Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Berlin: Springer-Verlag.

# 5. Experiments

The experiments described in this chapter aimed to provide appropriate data by which the model (Chap. 4) could be developed, tested and fine-tuned. In the *multiplicity* experiment, listeners estimated the number of tones in musical sounds (full complex tones, octave-spaced tones, bell-like sounds and chords). The *pitch analysis* experiment measured the saliences of particular tones in musical sounds by a probe tone technique. Four further experiments investigated the apparent *similarity* of sequential pairs of tones and chords of similar loudness, and its dependence on timbral similarity, pitch distance and pitch commonality. Experimental results were used to establish optimal values for three of the four free parameters of the model. These values were subsequently used in music-theoretical applications of the model (Chap. 6).

## 5.1 General Method

### 5.1.1 Results and Modelling

Each experiment consisted of a fixed number of trials, and each experimental run (by a particular listener) produced one piece of experimental data for each trial. No results were rejected or eliminated.

The mean response to each trial in an experiment was initially calculated over all runs of the experiment, and 95% confidence intervals were calculated for each mean response. Confidence intervals for different trials normally covered about the same range, so the mean results of trials in an experiment were "significantly different" according to the standard *t test* ($p < 0.05$) if they differed by more than the half-width of a 95% confidence interval divided by $\sqrt{2}$.

Results were also calculated for particular groups of listeners in each experiment. The responses of different groups to a particular trial usually had similar deviations, so the size of 95% confidence intervals for individual groups was usually roughly proportional to the reciprocal of the square root of the number of listeners in the group. For example, the confidence interval for a group of 10 was about twice as wide as that for a group of 40. This rule of thumb allowed the significance of differences between results of different groups to be estimated from graphed results (Figs. 5.2–6, 5.8).

Out of six experiments, two (*similarity of piano tones*, *similarity of synthetic tones I*) primarily aimed to produce *qualitative* data. These results inspired the extension of Terhardt's model [Terhardt et al. 1982b] to include formulations of pitch commonality and pitch distance (Sect. 4.6). The results of the other four experiments (*multiplicity*, *pitch analysis*, *similarity of synthetic tones II*, *similarity of chords*) were compared with calculated results according to the model. This allowed the model to be tested and improved.

The correlation between calculations and results was optimized by independent adjustment of the model's four free parameters. Calculations were fitted first to mean results of all listeners, and then to mean results of each group. Each parameter was supposed to reflect how analytically listeners responded to their experimental tasks at a particular level (Sect. 4.1.2). Due to the arbitrariness of the mathematical formulations in the model, and its failure to account for important cultural effects, optimal parameter values should be interpreted with considerable caution. This is particularly true in the case of individual groups, as the mean results of individual groups are based on smaller data samples.

The calculated result for a trial was said to deviate significantly from the experimental result if it lay outside the 95% confidence interval of the mean result for that trial. Such deviations were explicable in terms of the following. Either (i) the experimental result had a cultural or arbitrary component; (ii) there was something wrong with the experimental design; or (iii) psychoacoustical theory had been wrongly applied or inappropriately interpreted. (The same applies for disagreement between calculations and music-theoretical conventions in Chap. 6.) Perhaps the most difficult challenge of this study was to design the experiments and the model in such a way that the probability of options (ii) and (iii) would be minimized.

For reasons of space, not all experimental results are presented here. For further details (including raw data) see Parncutt [1987a].

### 5.1.2 Cultural Effects

The main effect of conditioning on the results of the experiments was assumed to be familiarity with the harmonic pitch pattern described by the audible harmonics of speech vowels. Those experimental results which could be explained in terms of this auditory universal were described as "sensory". It was not generally possible to distinguish between experimental effects which were "directly sensory" and those which were "indirectly sensory", i.e. conditioned by exposure to aspects of a musical style which in turn had developed under the influence of specific sensory constraints (Sect. 3.1.1). Both "directly" and "indirectly" sensory effects may be described as "sensory in origin".

All experimental participants had had considerable exposure to Western music. The experiments were designed to minimize the effect of such cultural conditioning, in particular through the recognition of musical intervals and chords, so as to isolate sensory effects as far as possible. Techniques to reduce cultural effects involved (i) avoidance of familiar musical contexts, (ii) use of simple experimental tasks not involving musical terminology, (iii) brief presentation of a wide range of sounds in each experiment, and (iv) asking listeners to respond spontaneously. Regarding (ii), musical terminology was avoided in instructions to listeners for all experiments, and musically trained listeners were discouraged from interpreting instructions in musical terms. Regarding (iv), large numbers of trials were needed to reduce the uncertainty and variability ("noise") associated with spontaneous responses.

Effects of individual differences in musical experience were minimized by averaging results across listeners with a range of musical backgrounds. The *tone similarity* experiments (Sects. 5.4–6) were performed in 1983 at the Institute of Electroacoustics, Technical University of Munich, by Western adults with and without musical training. Musicians were mainly students at the Munich Musikhochschule; non-musicians, students and staff at the Institute of Electroacoustics. The *multiplicity*, *pitch analysis* and *chord similarity* experiments (Sects. 5.2, 3, 7) were performed in 1985 at the Department of Psychology, University of New England, Australia. Participants included not only Western adults with and without musical training, but also musically trained Eastern adults (members of a local *gamelan* orchestra) and musically trained Western children (from the local community). With the exception of the children, most participants were students or staff of the University of New England. In all experiments, listeners were designated "musicians" if they played a musical instrument regularly at the time of the experiment.

## 5.2 Multiplicity

### 5.2.1 Introduction

Complex sounds presented out of context generally have ambiguous or multiple interpretations (Sects. 2.3.4, 2.4.4). The present experiment was concerned with apparent *multiplicity*, i.e. the number of tones simultaneously noticed, in short, isolated presentations of musical sounds. Apparent multiplicity differs significantly from the actual number of tones, in a similar way that the apparent number of tones in a melody differs from the actual number [Kowal 1987].

Thurlow and Rawling [1959] investigated the accuracy with which listeners could estimate the number of tones in simultaneities comprising one, two or three pure tone components. Responses were found to correlate badly with actual numbers of components. This result is not surprising, as complex tones in speech and music are normally perceived as single entities, even though they contain several audible components (harmonics).

According to Stumpf's [1898] theory of *fusion*, "tones tend to fuse, to interpenetrate each other"; and "degree of fusion is a function of the vibration ratio of the components" [Boring 1942, pp. 360, 361]. In other words, fusion depends on the presence of strong harmonic intervals such as the octave, fifth, fourth and thirds. The above quotations were copied from a publication by DeWitt and Crowder [1987], concerning experiments on the apparent multiplicity of simultaneities of one, two or three *complex* (i.e. musical) tones. Stumpf's theory was confirmed by an analysis of reaction rimes (latencies).

The present experiment aimed to obtain data on the number of tones simultaneously noticed in a wider range of musical sounds. These data were then used to check the validity of eq. (4.12–24), and to establish a rough, general value for the free parameter $k_S$ in (4.24).

Thurlow and Rawling [1959] found that estimates of multiplicity depend on the number of allowed response categories: the more tones listeners are "invited" to hear, the more tones they report. In the present experiment, four response categories were used. This number was expected to be high enough to reflect multiplicities of different sounds relative to each other with reasonable precision, but not so high as to encourage an unusually analytical listening attitude.

### 5.2.2 Method

**Listeners.** There were 39 participants in the experiment. Ten were adults with four or more years of training in Western classical music; they were designated "Western musicians". Eight were "Eastern musicians": Indonesian *gamelan* players, who played no Western musical instruments. Nine were "Western non-musicians": adults unable to play any musical instrument, but well acquainted with the sounds of Western popular music. The remaining 12 listeners were "Western children" aged between 10 and 14 years, most of whom were receiving or had received private tuition in classical Western music.

**Equipment.** Waveform samples were calculated using a Digital PDP 11–23 computer and stored on disc as 12-bit integers. Each sound contained 2000 samples. A 12-bit digital-to-analog (D/A) converter realized the sounds at a rate of 10000 samples per second.

The signal was low-pass filtered using a General Radio Universal filter, type 1952. The filter had a 3-dB cutoff frequency of 2.5 kHz ($D_7$) and a spectral gradient of −30 dB per octave. A single Pioneer CS-T3 loudspeaker (with tweeter disconnected) played the sounds inside a 4 m × 3 m × 3 m semi-anechoic chamber. The loudspeaker was placed close to one wall at the height of the (seated) listener's head. The frequency response of the filter, amplifier, speaker and room was accounted for when modelling the results by adding the response characteristic of the filter/amplifier/speaker/room system (measured over the frequency range of the sounds using a Brüel & Kjaer sound level meter) to the spectra by which waveshapes had been calculated.

**Sounds.** The ten sounds investigated both here and in the next (pitch analysis) experiment are notated musically in Fig. 5.1. The frequencies of all pure tone components of all sounds were tuned to the standard equally tempered scale. No component lay outside the range 50–4200 Hz ($G_1$ to $C_8$).

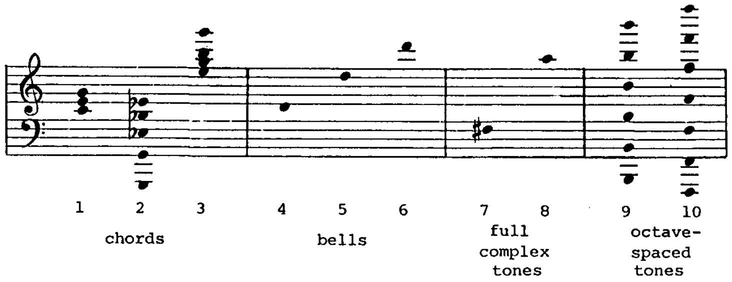


**Fig. 5.1.** Music notation of sounds presented in the multiplicity and pitch analysis experiments.

The first three sounds were musical chords. Sound number 1 was a C major triad in root position in the middle pitch register, sound 2 was an E flat major triad in first inversion in the bass region, and sound 3 was an E minor triad in root position in the treble. Each note was realized as a complex tone with all harmonics up to the tenth (or the upper cutoff at $C_8$), and a spectral envelope gradient (before amplification) of $-10$ dB per octave. The overall sound pressure levels (SPLs) of the complex tone components had been adjusted so that their separate loudnesses (in isolation) were approximately equal. SPLs of coinciding components were combined by adding intensities (4.11).

The next three sounds were bell-like in spectral content. (They didn't sound very much like bells, however, due to their uncharacteristically smooth, noiseless envelopes.) They each had 8 pure tone components at intervals of 0, 12, 15, 19, 24, 28, 31 and 36 semitones above the pitches $D_4$, $D_5$ and $D_6$ (respectively), imitating the superposed major and minor triads in the spectrum of an "ideal" bell [Lloyd 1954; Lehr 1986]. Component SPLs were equal (before amplification).

Sounds 7 and 8 were single harmonic complex tones with musical note names $F\#_3$ and $A_5$, with spectral envelope gradients (before amplification) of $-10$ dB per octave. Sound 7 contained all harmonics up to the tenth; sound 8 had only four harmonics below the upper cutoff at $C_8$. Sounds 9 and 10 were octave-spaced tones on B and F, with flat spectral envelopes (before amplification).

Sounds had equal maximum pressure amplitudes before filtering and amplification. Overall SPLs of the sounds (in order from 1 to 10) were 64, 61, 64, 65, 64, 62, 62, 70, 64 and 64 dB. Differences in loudness were small: for example, sound 8 had the highest SPL (70 dB), but was not noticeably louder than the other sounds, as it comprised only four pure tone components and so covered relatively few critical bands (Sect. 4.3.1).

Sounds had total durations of 0.2 s. The first and last 0.02 s of the sounds were shaped to prevent onset and offset clicks.

**Procedure.** In each trial, one of the ten sounds (duration: 0.2 s) was presented twice, with a pause of 0.6 s between presentations. After this, listeners could take as long as they wished to respond. They were asked, however, to respond quickly and spontaneously. The task was to answer the question: "*How many tones do you hear in the sound?*", and possible responses were 1, 2, 3 or 4. No feedback was given.

The sound presented in each trial was shifted up or down as a whole through a randomly chosen distance in the continuous range $-2$ to $+2$ semitones. This was done by adjusting the rate at which sample values were realized by the D/A converter. The size and direction of the shift was determined independently for every trial, and shifts were different in every experimental run. They were intended to confuse any listeners with latent or unrecognized perfect pitch [see Terhardt and Seewann 1983].

The experiment contained only 20 trials: two presentations of each of the ten sounds, so the probability that listeners became familiar with (and subsequently recognized) any of the sounds was low. The trials were presented in a random order which was different for each experimental run (i.e. for each listener).

### 5.2.3 Results

Results are summarized in the upper panel of Fig. 5.2. Sound 2, the low five-note chord, had the highest apparent multiplicity (between 3 and 4). Sound 8, the high single tone, had the lowest (between 1 and 2). The three chords (sounds 1 to 3) had significantly different multiplicities from each other. Among the bell-like sounds (sounds 4 to 6) and the full complex tones (sounds 7 and 8), there was a tendency for high-pitched sounds (sounds 6 and 8) to have lower multiplicity than similar sounds in middle or low ranges. Results for the octave-spaced tones (sounds 9 and 10) were not significantly different from those of sound 3 (the high chord) and sounds 4 and 5 (the low and mid-range bells).

The mean responses for each group of listeners are plotted in the lower panel of the figure. Western musicians (triangles) covered the widest range in their responses. This is typical in subjective ratings of musical sounds [cf. Roberts 1986], and reflects musicians' greater confidence in their responses. Western musicians also gave the highest overall mean result, reflecting their relatively analytical listening attitude.

### 5.2.4 Modelling

Calculations according to (4.24) were fitted directly to the mean responses for each group. Values of the free parameters were adjusted for optimal fit between calculations and results (Sect. 5.1.1). The parameters $k_M$ and $k_S$ in (4.14, 24) took fairly typical values: 29 and 0.51 respectively. The tone perception parameter $k_T$ in (4.19) was quite high (5.6 in comparison to the more usual

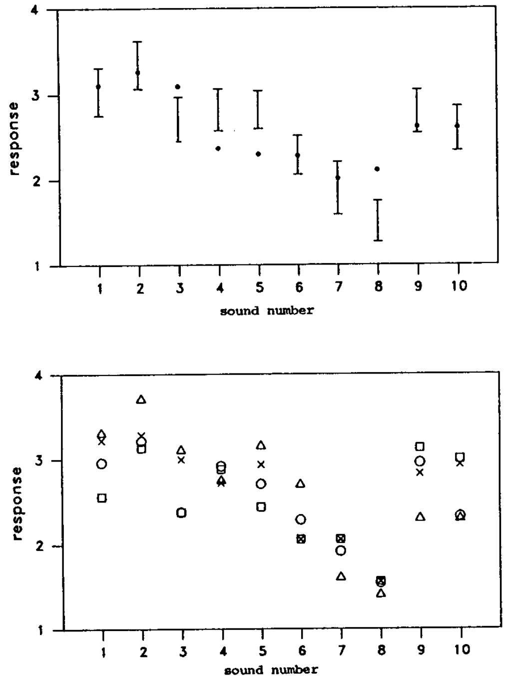


**Fig. 5.2.** Results of the sensory multiplicity experiment. *Upper panel*: 95% confidence intervals of mean responses of all 39 listeners, 2 data each (*bars*); calculated results according to (4.24) with $k_M = 29$, $k_T = 5.6$ and $k_S = 0.51$ (*points*). *Lower panel*: mean results of 10 Western musicians (*triangles*), 8 Eastern musicians (*squares*), 9 Western non-musicians (*crosses*) and 12 Western children (*circles*)

value of 3; see Sect. 5.3.4), suggesting that pure tone components were "heard out" during the experiment, perhaps because of the analytical bias of the experimental task.

Free parameters were also adjusted to fit the results of individual groups (Sect. 5.1.1). The masking parameter $k_M$ was set higher (to a value of 50) for Western musicians than for the other listeners, possibly reflecting an ability to infer the pitches of inaudible (i.e. masked) components of a familiar musical sound. The simultaneity perception parameter $k_S$ was set to slightly higher values for Western musicians (0.52) and non-musicians (0.53) than for the other listeners, in order to fit their slightly higher responses.

Calculations and results are compared for all listeners in the upper panel of Fig. 5.2. An initial observation is that the range of the calculations is smaller

than that of the responses. This may be due either to the imprecision of the model, or to the tendency of listeners to adapt their responses to fill the allowable range (in this case, 1 – 4). Calculated multiplicities tended to be too high for high-pitched sounds (sound 3, 6 and 8), apparently because of the failure of (4.17, 19) to account for spectral dominance (Sect. 6.1.2). Responses tended to be higher than calculations for the bell-like sounds (4, 5 and 6): listeners may have concluded from their relatively strange sound that relatively many components had been needed to synthesize them.

### 5.2.5 Conclusions

In agreement with Thurlow and Rawling [1959], the mean apparent number of simultaneous tones (multiplicity) of musical sounds was found to differ from the number of pure tone components. It also differed from the number of (full, harmonic) complex tone components. For example, the result for a single complex tone in register 3 (sound 7) was surprisingly high (almost 2), both experimentally and according to the model.

The formulation for multiplicity used in the model failed to account for some experimental effects (e.g. spectral dominance). The model nevertheless seems sufficiently logical and accurate for more qualitative music-theoretical applications.

The simultaneity perception parameter $k_S$ in (4.24) was estimated at 0.5. This value is used in music-theoretical applications of the model (Chap. 6).

## 5.3 Pitch Analysis

### 5.3.1 Introduction

The previous experiment was concerned with how many tones an average or typical listener notices simultaneously in a musical sound. Due to the pitch ambiguity of musical sounds (Sect. 2.4.4), different listeners can notice different tones or sets of tones in the same sound, and the same listener can notice different sets of tones when the sound appears at different times and in different contexts. Clearly, it is impossible to predict exactly which tones a particular listener will notice in a particular situation. The best one can do is predict the *probability* that a given tone in a given sound will be perceived. The present experiment aimed to estimate such probabilities (or *saliences*) as a function of pitch, for the ten sounds of the previous experiment.

Terhardt et al. [1986] investigated the pitch ambiguity of complex tones by a pitch-matching procedure: listeners were asked to adjust the frequency of a pure tone until its pitch coincided with that of a complex sound. This method has the advantage that listeners are free to choose probe pitches, so responses are not biased in favour of particular pitches.

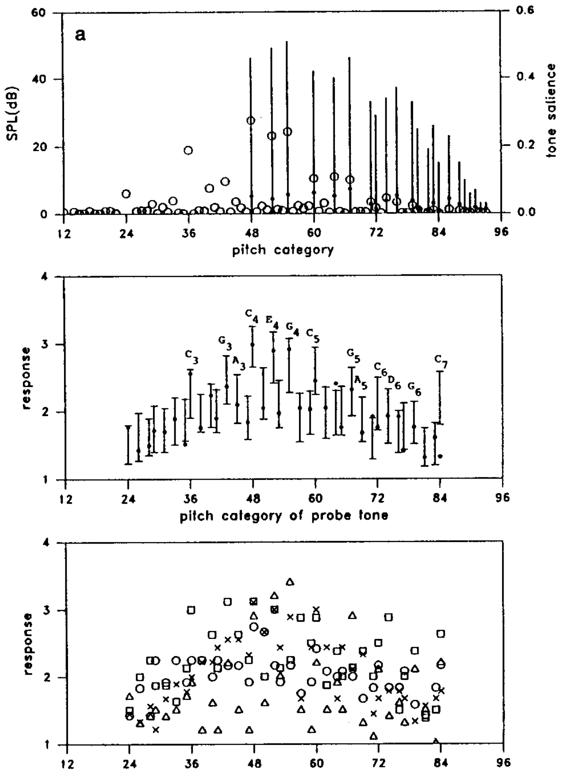


Fig. 5.3a, b, c, d. Results of the pitch analysis experiment. *Upper panels*: Pure tone components (*vertical lines, left scale*), pure tone sensations (*points, right scale*) and complex tone sensations (*circles, right scale*). *Centre panels*: 95% confidence intervals of mean responses of all 39 listeners (*bars*); calculations according to (4.35) with $k_M = 45$, $k_T = 3.5$, $k_S = 0.4$, and $k_R = 0.65$. *Lower panels*: Mean results of 10 Western musicians (*triangles*), 8 Eastern musicians (*squares*), 9 Western non-musicians (*crosses*) and 12 Western children (*circles*). (**a**) Sound 1: Chord $C_4-E_4-G_4$

The present experiment used a slightly different method. The frequencies of probe tones were set in advance, and listeners were asked whether or not each probe sounded like it was part of the sound which preceded it. This allowed salience values to be compared across different sounds.

The *pitch weights* of a complex sound according to Terhardt's model [Terhardt et al. 1982b] are relative estimates of salience: they may only be compared within sounds, not across sounds. For example, the pitch salience of a pure tone is slightly greater than that of a complex tone [Fastl and Stoll 1979]. According to Terhardt's model, however, the calculated pitch weight of a clearly audible pure tone is 0.5 (when spectral pitch weights are reduced to 50% so they become comparable with virtual pitch weights), while the calculated weight of the main pitch of a full complex tone in the pitch range of speech lies in the range 2.5 – 3. The present experiment's method of setting probe frequencies in advance enabled the development and testing of a model by which saliences are estimated as absolute values, see (4.25).

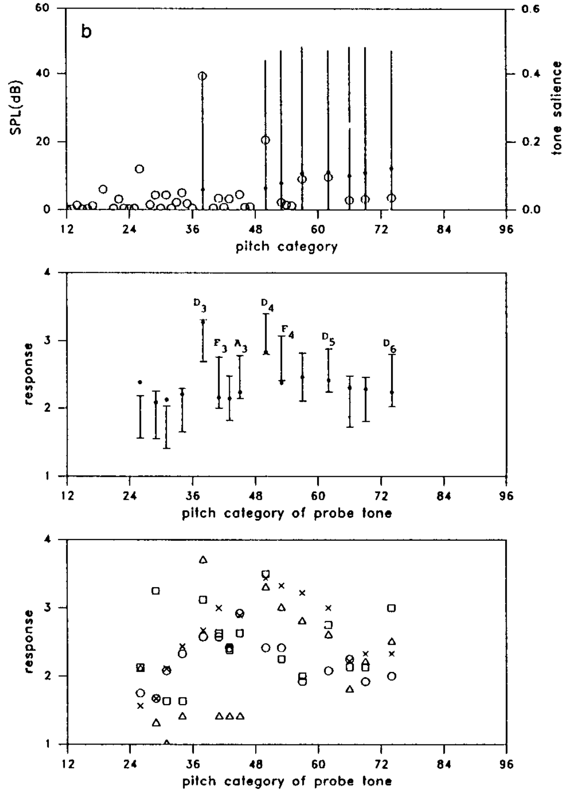


Fig. 5.3 *(continued)*. (**b**) Sound 4: Bell sound on $D_3$

### 5.3.2 Method

**Listeners** and **equipment** were the same as in the multiplicity experiment (Sect. 5.2.2).

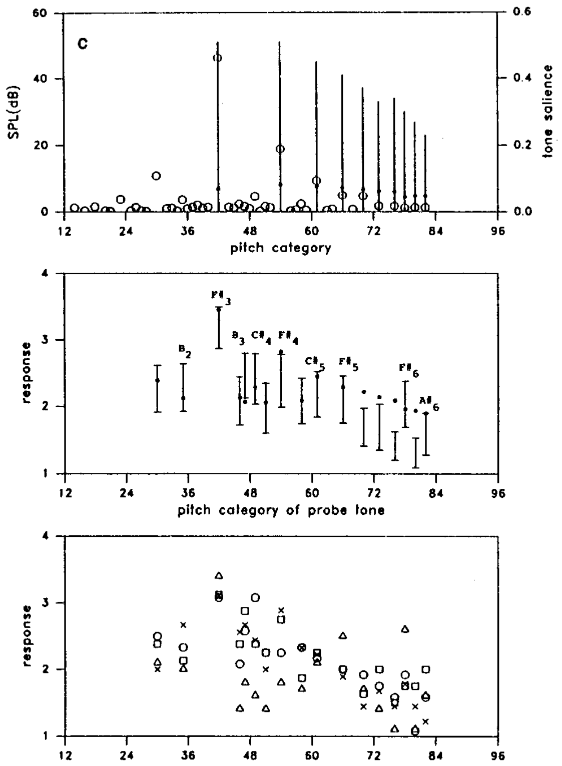


**Fig. 5.3** *(continued).* **(c)** Sound 7: Full complex tone on F$\#_3$

**Sounds.** The same ten complex sounds were investigated as in the previous experiment. The sounds had been composed such that their complex tone components were fairly evenly distributed across the central pitch range and around the chroma cycle; otherwise, undue emphasis on particular pitches or chroma during the experiment could have influenced the results via serial (context) effects.

In addition, pure *probe tones* were presented. Probe tones had frequencies in the range 65 Hz–2.1 kHz (C$_2$–C$_7$). They corresponded to tone sensations that had been predicted by an early version of the model to be evoked with saliences greater than some arbitrary minimum value. Sound 1 was additionally compared with probe tones whose pitches were not, according to the model, evoked at all (Fig. 5.3a): in all, it was compared with 36 different probe tones, representing five octaves of the C major scale. The numbers and pitches of probe tones presented in conjunction with sounds 4, 7 and 9 can be seen from Figs. 5.3b–d.

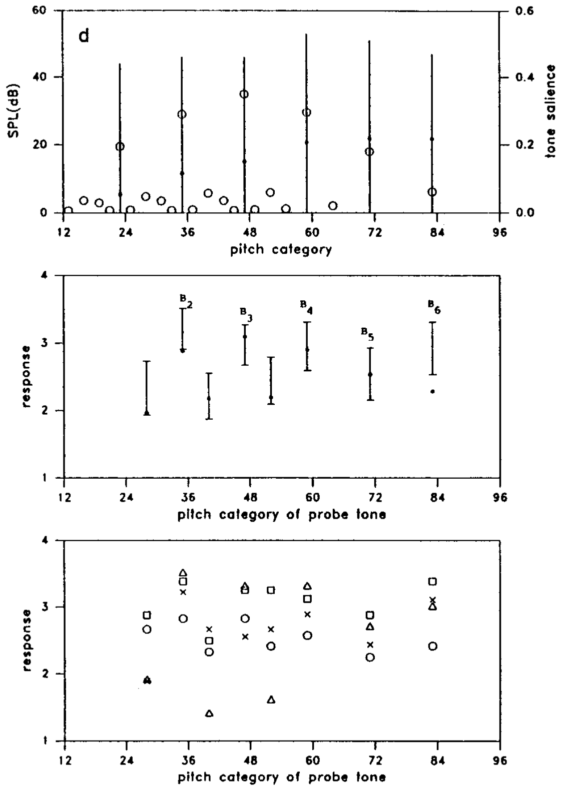


**Fig. 5.3** *(continued).* **(d)** Sound 9: Octave-spaced tone on B

The SPLs of probe tones were adjusted for approximately equal loudness. Probe tones were consistently quieter than the sounds being analyzed, but still clearly audible, in order to make it easier for listeners to imagine a probe tone to be "part of" a sound.

The frequency of each probe tone was tuned by the author, by comparing it with the complex sound with which it was to be presented. The adjustment process was performed in 0.1 semitone steps. Some probe tones had to be adjusted away from their original "equally tempered" frequencies by a semitone or even more. Pitch shifts of this magnitude are not unusual [Walliser 1969c]. In retrospect, it would have been preferable to ask a number of different musicians to tune the probe tones and average the results, as pitch shifts vary from one person to another.

**Procedure.** Each trial began with a complex sound (0.2 s), followed by a pause (0.3 s), then a probe tone (0.2 s), a longer pause (0.6 s), and the same sound-pause-probe group repeated (0.2+0.3+0.2 s). After this, listeners could take as long as they wished deciding on their responses, but were encouraged to do so quickly and spontaneously. Each sound-probe pair was shifted through a random interval in the range −2 to +2 semitones, in order to minimize the chance of perfect pitch effects influencing the results.

Listeners were asked "*Does the tone* (i.e. the second sound) *sound like it is part of the first sound?*" They responded by pressing one of four buttons, labelled (from left to right): "No", "No (not sure)", "Yes (not sure)", and "Yes". These four responses were mapped onto the numerical values 1, 2, 3 and 4 (respectively) for the purpose of recording responses and graphing data. Listeners were free to use as many or as few buttons as they wished. It was stressed that responses would be interpreted as opinions – that there were no absolutely right or wrong answers with which responses could be compared.

The experiment contained 158 different trials, presented in a unique random order for each listener. Listeners took a short break after the 80th trial.

### 5.3.3 Results

Overall results are displayed in the centre panels of Figs. 5.3 for sound 1 (C major triad), sound 4 (bell on $D_3$), sound 7 (full complex tone of $F\#_3$) and sound 9 (octave-spaced tone on B). In general, responses tended to be highest at pitches corresponding to actual (pure or complex) tone components, and to fall off with increasing pitch distance from actual tone components. This suggests that the mean response for a particular trial depended not only on the salience of the tone sensation at the pitch of the probe but also on the proximity of other tone sensations, even though pitch proximity was not implied by the question asked of the listeners. The musical note names in the centre panels of the figures indicate tone sensations in the sounds whose psychological reality was demonstrated ($p<0.05$) by comparing results for neighbouring probe tones (Sect. 5.1.1).

The mean responses for each group of listeners are plotted in the lower panels of Fig. 5.3. As in the multiplicity experiment, the responses of the Western musicians covered the greatest range, reflecting their confidence. The Western and Eastern musicians often disagreed with the other listeners (and with each other), presumably due to their musical training (i.e. cultural influences). Responses to trials testing the existence of "residue tone sensations" (complex tone sensations not corresponding to actual pure tone components) were generally lowest for the Western musicians. This was to be expected, as the musical sounds with which Western musicians are familiar usually comprise notes with audible fundamentals.

Eastern musicians (squares) gave significantly higher responses than the other groups (Sect. 5.1.1) for many of the tested residue tone sensations (e.g. $C_3$ and $G_3$ in sound 1, $F_2$ in sound 4, and $E_4$ in sound 9). This may have been due to their experience with non-harmonic sounds such as *gamelan* bells. The spectrum of a bell-like sound (Sect. 5.2.2) typically contains several components which conform fairly closely to a harmonic series of frequencies, as well as a few components which do not. For example, most of the pure tone components of the C major triad (sound 1) conformed approximately to a harmonic series on $C_3$; only the components at $E_4$ and $B_5$ did not. So it is possible that sound 1 seemed more bell-like than chord-like to the Eastern musicians.

The three musical chords (sounds 1–3) were found to evoke tone sensations both higher and lower in pitch than the actual notes of the chords. Lower-pitched tone sensations were always of the complex variety, and corresponded to possible roots of the chords in music theory. Higher-pitched tone sensations were either pure (heard-out harmonics) or complex.

The three bell-like sounds (sounds 4–6) and the two full complex tones (sounds 7 and 8) each evoked more than one tone sensation, as expected from the results of the multiplicity experiment. If only one tone had been noticed in these sounds, then its pitch would have been ambiguous. The two most salient tone sensations were generally octave equivalent.

The octave-spaced tones (sounds 9 and 10) were found to evoke tone sensations at all pitches corresponding to actual tone components. Due to the limited selection of probe tone pitches, the prediction (Sect. 6.1.5) that subfifth tone sensations (E in sound 9, B*b* in sound 10) are also evoked was not directly tested. However, the result for the Eastern musicians at $E_4$ in sound 9 suggests that at least some of the listeners noticed subfifths in octave-spaced tones.

### 5.3.4 Modelling

Results were modelled according to (4.12–35). Equation (4.35) transformed the set of 158 calculated responses in such a way that their overall mean and standard deviation were set equal to the overall mean and standard deviation of the mean experimental responses. Optimal values of the four free parameters in the model were as follows. The masking parameter $k_M$ in (4.14) was relatively high at 45 (instead of a more usual 25), suggesting either that listeners detected pure tone components that, according to Terhardt [1979a], are inaudible, or that relatively high values were needed in order to compensate for the inaccuracy of the model's masking algorithm, in comparison to Terhardt's. The value of $k_T$ in (4.19) was 3.5: this was probably slightly higher than normal, due to the analytical bias of the experimental task. The value of $k_S$ in (4.24) was a little low (0.3 as opposed to 0.5). The value of $k_R$ in (4.34) was relatively high (0.65), reflecting the emphasis on pitch commonality in the experimental task.

Parameter values were also adjusted to fit the results of each listener group. The tone perception parameter $k_T$ was highest (6) for the Western musicians (implying that they had the greatest tendency to hear out pure tone components), and lowest (2.5) for the Western children. The simultaneity perception parameter $k_S$ was also highest (1.1) for the Western musicians (reflecting their ability to notice several tone sensations at once in a musical chord), and lowest (0.1) for the Western children. The pitch relationship parameter $k_R$ was highest (0.8) for the Western musicians (i.e. they were most able to ignore pitch distance and concentrate on pitch commonality); it was lowest (0.4) for the Western non-musicians.

Calculations failed to fall within 95% confidence intervals of overall mean responses in 31 of 158 trials. If the model had adequately accounted for all relevant effects, then only about 8 calculated responses (5% of 158) would have been expected to fall outside the 95% confidence intervals. Clearly, the model did not account for all significant experimental effects.

Possible reasons why mean responses to certain trials were significantly *lower* than calculated responses are: that listeners knew from experience (especially of music) that a probe tone was not "really" in a sound, even though it "sounded like" it could be; and that the probe tone sounded out of tune with the corresponding tone sensation evoked by the sound. The above effects occurred most regularly for the Western musicians, as expected from their experience and training. Possible reasons why some mean responses were significantly *higher* than calculated responses are: that the pure tone component at the pitch of the probe tone was "heard out"; that the probe tone was heard to belong to the same category as a salient tone sensation one semitone higher or lower; that the probe tone was octave equivalent to a salient tone sensation; and that the probe tone corresponded to the lowest or most prominent pure tone component normally heard in a familiar, non-harmonic sound (e.g. a bell), but which was not necessarily present in the actual sound.

### 5.3.5 Conclusions

In agreement with Terhardt et al. [1986], typical musical sounds were found to evoke tone sensations corresponding neither to musical notes nor to pure tone components. Such tone sensations generally stand in strong harmonic relationships with actual notes and pure tone components. This finding supports Terhardt's theory that harmonic relationships in music theory have a sensory basis.

Formulas for "pitch weight" in Terhardt's model [with minor alterations, (4.17 and 4.19)] adequately reflected the relative relative audibilities of (pure and complex) tone components within complex sounds. A new addition to the model (4.25) enabled the saliences of tone sensations to be expressed as absolute values and compared across different sounds. Overall pitch distances between probe tones and experimental sounds (4.33) also influenced the results of the experiment. Agreement between the calculations of the model and mean experimental responses was satisfactory, considering that many effects that appear to have influenced the results were not accounted for at all by the model.

The results of the modelling procedure suggested a typical value of 3 for the tone perception parameter $k_T$. This value is used in music-theoretical applications of the model (Chap. 6).

## 5.4 Similarity of Piano Tones

### 5.4.1 Introduction

Terhardt [1983] suggested a sensory basis for the harmonic relationship between sequential complex tones at musical intervals such as octaves, fifths and fourths (Sect. 3.2.3). He proposed that such tones were to some extent "confusable" with each other due to the octave- and fifth-ambiguity of their pitch. If this is true then sequential tones at these intervals should sound more *similar* to each other than sequential tones at slightly different intervals.

Thurlow and Erchul [1977], in their "piano octave-similarity tests", asked listeners to rate the similarity of sequential piano tones. The difference between the similarity of sequential tones an octave apart and the similarity of tones spanning minor sevenths and major ninths was only significant for 3 out of 9 listeners, 8 of whom could recognize octaves. This result does not necessarily argue against Terhardt's proposal. According to music theory, tones a minor seventh or major ninth apart are not harmonically remote from each other: they lie only two steps apart on the cycle of fifths. So they, like tones an octave apart, may have some harmonic affinity.

In the present experiment, similarity of piano tones was compared for octaves and *major* sevenths (12 and 11 semitones), corresponding to spans of zero and five steps respectively on the cycle of fifths. The experiment therefore provided a more sensitive test of the octave-similarity of piano tones that of Thurlow and Erchul. In addition, the fifth relationship was tested, by comparing similarity ratings of fifths and tritones.

### 5.4.2 Method

**Listeners.** A total of 22 musicians and 12 non-musicians took part, see Sect. 5.1.2.

**Equipment.** The tones presented in the experiment had been played by the author on a Steinway grand piano ($A_4 = 442$ Hz) and tape recorded. The piano was about three years old and in excellent condition. No pedal was used. Tones were played as far as possible with the same loudness. Timing was controlled by means of an electronic metronome with an earplug. A condenser microphone (Peerless MBC540) and reel-to-reel tape recorder (Revox A77, 19 cm/s) were used to make a monophonic recording of the tones.

The timbre of selected tones on the tape was changed by electronical filtering. One filter (built at the Institute of Electroacoustics, Technical University of Munich) had a lowpass characteristic, cutting off very steeply (roughly 200 dB/octave) at 840 Hz ($Ab_5$). The other had a bandpass characteristic with 36 dB/octave flanks at 600 Hz ($D_5$) and 2 kHz ($B_6$). Levels were adjusted so that filtered and unfiltered tones had roughly the same loudness.

The listener sat in a sound-isolated both, and heard the tones diotically (same in each ear) over electrodynamic headphones (Beyer DT48) whose frequency response was compensated for by a free-field equalizer [Zwicker and Feldtkeller 1967].

**Sounds.** Tones on the final tape were either "unfiltered", "low-pass filtered", or "high-pass filtered" (actually, band-pass filtered, as described above). The levels of the tones were adjusted so that all were moderately quiet, but still clearly audible (approximately 60 dB SPL).

Sixty-two tone pairs were tested for similarity. Of the 62 tone pairs, 16 consisted of two unfiltered tones encompassing the intervals 1, 4, 6, 7, 11, 12, 16 and 19 semitones, of which either the upper or the lower tone was middle C; 6 consisted of two middle C's of different timbre; and the remaining 40 involved combinations of the above intervals and timbres.

**Procedure.** The tone pair in each trial was presented twice in succession. Each tone had a duration of about 0.4 s. The time interval between the tones in a pair was about 0.4 s; between pairs, 1.0 s; and between trials, 4.6 s.

During the time interval between trials, the listeners rated the *similarity* of the two tones on a 4-point scale, labelled "sehr unähnlich ... sehr ähnlich" ("very dissimilar ... very similar"). They were asked to regard similarity as a general term concerning all properties of the tones. Responses were written on a prepared form.

Each trial was presented twice, in two different orders. Altogether, 124 trials were presented, in a random order that was the same for each listener.

### 5.4.3 Results

Results for the 16 pairs of unfiltered tones are illustrated in Fig. 5.4. The responses for each pair are averaged over two presentations (in two different orders) to each listener.

Consider first the mean results of all 34 listeners. They indicate clearly that tones spanning an octave (12 semitones) sound more similar to each other than tones spanning a major seventh (11). In addition, tones spanning a perfect fifth (7 semitones) were rated more similar than tones spanning a tritone (6). The latter effect was smaller, but still significant.

For non-musicians, the result for the falling major seventh was not significantly different from the result for the falling octave (Sect. 5.1.1). On the other hand, the results for the rising major seventh and octave *were* significantly different. A possible reason for this involves the effect of context on pitch ambiguity. The first of a pair of tones in an experimental trial provides a context for the perception of the second tone. This reduces the pitch ambiguity of the second tone, i.e. it reduces the salience of its subsidiary pitches. The subsidiary pitches of the first tone in a pair should therefore contribute more to the similarity of the pair than those of the second tone. The main subsidiary pitch of a complex tone in the low and middle musical pitch ranges lies an octave above its main pitch [Terhardt et al. 1986] (Sect. 6.1.5). Therefore pitch commonality can be expected to be stronger for rising octave leaps than for falling octaves. This order effect was not accounted for in the model.

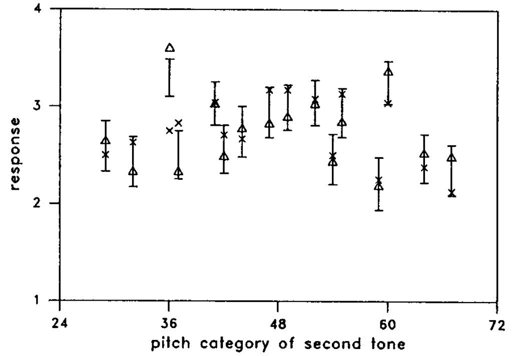


**Fig. 5.4.** Results of the similarity of piano tones experiment for unfiltered tones. The 95% confidence intervals of mean responses of all 34 listeners, 2 data each (*bars*); mean responses of 22 musicians (*triangles*) and 12 non-musicians (*crosses*). The first tone was always $C_4$ (48)

The non-musicians apparently did not recognize the octave interval, and so could not distinguish it from the major seventh (although some of them may have succeeded in recognizing octaves of they had been *required* to do so). If they did not recognize the falling octave it is unlikely that they recognized any other interval. It may be concluded that the experimental effects at the rising octave and the rising and falling fifths for non-musicians were predominantly or completely sensory in origin (i.e. due to pitch commonality).

The size of the effect at the fifth was about the same for both groups of listeners. This is consistent with the theory that the responses of the musicians to these trials were determined by the same sensory effects which determined the responses of the non-musicians. The octave effect, by contrast, was bigger for musicians. This difference may be attributed to the cultural component of octave equivalence in music (Sect. 3.3.1). Culturally conditioned octave equivalence apparently only influenced the results of musicians, because – presumably – only they could recognize octaves during the experiment.

The tones of falling intervals were rated significantly more similar than those of rising intervals, for both musicians and non-musicians. This effect may have been due to pitch (pitch distances in the bass are slightly smaller than in the treble, even for complex tones, see Sect. 2.5.2) or timbre (timbre of piano tones does not necessarily vary linearly with pitch). Alternatively, it could have been due to cultural conditioning: descending intervals significantly outnumber ascending intervals in Western popular and art melodies [Jeffries 1974].

Only the results for the 16 pairs of unfiltered tones are presented graphically, as only these are relevant to the aim of the experiment. The other tone pairs were included only to encourage listeners to attend to timbral variations, and thereby to reduce the likelihood that they would recognize intervals and respond according to musical knowledge.

### 5.4.4 Grouping

The 34 listeners were classified into groups according to the extent to which harmonic relationship, pitch proximity and timbral similarity influenced their responses. (No such objective classification has been possible in the multiplicity and pitch analysis experiments, due to lack of simple grouping criteria.) Listeners were assigned to a particular group if they rated the similarity of certain tone pairs to be, on average, at least one rating scale category higher than the similarity of certain other tone pairs. A *harmonic relationship* group was established by comparing responses for octaves and fifths with responses for major sevenths and tritones, between unfiltered tones; 10 musicians and 1 non-musician were allocated to the group. A *pitch proximity* group was established by comparing responses for twelfths and tenths with responses for fifths and thirds (again between unfiltered tones); the group contained 5 musicians and 5 non-musicians. A *timbral similarity* group was established by comparing the responses for the 16 unfiltered pairs those for the 46 filtered pairs; it contained 11 musicians (i.e. all but one) and 5 non-musicians. Of the three groups, the harmonic relationship group had the highest ratio of musicians to non-musicians, due to the particular emphasis placed on octaves by the musicians. The timbral similarity group was the largest of the groups, suggesting that timbral similarity was the most salient effect in the experiment, and supporting the claim that relatively few listeners recognized musical intervals.

### 5.4.5 Conclusions

Similarity ratings of pairs of tones of different pitch and timbre were determined mainly by similarity of timbre, and – to a lesser extent – by pitch distance and harmonic relationship. Pairs of musical tones an octave or a fifth apart were usually more similar ($p<0.05$) than pairs covering chromatically neighbouring intervals, even for non-musicians, who apparently did not recognize musical intervals at all during the experiment. This is consistent with the theory that the similarity of such tones is sensory rather than cultural in origin. On the other hand, the large difference in the similarity of tones spanning octave and major seventh intervals as perceived by musically trained listeners appeared to be mainly due to cultural conditioning.

## 5.5 Similarity of Synthetic Tones I

### 5.5.1 Introduction

Kallman [1982] compared *pure* tones in musical intervals for similarity. In general he found no effect at the octave. Similarly, Thurlow and Erschul [1977], in their "tone octave-similarity test", found that listeners either recognized octaves between pure tones and rated similarity on that basis, or else they showed no significant octave effects at all. These results contrast with the results of experiments on similarity of *complex* tones [Stoll and Parncutt 1987] (Sect. 5.4), in which significant effects at the octave were observed even in the (apparent) absence of octave recognition.

The present experiment aimed to test whether the above distinction between the similarity of pure and complex tones is general and independent of experimental method. To achieve this aim, pairs of pure tones and pairs of complex tones were presented at different (random) points in a single experimental sequence.

Previous experiments on tone similarity (including Stoll and Parncutt [1987]) had tested a restricted number of musical intervals. In the present experiment, all chromatic intervals from 0 to 13 semitones were investigated. This was intended to reduce the probability of intervals being recognized, and to produce more comprehensive results of similarity as a function of interval.

### 5.5.2 Method

**Listeners.** The aim of the experiment was to explore the sensory basis of interval relationships; results in the absence of musical interval recognition were of particular interest. Therefore, more non-musicians than musicians were tested. In all, 9 musicians and 21 non-musicians took part.

**Equipment.** Waveform samples were generated by a microprocessor system, and converted to analogue signals by a 12 bit D/A converter. They were then low-pass filtered to a 3-dB cutoff frequency of 3.7 kHz ($B\flat_7$) with a filter slope of about 200 dB/octave, using a filter built at the Institute of Electroacoustics, Technical University of Munich. The listener was seated in a sound-isolated booth equipped with a response keyboard, and heard the tones diotically (same in each ear) over headphones via a free-field equalizer.

**Sounds.** Tones were either pure (with sinusoidal waveforms) or complex (with sawtooth or rectangular waveforms). (The spectra of both sawtooth and rectangular waveforms have spectral envelope gradients of $-6$ dB per octave; sawtooth waveforms have all harmonics, while rectangular waveforms have only odd-numbered harmonics.) The bandwidth of the complex tones was limited to the lower 16 harmonics, or to the cutoff frequency of the filter (3.7 kHz). Waveforms were sampled at a rate of 18 kHz. The SPL of the pure tones was about 70 dB. Tones were adjusted for roughly equal loudness by setting the SPL of sawtooth tones 12 dB lower, and of rectangular tones 15 dB lower, than that of the pure tones.

Altogether, 60 different tone pairs were tested, of which 3 had the same pitch and the same timbre (i.e. the same waveform), 6 had the same pitch and different timbre, 26 had different pitch and similar timbre (i.e. the same waveform), and 25 had different pitch and different timbre. Fundamental frequencies were tuned to the equally tempered scale with $A_4 = 440$ Hz, and frequencies of harmonics were exact multiples of fundamental frequencies. Of the 26 trials with tones of different pitch and similar timbre, 13 incorporated pairs of pure tones (over intervals $1-13$ semitones) and 13 incorporated pairs of sawtooth tones (over the same intervals). Fundamental frequencies ranged from 131 Hz ($C_3$) to 523 Hz ($C_5$).

In the 25 trials with different pitch and timbre, one tone was always a square wave, the other was either sinusoidal or sawtooth, and the interval between the tones was chosen at random from the range $1-13$ semitones. As before (Sect. 5.4.2), these trials were included only to increase emphasis on timbre and thereby reduce the likelihood that intervals would be recognized.

**Procedure.** Each trial consisted of a repeated tone pair, i.e. of four sequential tones. Each tone had a duration of 0.2 s. The time interval between tones in a pair was 0.15 s; between pairs in a trial, 0.35 s. Listeners had as long as they wished to respond, but they were asked to do so quickly and spontaneously. They rated the similarity of the tones by pressing one of six buttons, labelled "very similar ... very dissimilar". Responses were recorded as digits in the corresponding range $1-6$.

Each experimental run contained 120 trials, in which the 60 different tone pairs were presented in two different orders. The overall order of the experiment was random, and differed for each experimental run. Each listener performed the experiment twice. In one run, pairs of tones were anchored to a particular pitch. In anchored runs, the pitch of one of the tones in a pair was always $C_4$ (for pure tones) or $F\#_3$ (for sawtooth tones). In the other run performed by each listener, the pitches of both the tones in a pair changed from one trial to the next. In the 9 trials with tones of the same pitch, that same pitch varied from $F\#_3$ to $F\#_4$, and in the 51 trials comprising tones of different pitch, the lower tone varied from $C_3$ to $B_3$. Half the listeners performed the anchored experiment first; the others started with the unanchored experiment.

### 5.5.3 Results

Results for pairs of tones with similar timbre are presented in Fig. 5.5. The plots compare results for pairs of pure (sinusoidal) and complex (sawtooth) tones. Results for pairs of different timbre are not graphed, as these were not relevant to the aims of the experiment; these pairs were only included to encourage listeners to concentrate on timbral similarity, and thereby to reduce the probability that they would recognize musical intervals.

The results for tone pairs which were identical in both pitch and timbre are denoted in the figures by "interval = 0". Almost all listeners responded "very similar" (6) to these trials. This may be regarded as a trivial case of categorical perception (the tones clearly fall in the same pitch category) and of pitch commonality [the tones, according to (4.29), have a pitch commonality of 1].

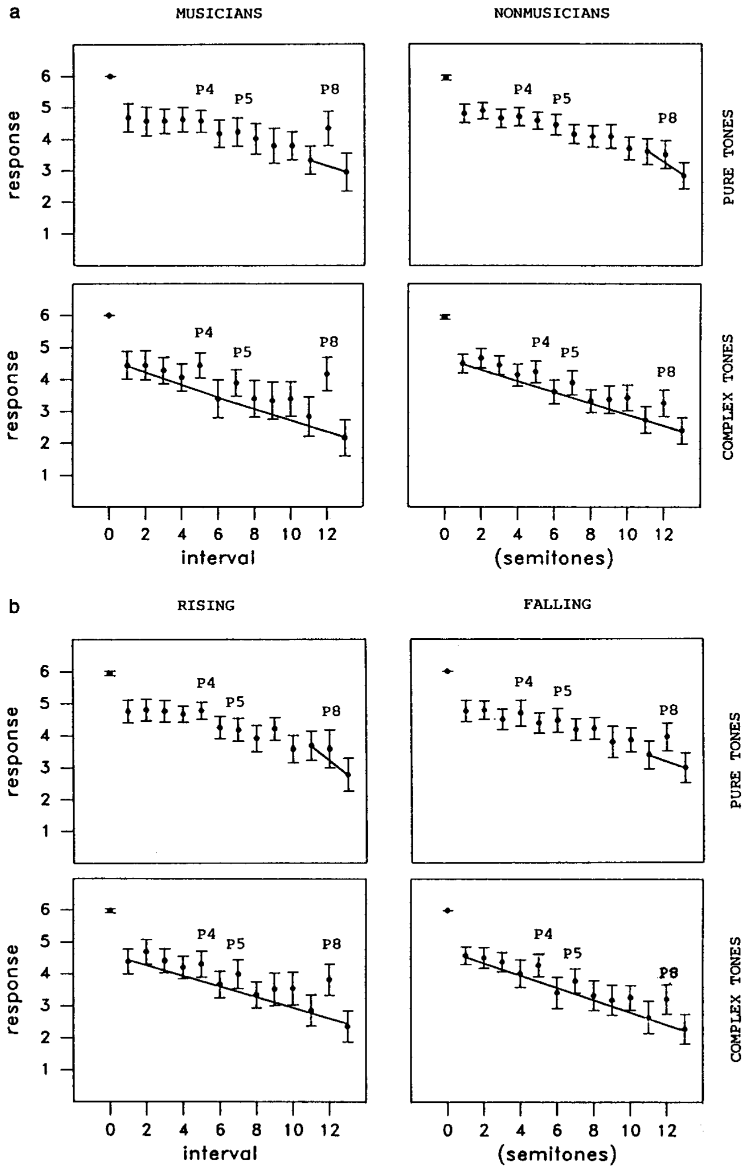
A feature of all panels of Figs. 5.5 is the effect of pitch distance (interval size). Similarity tends to fall with increasing pitch distance. Responses fall particularly suddenly as pitch distance increases from 0 to 1 semitone, i.e. as pitch proceeds from "same" to "different". The position of this category boundary is determined more precisely in the next experiment (Fig. 5.6a, Sect. 5.6.3).

In Fig. 5.5a, responses for 9 musicians are compared with those for 21 non-musicians. No significant effect occurs for non-musicians and pure tones at any interval except unison, in agreement with Kallman [1982]. This suggests that the effect of pitch commonality on similarity of pure tones (due to weak subharmonic pitches, see Sect. 2.4.5) is negligible. For musicians, on the other hand, there was a strong effect at the octave: they presumably recognized octaves, and rated them according to their "equivalence".

The responses for complex tones have more structure than those for pure tones. Consider first the responses of the non-musicians. The mean responses for intervals of 1, 6, 8, 11 and 13 semitones lie very close to a straight line. The mean responses for intervals of 2, 3, 5, 7, 10 and 12 semitones are significantly higher than the line (Sect. 5.1.1). This suggests that pitch commonality contributes to the pitch relationship between sequential tones spanning the latter intervals. The strongest effect occurs at P 8 (the octave); P 4 and P 5 follow; and M 2, m 3, and m 7 show small but significant effects. The effects at the M 3 and M 6 are insignificant, suggesting that the relationship between sequential tones an M 3 or M 6 apart in music is mainly learned (e.g. from regular exposure to major and minor triads).

The responses by musicians for complex tones are similar to those by non-musicians, except that they show an exaggerated positive emphasis on the octave and negative emphasis on the tritone, apparently due to interval recognition and cultural conditioning. The increased range of the confidence intervals for musicians is due to their smaller number (9, as against 21 non-musicians); it does not reflect any uncertainty in their responses.

Figure 5.5b compares similarity ratings of rising and falling intervals. (Note that all results include data from both musicians and non-musicians.) The octave effect for *complex* tones was significantly greater for *rising* intervals, as in the previous (piano tones) experiment (Sect. 5.4.3). Again, this may be understood in terms of the effect of context on the pitch ambiguity of a complex tone, taking into account that the main susidiary pitch of a complex tone in the central pitch range normally lies an octave above the main pitch.



The octave effect for *pure* tones was significantly greater for *falling* intervals than for rising intervals. This appears to be because the main subsidiary pitch of a pure tone lies an octave *below* rather than above its main pitch (Sect. 2.4.5).

### 5.5.4 Grouping

Listeners (9 musicians and 21 non-musicians) were classified into groups on the basis of their responses. Those whose responses to pairs of complex tones at intervals of 12 and 7 semitones exceeded responses to pairs at intervals of 11 and 6 semitones by an average of one response category or more were assigned to a *harmonic relationship* (*complex tones*) group; this group contained 5 musicians and 7 non-musicians. By an analogous method, 3 musicians and 2 non-musicians were assigned to a *harmonic relationship* (*pure tones*) group. Those 8 musicians and 14 non-musicians whose response to pairs of tones of similar timbre at intervals of 1 or 2 semitones exceeded responses for 13 or 10 semitones by an average of one response category or more were assigned to a *pitch distance* group. The 6 musicians and 19 non-musicians whose responses for pairs of the same pitch and similar timbre exceeded responses for pairs of the same pitch and different timbre by an average of one or more were assigned to a *timbral similarity* group.

The grouping analysis shows that timbral similarity and pitch proximity were generally more important than harmonic relationship, and that harmonic relationship between complex tones was generally more important than harmonic relationship between pure tones. The fact that only 2 of 21 non-musicians were assigned to the harmonic relationship (pure tones) group squares with the assumption that octaves were recognized only seldom by non-musicians.

### 5.5.5 Conclusions

Similarity ratings were mainly influenced by the sensory parameters timbral similarity and pitch distance. Harmonic effects were weaker, but still significant ($p < 0.05$) in many cases. The existence of both sensory and cultural contributions to harmonic aspects of the similarity of both pure and complex tones covering octave intervals was suggested.

The weakest harmonic effect was the sensory contribution of the similarity of pure tones an octave apart. It was only significant in the case of failing in-

**Fig. 5.5a, b.** Results of similarity of synthetic tones experiment I. The 95% confidence intervals of mean results for pairs of pure tones (*upper panels*) and complex tones (*lower panels*); diagonal lines indicate the assumed pattern of results in the absence of harmonic effects (pitch commonality). (**a**) Results of 9 musicians, 2 data each (*left panels*) and 21 non-musicians, 2 data each (*right panels*). (**b**) Results of all 30 listeners for rising tone pairs (*left panels*) and falling tone pairs (*right panels*) tervals. A possible explanation is that only the first in a sequential pair of pure tones has perceptible subharmonic pitches: the pitch of the second tone is disambiguated by the context created by the first. Similarly, the harmonic effects for complex tones on octave apart was significantly larger for rising intervals than for falling intervals, apparently because the first in a pair of complex tones has a more salient pitch at its upper octave than the second in the pair. This effect was found in the previous experiment (using piano tones) and confirmed in the present experiment (using synthetic tones). For complex tones, sensory harmonic effects (i.e. harmonic effects in the absence of musical interval recognition) were strongest at the octave, weaker at the fourth and fifth, and weaker still (but still significant) at the major second, minor third, major sixth and minor seventh.

The pattern of similarity ratings as a function of pitch distance was similar to ratings of familiarity and frequency of occurrence of melodic intervals [Jeffries 1972], and to rates of correct recognition of melodic intervals [Terhardt et al. 1986]. All these effects may be either sensory (due to pitch distance and pitch commonality) or cultural (due to the relative numbers of intervals in actual melodies [Jeffries 1974]). The cultural effect, if it is important, apparently depends on the sensory effect anyway, via music history and cultural conditioning (Sect. 3.1.1). It may therefore be described as "indirectly sensory" (Sect. 5.1.2).

## 5.6 Similarity of Synthetic Tones II

### 5.6.1 Introduction

In an experiment to see how well pure probe tones "fit in, musically" with a previously heard diatonic scale (also consisting of pure tones), Jordan [1987] obtained tone profiles with peaks corresponding to diatonic scale steps. The peaks had half-widths of about half a semitone. It may be concluded that sequential tones less than half a semitone apart in a musical context are assigned to the same pitch category. This is consistent with theory and data on the categorical perception of relative pitch [Burns and Ward 1978] (Sect. 2.5.3).

One of the aims of the present experiment was to test how categorical perception influences the results of similarity ratings. Some non-chromatic (microtonal) intervals were included, in order to investigate: how far apart pure tones must be before they are heard to be "different" rather than "the same" in an experiment of this kind; whether complex tones spanning non-chromatic intervals are less similar than complex tones spanning ordinary chromatic intervals; and whether there is a sensory basis for the similarity of complex tones at the major third (as no significant effect had been found for this interval in the previous experiment).

Previous experiments (Sects. 5.4, 5) tested the similarity only of pure and full complex tones; other tone types (square waves, residue tones) had been included only to encourage listeners to concentrate on timbral variations. In the present experiment, results were analyzed not only for pairs of ordinary pure and complex tones but also for pairs of residue and octave-spaced tones. Of interest were the following possibilities: that residue tones might show weaker harmonic effects than complex tones due to the physical absence of the fundamental; and that octave-spaced tones might show a harmonic effect at the interval of a fourth, due to subfifth tone sensations predicted by the model (Sect. 6.1.5).

Previous experiments had established in a qualitative way that pitch commonality influences similarity ratings of sequential tones. In the present experiment, this was tested quantitatively by comparing results with calculations according to a model of pitch relationship (Sect. 4.6).

### 5.6.2 Method

**Listeners.** 8 musicians and 12 non-musicians took part in the experiment (Sect. 5.1.2).

**Equipment** was the same as for the previous experiment (Sect. 5.5.2).

**Sounds.** Spectra of representative tones are shown in the upper panels of Figs. 5.6. Full complex tones had 16 harmonics with a spectral envelope gradient of $-6$ dB/octave (as in the previous experiment). The pure tone components of residue tones were confined to the range $C_4$ (middle C)$-C_7$; their amplitudes were constant in the range $C_5-C_6$, and fell linearly to zero in the sidebands $C_5-C_4$ and $C_6-C_7$. Octave-spaced tones contained octave-spaced pure tone components in the range $C_2-C_8$; their amplitudes were constant in the central range $C_4-C_6$ and fell linearly to zero in the sidebands $C_4-C_2$, and $C_6-C_8$. Waveforms were sampled at a rate of 18 kHz. Levels were adjusted for roughly equal loudness.

In all, 98 different tone pairs were presented in the experiment. No anchor pitches were used. Of the tone pairs, 22 consisted of pure tones in the range $C_4-C_5$ covering the intervals $\pm$(0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 11 and 12) semitones; 38 pairs consisted of full complex tones in the range $A_2-C_4$ covering intervals $\pm$(0, 0.5, 1, 1.5,..., 7, 7.5, 8, 11, 12) semitones; 14 pairs consisted or residue tones with (missing) fundamental frequencies in the range $C_3-C_4$ covering the intervals $\pm$(0, 1, 5, 6, 7, 11, 12) semitones; 24 pairs consisted of octave-spaced tones covering the intervals 0, 6, $\pm$(0.5, 1, 1.5,..., 5, 5.5) semitones.

Timbre (i.e. tone type) was the same within each trial. It was considered unnecessary to include additional trials with different timbre (as in previous experiments) as the experiment already contained a lot of timbral variation.

**Procedure.** Each trial consisted of a repeated tone pair. Each tone had a duration of 0.2 s. The time interval between tones in a pair was 0.2 s, and between pairs in a trial, 0.45 s. Listeners had as long as they wished to respond, but they were asked to do so quickly and spontaneously. They rated the

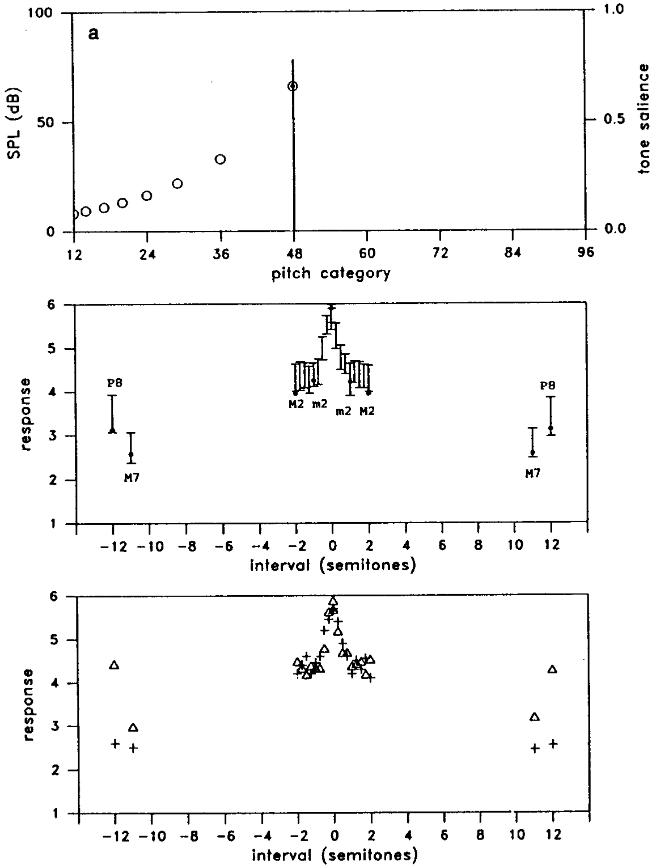


**Fig. 5.6a, b c, d.** Results of similarity of synthetic tones experiment II. *Upper panels*: Pure tone components (*vertical lines, left scale*), pure tone sensations (*points, right scale*) and complex tone sensations (*circles, right scale*). *Centre panels*: 95% confidence intervals of mean responses of all 20 listeners, 2 runs each (*bars*); calculations according to (4.37) with $k_M = 20$, $k_T = 1.0$, $k_S = 0.6$, and $k_R = 0.25$. *Lower panels*: Mean responses of 10 OE listeners, 2 runs each (*triangles*), and 10 non-OE listeners, 2 runs each (*crosses*). (**a**) Pairs of pure tones

similarity of the tones on a 6-point scale. The response keyboard was labelled "very dissimilar" (at the left) and "very similar" (at the right).

Trials were presented in a random order, which differed for each run of the experiment. Each listener performed to experiment twice.

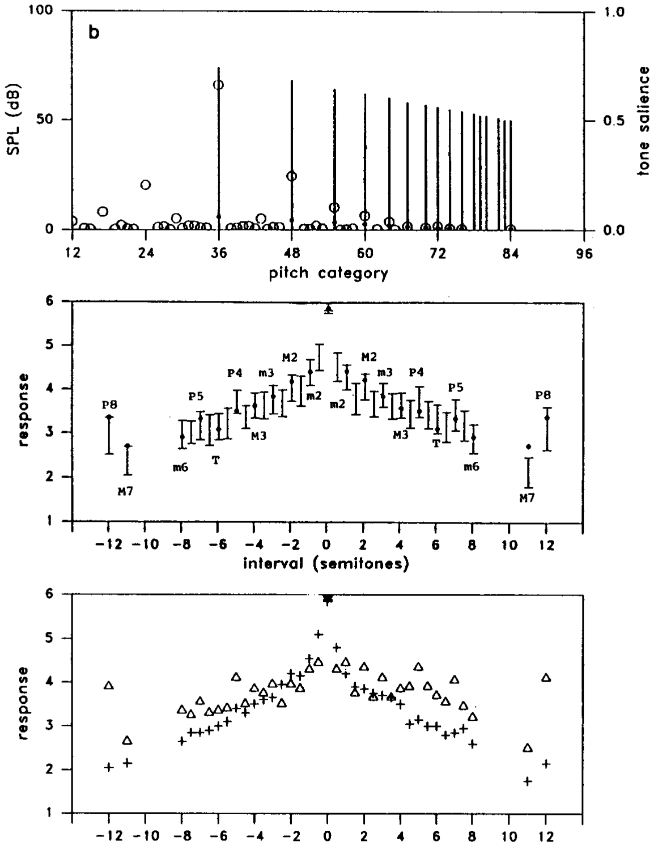


**Fig. 5.6** *(continued)*. (**b**) Pairs of full complex tones

### 5.6.3 Grouping and Results

The degree to which responses for octaves exceeded responses for major sevenths was calculated, taking into account responses for pure, full complex and residue tones. The 10 listeners who emphasized octaves to the greatest degree were assigned to an "octave equivalence" (OE) group. In the light of previous experiments, the ratio of musicians to non-musicians in this group was expected to be higher than in the other (non-OE) group. Surprisingly, these

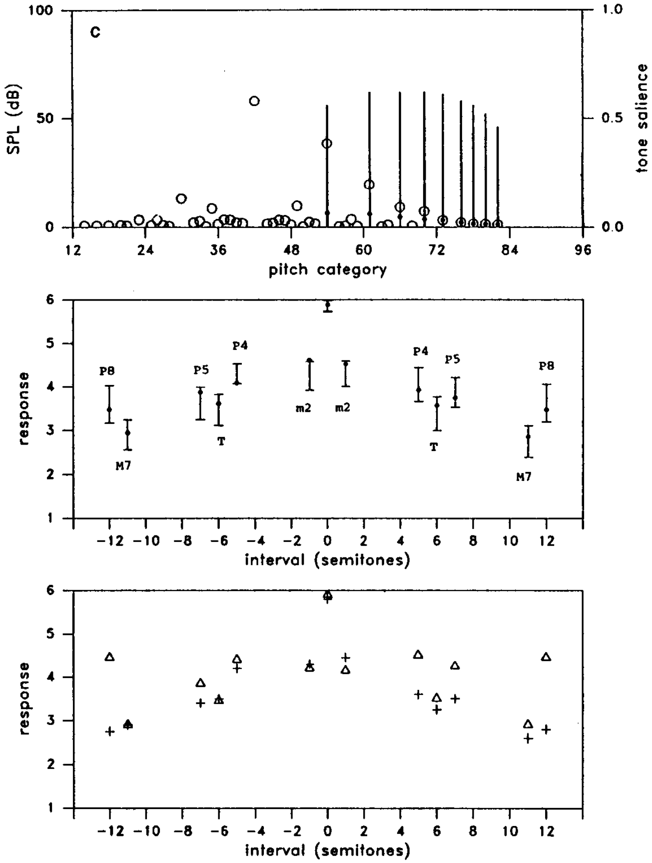


**Fig. 5.6** *(continued).* (c) Pairs of residue tones

ratios turned out to be exactly the same: each group contained 4 musicians and 6 non-musicians.

Results appear in Fig. 5.6. The lower panels illustrate differences in response strategies by a comparison of the results of the OE and non-OE groups. In general, members of the non-OE group showed no harmonic effects at all. Some possible interpretations of this are (i) that the pitch commonality hypothesis is false, and harmonic effects are merely learned from musical experience; (ii) that the effect of pitch commonality is weak and may, in some circumstances, be completely dominated by the effect of pitch distance; and

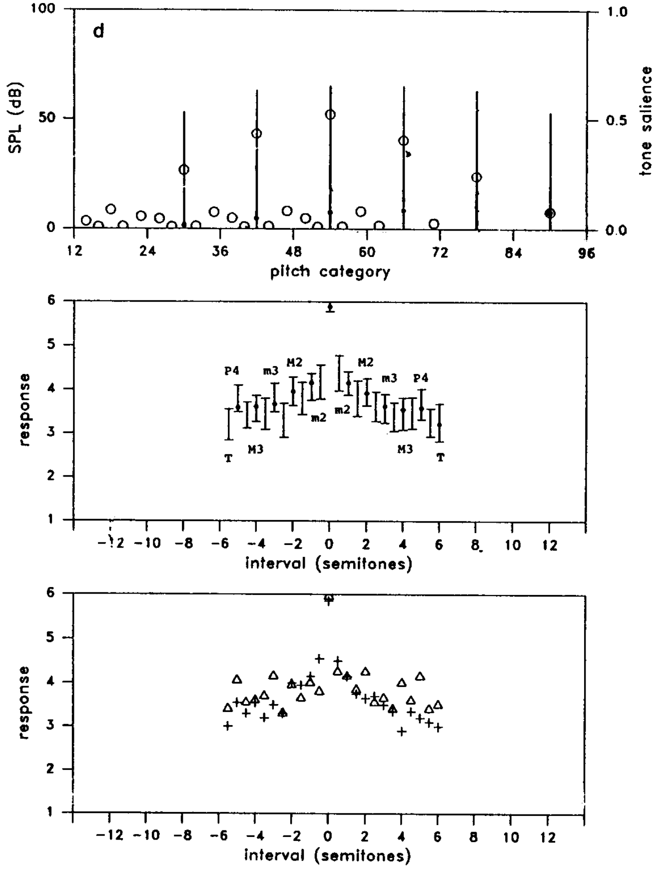


**Fig. 5.6** *(continued).* (d) Pairs of octave-spaced tones

(iii) that the lack of harmonic effects among the non-OE listeners is a direct consequence of the procedure by which the group was selected. Evidence from other experiments in this study favours the second and third possibilities.

According to Fig. 5.6a, sequential pure tones of 0.2 s duration and spanning up to about 0.5 semitones were interpreted as identical, by both OE and non-OE listeners. This may be regarded as a kind of categorical perception (Sect. 2.5.3). The category width of 0.5 semitones is about ten times greater than the local difference threshold, see Fastl and Hesse [1984]. Possible reasons for this include (i) context: listeners became used to the relatively large pitch distances in the present experiment; and (ii) task: in the present experiment listeners were asked to rate similarity, not identity.

The overall results for pairs of complex tones (Fig. 5.6b) generally confirm previous conclusions on the sensory basis of harmonic relationships between sequential tones, although in several cases insufficient trials had been carried out to establish significance. Significant effects (Sect. 5.1.1) were found for the intervals M2, P4 and P8 for both rising and falling intervals, and for the m3 (rising) and P5 (falling). There was no significant effect at the M3.

Significant harmonic effects were found for residue tones at the intervals P4, P5, and P8 (Fig. 5.6c) in spite of their missing fundamentals. This may be understood in terms of Terhardt's theory of pitch, according to which a residue tone evokes tone sensations in essentially the same pitch pattern as a full complex tone.

The results for octave-spaced tones (Fig. 5.6d) include significant effects at rising and falling perfect fourths. These are consistent with the model's prediction that octave-spaced tones evoke subfifth tone sensations (upper panel). The lack of a significant effect at the major third suggests that subthird tone sensations of octave-spaced tones are imperceptible. The effect at the falling major second may be cultural, due to the pervasiveness of this progression (about a quarter of all intervals in popular melodies are falling M2s [Jeffries 1974]). Alternatively, it may be sensory, due to pitch commonality.

The effect at the falling minor third in Fig. 5.6d is unlikely to be sensory in origin. Given that the calculated pitch properties of octave-spaced tones (upper panel) are valid, such a sensory effect could only be due to the coincidence of (weak) subfifth pitches in the first tone of the trial with (very weak) subthird pitches in the second. The effect may be related to the importance of the minor third in children's songs, at least in the West; but exactly how is not clear. Critical bandwidth (Sect. 4.3.1) equals three semitones in the most important frequency region for spectral pitch (the pitch dominance region), however, this is unlikely to be relevant, as it only affects simultaneous sounds.

### 5.6.4 Modelling

Equations (4.12–37) were used to model responses for those 58 pairs of tones which spanned chromatic intervals (i.e. whole numbers of semitones). Values of the free parameters producing optimal fit between calculations and responses were $k_M = 20$, $k_T < 1$, $k_S = 0.6$ and $k_R = 0.25$. Of these, the values for $k_M$, $k_S$ and $k_R$ are fairly typical. The low value of $k_T$ ($<1$, instead of about 3) may be due to effects of musical conditioning (e.g. interval recognition) not accounted for by the model.

The difference between the optimal value of $k_R$ for the OE group (0.42) and for the non-OE group (0.10) was a direct consequence of the way the two groups were selected. Otherwise, parameter values were much the same for the two groups. Calculations fit mean responses more closely for the non-OE group than for the OE group, suggesting that members of the OE group were influenced more than members of the non-OE group by cultural effects not accounted for in the model.

The same model simulated results for pairs of pure, full complex, residue and octave-spaced tones. Calculations agreed with mean responses over all listeners in all but 5 of the 58 modelled trials. Discrepancies were expected on the basis of statistical fluctuations in 5% of 58, or 3 trials. The success of the modelling procedure in this experiment was partly due to the relatively large spread in the results: the results would have improved, and the number of discrepancies between calculations and mean responses (presumably) increased, had there been more listeners.

### 5.6.5 Conclusions

Despite the mixture of sensory and cultural effects contributing to the results, it appears that the sensory contribution to the harmonic relationship perceived between sequential tones was lower for non-chromatic intervals than for their chromatic neighbours. Most listeners heard pure and complex tones a half a semitone or more apart to be different, implying that the pitch commonality of tones at intervals mistuned by half a semitone is lower than the pitch commonality of well-tuned intervals. (In actual music, the flow of perceptual information can be much higher than it was in the present experiment, and larger mistunings between sequential tones can go unnoticed, see Burns and Ward [1978].)

## 5.7 Similarity of Chords

### 5.7.1 Introduction

The algorithm used to model the results of the previous experiment (Sect. 4.6) quantitatively predicts the strength of harmonic and voice-leading relationships between musical chords (Sects. 3.2.3, 6.2.1,2). The present experiment aimed to test the accuracy of the model in the case of pairs of major triads in close position, by means of similarity ratings. It was assumed that similarity ratings of musical chords out of context would reflect the sensory basis of chord relationships in context.

Krumhansl et al. [1982] presented listeners with pairs of chords made up of octave-spaced tones, and asked them to rate on a 7-point scale "how well the second chord followed the first". In their multidimensional scaling solution of the results, pairs of chords in strong harmonic and melodic relationships and of similar consonance tended to be close together. Examination of the raw data (kindly supplied by Dr. Bharucha) showed the presence of an additional strong effect: dissonant chords followed by consonant chords were given high ratings (i.e. "followed well"). For example, apparent similarity was enhanced if a diminished triad resolved satisfactorily onto a major or minor triad.

The present experiment aimed to investigate the contributions only of harmonic relationship (i.e. pitch commonality) and pitch distance to pitch relationships between musical chords. The effect of resolution was avoided by holding consonance relatively constant (by using only major triads).

### 5.7.2 Method

**Listeners** and **equipment** were the same as for the multiplicity and pitch analysis experiments (Sect. 5.2.2).

**Sounds.** These are notated musically in Fig. 5.7. They include all major triads playable using the twelve tones $C_4-B_4$ inclusive. The frequency spectrum of chord 1 was the same as that of sound 1 in the multiplicity and pitch analysis experiments (upper panel of Fig. 5.3a). The complex tones from which the chords were built had roughly equal loudness when heard separately. Each chord lasted 0.2 s. Amplitude envelopes were shaped to remove onset of offset clicks.

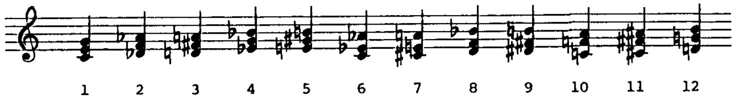



**Fig. 5.7.** Music notation of sounds presented in the similarity of chords experiment

**Procedure.** Each trial consisted of a repeated pair of chords. Each chord lasted 0.2 s. The time interval between chords was 0.3 s, and between repetitions of pairs of chords, 0.6 s. The spectra of the chords in each trial were shifted through in randomly chosen distance of −2 to +2 semitones. The shifts, which were determined separately for each trial and each run, were intended to confuse any listeners with perfect pitch.

Listeners were allowed as long as they wished to respond, but were asked to do so quickly and spontaneously. They rated each pair of chords by pressing one of four buttons labelled "very dissimilar", "sort of dissimilar", "sort of similar", and "very similar". Responses were recorded as numbers 1 – 4 (respectively).

The experiment comprised 132 trials: each of 12 chords was compared with each other chord (in two orders of presentation). The trials were presented in a random order which was different for each experimental run. Listeners took a break after the 66th trial.

### 5.7.3 Results

Mean responses for pairs of chords in which the first chord was chord 1 (the C major triad) are graphed in Fig. 5.8. In general, responses were higher for chords whose voices led well according to music theory; for chords which had a tone in common (N.B. no chords had more than one tone in common); and for chords whose roots were close on the cycle of fifths. In the figure, responses are highest when chord 1 (C) is compared with chords 6 (A*b*), 7 (A) and 10 (F). These are the only chord relationships in the set combining good voice-leading and a tone in common. Responses are lowest when the C chord is compared with chords 4 (E*b*), 5 (E) and 9 (B). Voice-leading for these progressions is weak. Voice-leading in the progression from C (1) to G (12) is also weak, but the effect of this appears to be offset by closeness on the cycle of fifths.

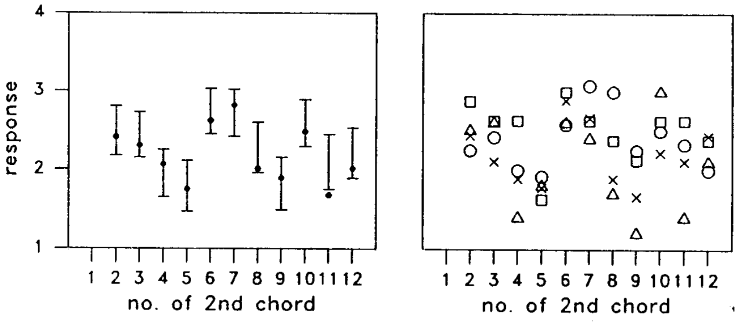


**Fig. 5.8.** Results of the similarity of chords experiment for pairs in which the first chord was chord 1. *Left panel*: The 95% confidence intervals of mean responses of all 39 listeners (*bars*); calculations according to (4.37) with $k_M = 28$, $k_T = 3.5$, $k_S = 0.3$, and $k_R = 0.3$. *Right panel*: Mean results of 10 Western musicians (*triangles*), 8 Eastern musicians (*squares*), 9 Western non-musicians (*crosses*) and 12 Western children (*circles*)

### 5.7.4 Modelling

Calculations of the model (4.37) were fitted to mean responses to all 132 trials for each group of listeners. Values of the free parameters producing optimal fit between calculations and responses were $k_M = 28$, $k_T = 3.5$, $k_S = 0.3$ and $k_R = 0.3$. Values of the parameters $k_M$, $k_T$ and $k_R$ were consistent with previous findings. The simultaneity perception parameter $k_S$ was lower than its typical value of 0.5, apparently because of the relatively holistic orientation of the experimental task.

The optimal value of the pitch relationship perception parameter $k_R$ was highest, as usual, for the Western musicians (0.55), indicating that they were influenced by pitch commonality more than were the other listeners. The Western musicians may have been responding to a large extent on the basis of harmonic relationships learned from music. The optimal value of $k_R$ was lowest (0) for the Eastern musicians, for whom pitch distance was more important than pitch commonality. Perhaps these listeners tended to hear chords as single, non-harmonic bell-like sounds, due to their familiarity with *gamelan* sounds. $k_R$ was also fairly low for the Western non-musicians and children (0.25): "non-musicians seem to be most influenced by considerations of pitch height and not by the complex hierarchical and ratio relations among pitches that a learned tonal system engenders" [Monahan et al. 1987, p. 599, in reference to Krumhansl and Kessler 1982]. Optimal values of the other free parameters ($k_M$, $k_T$ and $k_S$) varied relatively little among different groups.

Calculated responses disagreed with mean responses in 17 of 132 cases. This exceeds the expected number of about 7 (5% of 132, due to the statistical spread about 95% confidence intervals). Clearly, the model does not account for all experimental effects. Most of the discrepancies were found to be explicable in terms of familiarity with music; for example, pairs of major triads whose roots are a fifth apart are familiar (e.g. from perfect cadences) while pairs whose roots are a minor third apart are unfamiliar.

### 5.7.5 Conclusions

From a music-theoretical perspective, similarity of musical chords was explicable in terms of voice-leading, tone commonality and the cycle of fifths. The sensory basis for voice-leading was assumed to be pitch distance; for tone commonality and closeness on the cycle of fifths, pitch commonality. A model based on pitch distance and pitch commonality (in the absence of musical experience) successfully modelled the results of the experiment.

Consistent discrepancies between calculations and results were mainly due to inadequacies of the model and to musical conditioning. It was assumed in the model that the only familiar perceptual pattern influencing the perception of musical chords is the pattern of pure tone sensations in a typical complex tone (e.g. a speech vowel). The model does not account for familiarity with patterns of complex tone sensations in musical chords, or with particular chord progressions.

## 5.8 Discussion

### 5.8.1 Modelling

Values of parameters producing best fit between calculations and mean responses are summarized in Table 5.1. The table shows that parameter values varied considerably, depending on the type of sounds presented in an experiment and the task performed by the listeners. For example, the tone perception parameter $k_T$ was unusually high (5.6) for the multiplicity experiment, apparently because the manner of presentation of sounds and the experimental task encouraged the hearing out of pure tone components; and the pitch relationship perception parameter $k_R$ was especially high (0.65) for the pitch analysis experiment, as the task explicitly involved pitch commonality rather than pitch distance. Extreme values of individual parameters in the table (e.g. $k_M = 45$ for the pitch analysis experiment; $k_T < 1$ for similarity of synthetic tones II) are hard to interpret in terms of listeners' response attitudes and strategies. Perhaps these values compensated somehow for the failure of the model to account for cultural effects. The "music theory values" in the table are the values of the free parameters chosen for music-theoretical use (Chap. 6).

**Table 5.1.** Modelling results of experiments

| Experiment | $N_t$ | $N_d$ | $k_M$ | $k_T$ | $k_S$ | $k_R$ |
|---|---|---|---|---|---|---|
| Multiplicity | 10 | 78 | 29 | 5.6 | 0.51 | – |
| Pitch analysis | 158 | 39 | 45 | 3.5 | 0.3 | 0.65 |
| Similarity of tones | 58 | 40 | 20 | $<1$ | 0.6 | 0.25 |
| Similarity of chords | 132 | 39 | 28 | 3.5 | 0.3 | 0.3 |
| Music theory values | | | 25 | 3 | 0.5 | – |

$N_t$: Number of trials
$N_d$: Number of data per trial
$k_M$: Masking parameter (4.14)
$k_T$: Tone (pure/complex) perception parameter (4.19)
$k_S$: Simultaneity (multiplicity) perception parameter (4.24)
$k_R$: Relationship (commonality/proximity) perception parameter (4.34)

The use of no less than four free parameters in the model may seem somewhat extravagant: *any* experimental results may be modelled successfully if enough free parameters are introduced. However, the minimum number of data points which may be *exactly* fitted using four free parameters is only six, if two degrees of freedom are counted for the linear regression between calculations and measurements. Four free parameters may successfully fit more than six random points if limited deviations between experimental and calculated values are allowed. The allowable deviations in the experiments of this study – the 95% confidence intervals – were relatively large, but the number of points to be fit simultaneously in each experiment was also high: the three experiments modelled by adjusting four free parameters included 158, 98 and 132 data points respectively. In the multiplicity experiment (Sect. 5.2.4), three free parameters were adjusted to produce a rather bad fit to only ten data points. It would be useful in future work to repeat this experiment, using more varied sounds, in order to test more thoroughly (and possibly improve) the multiplicity algorithm (4.23, 24).

### 5.8.2 Musical Universals?

The study would have benefited greatly from the introduction of a proper control group into the experiments – a group of listeners with minimal experience of Western music. The attempt to introduce such a group (the Indonesian musicians) was largely unsuccessful, because of their small number, and because the effect of exposure to Western music on their responses was uncertain.

Would it be possible for a group of people with negligible experience of Western music to participate meaningfully in experiments of the kind described here? The experimental philosophy of a study such as this is itself culture-specific: it is biased in favour of an analytical mode of solving problems, as well as of listening to and performing music. This way of thinking and making music would make little sense in some non-Western societies. Further, instructions about Western-style experimental tasks may be difficult to translate into languages of remote cultures. Translations, where possible, would be marred by problems of interpretation, including interpretation of the nature of music itself. For any of these reasons, it may be meaningless to try to compare experimental results of remote cultural groups.

I attempted in this study to separate sensory and cultural effects in music perception by comparing experimental results with the calculations of a sensory model. Another approach might be to compare the musical styles of two originally independent cultures. Ideally, such a study would be conducted independently of the music theories of the two cultures, as music theories are normally quite ethnocentric. The researcher would need to become deeply involved in the music of both cultures, and at the same time maintain scientific objectivity; and any objective methods of analysis (including quantitative models) would need to include a balance of elements from the different cultures.

In any case, similarities in musical styles, or in experimental results of people from different cultural backgrounds, do not necessarily indicate perceptual universals. For example, octave relationships in Western music have both sensory and cultural components (Sect. 3.3.1): the similarity of tones an octave apart is exaggerated by musical conditioning (see the responses of the musicians in Fig. 5.5a). Presumably, this is also true in other musical cultures in which octaves are important. Thus, exaggerated estimates of octave similarity by people from different musical cultures reflect (non-universal) similarities in their musical cultures, not (universal) sensory effects.