Perception, 1997, volume 26, pages 645–665

1658

Spectral envelope and context effects in the tritone paradox

Bruno H Repp
Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695, USA;
e-mail: repp@haskins.yale.edu
Received 18 April 1996, in revised form 18 March 1997

Abstract. In previous studies of the 'tritone paradox' Deutsch has suggested that, when listeners are presented with pairs of octave-complex tones that are equal in average log frequency but differ in chroma by 6 semitones (a tritone), they perceive the direction of the chroma difference according to an individual pitch-class template. However, it has also been found that the perceived direction changes for many listeners when the spectral envelope of the tones is shifted along the frequency axis. Reanalysis of these data indicates a strong tendency to perceive the pitch class corresponding to the frequency on which the spectral envelope is centered as subjectively lowest. In experiment 1 this spectral-envelope effect was replicated with tone pairs presented in isolation, at the rate of one a day, which rules out artifacts of test format. In experiment 2, involving another context-free format, envelope center frequency was varied over a wide range and it was shown that some individuals are totally envelope dependent, whereas others rely more on pitch class, and yet others show mixed patterns. Experiment 3 demonstrated that listeners' judgments of tritone pairs can be swayed easily by preceding context. Finally, experiment 4 showed that strong envelope effects are also obtained with Deutsch's own tritone test (issued on CD). The subjective relative pitch height of octave-complex tones thus depends on several competing factors, only one of which is pitch class.

1 Introduction

The 'tritone paradox' is now widely known, thanks to a large number of publications by Diana Deutsch and her associates (Deutsch 1986, 1987, 1991, 1994a; Deutsch et al 1987, 1990; Ragozzine and Deutsch 1994). In a typical tritone test, listeners are presented with pairs of octave-complex tones (Shepard 1964) drawn from a set representing the twelve musical pitch classes in semitone (st) steps (C, C#, ..., B). Each individual tone is composed of a number of octave-spaced partials whose amplitudes are governed by a fixed, bell-shaped spectral envelope over the log-frequency axis. Two such envelopes, each defining a different set of twelve tones, are shown in figure 1. Since all tones in an envelope set have the same average log frequency, corresponding to the subjective dimension of pitch height, it seems that relative pitch judgments can be made only on

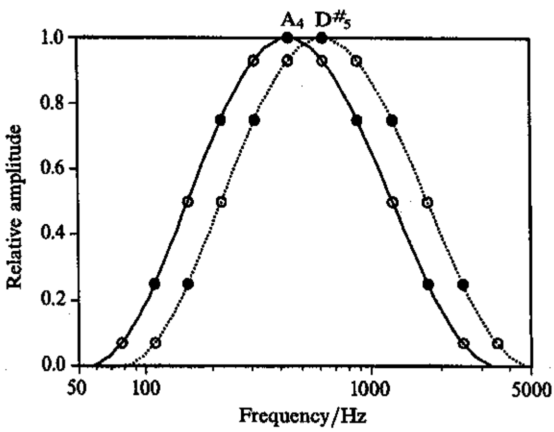

Figure 1. Two spectral envelopes, centered on A_4 (440 Hz) and $D\#_5$ (622 Hz), respectively, as used in Repp (1994). For each envelope, filled circles represent the amplitudes of the octave-spaced partials of the tone representing the pitch class of the envelope center, whereas open circles represent the amplitudes of the partials of the tone whose pitch class is 6 st removed from the envelope center.

the basis of proximity in the pitch class (chroma) circle (Shepard 1964). Tones separated by a tritone (half-octave, 6-st) interval, which are on opposite sides of the pitch-class circle, are maximally ambiguous as to the direction of their pitch difference. Nevertheless, when the twelve possible tritone pairs (C–F#, C#–G, ..., B–F) are presented in a random sequence, most listeners report hearing a clearly rising or falling pitch in any individual pair (Deutsch 1986). A test containing several randomizations typically yields an orderly response function (percentage "down" judgments as a function of the pitch class of the initial tone in a pair) indicating that some pitch classes are perceived as subjectively higher than others. This result has received much attention because it suggests that even ordinary listeners, who cannot identify musical pitches by name, have an internal pitch-class template that specifies that some pitch classes are 'higher' than others. However, there are large individual differences in the subjectively highest pitch class, and this is part of the tritone paradox (Deutsch et al 1990).

In addition to pitch class, the spectral envelopes of the tones could also influence listeners' perception. Although the theoretical envelope for a whole set of tones is a continuous function, individual tones have discrete spectral envelopes (ie the amplitudes of their octave-spaced partials) whose shape differs from tone to tone. Figure 1 illustrates that, in any envelope set, there are two special tones: one whose central partial coincides with the envelope peak, so that it represents the pitch class on which the envelope is centered (filled circles), and another whose two central partials straddle the envelope peak symmetrically, so that it represents the pitch class 6 st removed from the envelope center (open circles). The former has the most peaked and narrowest spectral envelope in the set, while the latter has the flattest and widest envelope. Under an envelope centered on A_4 (440 Hz), the tone with the narrowest spectrum represents pitch class A and the tone with the widest spectrum represents pitch class D#, whereas under an envelope centered on $D\#_5$ (622 Hz), it is just the other way around. However, it is not immediately clear why these envelope differences should affect subjective pitch height, and what the direction of such an effect might be.

To investigate this issue, Deutsch (1987) employed stimulus sets with different spectral envelope centers, ranging from C_3 (131 Hz) to A_5 (880 Hz) in 3-st steps. None of four selected listeners showed any large response shifts as a function of envelope.(1) Essentially, their response functions depended on pitch class alone, and this invariant pitch-class effect was considered an important defining property of the tritone paradox.(2) Nevertheless, Deutsch routinely used several stimulus sets with different spectral envelope centers in all her subsequent studies. The envelopes were usually spaced 6 st apart, and listeners' responses were combined across these sets to obtain a single average response function. The intention was to average out any potential spectral-envelope effects, leaving only effects of pitch class. The magnitude of envelope effects was no longer reported, except in one paper where the data of a single listener ("a particularly clear case") who was completely immune to 6-st shifts in envelope center (Deutsch et al 1987, figure 7) were displayed. Deutsch thus has suggested that envelope effects are generally negligible and that pitch class is the overriding determinant of listeners' responses in the tritone paradox.

However, this is not what I found (Repp 1994) when I conducted a study with two sets of tones whose spectral envelopes were centered on A_4 and $D\#_5$ respectively (as shown in figure 1). Of forty-three listeners tested, only fourteen showed similar

(1) These listeners had been "selected for showing a clear influence of pitch class on perceived pitch height in a brief preliminary experiment employing tones generated under a small subset of the envelopes employed" (Deutsch 1987, page 566).

(2) I (Repp 1994) used the term 'pitch-class effect' to designate simply an orderly response function for a given envelope set. However, the term is better reserved for a pitch-class-based response tendency that is independent of envelope center frequency, and it will be used in this fashion here.

(pitch-class-based) response functions in the two tests, whereas twelve effectively reversed their responses, and the rest showed intermediate or unclear patterns. I tested three groups of listeners from different countries, about which more will be said later in this paper. Figure 2 shows the average response functions of the American listeners ($n = 17$) who were most strongly affected by the spectral envelope.(3) It is plain that the two response functions are very nearly inverses of each other. Moreover, the functions suggest that the pitch class on which the envelope is centered is perceived as subjectively low, whereas the pitch class 6 st removed is perceived as the subjectively highest.

When the two functions in figure 2 are averaged, a very nearly flat function results which represents the average pitch-class effect. However, this negligible average effect does not imply the absence of individual pitch-class effects; rather, it may be due to large individual differences in pitch-class templates. What the data do imply is that pitch class was not the only and perhaps not even the primary determinant of most listeners' responses, and that individual differences were much smaller with regard to the envelope effect, or else the average trends would not have been so pronounced. We will return to these data later in this paper.

In a commentary on that study, Deutsch (1994b) pointed to some methodological features that could have provoked or enhanced spectral-envelope effects. They include relatively short intervals between tritone pairs, blocked presentation of all stimuli in each envelope set, and the fact that each block was preceded by an ascending scale formed from the tones in that set. Although none of these explanations seems very plausible [and I have already presented counterarguments (Repp 1994)], I also reported some informal evidence for strong effects of sequential context in the tritone test, which suggests that presentation format may play a role. Surprisingly, context effects have not been mentioned in any previous studies on the tritone paradox (but see Shepard 1983), though they are hardly surprising in view of the theoretical ambiguity of all tritone pairs. Context may account for the observation (Shepard 1964; Repp 1994) that, for many listeners, tritone pairs are rarely subjectively ambiguous: perhaps they are disambiguated by the preceding pair of tones. When a tritone pair is preceded by a pair that is perceived as rising (for whatever reason), then it will be more likely to be perceived as rising when it is similar to the precursor pair than when it is dissimilar (Repp 1994). Such context effects may propagate through a random sequence of test stimuli and may constrain the resulting response functions. But could they have produced or enhanced the envelope effects in my study?

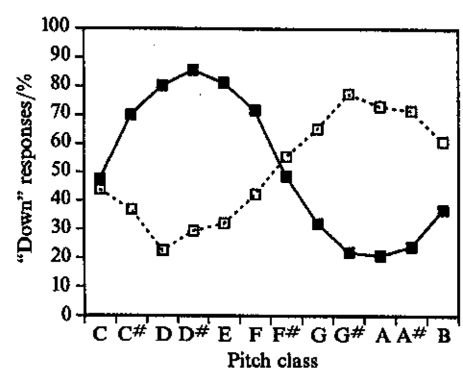

Figure 2. Average response functions (percentage "down" responses as a function of initial pitch class of tritone pair) of seventeen American listeners for two envelope sets (data from Repp 1994).

(3) These average response functions are shown here for the first time; in Repp (1994) only individual data were shown.

The present study comprises four experiments in which effects of spectral envelope and sequential context in the tritone paradox are examined. In experiment 1 context is done away with altogether, with individual tritone pairs presented on separate days. If the spectral-envelope effects obtained previously (Repp 1994) are replicated, then context effects or other details of test format cannot be held responsible. In experiment 2 spectral-envelope position is varied over a wide range, again within a paradigm that minimizes context effects, in an attempt to replicate the results of Deutsch's (1987) similar study, in which envelope effects had been minimal. Experiment 3 provides a direct demonstration of context effects, which were merely assumed to exist in the earlier experiments. In experiment 4 effects of pitch class and spectral envelope are examined in a larger group of listeners who are presented both with Deutsch's own tritone test (recently issued on CD) and with my original stimuli. At the end of the paper, we will have another look at my data (Repp 1994) in the light of the foregoing results.

2 Experiment 1

In this experiment, contextual effects were eliminated by presenting listeners with a single tritone pair per day. It seems unlikely that the pitch classes of the two tones could be remembered accurately from one day to the next. Under optimal conditions, auditory memory for pitch seems to last about three minutes (Rakowski and Morawska-Büngeler 1987; Rakowski 1993). Although there is evidence that ordinary listeners—and, of course, musicians with perfect pitch—have some long-term pitch memory (Terhardt and Seewann 1983; Halpern 1989; Levitin 1994), this is not the sensory memory that presumably is responsible for context effects.

The main question of interest was whether the large envelope effect found in my earlier study (Repp 1994, figure 2) would be replicated. Again, two envelope sets were used, spaced 6 st apart, but the envelope centers differed from those used earlier, so that effects of envelope shift could be assessed not only within experiment 1 but also relative to the earlier data. Moreover, some of the earlier listeners were still available, which made it possible to examine individual consistency across different presentation formats and a temporal interval of about one year. Actually, it was not certain that orderly individual response functions would be obtained at all in this context-free format, because only a single judgment of each tritone pair was obtained from each listener. However, neither effects of pitch class [which, according to Deutsch (1986), are based on a mental pitch-class template] nor effects of spectral envelope (which presumably are psychoacoustic in origin) presuppose any context, so there was no particular reason to expect any anomalies.

2.1 *Methods*

2.1.1 *Stimuli.* The stimuli were derived from the ones I employed previously (Repp 1994).[4] The spectral envelopes of the original tones (shown in figure 1) were centered on A_4 (440 Hz) and $D\#_5$ (622 Hz), respectively, and followed Deutsch's specifications (see, eg, Deutsch et al 1987).[5] The twelve tones in each set were composed of six octave-spaced partials and were 500 ms in duration, with 10-ms amplitude ramps at beginning and end. The tones in a tritone pair followed immediately upon each other. The twenty-four original tritone pairs (sampled at 48 kHz) were transferred from digital audio tape to a Macintosh Quadra 700 computer. For playback, however, a sampling rate of 44.1 kHz was specified. This amounted to a decrease in playback

[4] This is true for the following experiments as well. The original tones had been synthesized during a visit to the Institute for Perception Research (IPO) in Eindhoven (see Repp 1994).

[5] They look different from the envelopes displayed in Deutsch's papers because she displays amplitude in dB, whereas figure 1 shows linear amplitude.

speed which lowered all spectral frequencies by 8.1%, without any distortion, and increased tone duration by 8.8%, to 544 ms. The change in duration was considered inconsequential. The new spectral envelopes had the same log-frequency shape as the original envelopes but were centered on 404 Hz (between G_4 and $G\#_4$) and 571 Hz (between $C\#_5$ and D_5), respectively. These new sets will be called the G_4+ and $C\#_5$+ sets, respectively. All tritone pairs were stored in computer files for individual access and binaural playback via a Realistic SA-155 amplifier and Sennheiser HD 540 II earphones.

2.1.2 *Listeners.* Fifteen members of the research staff at Haskins Laboratories served as unpaid volunteers. They comprised seven women and eight men, ranging in age from 19 to 66 years. Seven of them had participated in the experiment about one year earlier (Repp 1994). (They included me; I had served as a pilot subject only.) The listeners had little or no musical training, with the exception of me (an amateur pianist) and my research assistant (a violinist with perfect pitch). The majority were American born, though they also included two Chinese graduate students, one Austrian, I, and one son of Russian immigrants.(6) Eight participants listened to both envelope sets (on twenty-four different days), first to one and then to the other, with their order roughly counterbalanced across listeners. Seven participants, owing to their lesser availability, listened to only one of the two sets (on twelve different days). The order of the tritone pairs within a set was random and different for each listener.

2.1.3 *Procedure.* Participants came to listen to a single tritone pair whenever this was convenient. The pair was played three times in succession before a judgment was made, and additional repetitions were provided if the listener had difficulty reaching a decision. I judged the listener's degree of uncertainty according to the following objective criteria and scored the responses accordingly: 1 = confident "down" (prompt response after three repetitions), 0.8 = "down" (some uncertainty expressed—"I guess", "probably"—or additional repetitions requested), 0.6 = uncertain "down" (long deliberation, great uncertainty expressed, or reversal of initial judgment after repeated listening), 0.5 = undecided, 0.4 = uncertain "up", 0.2 = "up", 0 = confident "up".

2.2 *Results and discussion*

2.2.1 *Average results.* Figure 3 shows the average response functions for the G_4+ and $C\#_5$+ stimulus sets (eleven and twelve listeners, respectively, with eight listeners in common).(7) Although the data were predictably more noisy than those shown in

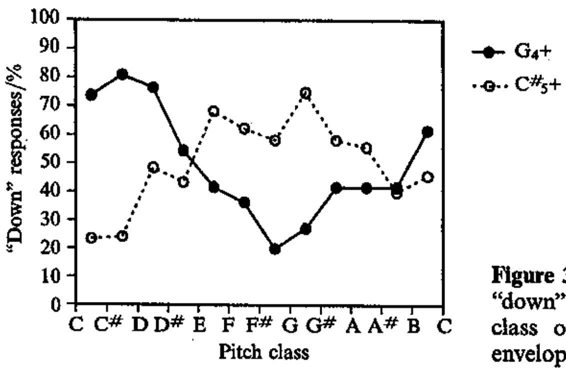

Figure 3. Average response functions (percentage "down" responses as a function of initial pitch class of tritone pair) in experiment 1 for two envelope sets.

(6) This information is provided in view of Deutsch's (1991) hypothesis of a connection between the tritone paradox and language experience. However, the present experiments did not specifically address this issue.

(7) The results are quite similar if my confidence ratings are ignored and the responses are scored categorically.

figure 2 because they were based on a much smaller number of responses, the similarity of the results is obvious. Again, the average response functions are roughly inverses of each other, and their peaks are very nearly opposite to (ie, 6 st removed from) the envelope center. With respect to the functions in figure 2, those in figure 3 are shifted to the left (counterclockwise in the pitch-class circle) by about 1.5 st—the difference in envelope centers due to the change in playback speed.[8]

A direct comparison between the old and new results is presented in figure 4, which also defines the envelope effect more precisely. Here the data of figures 2 and 3 have been reduced further by averaging each pair of response functions after aligning them according to their envelope centers (ie, shifting them by 6 st relative to each other). The dimension on the abscissa is now clockwise distance from the envelope center, and the average response functions of the two studies are clearly similar and have their peaks at a distance of 6 st from the envelope center. Below these functions, a further reduction of the data is presented: each response percentage has been subtracted from the 'opposite' percentage on the pitch class circle (6 st removed), to take into account both the maxima and the minima in "down" responses. The result is a function consisting of six positive and six negative difference scores (reversed symmetric around the abscissa), of which only the positive scores are shown. The central two data points (arrow in figure 4) define the subjectively highest point(s) on the pitch-class circle, and their average difference score is a measure of the magnitude of the envelope effect (equivalent to the distance between the maxima and the minima of the average response functions shown in figure 4).[9] Figure 4 thus indicates that, in both experiments, the subjectively highest pitch classes were 5–6 st removed from the envelope center, and the magnitude of each envelope effect was in the vicinity of 50% (100% being the maximum possible). The very same method may be used to determine the subjectively highest pitch class and the magnitude of the pitch-class effect from the simple average of the original (unshifted) response functions. However, it is clear from figure 3 that there was hardly any average pitch-class effect.

In summary, these average data show that envelope effects are just as strong for context-free stimuli as for tritone pairs in a conventional randomized test. Deutsch's

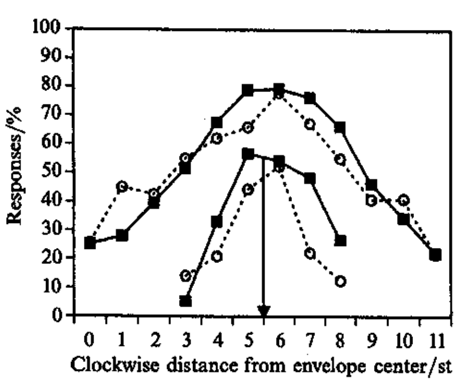

Figure 4. Upper functions: average percentages of "down" responses as a function of clockwise distance from the envelope center [squares, American listeners (from Repp 1994); circles, experiment 1]. Lower functions: difference scores derived from the functions above, illustrating determination of subjectively highest pitch class (relative to envelope center) and envelope effect size (see text).

[8] Note that the abscissa labels are aligned with ticks in figure 3 but with spaces between ticks in figure 2.

[9] Alternatively, the single point with the highest score could be determined, or some more complex weighting procedure could be used to take into account asymmetries. The present procedure was preferred because it resembles that used in earlier studies. Occasionally, it may underestimate the size of an effect. Because it is usually based on two data points, the estimated distance from envelope center will usually be halfway between integral semitones (5.5 st, or 5–6 st, in figure 4). Sometimes there is only a single central data point in the reduced response function (ie when there are only five positive difference scores), and in that case there will be an integral estimate.

(1994b) suggestion that details of presentation format in my earlier study (Repp 1994) may have been responsible for the large envelope effects can now be dismissed.

2.2.2 *Individual results.* As already observed (Repp 1994), listeners in this task tend to be of two types, analytic and synthetic (cf Terhardt et al 1982; Houtsma and Fleuren 1991), or rather they fall on a continuum extending between these extremes. When presented with a tritone pair, synthetic listeners hear a clear pitch change in one or the other direction and are usually very confident about their response. Analytic listeners hear two or more components in each tone which may seem to change in opposite directions; thus these listeners are often uncertain about their response and may consider their decision arbitrary. To judge from their behavior and comments, most listeners in the present study seemed to be of the synthetic type; only two complained about hearing several simultaneous tones and had great difficulty making up their minds. A few others were mildly analytic; although they could hear several component tones, they said they found one of them more salient than the others. The fact that many (synthetic) listeners were able to make rapid and confident decisions indicates that isolated tritone pairs are no more ambiguous than pairs presented in context. This supports a role of stable listener-internal criteria in making relative-pitch judgments, be they based on pitch class or on envelope shape.

Some findings, however, suggest that these criteria are not always sufficient to yield orderly response functions for isolated stimuli. Of course, the individual response functions must be regarded with some caution, as their reliability is unknown. I took it upon myself to listen to all twenty-four tritone pairs for a second time (on twenty-four additional days), and changed my judgment only once. I and six other listeners, however, had participated in the earlier study (Repp 1994), and these earlier results could be inspected for comparison. With some variability allowed for and the expected 1.5-st shift in response functions taken into account, there was good agreement for five listeners. A sixth listener had shown very similar response functions for the A_4 and $D\#_5$ envelope sets in the earlier study, indicating a strong pitch-class effect. However, his responses to the context-free stimuli showed a large envelope effect instead; his responses to the A_4 set (now the G_4+ set) had effectively been reversed.

The other listener who showed a discrepancy was I myself. As a pilot subject in the earlier study, I had shown rather flat response functions and strong influences of preceding context (see Repp 1994, appendix). Thus I showed neither a strong pitch-class effect nor a strong envelope effect, but I did give approximately equal percentages of "up" and "down" judgments. In the present experiment, however, I heard all tritone pairs in the $C\#_5$+ set and all but two in the G_4+ set as falling in pitch.[10] This kind of perceptual bias was not unique. Four other listeners, including the only one with perfect pitch, heard all pairs in a set as either rising or falling. (None of them had been in the earlier study.) This is hardly ever observed in standard randomized tests, though it may occur in a single block of such a test. Sequential context may help establish the balanced distribution of "up" and "down" responses that is typically found. Responses of a single kind suggest an ineffective mental pitch-class template as well as an insensitivity to envelope differences. But what, then, tells the listener's brain that the pitch is clearly going up or down?

[10] When, immediately after listening to a tritone pair (eg C−F#), I listened to the same tones in the opposite order (F#−C) I certainly heard the pitch going up. When the very same pair (F#−C) was heard first on another day, however, the pitch clearly went down. The effect is truly perceptual, not a simple response bias. For a synthetic listener such as myself, it is difficult to reverse at will the perceived direction of the pitch change in a given tritone pair, as there is no subjective ambiguity at all.

Another peculiar finding was that individual response functions for the C$\#_5$+ set often showed irregularities, whereas those for the G_4+ set did not. For example, one evidently synthetic listener, a veteran of the previous study, heard three *nonadjacent* tritone pairs in the C$\#_5$+ set as clearly falling, the others as rising. Other listeners showed less striking irregularities. This finding may be related to the finding (Repp 1994) of larger individual differences and variability of response functions for the D$\#_5$ set than for the A_4 set. It may reflect a conflict between two response tendencies, one based on pitch class and the other on envelope shape. Since, according to Deutsch et al (1987), American (or at least Californian) listeners tend to perceive pitch classes between C and D# as subjectively highest, such a pitch-class effect would be competing with the envelope effect in sets whose envelope is centered in this region, whereas it would be congruent with the envelope effect in sets whose envelope is centered in the opposite region of the pitch class circle (between F# and A).

Did the individual listeners in experiment 1 show a pitch-class effect? The question can be posed only for the eight participants who listened to both envelope sets. Of them, four did not exhibit any clear pitch-class effect.[11] The other four showed effects with magnitudes between 30% and 65% and subjectively highest pitch classes near B, C#, D#, and G, respectively. The first three are in the region considered typical of American (Californian) listeners; the last one is not. (Interestingly, but perhaps misleadingly, that listener was Chinese.) Seven of the eight listeners, however, showed envelope effects, with magnitudes ranging from 31% to 58% and highest pitch classes between 4 and 7 st from the envelope center.

3 Experiment 2

The purpose of experiment 2 was to explore envelope effects further by presenting a few listeners with a much wider range of stimulus translations along the frequency axis. In that respect, the experiment was similar to Deutsch's (1987) study, which purported to show that envelope center frequency has little effect. In contrast to Deutsch's study, the six listeners in the present experiment were not selected on the basis of pretest results, and it was expected that at least some of them would show large envelope effects. The question was how these effects would be patterned over a range of several octaves. Deutsch (1987) noted that some of her listeners were affected more by an octave shift in envelope center than by a 6-st shift within an octave, though all effects were small. Also, in a recent reanalysis of earlier data, Deutsch (1994a) found that American listeners showed shallower response functions and gave more "up" responses in the fifth octave compared with the fourth. There is more to be learned about the role of absolute pitch height in this paradigm.

Experiment 2 differed from Deutsch's (1987) study in the method of stimulus presentation. Rather than blocking the tritone pairs by envelope and randomizing their order within blocks, they were blocked here by pitch-class content and were presented in a nonrandom sequence, as explained below. This was another way of minimizing influences of preceding context.

3.1 *Methods*

3.1.1 *Stimuli.* The stimulus sets from Repp (1994), with envelope centers at A_4 and D$\#_5$, were reinput from the original digital audio tape to a Macintosh Quadra 700 computer and then were transferred to a Quadra 660AV. All stimuli were generated on-line from these original tritone pairs.

[11] The minimal requirement for an effect of either pitch class or envelope is a contiguous region of at least five (usually six) positive difference scores with a single peak, as illustrated in figure 4.

3.1.2 *Listeners.* Six musically trained listeners participated. They included two female research assistants who had doctorates in music theory and musicology, respectively (LP, LR), one female graduate student in musicology (RB), two male Yale undergraduates who played musical instruments (DZ, NY), and myself (BR). The students were paid for their services. RB was British, BR Austrian, the others American.

3.1.3 *Procedure.* Each listener completed two self-paced sessions, each lasting about 3 h. The two sessions were separated by intervals ranging from a few days to several months. Those participants who had not heard octave-complex tones previously (LR, RB, DZ, NY) first listened to the original stimulus tape (Repp 1994) for familiarization. (The results will not be discussed here.) In the main task, the listener sat in front of the Macintosh Quadra 660AV computer and listened to its sound output binaurally over Sennheiser HD 540 II earphones. SOUNDEDIT16 software was used for playback. This program has a 'Play as Instrument' option which allows a digitized sound to be played back at various speeds, without any distortion. This makes it possible to transpose tritone pairs on-line, instead of synthesizing and sequencing a very large number of individual stimuli. A sound file is played at different speeds by clicking on the keys of a miniature piano keyboard displayed on the computer screen. Middle C (C_4) corresponds to the original sound, C_3 to half the speed, C_5 to twice the speed, and so on. The speeds used ranged from slower by a factor of 4 to faster by a factor of 4 than the original, so that the durations of individual tones ranged from 2000 to 125 ms. Tone duration thus was confounded with absolute pitch height in this study, which seemed a reasonable price to pay for the convenience of the procedure. It was considered unlikely that tone duration as such would interact with the perception of relative pitch.

The listener selected a stimulus pair and played it on different keys of the miniature keyboard, according to a prescribed sequence. In the first session, this sequence of keys went from C_2 to C_6 and back three times in 6-st steps; in the second session, it went from $D\#_2$ to A_5 and back three times in 6-st steps. Thus each stimulus pair was heard six times at each playback speed. At any given speed, the pair could be repeated as often as the listener found necessary to reach a decision, but deviations from the prescribed sequence were not permitted. The response (an up or down arrow) was entered on a prepared answer sheet. In the course of the two sessions, each stimulus pair was heard in spectral translations ranging from two octaves below to two octaves above the original in 3-st steps, though the sequence within each session proceeded strictly in 6-st steps. This step size had the effect of keeping the pitch classes constant but reversing their order from one translation to the next.[12] A listener with an invariant response to pitch class should hear the direction of the pitch change reverse from one step to the next, whereas a listener subject to strong influences of envelope position should continue to hear the pitch changing in the same direction.

In each session, all twenty-four original tritone pairs were heard in all prescribed translations. The original pairs were arranged in six groups of four, so that each group had the same pitch-class content. For example, one group might include D#–A (A_4 set), A–D# ($D\#_5$ set), A–D# (A_4 set), and D#–A ($D\#_5$ set). Listeners were encouraged to take a break after each group, ie, about every 30 min. Within those 30 min, only two pitch classes were heard.

The responses were scored by assigning '1' to every "down" response and '0' to every "up" response and then averaging across the six responses to each specific stimulus. Since any pair from the A_4 set was identical with the reverse-order pair from the $D\#_5$ set

[12] For example, if the original pair selected was D–G#, the listener would start in the first session with D–G# (24 st lower), then would hear G#–D (18 st lower), then again D–G# (12 st lower), and so on. In the second session, (s)he would hear F–B (21 st lower), then B–F (15 st lower), then F–B again (9 st lower), and so on.

translated down by 6 st (except for a difference in duration), the results for these matching pairs were also averaged, so that there were in fact twelve judgments for most pairs.

3.2 *Results and discussion*

The results for the six listeners are shown in figure 5 in the form of contour diagrams. To make the patterns more visible, the data have been replicated once along the abscissa, going twice around the pitch-class circle. Dark areas represent "down" responses. Vertical bands in these graphs indicate a pitch-class effect, whereas rising diagonal bands reveal spectral-envelope effects. These effects were quantified by the method illustrated in figure 4.

LP showed the most nearly invariant pattern. Her pitch-class effect was large (80%) and was centered on A#/B as the subjectively highest pitch classes. However, she did have some diagonal trends, especially in the lower two octaves, indicating some envelope dependency. Her overall envelope effect was 18%, with the highest pitch classes being 6–7 st from the envelope center.

LR also showed indications of a pitch-class effect, but it was much smaller (22%) and centered on C/C#. She showed a striking asymmetry in her reduced response functions, however, indicating that D should really be considered as her single subjectively highest pitch class, in which case the magnitude of the effect rises to 72%. (The alternating response pattern for the adjacent C# in figure 5 indicates a systematic change in responses across sessions.) Diagonal trends were discernible in the lower octaves, but the envelope effect was not very clear. This listener heard the lowest-pitched tritone pairs mostly as falling in pitch.

RB showed predominantly diagonal patterns that repeated from octave to octave. She showed a clear envelope effect (46%), with the highest pitch classes 5–6 st away from the envelope center. She also showed a small pitch-class effect (27%), with the highest pitch classes being G#/A. This is indeed in accord with the results Deutsch (1991) has reported for typical British listeners. Like LR, RB heard the lowest-pitched pairs as falling.

The most striking diagonal pattern was shown by DZ. His responses were governed entirely by envelope position rather than by pitch class. In addition, however, he was strongly influenced by absolute pitch height: he perceived high-pitched pairs invariably as rising and low-pitched pairs as falling. Owing to these perceptual biases, DZ's envelope effect was only of moderate size (38%) when all data were included, but it became very large (70%) when only the central two octaves were considered. His highest pitch classes were once again 5–6 st from the envelope center.

These four listeners thus lie along a continuum from predominantly pitch-class-determined to purely envelope-determined responses, with a concomitant increase in perceptual biases at extreme pitch heights.

Of the remaining two listeners, NY showed a very unclear pattern, with faint diagonal trends and a strong bias to hear rising pitch throughout. His comments suggested that he heard several tone components and thus was an analytic listener. Last, I (BR), like DZ, showed a very strong effect of pitch height that actually led me to omit the lowest octave of envelope positions after it became obvious that these stimuli all sounded falling to me. This trend is consistent with my responses in experiment 1, where I heard nearly all pairs in the $C\#_5+$ and G_4+ sets as falling. Despite the absence of strong patterns in NY's and BR's response maps, overall pitch-class and envelope effects could be determined for both. They were naturally weak (24% or less) but otherwise similar: for both, the highest pitch classes were, respectively, at B/C and 4–5 st removed from the envelope center. Of course, the smaller the effects, the less reliable are these estimates.

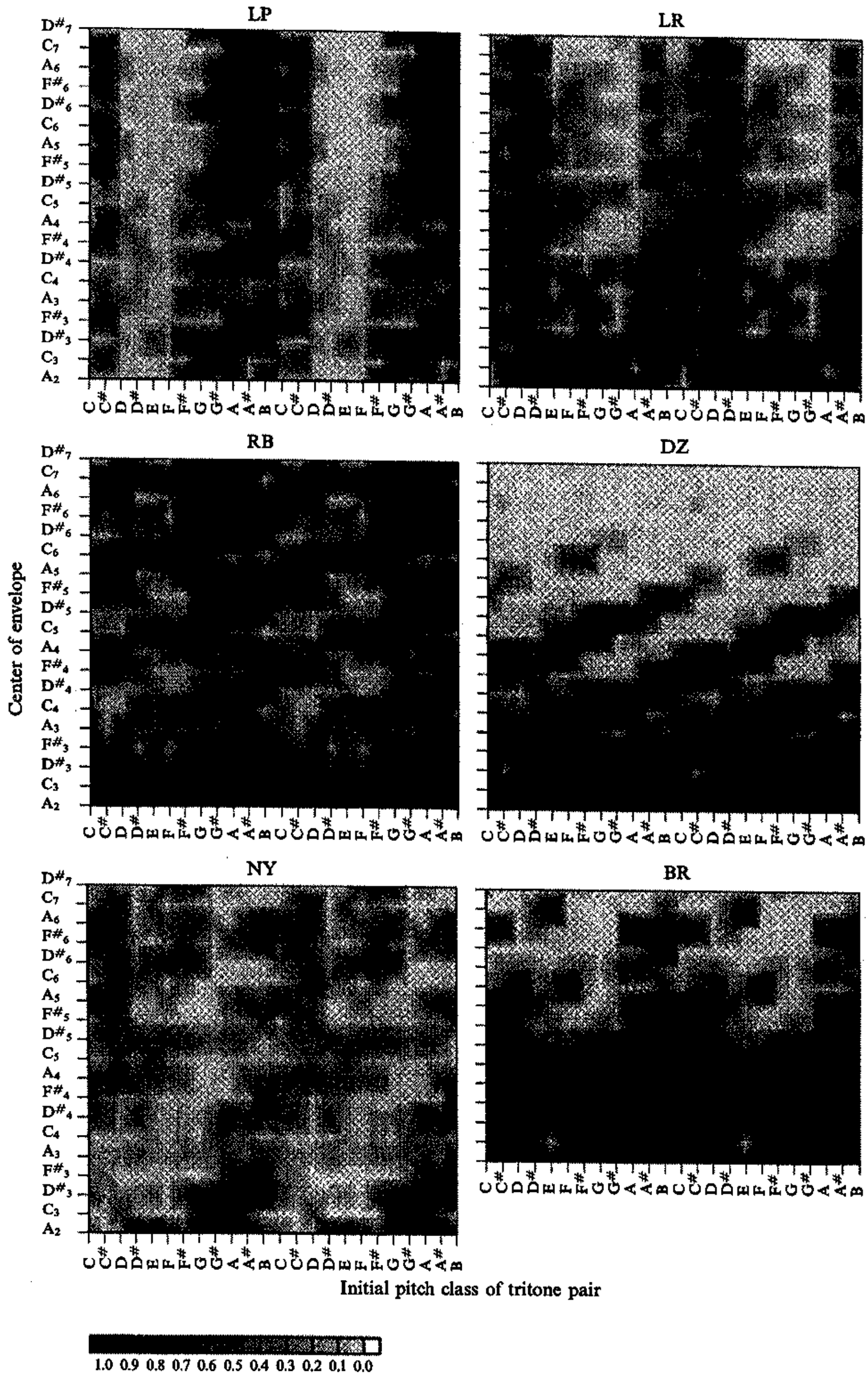

Figure 5. Contour plots of results (proportions of "down" responses) for six individual listeners in experiment 2. The data are plotted twice along the abscissa.

These results, in conjunction with the results of experiment 1 and those of the previous study (Repp 1994), confirm that a pure pitch-class effect is rare. Most listeners are affected by the spectral envelopes of the stimuli, and some listeners' responses depend entirely on envelope rather than on pitch class. While the invariant pitch-class-based response functions obtained by Deutsch (1987) remain an interesting phenomenon in need of explanation, additional questions now arise as to why only relatively few listeners show this effect and what abilities or experiences may lead to it. The perceptual biases at extremes of pitch height (and duration) are another phenomenon whose causes remain to be explored.

4 Experiment 3

In both the two preceding experiments an attempt was made to control context effects: in experiment 1, context was eliminated entirely; in experiment 2 it was held constant in terms of pitch classes. The underlying assumption, of course, was that sequential context does have an effect on the perception of tritone pairs. However, apart from an informal demonstration (Shepard 1983), casual observations (Repp 1994), and an unpublished conference paper (Tuller and Giangrande 1991), there has not been a clear demonstration of contextual effects on tritone perception.(13) Experiment 3 was conducted to provide such a demonstration.

In Shepard's (1964) classic study, octave-complex tones from the same envelope set, separated by a variable number of semitones, were presented in pairs and listeners had to tell whether the pitch went up or down. Their responses were largely determined by pitch-class proximity: for example, the pair C–A was nearly always heard as falling by 3 st, not as rising by 9 st, whereas C–D# was heard as rising by 3 st, not as falling by 9 st. A similar principle is likely to underlie sequential context effects: an adaptive perceptual system will try to minimize the pitch-class distance among successive octave-complex tones, even if their number exceeds two. If so, then it should be possible to sway the perception of tritone pairs by preceding them with a single precursor tone. These effects should be largest for listeners with weak pitch-class and envelope effects, such as myself.

In this experiment, there were three conditions. In condition 1, each tritone pair (eg C–F#) was immediately preceded by a tone whose pitch class bisected the tritone interval (eg, D#–C–F# or A–C–F#). According to the principle of pitch-class proximity, the direction of pitch change between the first two tones in such a triplet should be rather unambiguous; that is, listeners should hear a pitch step of 3 st rather than 9 st, so that D#–C is perceived as falling and A–C as rising. The same principle predicts that the direction of change of the following tritone step will tend to be perceived as going in the opposite direction. That is, a triplet such as D#–C–F# should be perceived as falling–rising, which keeps the total pitch range to 6 st. If it were perceived as falling–falling, the pitch distance between the first and third tones would be 9 st, which is contrary to the proximity principle. Conversely, the triplet A–C–F# should be perceived as rising–falling, not as rising–rising. Thus, regardless of how the tritone pair C–F# is perceived in isolation, listeners should perceive it more often as falling when it is preceded by A than when it is preceded by D#.(14)

In condition 2, each tone triplet was further preceded by a presentation of the tritone pair (the second and third tones) alone. The question of interest here was whether prior exposure to the tritone pair would make listeners resistant to the subsequent

(13) An effect of adaptation to spectral motion, possibly mediated by immediate context, is reported in a recent paper by Dawe et al (in press).

(14) A related effect was demonstrated informally by Shepard (1983). He preceded a tritone pair such as C–F# with the triplets C–D#–F# or C–A–F#, which were perceived as rising–rising and falling–falling, respectively. The following tritone pair was perceived accordingly as rising in the first context and as falling in the second context.

effect of the precursor tone. Thus, if C–F# were first heard as rising, would this judgment still be reversed a few seconds later in the context of a preceding A? If so, this would be particularly strong evidence for an effect of context.

In condition 3, each tone triplet was preceded by the first two tones of the triplet. There was no reason to expect this form of preexposure to lead to results different from condition 1, but it was included because it did have an unexpected effect on my own perception.

4.1 *Methods*

4.1.1 *Stimuli.* The stimuli were constructed from the tritone pairs of the two original series, whose spectral envelopes were centered on A_4 and $D\#_5$, respectively. By using a waveform editing program, each of the twelve pairs in each set was prefixed with each of two precursor tones from the same set, which were either 3 st lower or 3 st higher than the first tone of the tritone pair. This resulted in twenty-four tone triplets per stimulus set. The precursor tone immediately preceded the tritone pair, without any intervening silence.

4.1.2 *Listeners.* Five listeners participated, four of whom (LP, LR, RB, and BR) had also been in experiment 2. The fifth individual, DR, was a female volunteer from the staff of Haskins Laboratories who had participated in experiment 1.

4.1.3 *Procedure.* The digitized stimuli were played back on a Macintosh Quadra 660AV computer by means of the SOUNDEDIT16 program, and the listeners heard the stimuli binaurally over Sennheiser HD 540 II earphones at a comfortable intensity. Conditions 1, 2, and 3 were presented in that order. Stimulus presentation was controlled by me: I sat at the computer and also wrote down the listener's verbal responses. The stimulus sequence was quasi-random and different for each listener, except that stimuli were drawn alternately from the A_4 and $D\#_5$ sets. Intertrial intervals were variable (roughly between 5 and 10 s). In condition 1, the listeners' response choices were "up–down", "up–up", "down–up", and "down–down". In conditions 2 and 3, listeners first made an "up" or "down" decision on the tone pair presented and then reported their perception of the subsequent triplet. A few seconds intervened between the pair (selected from the triplet waveform by means of a mouse-controlled cursor) and the triplet. If either a pair or a triplet sounded ambiguous to the listener, it was repeated immediately until a decision could be reached. The session lasted about 75 min.

4.2 *Results and discussion*

4.2.1 *Condition 1.* The results will be discussed in terms of average percentages of "down" responses to the twelve tritone pairs in each envelope set following each type of precursor. These percentages are listed in table 1. It was predicted that listeners would give more "down" responses after a −3 st precursor than after a +3 st precursor. It is evident that all five listeners exhibited such a context effect in both stimulus sets. LR and BR showed the most striking effects; their perception of the tritone pairs was governed entirely by the preceding context. The other three listeners showed some partial resistance to the effect of the precursor. However, even LP—a listener with a strong pitch-class effect (see figure 5)—was not immune to context, especially not in the A_4 set.

Three of the listeners (LP, LR, and BR) always perceived the pitch interval between the first two tones of a triplet as a 3-st step, as expected. The other two listeners, however, sometimes perceived it as a 9-st step in the opposite direction; in the forty-eight trials, this happened three times for RB and seven times for DR. In all of these instances, the pitch of the following tritone pair was heard as changing in the opposite direction, so that the total pitch range was 9 st, never 15 st (which is logically impossible in terms of pitch classes and probably perceptually impossible with octave-complex tones).

Table 1. Average percentages of "down" responses in the three conditions of experiment 3, as a function of envelope set and precursor tone.

Listener	Condition 1				Condition 2				Condition 3			
	A_4		$D\#_5$		A_4		$D\#_5$		A_4		$D\#_5$	
	+3 st	−3 st	+3 st	−3 st	+3 st	−3 st	+3 st	−3 st	+3 st	−3 st	+3 st	−3 st
LP	8	83	33	58	25	42	42	42	0	42	50	42
LR	0	92	0	100	0	83	0	83	0	100	0	100
RB	67	100	58	83	58	92	50	50	33	83	42	75
DR	42	83	42	75	25	58	42	50	25	83	42	58
BR	0	100	0	100	58	50	50	50	0	0	0	0
									[8	75	8	50][1]

[1] Data from a later repeat of this condition.

4.2.2 *Condition 2.* In this condition the listeners first responded to the tritone pair without the precursor and then listened to the triplet. As can be seen in table 1, this resulted in a substantial reduction of the context effect. Most strikingly, my tritone perception, which had been completely determined by the precursor in condition 1, became completely immune to context after pre-exposure! LR, on the other hand, was still almost totally context dependent; that is, she frequently reversed her judgments of the tritone pairs. LP, RB, and DR still showed small effects of context, but only in the A_4 set.

4.2.3 *Condition 3.* This condition, in which listeners were preexposed to the first two tones of each triplet, was expected to yield results similar to condition 1. This was essentially the case, though context effects were somewhat smaller than in condition 1. Only LR was still completely context dependent. The only listener for whom there was a substantial change was I myself: I now heard all tritone pairs as rising, regardless of precursor! This was particularly paradoxical in view of my results in experiment 1, where I perceived nearly all pairs as falling. Three months later, I listened to this condition for a second time. This time, my results (in brackets in table 1) were more similar to those of the other listeners, though I still showed an overall bias towards rising percepts.

4.2.4 *Response functions.* Individual response functions were generally orderly, even though only a single response had been obtained from each listener to each stimulus combination. To determine pitch-class and envelope effects, each listener's response functions were averaged across the three conditions and across the two precursor types within each condition, under the assumption that these manipulations affected an independent perceptual criterion, but did not interact with the two effects of interest. Indeed, LP showed a pitch-class effect similar to the one she had shown in experiment 2, centered here on A/A# (58%), as well as an envelope effect (50%) centered at 8 st (ie with the highest pitch class 8 st away from the envelope center). No such effects could be determined for LR. RB only showed a large envelope effect (75%), centered at 6–7 st. DR also showed an envelope effect (54%), centered at 7-8 st, as well as a pitch-class effect (37%) centered at a location not common among American listeners (G#/A). I (BR) showed a small pitch-class effect (33%) centered at G/G#, which did not match my even smaller effect in experiment 2, and an envelope effect (46%) centered at 8–9 st. It is not clear what caused the shifts of the envelope effects away from the expected location around 6 st; there may have been some interaction with context, after all. Suffice it to say that envelope effects were again common and typically larger than pitch-class effects.

5 Experiment 4

After experiments 1–3 had been completed, a CD with Deutsch's tritone test became available (Deutsch 1995). Although the results so far leave little doubt that the prevalence of spectral envelope effects in my previous study (Repp 1994) was not due to peculiarities of the presentation format, and although my stimuli were similar to those employed by Deutsch, a demonstration of similar findings with Deutsch's own test would be definitive. This was the purpose of experiment 4, which also employed a larger number of listeners than the preceding experiments.

5.1 *Methods*

5.1.1 *Stimuli.* The tritone test contained on Deutsch's CD comprises sixteen blocks of twelve tone pairs each, with interpair intervals of 5 s and interblock intervals of 30 s. The total test takes about 22 min. (The spoken introduction and the practice trials preceding the test on the CD were omitted.) The stimulus sequences are listed in a booklet accompanying the CD. However, the booklet does not provide any information about the spectral envelopes of the stimuli. Since a request for this information remained unanswered, the envelope characteristics were determined through spectral analysis of selected stimuli from each block. It emerged that four envelope center frequencies were represented, each in four blocks of stimuli: C_4 (blocks 2, 8, 10, and 14), $F\#_4$ (blocks 1, 5, 13, and 15), C_5 (blocks 3, 6, 12, and 16), and $F\#_5$ (blocks 4, 7, 9, and 11). The test thus appears to be identical with the one used in Deutsch's more recent studies (Deutsch et al 1987, 1990; Deutsch 1991), except that interblock intervals of 1 min were used there.

To familiarize listeners with the task and stimuli, and to test their ability to make consistent pitch judgments, a pretest was devised consisting of a random sequence of twenty-four pairs of octave-complex tones whose pitch classes differed by 3 st rather than 6 st. These pairs were constructed from the tones of the A_4 set from Repp (1994). In accord with the pitch-class-proximity principle, listeners were expected to hear these pairs 'correctly' as 3-st rather than 9-st pitch steps. My original $D\#_5$ and A_4 tests (Repp 1994), each containing twelve blocks of twelve tritone pairs, with interpair intervals of 2.5 s and interblock intervals of 5 s, were also administered.

5.1.2 *Listeners.* Thirty-two individuals participated. Nearly all of them were students of college age who had responded to an advertisement on Yale campus. They represented a wide range of musical experience and ethnic/linguistic backgrounds and were paid for their efforts. My research assistant (LR) participated as an unpaid volunteer.

5.1.3 *Procedure.* Each listener was initially given a standard hearing test, by means of a commercial audiometer in a sound-insulated cubicle. Pure-tone thresholds revealed that all participants had normal hearing in both ears in the range from 250 to 4000 Hz. Subsequently, the pretest was presented, played from a Macintosh Quadra 660AV computer binaurally over Sennheiser HD 540 II earphones at a comfortable intensity. Then the earphones were connected to an Optimus CD-3480 portable CD player, and Deutsch's CD (tracks 15–18) was played. Finally, after a short break, the original $D\#_5$ and A_4 tests from Repp (1994) were presented monaurally, in that order, from digital tape. For each test, the listener entered "up" and "down" arrows on prepared answer sheets; a decision was requested for each tone pair. During the break, participants completed a questionnaire about their own and their parents' native language(s) and about the region(s) in which they had grown up. All but three also filled out a questionnaire about their musical experience in connection with a different experiment.

5.2 *Results and discussion*

5.2.1 *Pretest.* As expected, the pretest stimuli were generally perceived 'correctly', ie according to the shorter pitch-class distance between the tones (3 st). Sixteen of the thirty-two listeners judged all twenty-four tone pairs correctly, eleven listeners between twenty and twenty-three pairs, four listeners between seventeen and nineteen pairs, and one listener thirteen pairs. Although a low pretest score may indicate an inability to make relative-pitch judgments, no listener's data were discarded on the basis of the pretest results.

5.2.2 *Average results.* The average response functions for the different stimulus sets are displayed in figure 6. Figure 6a shows the results for the Deutsch CD test. It can be seen that three of the envelope sets yielded average response functions with pronounced maxima and minima; only the C_4 set gave, surprisingly, a flat average function. The C_5 set showed an overall bias in favor of "up" responses, but its response function was the inverse of those for the two F# sets, which agreed well with each other. Moreover, the peaks of the F# response functions were in the vicinity of C, whereas the peak of the C_5 response function was at F#, exactly as expected. Overall, there was a spectral-envelope effect of 39%, with the subjectively highest pitch class being 7 – 8 st away from the envelope center. The data for the Repp stimuli, shown in figure 6b, resemble the earlier findings shown in figure 2. Here the average envelope effect was 46% and also was centered at 7 – 8 st. The shift away from the pitch class opposite the envelope center was unexpected and evidently was caused more by the minima than by the maxima in the average response functions.

In a reanalysis of earlier data (Deutsch 1991), Deutsch (1994a) found that a group of Californian listeners showed a flat response function for the C_4 and $F\#_4$ sets combined but a function resembling that of the present $F\#_5$ set for the C_5 and $F\#_5$ sets combined. The present results imply just the opposite—a flatter average response function in the fifth than in the fourth octave. The difference could be due to the different composition of the present group of listeners.

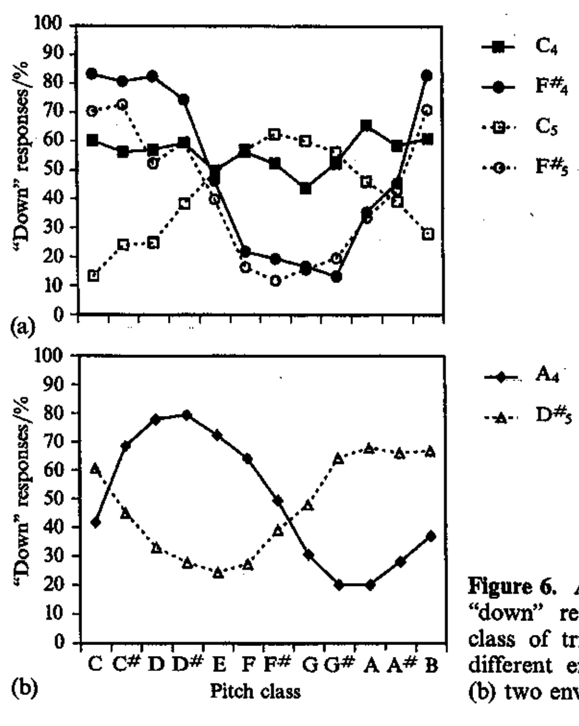

Figure 6. Average response functions (percentage "down" responses as a function of initial pitch class of tritone pair) in experiment 4 for (a) four different envelope sets in Deutsch's CD test and (b) two envelope sets in Repp's test.

5.2.3 *Individual results.* Individual pitch-class effects were determined for all listeners, separately for the Deutsch and Repp tests, by using the procedure illustrated in figure 4. For those twenty-two listeners for whom an effect could be determined on both tests, there was a good correlation between the subjectively highest pitch classes (numerically coded) in the two tests ($r_{20} = 0.76$, $p < 0.001$). For seventeen listeners, the highest pitch classes differed by 2 st or less; only for five there were larger differences. Moreover, there was no systematic average shift of subjectively highest pitch classes from one test to the other, even though different sets of envelopes had been employed. The two tests thus seemed to be measuring the same thing.

Each listener's response functions were then averaged across both tests, and pitch-class effects were recalculated. For five listeners no effect could be determined. For the other twenty-seven participants, figure 7a shows the distribution of subjectively highest pitch classes and the magnitudes of the effects. The distribution, especially of the larger effects, is in good agreement with the one obtained for a group of Californian listeners by Deutsch et al (1987), even though the present group of listeners was much more heterogenous.

There was a significant correlation between the magnitude of the pitch-class effect and musical training ($r = 0.63$, $p < 0.001$), quantified as the number of years of instruction on a (primary) musical instrument. All effects smaller than 25% were shown by individuals with less than 5 years of musical instruction. However, two of the five individuals who did not show any pitch-class effect had considerable musical training; if they are assigned an effect magnitude of zero and are included, the correlation is smaller but still significant ($r = 0.45$, $p < 0.01$). The pretest scores were not predictive of the magnitude of the pitch class effect, nor was there any significant relationship between musical training and pretest scores.

Figure 7b shows the distribution and magnitudes of individual envelope effects, which could be determined for all thirty-two listeners. With the exception of one very small effect (probably spurious), all subjectively highest pitch classes were 5–8 st removed from the envelope center frequency. It is not clear what caused the small but obviously systematic shift away from the 6-st point, but the data certainly reinforce the earlier observations that pitch classes roughly opposite the envelope center frequency tend to be perceived as high.

A negative correlation was expected between the magnitudes of individual pitch class and envelope effects, but it did not reach significance ($r = -0.22$). This was due

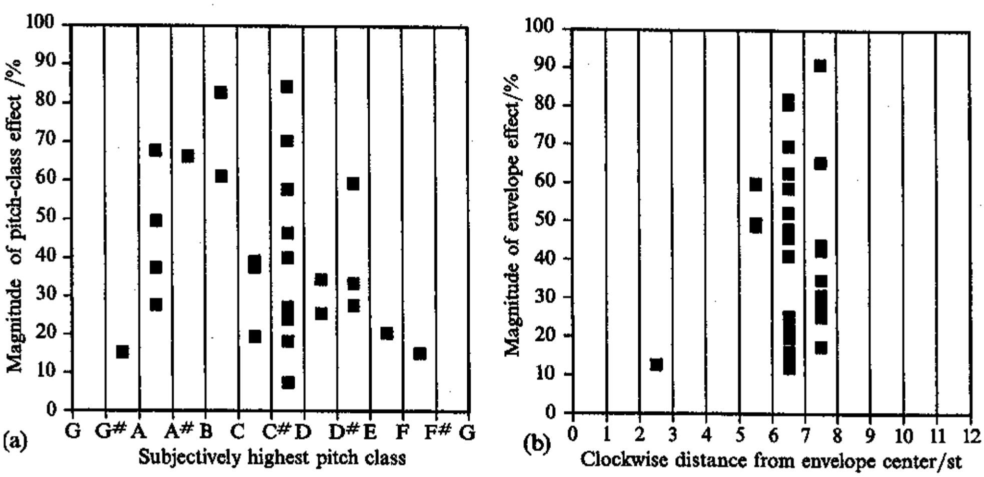

Figure 7. (a) Individual pitch class effects and (b) individual envelope effects in the Deutsch and Repp tests combined.

to a number of individuals for whom both effects were small. If only the fourteen listeners were considered who showed at least one effect greater than 50%, then the negative correlation was very clear ($r = -0.85$, $p < 0.001$). Obviously, it is impossible to have both a large pitch-class effect and a large envelope effect. Years of musical training were unrelated to the magnitude of the envelope effect. However, there was a positive correlation between the pretest results and the envelope effect ($r = 0.52$, $p < 0.01$): listeners who got fewer than twenty-one items 'correct' on the pretest all showed small envelope effects. If listeners had been preselected according to pretest results (as in several of Deutsch's studies), the evidence for envelope effects would have been even more striking.

6 General discussion

Deutsch in her studies of the tritone paradox either did not report spectral-envelope effects at all or focused on a few individuals who were particularly resistant to envelope effects (Deutsch 1987). From my previous (Repp 1994) and present experiments, it is now abundantly clear that such a 'pure' pitch-class effect is rare and that envelope effects are strong and ubiquitous.(15) This finding does not invalidate Deutsch's hypothesis of an underlying individual pitch-class template. Rather, it indicates that this underlying response tendency is in competition with another tendency due to spectral-envelope shape, and that these effects vary in absolute and relative strength. Any given individual's responses are a weighted mixture of these two tendencies. It has been assumed all along, by Deutsch as well as by me, that the two effects are additive, so that each can be assessed by 'averaging out' the other. However, this is not necessarily so, and investigations of tritone perception are needed that specifically examine this assumption.

At present, it is not clear how the effect of envelope position comes about. The effect seems to invite some kind of psychoacoustic explanation, perhaps in the framework of a pitch-processing theory such as Terhardt's (Terhardt et al 1982, 1986; Terhardt 1991). Listeners may judge the direction of pitch change by relying on the most salient partials (spectral pitches) in their auditory representation of the stimuli, or on a virtual pitch coinciding with one of these partials (Terhardt 1991; Cohen et al 1995). An auditory spectral-weighting function with a peak around 700 Hz (F_5) has been postulated by Terhardt et al (1982). Partials below 700 Hz lie on the ascending slope of this function, so that higher-frequency partials are more salient than lower-frequency partials of the same amplitude. However, any weighting function applied to all tones in a stimulus set merely distorts their spectral envelopes but does not explain why some tones are perceived as higher than others. Moreover, this whole issue is complicated by the results of experiment 4, which suggest that the subjectively highest tone is not always exactly 6 st removed from the envelope center. Another troubling finding (Repp 1994) is that the subjectively highest tone in a set of ten-component octave-complex tones under an envelope centered on $D\#_3$ appears to be invariant and 8–9 st (clockwise) away from the envelope peak pitch class. There are many unanswered questions here.

There is a third factor that impinges on the listener, namely preceding context. The results of experiment 3 showed that tritone perception is highly context sensitive, although individuals with a strong pitch-class effect seem to be partially resistant to context. The context effects were obtained with immediately preceding precursor tones, but there is little doubt that context effects also operate in the standard test format, where successive tritone pairs are separated by several seconds of silence (see Repp 1994).

(15) Cohen et al (1994), in an unpublished study, may have found a larger number of pure pitch-class effects. However, I have not seen a detailed report of this work. Dawe et al (in press) also seem to have found only small envelope effects, so that it is possible that the results of Deutsch (1987) are representative, after all. In that case, it remains an enigma why only I find large envelope effects.

Tritone perception seems to be a dynamic process in which the listener continuously redefines a pitch-class range or 'octave window' within which the stimuli are judged (see Cohen et al 1995; Leman 1995). Some listeners may have a default position of this window that is outside the stimulus range, and this may lead them to hear isolated tritone pairs as uniformly rising or falling (experiment 1). There may also be individual limits on the frequency range over which the window can move, and this may explain the biases observed at extreme pitch heights (experiment 2). These results are somewhat problematic for Deutsch's (1986) hypothesis of a frequency-independent mental pitch-class template. The template may operate only in a certain frequency range, and it may have to be 'transposed' from a default position into the current stimulus range. Deutsch herself has found some evidence that the template is not equally strong in different octaves (Deutsch 1994b), and experiment 4 replicated this general finding, though the specific pattern obtained was just the opposite of what she had found.

Little has been said so far about Deutsch's widely publicized hypothesis that a listener's internal pitch template derives from language experience (see Deutsch et al 1990; Deutsch 1991, 1992, 1994a, 1994b, 1996; Ragozzine and Deutsch 1994). Although I have previously expressed scepticism (Repp 1994), the theory remains an interesting proposal without serious rivals at present, and the present study adds little to that debate. Empirical support for the theory is still quite limited, however, and concerns primarily the difference between American (Californian) and British (southern English) listeners, who show complementary distributions of individual pitch-class effects (Deutsch 1991). As already noted, the pitch-class-effect results of experiment 4 replicate Deutsch's distribution for Californians, despite a much more diverse listener sample that included not only individuals from various regions of the USA but many with Asian, African, or European backgrounds. None of these characteristics seemed to be systematically related to individual differences in the pitch-class effect, though it was difficult to tell because of the diversity. However, very few listeners gave results like those reported by Deutsch (1991) for British listeners.(16)

This American–British difference, even though it was very striking in Deutsch's (1991) data, is in urgent need of replication, as I (Repp 1994) found only weak support for it. I tested three groups of listeners: fifteen Dutch, eleven British, and seventeen Americans. Unfortunately, I did not report separate measures of pitch-class and envelope effects. It seems appropriate to conclude this paper with a reanalysis of these data, using the methods applied in the present experiments. Two aspects of this reanalysis are of special interest: (1) was there really a difference in pitch-class effects between British and American listeners? (I found a significant difference only in their responses to the $D\#_5$ stimulus set, not to the A_4 set); (2) are envelope effects independent of language and nationality, as they should be if they are psychoacoustic in origin?

Figure 8a shows the recalculated pitch-class effects for the three listener groups. It is evident, first, that the American listeners had a very wide distribution of subjectively highest pitch classes. This could be due to their coming from various regions of the USA, though there is little evidence to date of systematic variation in the pitch-class effect within this country. (However, see Ragozzine and Deutsch 1994.) Second, the British listeners (all of whom came from southern England, like Deutsch's subjects) did not have their highest pitch classes in the region that Deutsch (1991) found to be characteristic for British listeners, namely between F and A. They were all located in the half of the pitch-class circle between C/C# and F#/G. The Dutch listeners, with one exception, occupied the same region. Thus, there were no clear differences among the three subject groups, which heightens the need for a replication of Deutsch's (1991) results.

(16) However, in a very recent study, Dawe et al (in press) obtained a British distribution of pitch-class effects for an equally heterogeneous group of students residing in Toronto. This creates another puzzle for Deutsch's hypothesis of a language connection.

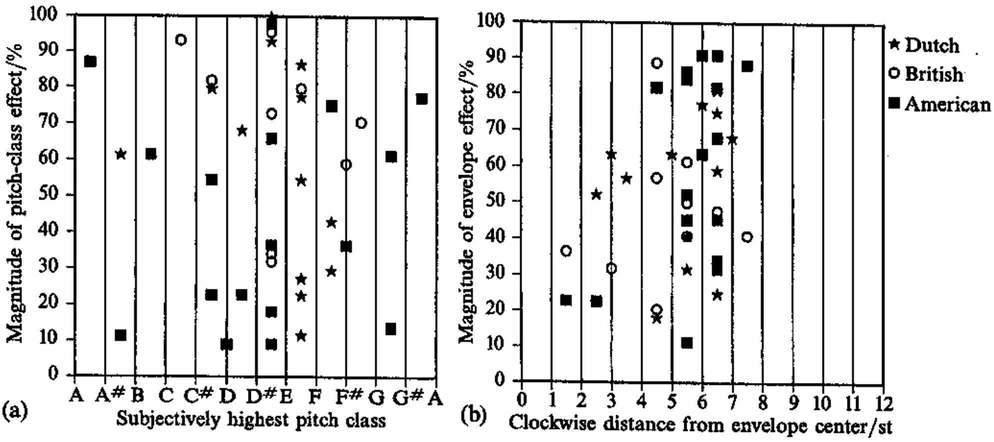

Figure 8. (a) Individual pitch class effects and (b) individual envelope effects in Repp's (1994) study.

The envelope effects for my (Repp 1994) listeners are shown in figure 8b. They showed their tightest concentration around the 6-st point, as expected, but there were several listeners whose subjectively highest pitch classes were closer (clockwise) to the envelope center, in contrast to the opposite tendency in experiment 4. The reason for this difference is not clear. There were no obvious differences among the three nationalities. In any case, differences must be interpreted with caution because it is possible that the pitch class and envelope effects are not independent of each other. A clarification of their relationship should be high on the agenda of future research in this area.

In conclusion, all these data show that octave-complex tones constructed under the same theoretical spectral envelope, intended to differ in chroma only, are not perfectly equated with respect to perceived pitch height: tones with flat-topped, wider envelopes tend to be perceived as subjectively higher than tones with peaked, narrower envelopes. This suggests that the auditory system gives greater weight to the higher than to the lower partials in computing an estimate of average pitch height or sharpness of timbre (Bismarck 1974). It has been noted in previous research on pitch judgment that there are large individual differences in the relative weights given to average pitch height and chroma when the two dimensions are varied independently (eg Ueda and Ohgushi 1987; Demany and Semal 1993). The present results are consistent with these findings, as well as with Deutsch's results suggesting that certain chromas are perceived as subjectively higher than others. However, many unanswered questions remain.

Acknowledgements. This research was partially supported by NIH Grant MH-51230. Thanks are due to Charles Nichols and Lisa Robinson for assistance, to Diana Deutsch, Marc Leman, Richard Parncutt, Anthony Watkins, and several anonymous reviewers for many helpful comments, and to the colleagues at Haskins Laboratories who lent their ears to this project.

References

Bismarck G von, 1974 "Sharpness as an attribute of the timbre of steady sounds" *Acustica* **30** 159–172

Cohen A J, MacKinnon K, Swindale N, 1994 "The tritone paradox revisited: Effects of musical training, envelope peak, and response mode" *Journal of the Acoustical Society of America* **95** 2937

Cohen M A, Grossberg S, Wyse L L, 1995 "A spectral network model of pitch perception" *Journal of the Acoustical Society of America* **98** 862–879

Dawe L A, Platt J R, Welsh E (in press) "Spectral-motion aftereffects and the tritone paradox among Canadian subjects" *Perception & Psychophysics*

Demany L, Semal C, 1993 "Pitch versus brightness of timbre: Detecting combined shifts in fundamental and formant frequency" *Music Perception* **11** 1–14

Deutsch D, 1986 "A musical paradox" *Music Perception* **6** 275–280

Deutsch D, 1987 "The tritone paradox: Effects of spectral variables" *Perception & Psychophysics* **41** 563–575

Deutsch D, 1991 "The tritone paradox: An influence of language on music perception" *Music Perception* **8** 335–347
Deutsch D, 1992 "Paradoxes of musical pitch" *Scientific American* **267**(2) 88–95
Deutsch D, 1994a "The tritone paradox: Some further geographical correlates" *Music Perception* **12** 125–136
Deutsch D, 1994b "The tritone paradox and the pitch range of the speaking voice: Reply to Repp" *Music Perception* 12 257–263
Deutsch D, 1995 "Musical illusions and paradoxes", compact disc available from Philomel Records Inc., PO Box 12189, La Jolla, CA 92039-2189
Deutsch D, 1996 "Mothers and their children hear a musical illusion in strikingly similar ways" *Journal of the Acoustical Society of America* **99** 2482
Deutsch D, Kuyper W L, Fisher Y, 1987 "The tritone paradox: Its presence and form of distribution in a general population" *Music Perception* **5** 79–92
Deutsch D, North T, Ray L, 1990 "The tritone paradox: Correlate with the listener's vocal range for speech" *Music Perception* **7** 371–384
Halpern A R, 1989 "Memory for absolute pitch of familiar songs" *Memory and Cognition* **14** 434–443
Houtsma A J M, Fleuren J F M, 1991 "Analytic and synthetic pitch of two-tone complexes" *Journal of the Acoustical Society of America* **90** 1674–1676
Leman M, 1995 *Music and Schema Theory* (Berlin: Springer)
Levitin D J, 1994 "Absolute memory for musical pitch: Evidence from the production of learned melodies" *Perception & Psychophysics* **56** 414–423
Ragozzine F, Deutsch D, 1994 "A regional difference in perception of the tritone paradox within the United States" *Music Perception* **12** 213–225
Rakowski A, 1993 "Short-term memory for pitch investigated with vocal matchings" *Archives of Acoustics* **18** 173–179
Rakowski A, Morawska-Büngeler M, 1987 "In search for the criteria of absolute pitch" *Archives of Acoustics* **12** 75–87
Repp B H, 1994 "The tritone paradox and the pitch range of the speaking voice: A dubious connection" *Music Perception* **12** 227–255
Shepard R N, 1964 "Circularity in judgments of relative pitch" *Journal of the Acoustical Society of America* **36** 2346–2353
Shepard R N, 1983 "Demonstrations of circular components of pitch" *Journal of the Audio Engineering Society* **31** 641–649
Terhardt E, 1991 "Music perception and sensory information acquisition: Relationships and low-level analogies" *Music Perception* **8** 217–240
Terhardt E, Seewann M, 1983 "Aural key identification and its relationship to absolute pitch" *Music Perception* **1** 63–83
Terhardt E, Stoll G, Schermbach R, Parncutt R, 1986 "Tonhöhenmehrdeutigkeit, Tonverwandtschaft und Identifikation von Sukzessivintervallen" *Acustica* **61** 57–66
Terhardt E, Stoll G, Seewann M, 1982 "Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions" *Journal of the Acoustical Society of America* **71** 671–678
Tuller B, Giangrande J, 1991 "Hysteresis in the perception of directional pitch change" *Journal of the Acoustical Society of America* **89** 1988
Ueda K, Ohgushi K, 1987 "Perceptual components of pitch: Spatial representation using a multidimensional scaling technique" *Journal of the Acoustical Society of America* **82** 1193–1200