragments, all of which met the following five selection criteria: (1) The rhythmic structures of the discant and bass parts within a setting had to be equal, that is, the successive tones of the higher part were played in synchrony with the tones of the lower part, which is characteristic of these homophonic settings (see Rasch, 1981). When a musical interval between the simultaneous tones exceeded the octave, the tone of the bass part was transposed by one, or, in a few conditions, by two octaves. The next requirements were that a fragment should contain (2) both fifths and minor and major thirds, (3) no major sixths, and (4) not too many octaves. As a result the harmonic intervals between the lower and the higher parts were minor third (22.1%), major third (19.5%), fifth (30.0%), minor sixth (2.1%), and octave (26.3%); for each interval the relative frequency of occurrence, averaged across the 24 fragments, is given between brackets. The last selection criterion (5) was that there had to be at least some melodic

Figure 4 Seven examples of the musical fragments used in the experiments. The fragments are based on the discant and the bass parts of the opening phrases of four-part chorale settings to be found in the Michael Praetorius's **Musae Sioniae, Part VI** (1609). The numbers correspond to the classification used in the Volume referred to.

movement in the parts. The numbers of the chorale settings finaly used were 9, 15, 18, 28, 30, 46, 54, 74, 78, 84, 92, 107, 108, 110, 124, 128, 134, 147, 148, 150, 155, 161, 175, and 177. The musical score of a selection of the fragments presented is given in Figure 4.

2. Musical tones

Both the discant and the bass parts were played with complex tones, consisting of 20 harmonics with amplitudes a_n proportional to 1/n. The spectral-envelope slope was therefore -6 dB/oct. The phases of the individual harmonics were chosen randomly. Because of these random phases the waveforms of the tones in the lower and the higher parts were different.

For all tempered intervals, level-variation depth of the beating harmonics was maximal. This occurs when the amplitudes of the nearly coinciding harmonics are equal. Depending upon the kind of musical interval, the overall level of the tones in the higher part was therefore 20log(q/p) dB lower than the tones in the lower part [see Table I for values of p and q]. Since in our experiment the octaves were always presented in their pure form, level-variation depth is not relevant for these intervals. For the octaves the level of the higher tone was only 3 dB lower than that of the lower tone. For all intervals, the overall level of the two simultaneous tones was about 73 dB(A). Rise and decay functions of the tones were half sinusoidal; the rise and decay times of the tones, defined as the time interval between 10% and 90% of the maximum amplitude, were 30 and 20 ms, respectively.

3. Apparatus

The experiment was run under the control of a PDP-11/10 computer. The tones were generated in the following way. One period of the waveforms of each tone was stored in 256 discrete samples (with 10 bit accuracy) in external revolving memories. These recirculators could be read out by digital-to-analog converters. Sampling rates were determined by pulse trains, derived from two frequency generators. After gating, each tone passed a low-pass filter with a cutoff frequency of 8 kHz and a slope of -48 dB/oct. The sound-pressure levels of the tones were controlled by programmable attenuators. After appropriate attenuation, the tones were mixed and fed to an amplifier set, to which four Beyer DT 48S headphones were connected. By means of these headphones the musical fragments were presented diotically (same signal to both ears). Subjects were seated behind a terminal in a soundproof room. They responded by pressing the keys of a keyboard. Part of the instructions was presented visually on a display.

4. Subjects

Twenty-four musically trained subjects participated in the experiment. Most of them were students, either of the Institute of Musicology or of the Conservatory at Utrecht. The subjects were tested in six groups of four subjects each over two sessions, both of which lasted about 3.5 h. The participants were paid for their services.

5. Experimental variables and measures of subjective acceptability

The material evaluated in this study consisted of the 24 different musical fragments, as described above. The fragments were presented in seven different 12-tone regular tuning systems (tempering, A, of the main fifths was 2.0, 0.0, -1.96, -3.91, -5.38, -7.17, and -10.0 cents) and at two different tempi (fast and slow). In experiment 1 subjective acceptability was measured by means of a rating scale and by means of paired-comparison judgments.

Rating. The 24 musical fragments were presented in three sets of eight fragments each. Within each set the fragments were performed according to the seven tuning systems, resulting in 56 stimulus items per block. These 56 musical events were played at the two tempi. This made a total of six blocks of eight fragments x seven tuning systems. Each block was presented twice for evaluation. Before starting with the first block the subjects were presented with a familiarisation block, comprising three different musical fragments which were performed according to the seven tuning systems. The subjects were told that the purpose of this block was to give them a frame of reference within which they had to rate the "overall acceptability" of the performance of the various fragments. They were encouraged to use the whole range of scale-values from "very unacceptable" (1) to "very acceptable" (7).

Paired-comparison judgments. For this task only five musical fragments were used, namely the numbers 18, 30, 147, 155, and 175 from the total set of 24 fragments. Each fragment was played in the seven tuning systems in a given tempo, fast or slow, making a block of (7)(6)/(2)=21 different pairs of performances of the same fragment. With five fragments and two tempi we obtained in this way 10 blocks of 21 pairs each. The subjects had to select the "more acceptable" performance from each pair.

6. General procedure

Before the frequencies of the 12 tones in the octave were computed, the tones were arranged in a row of fifths. The first tone in this row was the minor third above the tonic of a given musical fragment. As a result, the rest fifth (wolf) was the fifth on the augmented fifth above the tonic. We computed a new series of frequencies for each musical fragment. Because Praetorius never used enharmonics within his chorale-settings, the use of rest intervals could be entirely avoided. For each tuning system included, the temperings of a number of consonant intervals are given in Table I.

When the musical fragments were presented for rating, the frequency of A4 (the A above middle C) was set to 440 Hz. In the paired-comparison task, however, the frequencies of the tones in the second and all the successive fragments within a block were normalized in such a way that the first tone in the bass part had always the same frequency.

In the slow tempo total duration of each fragment was fixed at 9.5 s. The relative onset times of the successive tones depended on the rhythmic structure of the chorale settings. Since the silent intervals between the tones were about 0.2 s, the durations of the tones varied with the number of tones in the fragments. In the fast tempo total duration of each fragment was 4 s. Overall, tone duration in the fast tempo condition was about 75% shorter than that in the slow tempo condition.

Three groups of subjects were presented in their first sessions with the six blocks in which ratings had to be given first, followed by the paired-comparison judgments for five of the ten blocks. In the second sessions, these groups were presented again with the six blocks of stimuli for a second rating of the same stimuli, followed by the remaining five blocks of stimuli for the paired-comparison judgments. For the other three groups of subjects, the order of ratings and paired-comparison judgments was reversed. Presentation order of the different blocks for rating was balanced according to 6 x 6 Latin squares. Presentation order of the 56 stimuli within a block was randomized. The presentation order of the blocks of stimuli for paired-comparison judgments was also randomized.

B. Results

1. Ratings

Effect of tuning system. For the various tuning systems, the mean acceptability ratings are plotted in Fig. 5 for the fast and the slow conditions, separately. These ratings are averaged across the 24 subjects, the 24 musical fragments, and the two replications. Mean ratings were about the same

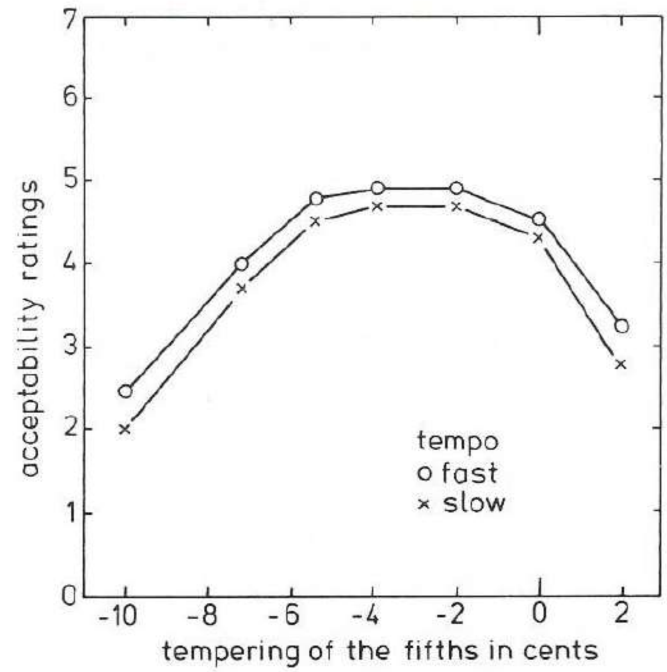

Figure 5 Mean acceptability ratings as a function of the tempering of the main fifths in regular 12-tone tuning systems, for the fast and the slow tempo.

for the meantone tuning (A=-5.4 cents), the Silbermann tuning (A=-3.9 cents), and equal temperament (A=-2.0 cents), whereas the mean rating slightly decreased for the Pythagorean tuning system (A=0.0 cents). A very strong decrease in the ratings was found for A=2.0 and A $\leq$ -7.2 cents. It can be seen in Fig. 5 that the effect of tuning system was the same for the two tempi, and that the mean ratings in the fast tempo were higher than in the slow tempo. Because tempo was not varied within the same blocks of trials, this difference cannot unequivocally be ascribed to the tempo difference. Therefore, we preferred subjecting the data to two separate analyses of variance (ANOVAs), one for each tempo, with tuning system and musical fragment as within-subjects variables. Because the ratings given in the first presentation were not significantly different from those given in the second presentation and, more importantly, the effects of tuning system and musical fragment were not significantly different for the first or the second ratings, replication was considered as a within-cell variable. In the ANOVAs, both subjects and musical fragments were treated as random effects (Clark, 1973), whereas tuning systems was treated as a fixed effect. In Table II, in which the results of the ANOVAs are summarized, it can be seen that both for the fast and the slow tempo, the effect of tuning system is very powerful. For the conditions in which the

Table II Summary table of two analyses of variance on the acceptability ratings. For both tempi we used a 24 [subjects] x 7 [tuning systems] x 24 [musical fragments] factorial design, repeated measures with replication as a within-cell variable; ω^2 is the percentage of estimated component of variance.

			Fast tempo				Slow tempo			
Source	df_1	df_2	MS	F-ratio	p	ω^2	MS	F-ratio	p	ω^2
Within-cell	4032		1.42			51.3	1.40			48.6
Between-cell										
Subjects (S)	23	529	37.07	9.5	0.00001	3.4	41.73	12.0	0.00001	3.8
Systems (A)	6	-*	1026.19	71.9	0.00001	27.8	1224.15	78.7	0.00001	31.2
Fragments (B)	23	529	30.71	7.9	0.00001	3.0	28.72	8.3	0.00001	2.7
A x B	138	3174	4.22	3.0	0.00001	1.9	3.74	2.7	0.00001	1.4
A x S	138	3174	11.45	8.1	0.00001	6.2	13.21	9.4	0.00001	7.4
B x S	529	4032	3.91	2.8	0.00001	6.4	3.47	2.5	0.00001	4.9
AxBxS	3174	4032	1.41	1.0	0.55	0.0	1.39	1.0	0.5	0.0

*Note: df_2 is the nearest integer value of $(MS_{AS} + MS_{AB} - MS_{ABS})^2/(MS^2_{AS}/df_{AS} + MS^2_{AB}/df_{AB} - MS^2_{ABS}/df_{ABS})$. For the fast tempo $df_2 = 189$; for the slow tempo $df_2 = 177$.

musical fragments were played in the fast tempo, a Newman-Keuls paired comparison test (Winer, 1970) showed that the mean rating at A=0.0 cents was significantly lower than the ratings at $-5.4 \leq A < -2.0$ cents at the 0.05 level only. Similar results were found for the slow tempo conditions, except that the mean ratings at A=-5.4 and A=0.0 cents were not significantly different.

Effects of musical fragment. Independent of tuning system, the acceptability ratings depended in a systematic way on the specific musical fragments (see Table II for significance levels). For the musical fragments that were played at the slow tempo, for example, the mean rating, M, was 3.84, while the standard deviation, σ, was 0.29. Musical fragments # 9, 15, 74, and 134 had mean ratings higher than $M+\sigma$; the fragments # 54, 108, 150, and 175 had mean ratings lower than $M-\sigma$. The four musical fragments that, on the average, received the highest ratings all started with an octave and three of these fragments again ended in an octave. The four musical fragments that received the lowest ratings all started with a fifth and ended in a major third.

If the final interval of a musical fragment had a particularly strong effect on the overall acceptability ratings, then the fragments ending in a major third would be affected differently by tuning system than the fragments ending in an octave, because the tempering of the major third is very high at, for example, $A \geq 0.0$ cents, but low, or even zero, at $-7.2 < A < -3.9$ cents, whereas the octaves are, by definition, pure in all conditions. Since the four fragments ending with a major third received lower ratings than the other fragments in nearly all conditions, evidence for a relatively large effect of the final interval in the fragments on the overall rating could not be found. Similarly, our data do not support the hypothesis of a relatively high influence of the first interval in a musical fragment (the fifths in the fragments that received the lowest ratings).

It seems that the main effect of musical fragment is simply related to the mean absolute tempering of the harmonic intervals (including the octaves) in the fragments. Averaged across the various tuning systems, this mean absolute tempering is 6 cents for the four fragments with the highest ratings and 8.5 cents for the four fragments with the lowest ratings. Significant relations between mean absolute tempering and overall acceptability are also obtained when all 24 musical fragments are taken into account: For the slow tempo conditions the correlation coefficient r is equal to -0.59 ($p < 0.005$); for the fast tempo conditions r is -0.60 ($p < 0.001$).

In addition to the main effects of tuning system and musical fragment, 1.4-1.9% of the variance in the ratings (Table II) is due to the interaction between these effects. This means that for a number of musical fragments, the effect of tuning system deviated from the mean effect, shown in Fig. 5. For each fragment the ratings that were given for the seven tuning systems were converted into deviations from the overall mean rating of that fragment. These 168 scores were related to the corresponding mean temperings of the fragments. Both for the slow and the fast tempo conditions, r was -0.51 ($p < 0.001$). This indicates that, similar to the main effect of musical fragment, the interaction effect can be partly explained by differences in the mean tempering of the fragments at the various tuning systems.

2. Paired-comparison judgments

For each fragment and tempo, that is, for each of the 10 blocks of 21 pairs separately, the preference scores obtained were analyzed under the assumption of Thurstone's case V (see Torgerson, 1958), yielding scale values of subjective acceptability. These are presented in Fig. 6 as a function of the tempering of the main fifths in the tuning systems. Again, the scale values were highly affected by tuning system. It can be seen in Fig. 6 that for most fragments acceptability is highest for $-5.4 \leq A \leq -2.0$ cents, some-

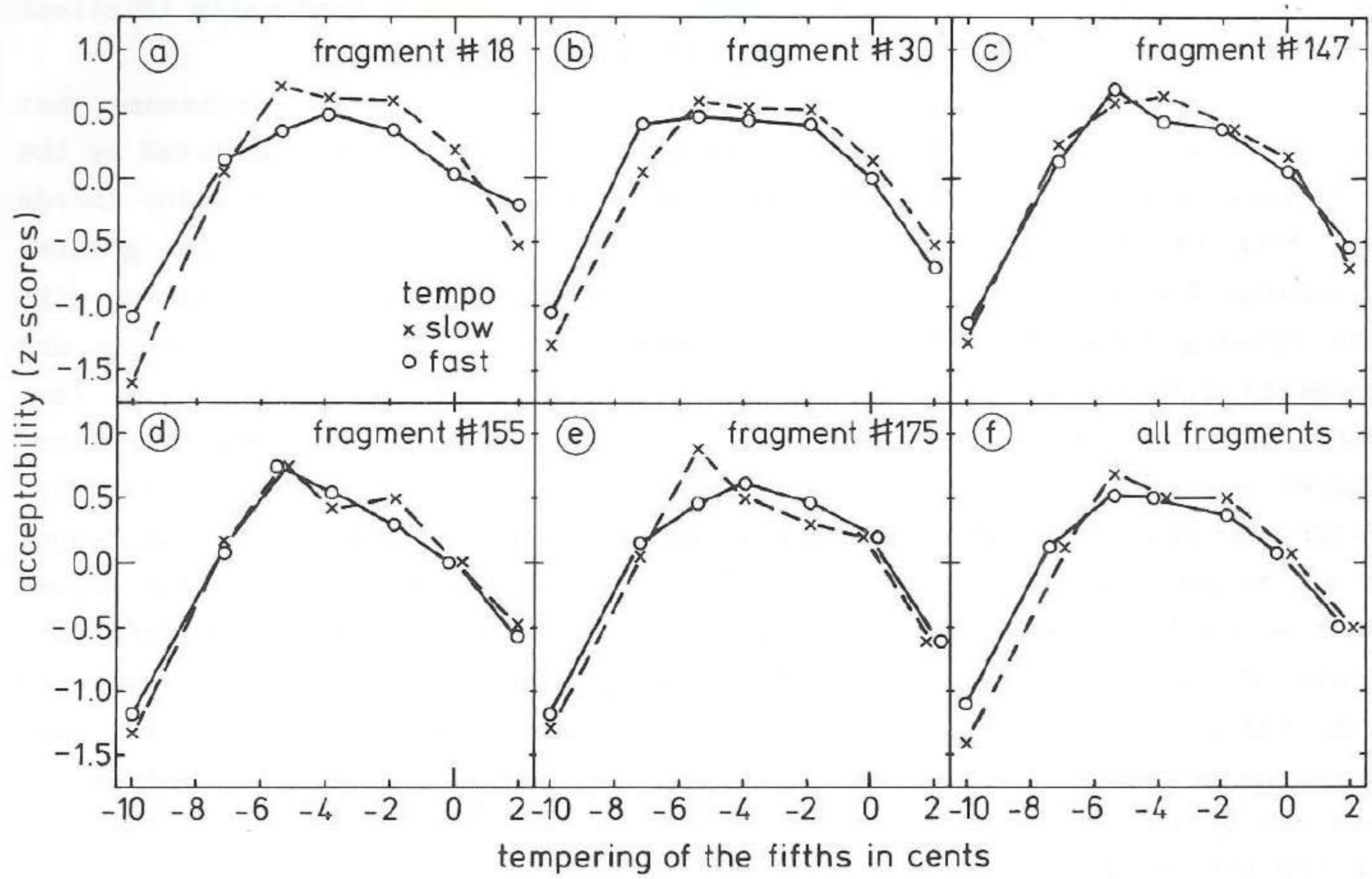

Figure 6 Scale values for subjective acceptability, plotted as a function of the tempering of the main fifths in regular tuning systems for the fast and the slow tempo, separately. The scale values were derived from paired-comparison preference data. In panels (a) - (e), values are given for single fragments, whereas in panel (f) the values are based on the combined preference scores.

what lower for A=0.0 and A=-7.2 cents, and very much lower for A=2.0 and A=-10.0 cents.

3. Overall acceptability of tuning systems and subjective purity of isolated intervals

To find possible determinants of overall acceptability, the mean overall acceptability ratings obtained (Fig. 5) were related to subjective purity/impurity ratings of isolated harmonic intervals obtained in previous experiments (Vos, 1986). The musical fragments comprised pure and tempered fifths, major thirds, minor thirds, and a few minor sixths. By means of the Minkowski power metric [Eq. (8)], we tried to predict the overall acceptability ratings from the subjective purity ratings of pure and tempered isolated intervals. This comparison is especially interesting because these purity ratings (see Fig. 1) are available for conditions in which the tones used are spectrally identical to those used in the musical fragments (Vos, 1986, Table II).

For lack of purity ratings of minor thirds and sixths, we assume that these ratings are close to those of the major thirds. This is supported by the experimental results obtained by Hall and Hess (1984). Stretched major thirds had been rated slightly purer than the compressed intervals. The greater tolerance for stretched than for compressed major thirds is consistent with the finding that both in solo and ensemble music performed on string and woodwind instruments, the intonation of the major third is stretched (see Ward, 1970). Since the latter research indicates that the intonation of minor thirds tends to be compressed, the exponential functions of the stretched major thirds will be used as estimates of the purity ratings of the compressed minor thirds; in the same vein the functions of the compressed major thirds will be used to estimate the ratings of the stretched minor thirds. In addition, it may be expected that the subjective differences between pure and tempered minor thirds are slightly smaller than those between pure and tempered major thirds. Therefore, the range of ratings of the minor thirds was reduced by 10%. For simplicity, the small number of minor sixths (2.1%) will be treated as minor thirds.

Two correlation analyses were carried out in which correlation coefficients, r, between observed and predicted overall acceptability were computed for a range of values of the exponent k [see Eq. (8)]. We preferred to relate overall acceptability to impurity rather than to purity (see Fig. 3). The values of the "impurity ratings" are simply the differences between the theoretical maximum purity ratings (= 10) and the observed purity ratings. One analysis was performed on the overall ratings of the musical fragments played in the fast tempo and the subjective impurity ratings for the intervals in which both tones consisted of the first 20 harmonics, the tones being presented at a duration of 0.25 s (see Vos, 1986, Table II). The other analysis was based on the acceptability ratings obtained in the slow tempo condition and the impurity ratings for the intervals with spectrally comparable tones presented at a duration of 0.5 s. The relevant ratings for the seven differenttuning systems and the two different tempo conditions are presented in Table III.

As a first approximation, the acceptability ratings used are mean values based on the total set of 24 fragments. Differences in frequency of occurrence of the various harmonic intervals were not taken into account, which is justified because, overall, the relative frequency of occurrence ranged only

Table III Mean overall acceptability ratings of musical fragments performed according to various regular 12-tone tuning systems, and mean subjective purity ratings for corresponding isolated fifths, major thirds, and minor thirds. Ratings of the minor thirds are estimates. In the fast tempo conditions, the acceptability ratings were related to impurity* ratings for intervals presented at a tone duration of 0.25 s. In the slow tempo conditions impurity ratings for intervals 0.5 s in duration were used as the predictors.

Tuning system (tempering of main fifths in cents)	Overall Acceptability Rating (X_1)		Subjective purity rating: Fifth (X_2)		Major third (X_3)		Minor third (X_4)	
	fast	slow	fast	slow	fast	slow	fast	slow
2.0	3.3	2.9	7.8	8.2	3.8	3.3	4.1	3.6
0.0	4.5	4.3	8.4	8.9	4.5	4.0	4.5	4.1
- 2.0	4.9	4.7	7.8	8.2	5.2	4.9	5.0	4.7
- 3.9	4.9	4.7	7.4	7.5	6.0	6.0	5.5	5.4
- 5.4	4.8	4.5	7.1	7.1	6.7	6.8	6.0	5.9
- 7.2	4.0	3.8	6.6	6.6	5.6	5.4	6.5	6.5
-10.0	2.4	2.0	5.9	5.9	4.2	3.9	5.4	5.2

*Note: The impurity rating is simply the difference between the theoretically maximum purity rating (= 10) and the observed purity rating.

between 20 and 30%. The results of more detailed analyses will be given in Section III C2 of this paper.

Surprisingly, inclusion of the estimated ratings of the minor third appeared to yield lower correlation coefficients than when only the ratings of fifth and major third were considered. Therefore, only the results from the conditions in which fifth and major third are taken into account will be reported here. Both for the fast and the slow tempo conditions we obtained rather symmetrical inverted U-shaped curves between the exponent k and the correlation coefficient. In both conditions, the best fitting k was 1.0, with r being -0.97 and -0.96 in the fast and in the slow tempo condition, respectively. Especially for $k < 0.75$ and $k > 1.25$, r^2 strongly decreased.

The scale values that were derived from the preference data collected by means of the method of paired comparisons [Fig. 6(f), slow tempo condition] and the scale values (converted into subjective impurity) for isolated fifths and major thirds at a duration of 0.5 s (Vos, 1986, Table II) were also subjected to this analysis. The best fitting k was 1.1 with r equal to -0.98; a strong decrease of r^2 was found for $k < 0.9$ and $k > 1.25$. Because the best fitting k was 1 or close to 1, overall acceptability seems to be composed of two independent dimensions, which means that overall acceptability can be predicted by a multiple linear regression equation.

Two multiple linear regression analyses showed that overall acceptability rating (X_1) can only to a limited degree be predicted from the impurity ratings of either fifth (X_2), major third (X_3), or minor third (X_4). For the fast tempo condition, correlation coefficients are r_{12}=-0.57, r_{13}=-0.70, and r_{14}=-0.20, respectively. Predictability of X_1 is very high when both X_2 and X_3 are taken into account: the multiple-correlation coefficient $r_{1.23}$ is 0.97. The gain of including the third predictor, X_4, is negligibly small: $r_{1.234}$ is 0.98. These findings hold both for the fast and the slow tempo conditions (see Table IV, in which r-values are summarized for the two conditions). We may conclude that for the range of tuning systems presented, overall acceptability

Table IV Correlation coefficients (r), regression weights (m), and constants (c), in the prediction of the overall acceptability rating (X_1) from subjective impurity ratings for fifths (X_2) and major thirds (X_3) by means of the multiple linear regression equation $X_1 = m_1X_2 + m_2X_3 + c$, or from the ratings for X_2, X_3, and minor thirds (X_4) by means of $X_1 = m_1X_2 + m_2X_3 + m_3X_4 + c$. Details on the fast and the slow tempo conditions are given in Table III.

Regression model	Tempo: Fast	Tempo: Slow
$X_1 = m_1X_2 + m_2X_3 + c$		
$r_{1.23}$	0.97	0.97
m_1	-0.77	-0.73
m_2	-0.73	-0.70
c	9.76	9.23
$X_1 = m_1X_2 + m_2X_3 + m_3X_4 + c$		
$r_{1.234}$	0.98	0.98
m_1	-0.94	-0.94
m_2	-0.57	-0.50
m_3	-0.30	-0.39
c	10.84	10.69
Simple correlations		
r_{12}	-0.57	-0.51
r_{13}	-0.70	-0.66
r_{14}	-0.20	-0.19

can be nicely explained by the combined effect of the (im)purity ratings of the fifths and the major thirds. Excluding the third predictor, the multiple regression equation is given by

$$X_1 = m_1X_2 + m_2X_3 + c. \qquad (9)$$

In both the fast and in the slow tempo conditions, the regression weights m_1 and m_2 are about equal, which indicates that for the prediction of overall acceptability of the whole set of 24 fragments together, the impurity ratings of the fifth and the major third are about equally important. When all three predictors are used, the sums of the regression weights of the major and minor thirds are about equal to the weights of the fifth (see Table IV).

C. Discussion

1. Ratings versus paired-comparison judgments

The main difference between the mean ratings given in Fig. 5 and the scale values plotted in Fig. 6 is that in Fig. 5 the acceptability for A=-7.2 cents was slightly but significantly lower than for A=0.0 cents, whereas in Fig. 6 the acceptability for A=0.0 and A=-7.2 cents was about the same. The ratings averaged across fragments # 18, 30, 147, 155, and 175 rendered the same results as those obtained with the paired comparisons. Therefore, the slight discrepancy between the acceptability patterns in Figs. 5 and 6 are the result of an effect of musical fragment rather than of an effect of experimental paradigm. It can be seen in Fig. 6 that the subjective distances between the various tuning systems are about the same for the musical fragments played in the fast and in the slow tempo. Only for fragment # 18 do the subjective distances in the fast tempo seem to be less pronounced than in the slow tempo.

2. Multiple linear regression analysis: a closer look

In a previous section we found that in the prediction of overall acceptability, the impurity ratings of the fifth and the major third were equally important. This analysis was a rather global one because differences in frequency of occurrence of the various harmonic intervals were not taken into account.

Even though our data do not provide a sufficient basis for the determination of the perceptual equivalent of frequency of occurrence of a certain interval in a musical fragment, it seems plausible to hypothesize that in a fragment with, for example, many fifths, the impurity ratings of the fifth receive a higher weighting in the prediction of overall acceptability than they do in a fragment with only one or two fifths. A positive relation between frequency of occurrence of intervals within a musical fragment and regression weights of the impurity ratings of these intervals may support the perceptual relevance of our regression analysis.

We used two different subsets of fragments to test the hypothesis just raised. The first subset consisted of fragments # 18, 28, 30, 147, and 155. The numbers of fifths and major thirds were equal within each fragment. In the slow tempo condition, the regression weights for fifth and major third, m_1 and m_2, were -0.62 and -0.77, respectively ($r_{1.23}$=0.97). In the fast tempo condition, m_1 and m_2 were -0.64 and -0.85 ($r_{1.23}$ = 0.98). Thus, if there are as many fifths as major thirds in a fragment, the impurity ratings of the major thirds receive a higher weighting than those of the fifths.

The second subset consisted of fragments # 46, 74, 107, 110, 124, 128, 148, 150, and 177. In these fragments the number of fifths exceeded that of major thirds by two or more. In the slow tempo condition, m_1 and m_2 were -0.74 and -0.58, respectively ($r_{1.23}$=0.95); in the fast tempo condition these values were -0.85 and -0.64 ($r_{1.23}$=0.96). From a comparison of the regression weights in these two different subsets [A third subset with fragments in which there were at least twice as many major thirds as fifths could not be composed], we may conclude that an increase in the frequency of occurrence of the fifth corresponds to a clear increase in the weighting of the impurity rating of the fifth. This correspondence supports the perceptual relevance of the multiple linear regression analysis.

IV. EXPERIMENT 2

A. Method

1. Stimuli and Apparatus

The stimuli were 8 of the 24 musical fragments that were used in experiment 1. From that set of 24 fragments we presented the numbers 9, 28, 54, 84, 108, 128, 148, and 161. Again, the discant and the bass parts were played with complex tones. However, there were two different spectral conditions. In the complete spectrum condition all tones consisted of a consecutive series of 20 harmonics. These tones were identical to those presented in experiment 1. In the reduced spectrum condition the even harmonics of the tones in either the discant or the bass parts were deleted so that these tones only comprised the first 10 odd harmonics (1, 3, 5,...19). The spectrum of the higher tone was reduced in intervals in which p was an even integer, as is the case in fifth and major third. In intervals in which p was an odd integer, as is the case in minor third, minor sixth, and octave, the spectrum of the lower tone was reduced. In all other respects, such as, e.g., the amplitudes of the various harmonics, the tones were identical to the tones used in experiment 1. The apparatus was the same as in experiment 1.

2. Subjects

Sixteen musically trained subjects were tested. All of them had also participated in experiment 1. They were paid for their services.

3. Experimental variables and procedure

The eight musical fragments were presented in five different 12-tone regular tuning systems (A was 0.0, -1.96, -3.91, -5.38, or -7.17 cents) and at two different spectral conditions (both the lower and the higher tones a complete spectrum or one of these tones a reduced spectrum). The resulting 80 combinations were given in one block, presentation order of the various fragments was randomized. Subjects were asked to rate the overall acceptability of the performance of the various fragments. The rating scale was the same as in experiment 1. The musical fragments were played in a slow tempo, and, similar to the slow tempo in experiment 1, total duration of each fragment was fixed at 9.5 s. Other details, such as, e.g., the way in which the frequencies of the 12 tones in the octave were computed were described in section III.A.

B. Results

1. Ratings obtained in experiment 2

The mean ratings, averaged across subjects and musical fragments, are plotted in Fig. 7 as a function of the tempering of the main fifths for the reduced and the complete spectral conditions. It can be seen that for all tuning systems overall acceptability was considerably higher in the reduced than in the complete spectral conditions. An ANOVA [a 16 (subjects) x 2 (spectral conditions) x 5 (tuning systems) x 8 (musical fragments) factorial design, all repeated measures; subjects and fragments random, spectrum and system fixed effects], performed on the ratings showed that the difference

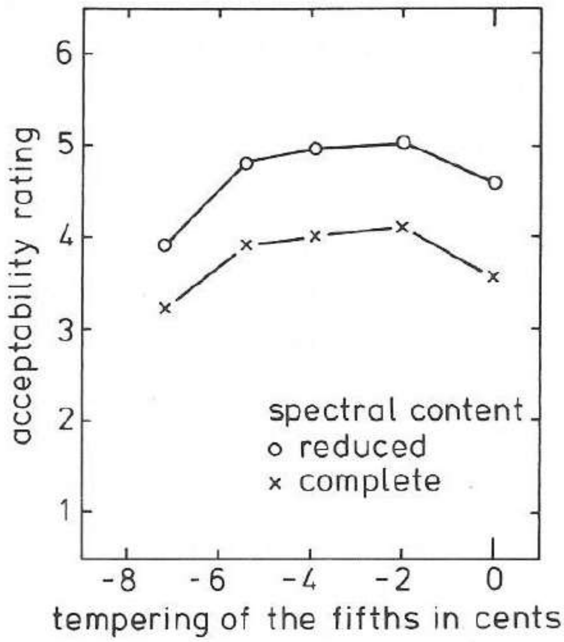

Figure 7 Mean acceptability ratings as a function of the tempering of the main fifths in regular tuning systems for the reduced and the complete spectral conditions.

between these two spectral conditions was significant [$F(1,16)=18.4$, $p < 0.001$]. The slight reduction in difference at A=-7.2 cents was not significant ($p=0.72$). Again, tuning system had an effect on the ratings [$F(4,59) = 6.4$, $p < 0.0005$]. Mean ratings were about the same for $-5.4 \leq A \leq 0.0$ cents, whereas the ratings for the Salinas tuning system (A=-7.2 cents) were lower. A Newman-Keuls paired-comparison test showed that in the reduced spectrum conditions the mean rating at A=-7.2 cents was not only significantly lower than the mean ratings at $-5.4 \leq A \leq -2.0$ cents ($\alpha=0.01$), but also lower than the mean rating at A=0.0 cents ($\alpha=0.05$). In the complete spectrum conditions the mean rating at A=-7.2 cents was only significantly lower than the ratings at $-5.4 \leq A \leq 2.0$ cents ($\alpha=0.05$).

2. Comparison of the results from experiment 1 and experiment 2

In experiment 1 the tuning systems ranged between A=2.0 and A=-10.0 cents, whereas in experiment 2 they ranged between A=0.0 and A=-7.2 cents. One could argue that in experiment 1 the differences in ratings at the more relevant systems ($-5.4 \leq A \leq 0.0$ cents) may have been depressed as a result of the presentation of the more extreme systems at A=2.0 and A=-10.0 cents (Repp, personal communication, 1985).

Because these more extreme systems were not included in experiment 2, it is interesting to compare the results of the two experiments, especially those obtained in the slow tempo condition in experiment 1 (Fig. 5) and the complete spectrum condition in experiment 2 (Fig. 7). A comparison of the mean results at $-7.2 \leq A \leq 0.0$ cents in these figures shows that, although the ratings in experiment 2 are all lower than those in experiment 1, the effect of tuning system is about the same in the two experiments. Yet, we preferred to compare the results in more detail because the data in Fig. 7 are based on a smaller number of subjects and musical fragments than the data in Fig. 5. Therefore, we reanalyzed the ratings given in experiments 1 and 2 for corresponding subjects and musical fragments. An ANOVA [a 2 (experiments) x 16 (subjects) x 5 (tuning systems) x 8 (musical fragments) factorial design, all repeated measures] performed on the ratings, showed that the general shift in the mean ratings of about 0.5 was significant [$F(1,15)=13.7$, $p < 0.003$]. For the tuning systems ordered from $A=0.0$ to $A=-7.2$ cents, the acceptability ratings from experiment 1, averaged across the reduced set of conditions, were 4.0, 4.7, 4.6, 4.3, and 3.6, respectively. The interaction effect of experiment and tuning system, however, was far from significant ($p = 0.89$), which suggests that perceptual differences between tuning systems do not change as a result of a decrease of the range of tuning systems presented within a block of trials.

3. Overall acceptability ratings and purity ratings of isolated intervals

The data of experiment 2 provide a second occasion to relate the mean overall acceptability ratings to subjective purity ratings of isolated intervals. Two multiple linear regression analyses were carried out. One analysis was performed on the overall ratings that were given to the musical fragments played in the complete spectrum condition and the subjective purity ratings for the intervals with spectrally comparable tones presented at a duration of 0.5 s (Vos, 1986, Table II). The other analysis was based on the acceptability ratings given in the reduced spectrum condition and the purity ratings for the intervals with tones (again 0.5 s in duration) whose spectra were in the same way reduced as the tones in the musical fragments. In Table V, the relevant ratings are given for the five different tuning systems and the two spectral conditions, separately.

As in experiment 1, overall acceptability (X_1) can only to a limited degree be predicted from the impurity ratings of either fifth (X_2) or major

Table V Mean overall acceptability ratings (experiment 2) of musical fragments performed according to various regular 12-tone tuning systems, and mean subjective purity ratings for corresponding isolated fifths and major thirds. The acceptability ratings were related to impurity* ratings for intervals with tones that both in the complete and in the reduced spectral conditions were spectrally identical to the tones used in the musical fragments. The duration of the isolated intervals was 0.5 s.

Tuning system (tempering of main fifths in cents)	Overall Acceptability Rating (X_1)		Subjective purity rating			
			Fifth (X_2)		Major third (X_3)	
	complete	reduced	complete	reduced	complete	reduced
0.0	3.6	4.6	8.9	9.6	4.0	5.6
- 2.0	4.1	5.1	8.2	8.9	4.9	6.4
- 3.9	4.0	5.0	7.5	8.4	6.0	7.3
- 5.4	3.9	4.8	7.1	7.9	6.8	8.0
- 7.2	3.2	3.9	6.6	7.4	5.4	6.7

*Note: The impurity rating is simply the difference between the theoretically maximum purity rating (= 10) and the observed purity rating.

Table VI Correlation coefficients (r), regression weights (m), and constants (c), in the prediction of the overall acceptability rating (X_1) from subjective impurity ratings for fifths (X_2) and major thirds (X_3) by means of the multiple linear regression equation $X_1' = m_1X_2 + m_2X_3 + c$. Further details are given in Table V.

	Spectral condition	
	Complete	Reduced
r_{12}	-0.34	-0.51
r_{13}	-0.28	-0.22
$r_{1.23}$	0.85	0.91
m_1	-0.50	-0.67
m_2	-0.41	-0.53
c	6.82	7.38

third (X_3). For the reduced spectrum condition, correlation coefficients are r_{12}=-0.51 and r_{13}=-0.22. Predictability of X_1 is high when both X_2 and X_3 are taken into account, $r_{1.23}$=0.91. Comparable results were obtained for the complete spectrum condition (see Table VI, in which r-values are listed for the two conditions). We may conclude that also in spectrally different conditions overall acceptability can be predicted from the combined effect of the (im)purity ratings of the fifths and the major thirds. The values of the regression weights and the constants [see Eq. (9)], are given in Table VI as well.

V. GENERAL DISCUSSION

A. Harmonic intervals

Mean overall acceptability of the performance of our two-part musical fragments was about the same for $-5.4 \leq A \leq 0.0$ cents, whereas for $A > 0.0$ and $A < -5.4$ cents, overall acceptability strongly decreased. This specific effect of tuning system was the same both in the two tempo conditions and in the two spectral conditions. The main reason to vary tempo, or tone duration, and spectrum was to reduce the contribution of beats and roughness in the perception of the tempered harmonic intervals. A decreased perceptibility of beats, it was thought, might reduce the differences in evaluation of the various tuning systems. Since reduced duration and spectral content of the tones resulted in an upward shift of the overall acceptability without affecting the differences in evaluation of the tuning systems, it seems that this hypothesis has to be rejected.

Our analysis has shown that overall acceptability is highly correlated with a linear combination of the purity ratings of isolated fifths and major thirds. In a previous set of experiments (Vos, 1986), we concluded that the strong differentation in rating of these isolated intervals may be ascribed mainly to the presence of beats or roughness. Despite the high multiple correlation between overall acceptability and subjective purity of isolated harmonic intervals and the correspondence between frequency of occurrence and value of the regression weight, noted in section III C2, it is not proven that our subjects have based their overall judgments on the subjective purity of the separate harmonic intervals. Some support for the argument that they may have based their overall judgments at least partly on the subjective purity of the melodic intervals is given below.

B. Melodic intervals

Melodic intervals are musical intervals between the successive tones in the discant part and between the successive tones in the bass part. For all tuning systems the temperings of these melodic intervals are identical to those of the harmonic intervals (see Table I). The melodic intervals were unison, minor second or diatonic semitone, major second or whole tone, minor

Table VII Relative frequencies of occurrence of the melodic intervals in the 24 two-part musical fragments, for the discant and the bass parts, separately.

Melodic interval	Frequency of occurrence in % Discant	Bass	Mean
Unison	18.0	7.4	12.7
Minor second	20.5	7.9	14.2
Major second	39.8	21.0	30.4
Minor third	13.3	7.1	10.2
Major third	2.4	3.6	3.0
Fourth	3.6	18.6	11.1
Fifth	2.4	19.8	11.1
Minor sixth	0.0	1.8*	0.9
Major sixth	0.0	3.0*	1.5
Minor seventh	0.0	3.0*	1.5
Major seventh	0.0	3.6*	1.8
Octave	0.0	3.2	1.6
Total	100.0	100.0	100.0

* Note: These intervals were exclusively the result of transpositions

third, major third, fourth, fifth, minor sixth, major sixth, minor seventh, major seventh, and octave. Relative frequencies of occurrence, averaged across the 24 fragments, are listed in Table VII.

1. Data from previous research

For successive sinusoidal tones, the perception of the sizes of all these intervals has been investigated by Rakowski (1976). Rakowski asked his musically trained subjects to adjust the frequency of either the lower or the higher tone until the interval size between the variable and the fixed tones corresponded to a specific musical interval. The median adjusted interval sizes are given in the second column of Table VIII. The deviations relative to the size of the pure intervals and to the size of the intervals in equal temperament are given in the third and the fourth columns, respectively.

From the values in these columns it can be concluded that the adjusted interval sizes are more strongly related to the equally-tempered sizes than to the sizes of the pure intervals: the mean absolute difference is a factor of two smaller when the first rather than the second reference is used. Further, it can be seen in Table VIII that, relative to the sizes of the corresponding equally-tempered intervals, there is a general tendency to compress intervals smaller than the fifth and to stretch the larger intervals.

The variability in the adjustments, defined as the difference between the adjusted interval sizes with percentiles of 25 and 75, ranged from about 20 to 40 cents (Table VIII, last column). On the assumption that the distribution of the adjustments between these percentiles is not too much different from a normal distribution, the standard deviation may have ranged between 30 and 60 cents. These estimated standard deviations seem to be very large, especially with regard to the observation that the difference limen found in adjustment experiments is smaller, sometimes even by a factor of two, than the value found with a binary-choice method (Cardozo, 1965; Wier et al., 1976).

Table VIII Perception of melodic intervals (method of adjustment; Rakowski, 1976). Median and variability (difference between the adjusted interval sizes with percentiles of 25 and 75) determined from the results for four frequencies of the fixed tone (125, 250, 500, and 1000 Hz), ascending and descending intervals, subjects, and replications. Values given in cents. Deviation given relative to the sizes of the physically pure and equally-tempered (ET) intervals.

Musical interval	Median adjusted interval size	Deviation re pure	Deviation re ET	Variability
Minor second	78	-34	-22	19
Major second	184	-20	-16	27
Minor third	293	-23	-7	35
Major third	397	11	-3	29
Fourth	498	0	-2	25
Fifth	700	-2	0	33
Minor sixth	804	-10	4	35
Major sixth	904	24	8	42
Minor seventh	1013	17	13	39
Major seventh	1107	19	7	39

Table IX Thresholds (79.4% correct in 2-AFC paradigm) for discrimination between pure and tempered melodic intervals [Vos and van Vianen (1986)]. Data from an experiment in which the subjective purity of pure and tempered melodic intervals was rated (Vos, 1986) are summarized in the last two columns. All temperings are given in cents relative to the pure intervals. T denotes the degree of tempering.

Musical interval	Spectral condition	Thresholds in cents			Tempering at which mean purity rating is lower ($p < 0.01$) than at T = 0 cents	
		compressed	stretched	mixed	compressed	stretched
Major third	sinusoid	-31.5	51.6	±47.9	-35	> 50
	complex	-26.0	42.6	±42.3	-	-
Fifth	sinusoid	-37.5	42.3	±45.6	-50	50
	complex	-30.8	38.1	±39.7	-	-

Burns and Ward (1978) presented their subjects with pairs of successive melodic intervals (sinusoids) and asked them to indicate which interval was wider. After moderate training stretched major thirds could be discriminated from compressed major thirds in 70.7% of the cases when the difference between the intervals was 30 cents. For intervals that approximated minor third and fourth, the mean thresholds were about 20 cents (Burns and Ward, 1978, Fig. 4).

Thresholds for discrimination between pure and tempered melodic intervals have also been determined by Vos and van Vianen (1986). These thresholds are summarized in Table IX for major third and fifth in various conditions. Especially for the major third, the thresholds depended on whether a block of trials consisted only of pure and compressed, pure and stretched, or pure and both compressed and stretched intervals. The asymmetry between the thresholds for compressed and stretched melodic major thirds may indicate that the "internalized" pure interval is closer to the equally tempered major third than to the pure major third. In the experiment of Vos and van Vianen (1986) this shift is from about 386 to 395 cents, which is in line with the findings of Rakowski (1976; see our Table VIII). It is important to note that the thresholds for the intervals (unison included) with complex tones were, on the average, 7 cents lower than for the intervals with sinusoidal tones. This spectral effect was not significantly affected by the other variables.

To our knowledge, experimental results on the relation between tempering and subjective purity of isolated melodic intervals have only been reported by

Vos (1986). In comparison with harmonic intervals, subjective purity of melodic intervals (with sinusoidal tones) is only weakly affected by tempering. For major thirds and fifths tempering had to be 35 cents or more to obtain ratings that were significantly lower than those obtained close to or at physical purity. The relevant data are summarized in the last two columns of Table IX and can be compared with the discrimination thresholds in the corresponding conditions.

2. Estimated importance of the melodic intervals

In the estimation of the role that the melodic intervals may have played in overall acceptability, the following observations are relevant:

a) The experiments of Rakowski (1976), Vos (1986), and Vos and van Vianen (1986) have shown that the sizes of melodic intervals are related more to the equally-tempered sizes than to the sizes of the pure intervals.

b) If tempering is expressed relative to the size in equal temperament, the temperings of a number of intervals are larger than indicated in Table I. For A=-7.2 cents the tempering of the minor second is 26 cents and that of the major third is -21 cents; for A=-10.0 cents, the temperings of these intervals are 40 and -32 cents, respectively, and the tempering of the minor third is 24 cents. For A=2.0 cents, finally, the tempering of the minor second is -20 cents.

c) The minor and major seconds, the fourth, and the major third have lower difference limens than the other intervals (Rakowski, 1976).

d) Sensitivity to tempering is about 7 cents higher for melodic intervals with complex tones than with sinusoidal tones (Vos and van Vianen, 1986).

e) The level at which intervals can be discriminated or recognized depends on procedural or stimulus complexity. This has been demonstrated by, e.g., Zatorre and Halpern (1979), who found for their subjects that consistency of labeling data was entirely dependent on the constancy of the lower tone. Sensitivity to tempering may also depend on stimulus complexity. In the experiment of Burns and Ward (1978), all stretched and compressed major thirds had the same central frequency, which means that when, e.g., a compressed interval was presented before the stretched version, the frequency of the lower tone was always decreased and the frequency of the higher tone was always increased. In the experiment of Vos and van Vianen (1986), stimulus complexity was much higher because even within one trial the central frequencies of the pure and tempered intervals were different in 90% of the cases.

Sensitivity to tempering of the melodic intervals may have been higher in our musical fragments than in isolated intervals presented in a context characterized by a more or less high stimulus complexity. The melodic intervals may have been processed more adequately in the musical context. In our fragments the number of different tones within the octave ranged between 5 and 7 only. This limited number of tones might have facilitated the formation of a stable frame of reference within which the various intervals could be evaluated. Expressed differently, the musical context may have provided more relational cues than the only one available in isolated intervals (see also Dewar et al., 1977).

f) Considering the observations summarized in (d) and (e), we hypothesize that when the temperings of the melodic intervals in our musical fragments exceed a value of about 20 cents (relative to equal temperament), the subjective purity of these intervals is significantly decreased. This is supported by the results from a recent experiment on the perception of melodic and harmonic intonation in musical fragments that were selected from the same set of chorale settings used in our experiments (Rasch, 1985). At each level of harmonic tempering there were three conditions in which either the discant part, or the bass part, or both the discant and the bass part were tempered. In the condition with the most extreme harmonic tempering but one (the condition with a mistuning factor of '5'), Rasch (1985) found that the perceived quality of intonation was much lower when the discant part was tempered than when the bass part was tempered. In the discant part the mean absolute tempering of the minor second (relative frequency of occurrence is 23%) was 19 cents and that of the major third (relative frequency of occurrence is 6%) was 32 cents, whereas in the bass part the mean absolute temperings of the various melodic intervals were all lower than 10 cents. The temperings of the minor second and the major third in the discant part may have been responsible for the obtained difference in perceived quality of intonation.

From the above we may conclude that for tuning systems in which $A > 0.0$ and $A \leq -7.2$ cents, at least some melodic intervals, including minor second and minor and major third, may have been perceived to be relatively impure and that they therefore may have affected overall acceptability. For systems in which $-5.4 \leq A \leq 0.0$ cents, however, it is unlikely that the various tempered melodic intervals have played a significant role.

C. Tuning systems

Our musical fragments were based on four-part chorale settings by Praetorius. Since meantone tuning was the most important system in the 16th and 17th centuries, and since it is known that Praetorius himself advocated this system (Praetorius, 1619) we would not have been surprised to find that overall acceptability was high at the meantone tuning system. What we did find, however, is that the Silbermann tuning, equal temperament, and the Pythagorean tuning are about as acceptable as the meantone tuning system. In our experiments no enharmonically equivalent tones were used, which means that the high acceptability at equal temperament cannot be explained by the additional positive effect caused by the possibility of free modulation. Since free modulation is an important requirement in Western music of the last three centuries it would be imprudent to reject equal temperament, especially because, overall, the tempered intervals seem not to be worse than those in the other regular 12-tone tuning systems.

VI. GENERAL CONCLUSIONS

Musically trained subjects rated the overall acceptability of the performance of two-part musical fragments. As long as these fragments were performed according to regular 12-tone tuning systems in which the physical purity of both the fifths and the major thirds were optimized ($-5.4 \leq A \leq 0.0$ cents), overall acceptability was about equally high. For tuning systems beyond this range the mean ratings strongly decreased.

Essentially the same effect of tuning system was found when instead of using a seven-point equal-interval scale, acceptability was determined by means of the method of paired comparisons, yielding scale values at the level of an interval scale.

Reduction of the duration and the spectral content of the tones, by means of which the prominence of beats and roughness in the perception of the tempered harmonic intervals was diminished, resulted in an upward shift of acceptability without affecting the perceptual differences between the various tuning systems.

Overall acceptability could be accurately predicted from a linear combination of the purity ratings of the harmonic fifths and the major thirds. However, it cannot entirely be excluded that the subjects based their acceptability ratings partly on the subjective purity of the melodic intervals. This hypothesis has to be tested in future research.

Overall, the tempered intervals in equal temperament seem not to be worse than those in the other regular tuning systems at $-5.4 \leq A \leq 0.0$ cents. This may explain why equal temperament, in which free modulation is possible, has become widely accepted.

ACKNOWLEDGMENTS

Part of this research was supported by Grant 15-29-05 from the Netherlands Organization for the Advancement of Pure Research (ZWO). The author is grateful to Ino Flores d'Arcais, to Rudolf Rasch, and to Reinier Plomp for their comments on an earlier draft of this paper.

REFERENCES

Barbour, J.M. (1940). "Musical logarithms", Scripta Math. 7, 21-31.

Barbour, J.M. (1951). Tuning and Temperament (Michigan State College, East Lansing, reprinted 1972, Da Capo, New York).

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

Cardozo, B.L. (1965). "Adjusting the method of adjustment: SD vs DL", J. Acoust. Soc. Am. 37, 786-792.

Clark, H.H. (1973). "The language-as-fixed-effect fallacy: a critique of language statistics in psychological research", J. Verbal Learning Verbal Behavior 12, 335-359.

Cohen, H.F. (1984). Quantifying Music: The Science of Music at the First Stage of the Scientific Revolution, 1580-1650 (Reidel, Dordrecht).

Dewar, K.M., Cuddy, L.L., and Mewhort, D.J.K. (1977). "Recognition memory for single tones with and without context", J. Exp. Psychol.: Human Learning and Memory 3, 60-67.

Dupont, W. (1935). Geschichte der musikalischen Temperatur (Bärenreiter, Kassel).

Esbroeck, G. van, and Monfort, F. (1946). Qu'est-ce que jouer juste? (Manteau, Brussels).

Hall, D.E. (1973). "The objective measurement of goodness-of-fit for tunings and temperaments", J. Music Theory 17, 274-290.

Hall, D.E., and Hess, J.T. (1984). "Perception of musical interval tuning", Music Percept. 2, 166-195.

Kok, W. (1954). "Experimental study of tuning problems", Acustica 4, 229-230.

Kok, W. (1955). Harmonische orgels (Technological University, Delft) Doctoral thesis.

Kolinski, M. (1959). "A new equidistant 12-tone temperament", J. Am. Musicol. Soc. 12, 210-214.

Loman, A.D. (1929). De logische grondslagen der muziek (Alsbach, Amsterdam).

Martin, D.W., and Ward, W.D. (1961). "Subjective evaluation of musical scale temperament in pianos", J. Acoust. Soc. Am. 33, 582-585.

Meerling (1981). Methoden en technieken van psychologisch onderzoek. Deel 2: Data-analyse en psychometrie (Boon, Meppel).

Meinel, H. (1954). "Zur Stimmung der Musikinstrumente", Acustica 4, 233-236.

Meinel, H. (1957). "Musikinstrumentenstimmungen und Tonsysteme", Acustica 7, 185-190.

Pikler, A.G. (1966). "Logarithmic frequency systems", J. Acoust. Soc. Am. 39, 1102-1110.

Praetorius, M. (1609). Musae Sioniae Sechster Theil (Wolfenbüttel: Möseler Verlag. Revision by F. Reusch, Gesamtausgabe der Musikalischen Werke von Michael Praetorius, Musae Sioniae, Band VI).

Praetorius, M. (1619). Syntagma Musicum, Tomus Secundus: De Organographia (Holwein: Wolfenbüttel, Germany).

Railsback, O.L. (1938). "A study of the tuning of pianos", J. Acoust. Soc. Am. 10, 86.

Rakowski, A. (1976). Tuning of isolated musical intervals. Unpublished paper presented at the 91st Meeting of the Acoustical Society of America, Washington, D.C.

Rasch, R.A.(1981). Aspects of the perception and performance of polyphonic music (Elinkwijk BV, Utrecht) Doctoral thesis.

Rasch, R.A. (1983). "Description of regular twelve-tone musical tunings", J. Acoust. Soc. Am. 73, 1023-1035.

Rasch, R.A. (1984). "Theory of Helmholtz-beat frequencies", Music Percept. 1, 308-322.

Rasch, R.A. (1985). "Perception of melodic and harmonic intonation of two-part musical fragments", Music Percept. 2, 441-458.

Schuck, O.H., and Young, R.W. (1943). "Observations on the vibrations of piano strings", J. Acoust. Soc. Am. 15, 1-11.

Terhardt, E., and Zick, M. (1975). "Evaluation of the tempered tone scale in normal, stretched, and contracted intonation", Acustica 32, 268-274.

Torgerson, W.S. (1958). Theory and methods of scaling (Wiley & Sons, New York).

Vêroli, C. di (1978). Unequal temperaments and their role in the performance of early music (Véroli, Buenos Aires).

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

Vos, J. (1984). "Spectral effects in the perception of pure and tempered intervals: Discrimination and beats", Percept. Psychophys. 35, 173-185.

Vos, J. (1986). "Purity ratings of tempered fifths and major thirds", Music Percept. 3, 221-257.

Vos, J., and Vianen, B.G. van (1985a). "Thresholds for discrimination between pure and tempered intervals: The relevance of nearly coinciding harmonics", J. Acoust. Soc. Am. 77, 176-187.

Vos, J., and Vianen, B.G. van (1985b). "The effect of fundamental frequency on the discriminability between pure and tempered fifths and major thirds", Percept. Psychophys. 37, 507-514.

Vos, J., and Vianen, B.G. van (1986). Thresholds for discrimination between pure and tempered melodic unisons, major thirds, and fifths (IZF-report 1986-3, Soesterberg, The Netherlands).

Ward, W.D. (1970). Musical perception. In J. Tobias (Ed.), Foundations of Modern Auditory Theory. Vol. 1, 407-447 (Academic Press, New York).

Ward, W.D., and Martin, D.W. (1961). "Psychophysical comparison of just tuning and equal temperament in sequences of individual tones", J. Acoust. Soc. Am. 33, 586-588.

Wier, C.C., Jesteadt, W., and Green, D.M. (1976). "A comparison of method-of-adjustment and forced-choice procedures in frequency discrimination", Percept. Psychophys. 19, 75-79.

Winer, B.J. (1970). Statistical principles in experimental design (McGraw-Hill, London), International student ed.

Young, R.W. (1939). "Terminology for logarithmic frequency units", J. Acoust. Soc. Am. 11, 134-139.

Young, R.W. (1954). "Inharmonicity of piano strings", Acustica 4, 259-262.

Zatorre, R.J., and Halpern, A.R. (1979). "Identification, discrimination, and selective adaptation of simultaneous musical intervals", Percept. Psychophys. 26, 384-395.

8
Epilogue

Basically, the aim of the research described in this study was to answer one central question: How important are the differences between the tuning systems proposed through the ages for the listener? In answering this question, many aspects are relevant. In fact, the psychological research on this problem in music perception is still in its infancy. The psychoacoustical experiments described in this study have to be interpreted as preparatory work. The differences between the various tuning systems are, in general, relatively small: investigation of the perception of intervals with simultaneous complex tones seemed to be most relevant. We reasoned that subjective evaluations of various musical intervals can be different only if these intervals can be discriminated. In the first chapters of this study, we therefore investigated discriminability between pure and tempered intervals. The present research is new, because no experimental results on discriminability between musical intervals with two simultaneous complex tones have been reported before.

In three of the four chapters on discrimination, we focused on the fifth and the major third. We determined the effects of physical parameters such as the spectral content, the fundamental frequency, and the duration of the tones on discriminability. In many conditions, the effects of the experimental variables on discrimination could be successfully related to the effects on, for example, perceived strength of beats in tempered intervals and on subjective purity. In addition, it could be concluded that the discrimination thresholds provided an additional quantitative measure for tonal consonance and that the thresholds were consistent with the role of the various intervals in tuning procedures.

Our study also suggested that the experimental findings could be applied to the intervals played on the organ and the harpsichord, and at least to the three or four central octaves of the piano. The experiments on subjective purity showed that the differentiation in ratings between the pure and the tempered intervals depended mainly on the sensation of beats and roughness. Because of this, the purity ratings provided a test for the predictive power of models on tonal or sensory consonance, such as the model proposed by Plomp and Levelt (1965).

In the previous chapter of this study we investigated the perception of pure and tempered intervals in a musical context. In the experiments the temperings of the various intervals corresponded to those in well-known tuning systems. The structure of the musical fragments was deliberately kept very simple. The results were surprising because they showed that playing in equal temperament, which has now been generally accepted for more than a century, needs in no way to be in conflict with the application of historical tuning systems by those who are in favor of authentic performances of old music.

The experimental results on the subjective acceptability of the various tuning systems in the musical fragments also emphasized the coherence of the design of the present study, because the acceptability ratings could be reduced to the subjective purity ratings of the separate harmonic intervals, while it has been demonstrated before that these purity ratings and the capacity to discriminate between pure and tempered intervals are related.

We conclude that this research has to be continued. One of the questions that are still open is, of course, to what extent the subjective purity of the melodic intervals contributes to the overall acceptability of the various tuning systems. In addition, it might be interesting to test in more detail the additivity model proposed with fragments that are different with respect to, for example, musical style and with fragments in which enharmonically equivalent intervals are used.

Another extension of the research on the perceptual relevance of differences between tuning systems could be that to three and four-part music. It is difficult to foresee whether the differences between the tuning systems will increase or decrease with the number of voices in a composition. As a last suggestion the rhythmical structure of the music may also be of significance. In the musical fragments used in this study, the tones in the lower and higher parts were strictly synchronized. Small degrees of asynchronization between the perceptual onsets of the tones, which are typical of performed music (Rasch, 1981; Vos and Rasch, 1981), may enhance the separate perception of the tones in the harmonic intervals and may therefore facilitate the detection of differences in interval size without reducing the interference between nearly coinciding harmonics. A similar facilitation of the perception of interval size may be found in many polyphonic compositions, where the onsets of the tones in the different voices coincide only partly or not at all (see Rasch, 1981). It is not known whether the resulting progression of harmonic intervals affects the sensation of beats and roughness. It is likely, however, that a more advanced model will be needed when overall acceptability of tuning systems in such complex musical stimuli has to be described and explained.

References

Plomp, R., and Levelt, W.J.M. (1965). "Tonal consonance and critical bandwidth," J. Acoust. Soc. Am. 38, 548-560.

Rasch, R. A. (1981). Aspects of the perception and performance of polyphonic music (Elinkwijk BV, Utrecht) Doctoral thesis.

Vos, J., and Rasch, R. A. (1981). "The perceptual onset of musical tones," Percept. & Psychophys. 29, 323-335.

Summary

In a pure consonant interval the ratio between the fundamental frequency f_1 of the lower tone and the fundamental frequency f_2 of the higher tone is equal to p:q, with p and q being small integers. For such an interval with simultaneous complex tones, the harmonics with frequencies of npf_2 and nqf_1 Hz (n=1, 2, ...) coincide. In the fifth (p=2; q=3), for example, the second (pth) harmonic of the higher tone coincides with the third (qth) harmonic of the lower tone. In a tempered interval ($f_1:f_2 \sim p:q$), these harmonics have different frequencies. The interference of these nearly coinciding harmonics results in level (=envelope) variations. The depth of the level variation depends on the relative sound levels of the interfering harmonics. For the nth pair of interfering harmonics, the frequency of the level variation is equal to $f_{bn}=|npf_2-nqf_1|$ Hz.

Perceptually, the level variation gives rise to beats or roughness at a beat strength that is related to level-variation depth and at a beat frequency, f_b, that in most cases will be equal to $|pf_2-qf_1|$ Hz. In addition to beats, changes in the size of these intervals may be detected.

In a musical context, both the sensation of beats or roughness and the detection of changes in interval size may affect the degree to which these tempered intervals are experienced to be "impure" or "out-of-tune". In tuning keyboard instruments such as the organ, the harpsichord, and the piano, it is inevitable that some of the intervals are tempered. The decision as to which intervals have to be tempered, and to what extent, depends on many factors. Accordingly, numerous tuning systems have been proposed in the history of music. In the discussion and the evaluation of the different tuning systems, it is important to know our sensitivity to tempering.

In Chapters 2, 3, 4, and 5 of this study, we therefore determined thresholds for discrimination between pure and tempered musical intervals. In the experiments we focused on intervals with simultaneous complex tones because it may be expected that with these intervals, the differences between the tuning systems are perceived most clearly. In the experiments the depth of the level variation was manipulated by the introduction of differences in sound level between the two tones. The discrimination threshold (DT) was expressed in terms of this difference or, in a direct manner, in terms of the depth of the level variation. The degree of tempering, expressed in f_b or in the logarithmic cent, was one of the independent variables. The experiments on discriminability were supplemented with studies in which pure and tempered intervals

were rated: In Chapter 6, we examined the relationship between tempering and subjective purity, and in Chapter 7, we studied subjective acceptability of tuning systems in musical fragments. In all experiments included in this study the participants were musically trained subjects.

In Chapter 2 we found that the DTs were lower (higher sensitivity to tempering) for fifths than for major thirds ($p=4$; $q=5$), and were relatively high for temperings up to about $\pm$ 4 cents. Especially for temperings up to about $\pm$ 16 cents, the DTs decreased with increasing tone duration (0.25, 0.5, and 1 s), whereas for temperings $>$ 32 cents and $<$ -32 cents, the DTs were about the same for the three tone durations included. It was concluded that at the higher degrees of tempering, DTs could be based more on the detection of differences in interval size than on the sensation of beats or roughness. A comparison of beat thresholds and thresholds for the identification of the direction of tempering with the DTs, however, revealed that in most conditions included, the DTs were mainly determined by sensitivity to beats.

In Chapter 3 the spectral content of the higher tones was varied to determine to what extent the DTs investigated in Chapter 2 were the result of the interference of nearly coinciding harmonics. The results showed that for fifths tempered by up to $\pm$ 30 cents, DTs are determined mainly by the interference of the first pair of nearly coinciding harmonics. For major thirds, however, the data suggest that the interference of other harmonics plays a role as well.

In Chapter 3 we also report the results of an experiment in which we determined both the dominantly perceived frequency and the perceived strength of beats in supraliminally tempered fifths and major thirds. In addition to musical interval, spectral content of the higher tone and degree of tempering were the independent variables again. In all spectral conditions included, we found that for both the fifth and the major third, the beat frequency of $|pf_2-qf_1|$ Hz was perceived more clearly than multiples of this beat frequency. From a comparison of the DTs and the perceived strengths of the beats for corresponding spectral conditions, it was concluded that especially for the fifth, perceived strength of beats and sensitivity to moderately tempered intervals are highly correlated.

In each of the three experiments reported in Chapter 4, DTs were determined for 13 intervals with different values of p and/or q. The experiments showed that there is a simple relation between frequency-ratio complexity and discriminability: DTs gradually increased with increasing $p+q$. The experiment in which tones with harmonics of equal amplitude were used indicated that level of the interfering harmonics was not responsible for this relationship. Yet, the third experiment, in which the spectral content of the tones was varied, clearly showed that for all intervals DT was determined by the interference between nearly coinciding harmonics. Because of this, the DTs for intervals that comprise tones with a complete series of harmonics should be expressed in terms of level-variation depth. Detailed analysis of the results revealed that the relation between ratio-complexity and DT might have been the result of masking.

In the experiments described in Chapters 2, 3, and 4, DTs were only determined for a restricted frequency range. For the fifths and the major thirds, for example, f_1 and f_2 had always been between middle C (C4 = 261.6 Hz) and the octave above middle C (C5). In Chapter 5 we determined to what extent previous findings might be generalized to the much wider frequency range between C2 and C6. The results showed that DTs do not depend on the frequency of the fundamentals, provided that the DTs are ordered according to f_b. When tempering is expressed in cents and only those conditions are considered in which the degree of tempering corresponds to that found in the relevant tuning systems, it can be concluded that perceptual differences between tempered and pure fifths are increasingly irrelevant for tones lower than about C4. Tempered major thirds, even those in a compromise tuning system such as that proposed by Silbermann (with a tempering of 5.9 cents), can be discriminated from pure major thirds for tones higher than about C3.

In Chapter 6 we found that both for isolated fifths and for major thirds (with complex tones), the relation between tempering and subjective purity can be described by exponential functions. Fifths compressed by up to about 20 cents or stretched by up to about 15 cents were rated to be purer than the corresponding major thirds. Especially for stretched intervals, the reverse tended to be the case with larger temperings. Deletion of the even harmonics of the higher tone, as a result of which interference of various pairs of nearly coinciding harmonics was canceled, resulted in higher purity ratings. This effect was most prominent for the major third. A further reduction of the interference between harmonics was still more effective: for simultaneous sinusoidal tones subjective differences between pure and tempered intervals were much smaller than for complex tones.

Because the strong differentiation in rating between the pure and the tempered intervals, as found for complex tones, may be ascribed mainly to the presence of beats or roughness, we compared the purity ratings with dissonance patterns predicted by models for tonal consonance/dissonance. Only in a few conditions do the patterns predicted by the model of Plomp and Levelt (1965) resemble the rating patterns obtained, and the dissonance patterns predicted by the model of Kameoka and Kuriyagawa (1969) are at variance with the purity ratings in all conditions. We suggested that the predictive power of the two consonance/dissonance models may be enhanced by adoption of the principles underlying the Zwicker and Scharf (1965) model on loudness summation.

In Chapter 7 we studied the subjective overall acceptability of various regular 12-tone tuning systems in two-part musical fragments. Included were Pythagorean tuning (tempering of the main fifths, A, equal to 0.0 cents), equal temperament (A=-2.0 cents), Silbermann (A=-3.9 cents), meantone (A=-5.4 cents), and Salinas tuning (A=-7.2 cents). In the first experiment two systems in which A=2.0 or A=-10.0 cents were also included. Mean acceptability ratings were about the same for $-5.4 \leq A \leq 0.0$ cents, whereas for $A > 0.0$ and $A < -5.4$ cents the ratings strongly decreased. The specific effect of tuning system was not affected by the tempo at which the fragments were played. The perception of beats was manipulated in the second experiment. The condition in which the interference of the nearly coinciding harmonics was canceled resulted in higher acceptability. In both experiments, acceptability could be accurately predicted from a linear combination of the purity ratings of the harmonic fifths and major thirds described in Chapter 6.

De waarneming van zuivere en ontstemde muzikale intervallen

Samenvatting

Zuivere intervallen worden gekenmerkt door een eenvoudige verhouding, p:q, tussen de grondfrequenties f_1 en f_2 van respectievelijk de laagste en de hoogste toon, zoals bijvoorbeeld 1:2 voor het octaaf, 2:3 voor de kwint en 4:5 voor de grote terts. Als twee gelijktijdig klinkende tonen, elk samengesteld uit een grondtoon en een serie harmonischen (boventonen), samen zo'n zuiver interval vormen, vallen bepaalde harmonischen van de hoogste toon [die met een frequentie van npf_2 Hz (n=1, 2, ...)] samen met harmonischen van de laagste toon [die met een frequentie van nqf_1 Hz]. Voor de zuivere kwint, bijvoorbeeld, is de frequentie van de tweede (p-de) harmonische van de hoogste toon gelijk aan die van de derde (q-de) harmonische van de laagste toon.

Bij onzuivere (of: ontstemde of getemperde) intervallen wijkt de verhouding f_1:f_2 iets af van p:q, waardoor ook de frequenties van de zojuist genoemde harmonischen iets van elkaar verschillen. De interferentie van deze bijna samenvallende harmonischen leidt tot variaties in de omhullende van de tonen in het ontstemde interval. De diepte van deze omhullendevariatie wordt bepaald door de amplitudes van de interfererende harmonischen: als de amplitudes even groot zijn kunnen ze elkaar afwisselend volledig versterken en volledig uitdoven, zodat de diepte van de omhullendevariatie maximaal is; als de amplitude van de ene harmonische veel sterker is dan die van de andere, is het verschil tussen versterking en verzwakking, en dus de diepte van de omhullendevariatie, gering. De frequentie van de omhullendevariatie is voor het n-de paar van interfererende harmonischen gelijk aan $|npf_2-nqf_1|$ Hz.

Voor de luisteraar leidt de omhullendevariatie tot de waarneming van zwevingen of ruwheid. De zwevingssterkte neemt toe met de diepte van de omhullendevariatie, terwijl in de meeste gevallen een zwevingsfrequentie van $|pf_2-qf_1|$ Hz in de waarneming zal overheersen. Naast zwevingen kan de met de ontstemming gepaard gaande verandering in de grootte van het interval worden opgemerkt. Bij niet-gelijktijdige tonen is dit zelfs het enige houvast voor het ontdekken van onzuiverheid.

In een muzikale context kunnen zowel de zwevingen of ruwheid, als de gewijzigde intervalgrootte van invloed zijn op de mate waarin deze ontstemde intervallen als onzuiver of vals worden ervaren. Bij het stemmen van toets-

instrumenten zoals bijvoorbeeld het orgel, het clavecimbel en de piano, is het onvermijdelijk dat er op z'n minst bij enkele intervallen onzuiverheden worden aangebracht. De keuze welke intervallen in welke mate ontstemd moeten worden hangt af van vele factoren, waaronder de muziek die moet worden gespeeld en het instrument dat daarvoor zal worden gebruikt. In de muziekgeschiedenis zijn dan ook vele stemmingssystemen voorgesteld.

De vraag die centraal staat in het hier beschreven onderzoek is hoe belangrijk de verschillen tussen deze stemmingssystemen nu eigenlijk voor de luisteraar zijn. Voor de discussie en de beoordeling van deze stemmingssystemen is het van belang te weten hoe gevoelig men voor ontstemming is. In de hoofdstukken 2, 3, 4 en 5 van deze studie gingen we allereerst na hoe goed men de zuivere intervallen kan onderscheiden van de ontstemde intervallen. We hebben ons hoofdzakelijk gericht op de zojuist beschreven intervallen met gelijktijdige samengestelde tonen, omdat de in het algemeen kleine verschillen tussen de stemmingssystemen bij deze intervallen nog het meest merkbaar zullen zijn. Tenzij anders vermeld, gebruikten we steeds tonen met 20 opeenvolgende harmonischen waarvan het geluidniveau gelijkmatig afnam met 6 dB per octaaf. Een dergelijke spectrale samenstelling lijkt op dat van tonen die door middel van een strijkstok in het lagere register van snaarinstrumenten ten gehore worden gebracht. In de experimenten werd de diepte van de omhullendevariatie gevarieerd door het aanbrengen van verschillen in het geluidniveau tussen de twee tonen. De onderscheidingsdrempel werd uitgedrukt in dit verschil of rechtstreeks in de diepte van de omhullendevariatie. De mate van ontstemming was een van de onafhankelijke variabelen en werd uitgedrukt in de eerder genoemde zwevingsfrequentie of in cents (de grootte van een octaaf is 1200 cents). De experimenten waarin de onderscheidingsdrempels werden bepaald werden aangevuld met onderzoek waarin zuivere en ontstemde intervallen werden beoordeeld: In hoofdstuk 6 onderzochten we het verband tussen ontstemming en subjectieve zuiverheid, en in hoofdstuk 7 bestudeerden we de subjectieve aanvaardbaarheid van stemmingssystemen in tweestemmige muziekfragmenten. Aan de experimenten namen uitsluitend muzikaal geschoolde proefpersonen deel.

In hoofdstuk 2 werd gevonden dat de onderscheidingsdrempel voor de zuiverheid van kwinten lager was dan voor grote tertsen (een lage drempel wil zeggen dat reeds een kleine diepte van de omhullendevariatie voldoende is voor het waarnemen van een verschil tussen een zuiver en een iets ontstemd interval; een lage drempel betekent dus een grote gevoeligheid voor ontstemming). In het algemeen was de onderscheidingsdrempel relatief hoog voor ontstemmingen tot ongeveer ± 4 cents. Voor ontstemmingen tot ongeveer ± 16 cents werd de drempel lager naarmate de toonduur (0.25, 0.5 en 1 s) toenam. Voor ontstemmingen groter dan 32 cents en kleiner dan -32 cents waren de drempels voor de drie onderzochte toonduren ongeveer even laag. Aangezien het aantal zwevingen bij een langere toonduur groter is dan bij een kortere toonduur, leken de zwevingen bij deze grotere ontstemmingen niet meer van belang te zijn en zou het onderscheidingsvermogen op de waarneming van verschillen in intervalgrootte gebaseerd kunnen zijn. Uit een vergelijking van zwevingsdrempels en drempels voor het benoemen van de richting van de ontstemming (in termen van groter of kleiner dan het zuivere interval) met de onderscheidingsdrempel bleek echter dat ook bij grotere ontstemmingen de gevoeligheid voor zwevingen in de meeste condities bepalend voor de onderscheidingsdrempel is geweest.

In hoofdstuk 3 onderzochten we tot op welke hoogte de in hoofdstuk 2 verkregen onderscheidingsdrempels werkelijk door de interferentie van de bijna samenvallende harmonischen zijn bepaald. Hiertoe varieerden we de spectrale inhoud van de hoogste toon. De resultaten lieten zien dat voor ontstemmingen tot ± 30 cents de drempels voor de kwinten grotendeels door de interferentie van het eerste paar (n=1) bijna samenvallende harmonischen is bepaald. Bij grote tertsen was de interferentie van andere harmonischen echter ook van belang.

In hoofdstuk 3 beschrijven we verder een experiment waarin we in bovendrempelige condities (beide tonen ongeveer even luid) voor maximaal tot ± 12 cents ontstemde kwinten en grote tertsen de frequentie en de sterkte van de duidelijkst waargenomen zwevingen bepaalden. Dit deden we door proefpersonen deze zwevingen met behulp van twee sinustonen te laten nabootsen. Zij konden de zwevingsfrequentie aanpassen door de frequentie van een van de sinustonen te veranderen. De zwevingssterkte kon gevarieerd worden door het geluidniveau van deze sinustoon te wijzigen. Naast muzikaal interval waren het spectrum van de hoogste toon en de mate van ontstemming weer de onafhankelijke variabelen. In alle onderzochte spectrale condities bleken zowel voor de kwint als de grote terts de zwevingen met een frequentie van $|pf_2-qf_1|$ Hz duidelijker te worden waargenomen dan veelvouden van deze frequentie. Verder vergeleken we de waargenomen zwevingssterkte en de onderscheidingsdrempel uit overeenkomstige spectrale condities met elkaar. Uit deze vergelijking bleek dat vooral bij de kwint de waargenomen zwevingssterkte en het onderscheidingsvermogen een duidelijk verband vertonen.

In de drie experimenten die in hoofdstuk 4 worden gerapporteerd, werd de onderscheidingsdrempel in ieder experiment bepaald voor 13 intervallen met verschillende waarden van p en/of q. Uit de resultaten bleek dat er een eenvoudige relatie bestaat tussen de complexiteit van de frequentieverhouding en het onderscheidingsvermogen. Dit vermogen nam geleidelijk af met de som van p en q. In de door ons gebruikte samengestelde tonen zijn de hogere harmonischen zwakker dan de lagere harmonischen. Uit het experiment waarin alle harmonischen in de gebruikte tonen even sterk waren, bleek dat het geluidniveau van de interfererende harmonischen niet voor de gevonden relatie verantwoordelijk was. Het derde experiment, waarin het spectrum van de tonen werd gevarieerd, liet duidelijk zien dat de drempel voor alle 13 intervallen door de interferentie tussen bijna samenvallende harmonischen werd bepaald. Voor intervallen die bestaan uit tonen met een volledige serie harmonischen zou de onderscheidingsdrempel daarom in de diepte van de omhullendevariatie uitgedrukt moeten worden. Een verdere analyse van de resultaten bracht aan het licht dat het verband tussen de complexiteit van de frequentieverhouding en de drempel waarschijnlijk het gevolg is van maskering van de interferentie tussen de bijna samenvallende harmonischen door met name de dichtstbijgelegen lagere harmonische.

In de experimenten die in de hoofdstukken 2, 3 en 4 worden beschreven, werd de onderscheidingsdrempel slechts voor een beperkt frequentiegebied bepaald. Bij de kwinten en de grote tertsen, bijvoorbeeld, lagen f_1 en f_2 tot nu toe altijd tussen de "c" van het ééngestreept (C4 = 261.6 Hz) en het tweegestreept octaaf (C5). In hoofdstuk 5 gingen we na in hoeverre eerdere bevindingen ook geldig zijn voor het veel grotere frequentiebereik tussen C2 en C6. De resultaten lieten zien dat als de ontstemming uitgedrukt wordt in de zwevingsfrequentie, de onderscheidingsdrempel niet afhangt van de grondfrequenties van de tonen. Als de ontstemming wordt uitgedrukt in cents en we ons alleen richten op die condities waarin de mate van ontstemming gelijk is aan de in de belangrijkste stemmingssystemen voorkomende ontstemming, blijken de waarneembare verschillen tussen ontstemde en zuivere kwinten in toenemende mate onbelangrijk te zijn voor tonen met een grondfrequentie lager dan ongeveer C4. Voor tonen vanaf ongeveer C3 en hoger kunnen ontstemde grote tertsen onderscheiden worden van zuivere grote tertsen. Dit geldt zelfs voor de tertsen uit de Silbermann-stemming, die slechts 5.9 cents te groot zijn.

In hoofdstuk 6 bepaalden we de relatie tussen ontstemming en subjectieve zuiverheid voor kwinten en grote tertsen. Deze intervallen, die ook nu weer uit de eerder genoemde samengestelde tonen bestonden, werden geïsoleerd, d.w.z. in een niet-muzikale context, aangeboden. De beoordeling van de zuiverheid geschiedde m.b.v. een 10-puntsschaal en d.m.v. paarsgewijze vergelijkingen. Voor zowel de kwinten als de grote tertsen bleek dat de relatie tussen ontstemming en subjectieve zuiverheid door exponentiële functies kon worden beschreven. Kwinten met ontstemmingen tussen ongeveer -20 en +15 cents werden als zuiverder beoordeeld dan de in dezelfde mate ontstemde grote tertsen. Voor grotere ontstemmingen werden de grote tertsen iets zuiverder gevonden dan de kwinten. Dit was het duidelijkst bij de vergrote intervallen. Hogere zuiverheidsbeoordelingen werden verkregen wanneer de interferentie tussen op zijn minst de bijna samenvallende harmonischen onmogelijk werd gemaakt, hetgeen bereikt werd door de even harmonischen van de hoogste toon te verwijderen. Deze verhoging was het sterkst bij de grote terts. Wanneer alle hogere harmonischen uit het spectrum werden verwijderd en de intervallen dus alleen uit twee gelijktijdige sinustonen bestonden, bleken de subjectieve verschillen tussen zuivere en ontstemde intervallen veel kleiner te zijn dan voor samengestelde tonen.

De grote verschillen in zuiverheidsbeoordelingen tussen de uit samengestelde tonen bestaande zuivere en ontstemde intervallen, kunnen grotendeels aan de aanwezigheid van zwevingen of ruwheid worden toegeschreven. Aangezien de bestaande modellen voor tonale consonantie/dissonantie ook op de waarneming van zwevingen en ruwheid zijn gebaseerd, hebben we de zuiverheidsbeoordelingen vergeleken met de dissonantie die voor de in onze experimenten gebruikte intervallen wordt voorspeld. De door het model van Plomp en Levelt (1965) voorspelde dissonantie komt niet in alle condities overeen met de verkregen zuiverheidsbeoordelingen. De voorspellingen door het model van Kameoka en Kuriyagawa (1969) komen in geen enkele conditie overeen met de beoordeelde zuiverheid. We brachten naar voren dat het voorspellend vermogen van de twee consonantiemodellen vergroot zou kunnen worden wanneer enkele principes die ten grondslag liggen aan bijvoorbeeld het luidheidssommatiemodel van Zwicker en Scharf (1965), in de modellen zouden worden opgenomen.

In hoofdstuk 7 bestudeerden we de subjectieve aanvaardbaarheid van regelmatige 12-toons stemmingssystemen in tweestemmige muziekfragmenten. Deze fragmenten werden gespeeld volgens de Pythagoreïsche (ontstemming van de hoofdkwinten, A, is hier 0.0 cents), de gelijkzwevende (A=-2.0 cents), de Silbermann (A=-3.9 cents), de middentoon (A=-5.4 cents) en de Salinas stemming (A=-7.2 cents). In het eerste experiment werden ook nog twee stemmingen met A=2.0 en A=-10.0 cents opgenomen. De proefpersonen gaven voor ieder fragment op een 7-puntsschaal aan in hoeverre ze de uitvoering in een bepaalde stemming in zijn totaliteit bezien acceptabel vonden. Uit het eerste experiment bleek dat de gemiddelde aanvaardbaarheidsbeoordelingen van de fragmenten in de

Pythagoreïsche, gelijkzwevende, Silbermann en middentoonstemming vrijwel even hoog waren, terwijl de beoordelingen veel lager uitvielen voor A=2.0 en A $\leq$ -7.2 cents. Dit werd bevestigd door de resultaten die voor een aantal fragmenten d.m.v. paarsgewijze vergelijkingen werden verkregen. De zojuist beschreven invloed van de verschillende stemmingssystemen op de aanvaardbaarheid was niet afhankelijk van het tempo (snel of langzaam) waarin de fragmenten werden gespeeld. In het tweede experiment werd de waarneming van zwevingen beïnvloed door in één van de twee condities de interferentie tussen op z'n minst de bijna samenvallende harmonischen onmogelijk te maken, wat werd bereikt door uit òf de laagste òf de hoogste toon van ieder interval de even harmonischen te verwijderen. Deze reductie in interferentie resulteerde in een verhoging van de aanvaardbaarheid. De aanvaardbaarheidsbeoordelingen in de verschillende stemmingen werden in verband gebracht met de in hoofdstuk 6 verkregen zuiverheidsbeoordelingen. Uitgaande van de volledige verzameling gepresenteerde muziekfragmenten, kon de aanvaardbaarheid uit zowel het eerste als het tweede experiment nauwkeurig worden voorspeld uit de som van de subjectieve zuiverheid van de uit gelijktijdige tonen bestaande kwinten en grote tertsen. De ontstemde intervallen in de gelijkzwevende stemming blijken in hun totaliteit bezien net zo acceptabel te zijn als die in de andere door ons onderzochte stemmingen met $-5.4 \leq A \leq 0.0$ cents. Dit zou kunnen verklaren waarom de gelijkzwevende stemming, die bovendien tegemoet komt aan voor de Westerse muziek van de laatste eeuwen zo belangrijke eis van onbeperkte modulatie, in de loop van de tijd algemeen is geaccepteerd.

De vruchtbaarheid van het hier beschreven onderzoek kan worden afgeleid uit de verbanden die we in een aantal fysisch verschillende condities gevonden hebben tussen de onderscheidingsdrempel voor zuivere en ontstemde intervallen, de zuiverheidsbeoordelingen van de geïsoleerd aangeboden intervallen en de aanvaardbaarheidsbeoordelingen van de in verschillende stemmingen uitgevoerde muziekfragmenten. In hoofdstuk 8 concludeerden we dat het zinvol is dit soort onderzoek voort te zetten. In dit hoofdstuk worden tevens enkele suggesties voor toekomstig onderzoek naar de waarneming van zuivere en ontstemde intervallen in een muzikale context gegeven. Een voorbeeld van zinvol geacht onderzoek is een nadere test van het in hoofdstuk 7 voorgestelde additiviteitsmodel voor het voorspellen van de aanvaardbaarheid van stemmingssystemen in meerstemmige muziekstukjes van uiteenlopend karakter.

Curriculum Vitae

Joos Vos werd 26 september 1949 te Vlissingen geboren. Van 1961 tot 1965 bezocht hij de ULO te Kortgene en behaalde daar het ULO-b diploma. Van 1965 tot 1970 studeerde hij aan de Rijkskweekschool te Middelburg. In 1969 verkreeg hij aldaar de akte van bekwaamheid als onderwijzer, in 1970 de akte van bekwaamheid als volledig bevoegd onderwijzer. Van 1970 tot 1972 was hij werkzaam als onderwijzer in het lager onderwijs te Waddinxveen. Vanaf 1972 studeerde hij psychologie aan de Rijksuniversiteit te Leiden. Tijdens deze studie verzorgde hij tot 1976 het muziekonderwijs aan enkele lagere scholen te Waddinxveen, van 1976 tot eind 1979 was hij student-assistent bij de Vakgroep Psychologische Funktieleer R.U. te Leiden. Het student-assistentschap hield in dat hij naast een onderwijstaak participeerde in onderzoek op het gebied van auditieve informatieverwerking (Dr. G. ten Hoopen). Tijdens zijn studie liep hij stages bij de afdeling psychologie van het Instituut voor Zintuigfysiologie TNO te Soesterberg (Prof.Dr. W.A. Wagenaar) en het Instituut voor Muziekwetenschap R.U. te Utrecht (Dr. R.A. Rasch). In 1979 behaalde hij het doctoraal examen cum laude met als hoofdrichting psychologische funktieleer (Prof.Dr. G.B. Flores d'Arcais) en als bijvak muziekwetenschap (Dr. R.A. Rasch). Van 1 februari 1980 tot 1 januari 1983 was hij ZWO-medewerker bij de afdeling Audiologie van het IZF-TNO, waar hij onder supervisie van Dr. R.A. Rasch en Prof.Dr.Ir. R. Plomp onderzoek verrichtte op het gebied van de waarneming van zuivere en getemperde consonante tweeklanken. Naast deze functie kwam hij in april 1981 op basis van een halve dagtaak als wetenschappelijk medewerker bij de eerder genoemde afdeling Audiologie in dienst van het IZF-TNO. In deze laatste functie, die in januari 1983 tot een volledige dagtaak werd uitgebreid, specialiseerde hij zich op het gebied van de geluidhinder, met als neveninteresse de waarneming van complexe signalen.

STELLINGEN

1

Kleine onzuiverheden die in muziekstukken met lang aangehouden tonen nog wel waarneembaar zijn, kunnen in snellere stukken aan de aandacht ontsnappen. Grotere onzuiverheden zijn zelfs door een snel tempo niet meer gemakkelijk te versluieren.

(Hoofdstuk 2, dit proefschrift)

2

Voor uit twee gelijktijdige samengestelde tonen bestaande intervallen wordt de waarneembaarheid van niet al te grote ontstemmingen hoofdzakelijk bepaald door de interferentie tussen bijna samenvallende harmonischen.

(Hoofdstuk 3 en 4, dit proefschrift)

3

Dat in een duet voor twee fluiten ontstemde kwinten iets minder opvallen dan bij een duet voor fluit en klarinet, kan verklaard worden door de aanwezigheid van een effectievere maskeerder van de interferentie tussen bijna samenvallende harmonischen in de eerste combinatie.

(Hoofdstuk 4, dit proefschrift)

4

De Pythagoreïsche, de gelijkzwevende, de Silbermann-, en de middentoonstemming zijn voor tweestemmige muziekstukken zonder restintervallen nagenoeg gelijkwaardig en superieur aan 12-toons regelmatige stemmingen zonder kwint- en grote-tertsoptimalisatie.

(Hoofdstuk 7, dit proefschrift)

5

De wijze waarop Kameoka en Kuriyagawa (1969) de waargenomen dissonantie van samenklanken in verband brengen met het in een gehoorzaal aanwezig achtergrondgeluid, is in tegenspraak met onze kennis over de waargenomen luidheid van een geluid in aanwezigheid van gedeeltelijk maskerende bijgeluiden.

Kameoka, A., en Kuriyagawa, M. (1969). "Consonance theory, Part II: Consonance of complex tones and its calculation method." Journal of the Acoustical Society of America, 45, 1460-1469. (Zie ook Hoofdstuk 6 van dit proefschrift.)

6

Het perceptief begin van muzikale tonen kan in verband worden gebracht met het tijdstip waarop de temporele omhullende van de tonen een bepaald drempelniveau overschrijdt.

Vos, J., en Rasch, R.A. (1981). "The perceptual onset of musical tones." Perception & Psychophysics, 29, 323-335.

7

Het verschil in hinder tussen wegverkeers- en impulsgeluid met hetzelfde A-gewogen equivalente geluidniveau, is kleiner bij hoge dan bij lage geluidniveaus.

Vos, J., en Smoorenburg, G.F. (1985). "Penalty for impulse noise, derived from annoyance ratings for impulse and road-traffic sounds." Journal of the Acoustical Society of America, 77, 193-201.

8

De door CHABA 84 (Galloway, 1981) verrichte herinterpretatie van Borsky's (1965) gegevens, leidt tot een ernstige onderschatting van de hinder van supersone knallen.

Borsky, P.N. (1965). Community reactions to sonic boom in the Oklahoma City area. Volumes I and II. AMRL-TR-65-37. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratory.

Galloway, W.J. (1981). Assessment of community response to high-energy impulsive sounds. Rapport van Working Group 84, Committee on Hearing, Bioacoustics, and Biomechanics. Assembly of Behavioral and Social Sciences. National Research Council. Washington D.C.: National Academy Press.

9

De door Hede en Bullen (1982) gedefinieerde algemene subjectieve reactie op impulsgeluid:

$$GR = (G + 0{,}32A + 0{,}64D + 0{,}65CD + 0{,}42)/4{,}22$$

kan worden vereenvoudigd tot

$$GR = \tfrac{1}{2}G.$$

Hede, A.J., en Bullen, R.B. (1982). "Community reaction to noise from a suburban rifle range." Journal of Sound and Vibration, 82, 39-49.

10

In gebieden met beboste zandgronden wordt bij een positieve windgradiënt en voor afstanden van ongeveer 1 tot 10 km, het A-gewogen en met een integratietijd van 35 ms gemeten geluidniveau (A,imp) van relatief zware vuurwapens ter plaatse van de ontvanger (immissieniveau, L_i) benaderd door

$$L_i = L_e - \{20 \log(D) + 60 + D(8.6e^{-0.19D})\},$$

waarbij
L_e = het op 1 m genormeerde bron- of emissieniveau in dB(A,imp)
D = de afstand tussen emissie- en immissiepunt in km.

11

In tegenstelling tot wat Terhardt en Zick (1975) beweren, correspondeert de door hen toegepaste gestrekte stemming voor de tonen in het groot en het klein octaaf (met grondfrequenties tussen ongeveer 65 en 260 Hz) niet met de gemiddelde gestrekte stemming die door Railsback bij piano's is gevonden.

Terhardt, E., en Zick, M. (1975). "Evaluation of the tempered tone scale in normal, stretched, and contracted intonation." Acustica, 32, 268-274.

12

Op basis van de resultaten van onderzoek naar de waarneming van melodische en harmonische intonatie van tweestemmige muziekfragmenten adviseert Rasch (1985) om minder ervaren musici niet de bovenste maar de onderste stem te laten spelen. Aangezien in het experiment van Rasch de gemiddelde melodische ontstemming in de bovenste stem een factor 1,4 tot 2,4 groter was dan die in de onderste stem, mist dit advies wetenschappelijke ondersteuning.

Rasch, R.A. (1985). "Perception of melodic and harmonic intonation of two-part musical fragments." Music Perception, Vol. 2, 441-458.

13

Het gebruik van enharmonisch gelijke tonen in Franse clavecimbelcomposities uit de 17e en het begin van de 18e eeuw suggereert dat er een stemming werd gebruikt waarin de kwintenrij open is.

14

Het trouw blijven aan de principes van het senario beperkt het scenario van de componist.

15

De zes pips waarmee de NOS-radio de hele uren aangeeft, zouden aanzienlijk welluidender klinken als de in- en uitklinktijden van de zes tonen enkele milliseconden verlengd zouden worden.

Stellingen, behorende bij het proefschrift

"The perception of pure and tempered musical intervals"

Joos Vos — Leiden, 28 januari 1987