Music Perception, 1991, ***8***, 335-347.

The Tritone Paradox: An Influence of Language on Music Perception

DIANA DEUTSCH
University of California, San Diego

The tritone paradox is produced when two tones that are related by a half-octave (or tritone) are presented in succession. Each tone is composed of a set of octave-related harmonics, whose amplitudes are determined by a bell-shaped spectral envelope; thus the tones are clearly defined in terms of pitch class, but poorly defined in terms of height. When listeners judge whether such tone pairs form ascending or descending patterns, their judgments generally show systematic relationships to the positions of the tones along the pitch-class circle: Tones in one region of the circle are heard as higher and those in the opposite region are heard as lower. However, listeners disagree substantially as to whether a given tone pair forms an ascending or a descending pattern, and therefore as to which tones are heard as higher and which as lower.

This paper demonstrates that the basis for the individual differences in perception of this musical pattern lies in the language spoken by the listener. Two groups of subjects made judgments of the tritone paradox. One group had grown up in California, and the other group had grown up in southern England. It was found that when the Californian group tended to hear the pattern as ascending the English group tended to hear it as descending, and when the Californian group tended to hear the pattern as descending the English group tended to hear it as ascending. This finding, coupled with the earlier results of Deutsch, North, and Ray (1990) that showed a correlate between perception of the tritone paradox and the pitch range of the listener's spontaneous speaking voice, indicates strongly that the same, culturally acquired representation of pitch classes influences both speech production and perception of this musical pattern.

THIS paper reports the first demonstration, to the author's knowledge, that perception of music can be influenced by the language spoken by the listener. It shows that two groups of listeners who grew up in different linguistic subcultures perceive the identical musical pattern in strikingly different ways.

Requests for reprints may be addressed to Diana Deutsch, Department of Psychology (1019), University of California, San Diego, La Jolla, California 92093.

The pattern used to demonstrate this relationship is known as the tritone paradox (Deutsch, 1986, 1987; Deutsch, Kuyper, & Fisher, 1987; Deutsch, North, & Ray, 1990). It consists of two successively presented tones that are related by a half-octave, or tritone. For example, C might be presented followed by F#, or D followed by G#, and so on. Each tone is composed of a set of harmonics that stand in octave relation, and whose amplitudes are scaled by a fixed, bell-shaped spectral envelope (Figure 1). The tones are therefore well-defined in terms of pitch class (C, C#, D, and so on) but are poorly defined in terms of height. When listeners determine whether such tone pairs form ascending or descending patterns, their judgments usually display systematic relationships to the positions of the tones along the pitch-class circle: Tones in one region of the circle are heard as higher and tones in the opposite region are heard as lower. However, there is striking disagreement among listeners as to which patterns are heard as ascending and which as descending, and therefore as to which tones are heard as higher and which as lower. For example, some listeners hear the pattern C#-G as ascending and the pattern G-C# as descending, so that for these listeners, pitch class G is heard as higher and

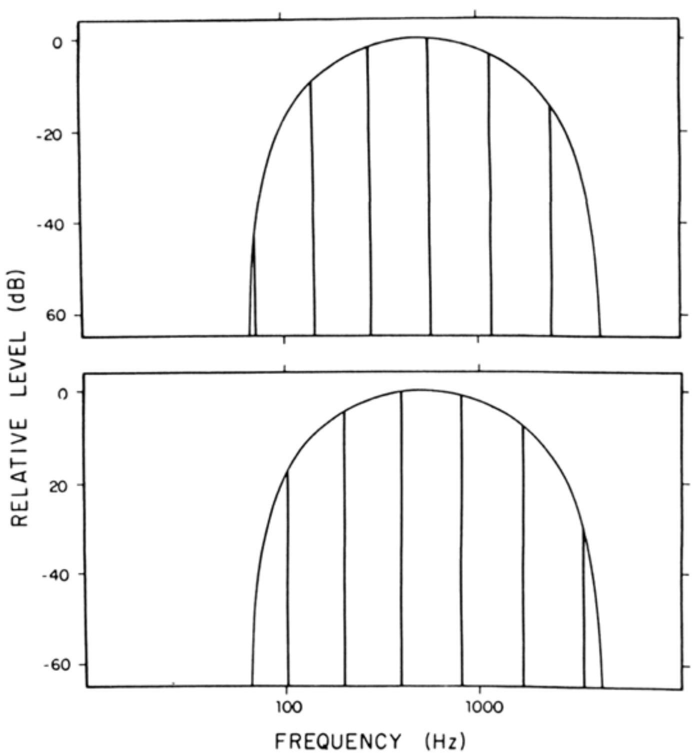

Fig. 1. Spectral composition of a tone pair producing the tritone paradox. Here the spectral envelope is centered at C_5. The upper graph represents a tone of pitch class D, and the lower graph represents a tone of pitch class G#.

pitch class C# as lower. However, other listeners hear the pattern C#-G as *descending* and the pattern G-C# as *ascending,* so that for these listeners the converse holds: pitch class C# is heard as higher and pitch class G as lower.

The tritone paradox has been found to occur in the large majority of subjects in a sizeable population, showing that the phenomenon is not confined to a few selected individuals (Deutsch et al., 1987). Within this population, no correlate with musical training was obtained, either in terms of the size of the effect, or its direction, or the probability of obtaining it. These findings indicate strongly that the phenomenon is not musical in origin. A number of studies have also ruled out explanations in terms of low-level characteristics of the hearing mechanism. For many subjects, the profiles relating pitch class to perceived height are largely unaltered when the position of the spectral envelope is shifted over a three-octave range (Deutsch, 1987). In addition, the profiles are unrelated to patterns of relative loudness for the harmonic components of the tones when these are compared individually (Deutsch, in preparation).

A number of informal observations led the author to hypothesize that perception of the tritone paradox might be related to the processing of speech sounds. Specifically, it was conjectured that the listener develops a long-term representation of the pitch range of his or her speaking voice, and that included in this representation is a delimitation of the octave band in which the largest proportion of pitch values occurs. It was further conjectured that the pitch classes delimiting this octave band for speech are taken as defining the highest position along the pitch class circle, and that this in turn determines the orientation of the pitch class circle with respect to height.

A study was undertaken to test this hypothesis (Deutsch et al., 1990, see also Deutsch, 1989). Subjects were selected who showed clear relationships between pitch class and perceived height in making judgments of the tritone paradox. A 15-min recording of spontaneous speech was taken from each subject, and from this recording the octave band containing the largest number of pitch values was determined. Comparing across subjects, a significant correspondence was indeed obtained between the pitch classes delimiting this octave band for speech and those defining the highest position along the pitch-class circle, as determined by judgments of the tritone paradox.

The findings from this experiment are in accordance with the hypothesis that perception of the tritone paradox is based on a representation of the pitch-class circle by the listener, whose orientation is related to the pitch range of his or her speaking voice. Two versions of this hypothesis may then be advanced. The first, and more restricted, version does not assume that the listener's vocal range for speech is itself determined by such an acquired template. The second, and broader, version assumes that this template is acquired developmentally through exposure to speech produced by others, and that it is used both to evaluate perceived speech, and also to constrain the listener's own speech output. The characteristics of this learned template would therefore be expected to vary across linguistic groups, in a fashion similar to other speech characteristics such as vowel quality. On this line of reasoning, the orientation of the pitch-class circle with respect to height, as reflected in judgments of the tritone paradox, should be similar for individuals within a linguistic group, but should vary for individuals across linguistic groups.

Evidence for the second hypothesis was provided in the earlier study of Deutsch et al. (1987). An orderly distribution of peak pitch classes[1] was found among a sizeable group of subjects who were undergraduates at the University of California, San Diego. As shown in Figure 2, C# and D occurred most frequently as peak pitch classes, the frequency of occurrence of the other pitch classes falling off on either side of these. Although no information was obtained concerning the linguistic backgrounds of these subjects, it can be assumed that the majority had grown up in California and were from the same linguistic subculture.

The present study was undertaken as a direct test of the hypothesis that listeners in a given linguistic subculture should tend to agree in terms of the orientation of the pitch-class circle with respect to height, and that listeners in different linguistic subcultures should tend to disagree. The two

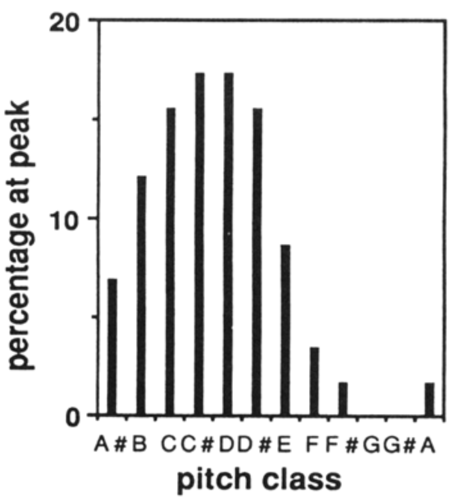

Fig. 2. Distribution of peak pitch classes within a subject population consisting of undergraduates at the University of California, San Diego. Redrawn from Deutsch et al. (1987).

1. The term "peak pitch classes" here refers to the two pitch classes that define the highest position along the pitch-class circle, as determined by judgments of the tritone paradox. See the Results section and Figures 3 and 4 for details.

groups chosen to test this hypothesis consisted of individuals who had grown up in California and those who had grown up in southern England. (This choice was motivated by the author's informal observation that individuals from these two backgrounds tended to hear the tritone paradox in opposite ways.) It was predicted that the first group would show a distribution of peak pitch classes similar to that obtained by Deutsch et al. (1987) in the study of Californian undergraduates, but that the second group would show a different distribution.

Method

SUBJECTS

Two groups of subjects participated in the experiment and were paid for their services. They were selected without regard for musical training, on the basis of obtaining no more than six errors out of a possible 48 in a preliminary experiment in which they judged whether sinusoidal tone pairs that were related by a half-octave formed ascending or descending patterns. All subjects were free of clinical hearing deficits, as determined by audiometry. The subjects in the first group (N = 24) had all grown up in California and had all spent the previous year in California. The subjects in the second group (N = 12) had all grown up in southern England, although most were now living in California. No subject in the Californian group had a parent who had grown up in England, and no subject in the English group had a parent who had grown up in California.

STIMULUS PATTERNS

The tones all consisted of six sinusoids that stood in octave relation and whose amplitudes were determined by a fixed, bell-shaped spectral envelope (Figure 1). The general form of the equation describing the envelope is as follows:

$$A(f) = 0.5 - .05\cos\left[\frac{2\pi}{\gamma}\log_{\beta}\left(\frac{f}{f_{\min}}\right)\right] \qquad f_{\min} \le f \le \beta^{\gamma} f_{\min}$$

where $A(f)$ is the relative amplitude of a given sinusoid at frequency f Hz, β is the frequency ratio formed by adjacent sinusoids (thus for octave spacing, $\beta = 2$), γ is the number of β cycles spanned), and $f_{\min}$ is the minimum frequency for which the amplitude is nonzero. Thus the maximum frequency for which the amplitude is nonzero is $\gamma\beta$ cycles above $f_{\min}$. Throughout, the values $\beta = 2$ and $\gamma = 6$ were used, so that the spectral envelope always spanned exactly six octaves, from $f_{\min}$ to 64 $f_{\min}$.

In order to control for possible effects of the relative amplitudes or loudnesses of the sinusoidal components, tone pairs were created under envelopes that were placed at four different positions along the spectrum. The envelopes were centered at 262 Hz (C_4), 370 Hz ($F\#_4$), 523 Hz (C_5), and 740 Hz ($F\#_5$), and so were spaced at half-octave intervals. We can observe that the relative amplitudes of the sinusoidal components of tones at any given pitch class when generated under the envelopes centered at C_4 and C_5 were identical to those at the pitch class a half-octave removed when generated under the envelopes centered at $F\#_4$ and $F\#_5$. (As an example, the sinusoidal components of the tones comprising the D-G# pattern, when generated under envelopes centered at C_4 and C_5, were identical to those comprising the G#-D pattern, when generated under envelopes centered at $F\#_4$ and $F\#_5$.) The averaging of results obtained by using these different spectral envelopes enabled the balancing out of possible effects of the relative amplitudes of the sinusoidal components of the tones.

Twelve tone pairs were generated under each of the four spectral envelopes, corresponding to the pitch-class pairings C-F#, C#-G, D-G#, D#-A; E-A#, F-B, F#-C, G-C#, G#-D, A-D#, A#-E, and B-F. There were therefore 48 tone pairs altogether. All tones were 500 msec in duration, with no gaps between tones within a pair. The tones were all of equal amplitude.

The tone pairs were presented in blocks of 12, each block consisting of tones generated under one of the spectral envelopes and containing one example of each of the 12 pitch-class pairings. Within blocks, the 12 tone pairs were presented in any of four orderings. The orderings were random, with the restriction that the same pitch class did not occur in any two consecutive trials. In this way, 16 blocks were created altogether, with each of the four orderings employed once for each of the four positions of the spectral envelope.

PROCEDURE

Subjects were tested in soundproof booths. On each trial, a tone pair was presented, and the subject judged whether it formed an ascending or a descending pattern. Within blocks, tone pairs were separated by 5-sec intertrial intervals, and there were 1-min pauses between blocks. There was a 5-min break between the eighth and ninth blocks. Each subject served in two sessions, and all 16 blocks were presented in each session. The data from the two sessions were averaged. Several practice trials were administered at the beginning of each session.

APPARATUS

The tones were generated on a VAX 11/780 computer, interfaced with a DSC-200 Audio Data Conversion System. They were recorded and played back on a Sony PCM-F1 digital audio processor. The output was passed through a Crown amplifier and presented to subjects binaurally through headphones (Grason-Stadler TDH-49) at a level of approximately 72 dB SPL.

Results

The percentage of judgments that a tone pair formed a descending pattern was plotted as a function of the pitch class of the first tone of the pair. The graphs in Figure 3 display the data obtained from six subjects, in each case averaged over two sessions. Three of the subjects were from England, and three were from California. As exemplified in these graphs, judgments were strongly influenced by the positions of the tones along the pitch-class circle. However, also as exemplified here, the direction of this influence varied substantially across subjects.

In order to investigate the form of relationship between pitch class and perceived height within each subject population, the following procedure was used. (This was identical to the procedure adopted earlier by Deutsch et al., 1987). For each subject, the pitch-class circle was bisected so as to maximize the difference between the averaged scores within the two

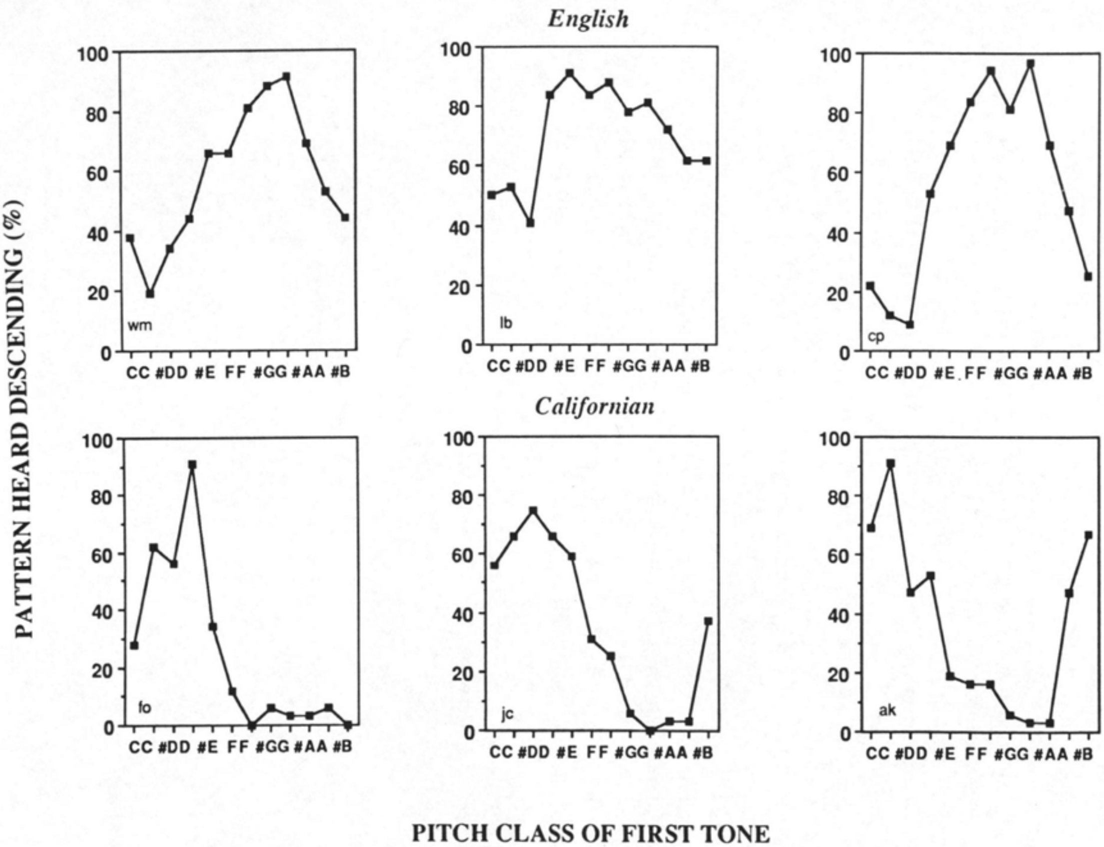

Fig. 3. Percentages of judgments that a tone pair formed a descending pattern, plotted as a function of the pitch class of the first tone of the pair. The three upper graphs show the results from English subjects, and the three lower graphs show the results from Californian subjects.

halves. Next, the circle was oriented so that the line of bisection was horizontal. The data were then retabulated, with the leftmost pitch class of the upper half of the circle taking the first position, its clockwise neighbor taking the second position, and so on. In this way, the peak pitch classes were defined as those that stood at the peak of the normalized circle (i.e., at the third and fourth positions as here defined). So, for example, from the graphs shown in Figure 3, the peak pitch classes for subject AK were C and C#, and those for subject CP were F# and G. Figure 4 depicts the two orientations of the pitch-class circle with respect to height derived from the data of AK and CP shown in Figure 3.

Next, the distributions of peak pitch classes were determined for the English and Californian groups separately. As shown in Figure 5, striking differences between the distributions emerged. For the English group, F#, G, and G# occurred most frequently as peak pitch classes, whereas for the Californian group, B, C, C#, D, and D# occurred most frequently instead.

In order to make a statistical comparison between the two groups, the hypothesis was tested that the Californian group would show a form of distribution similar to that obtained earlier by Deutsch et al. (1987) in the study on Californian undergraduates, but that the English group would show a different form of distribution. To this end, comparison was made between the number of subjects in each group for whom the peak position lay in the half of the pitch-class circle containing the larger number of peak positions in the earlier study. Twenty-one of the 24 Californian subjects fell into this category; however, only three of the 12 English subjects did so. This difference between the two groups was highly significant ($p < .001$ on a Fisher exact probability test).

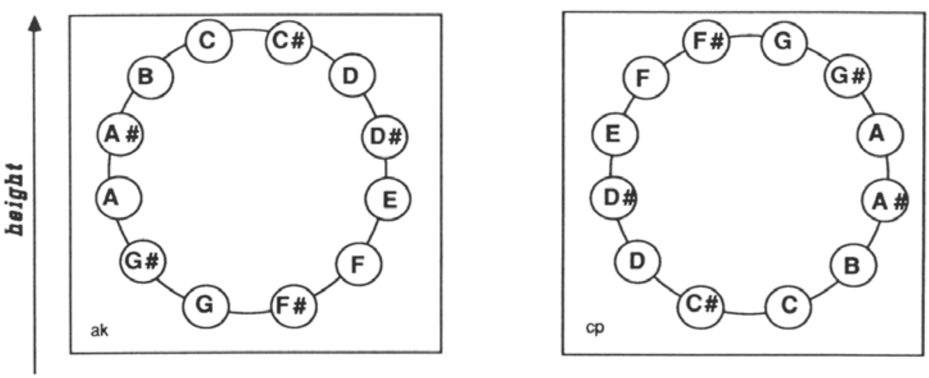

Fig. 4. Orientations of the pitch-class circle with respect to height, derived from the judgments of subject AK (from California) and subject CP (from England), whose data are shown in Figure 3. For subject AK the peak pitch classes were C and C#, and for subject CP the peak pitch classes were F# and G. It can be seen that the two subjects displayed opposite orientations of the pitch-class circle with respect to height.

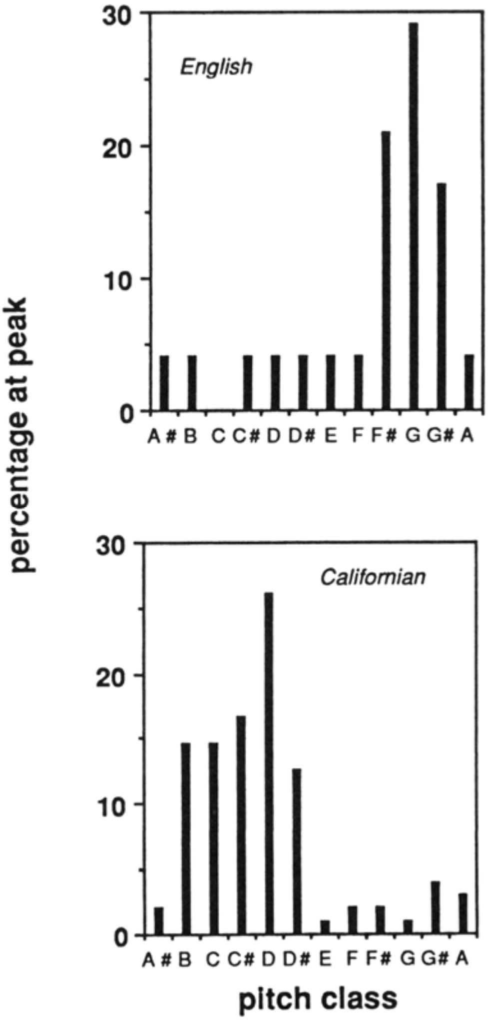

Fig. 5. Distributions of peak pitch classes within the English and the Californian subject populations.

In order to examine whether the phenomenon might be related to musical training, the Californian and English groups were each divided into those who had had more than 2 years of training, and those who had not. These subgroups were then compared by using the same criterion. No significant difference emerged on this measure, among either group ($p > .05$, on a Fisher exact probability test, in both cases). This is in accordance with the previous results of Deutsch et al. (1987), which showed no effects of musical training on perception of the tritone paradox. In order to examine whether there might be an effect of age, the Californian and English groups were each divided into those who were over 22 years of age and those who were under 22 years. Again, no significant difference emerged on this measure, among either group (p > .05, on a Fisher exact probability test, in both cases). Finally, comparison was made between the male and female subjects in both the Californian and the English groups, and again no significant difference emerged (p > .05 on a Fisher exact probability test, in both cases).

Discussion

The present findings provide strong support for the view that, through a developmental learning process, an individual acquires a representation of the pitch-class circle that has a particular orientation with respect to height. The form of this orientation is derived from exposure to speech sounds produced by others and varies from one linguistic subculture to another. From the present experiment we can conclude that for Californians, the agreed upon orientation of the pitch-class circle is such that the highest position occurs around C# and D. However, for people from southern England, the agreed upon orientation is such that the highest position occurs around G instead. It is assumed that such a template is employed both in the production of speech and in the interpretation of speech produced by others. We can observe that a template that is based on pitch class rather than pitch has the useful feature that it can be invoked by both male and female speakers, even though their voices are in different registers. Further evidence for this hypothesis was provided by the recent findings of Deutsch et al. (1990), described earlier, which showed a significant correspondence between a listener's orientation of the pitch-class circle with respect to height and the pitch classes delimiting his or her octave band for speech.

We may briefly speculate concerning the evolutionary value of such an acquired template. As one possibility, it could be of considerable advantage to determine the emotional state of a speaker through the pitch of his or her voice. A template such as this could serve to provide a common framework against which the pitch of a speaker's voice may be evaluated, so providing evidence concerning his or her emotional state. Such a template might also be involved in the communication of syntactic aspects of speech.

We now briefly discuss the implications of the present results for theories of pitch perception. It has been suggested by others that certain characteristics of pitch perception result from a learning process derived from exposure to complex sounds in the environment. For example, Whitfield (1967, 1970) suggested that patterns of neural activity resulting from exposure to combinations of harmonics are learned through continuous exposure to such sounds. Thus when presented with a harmonic series, we attribute the fundamental that in our experience has most frequently been associated with such a series. A more specific argument along these lines was made by Terhardt (1974). He proposed that, through exposure to speech sounds early in life, associative links are formed between the harmonic components of these sounds, so that ultimately when a harmonic series is presented, a fundamental (or "virtual pitch") is invoked by the listener. He also proposed that the same learning process accounts for our apprehension of certain intervallic relationships, such as the octave. Most recently, Terhardt (1991) suggested, in agreement with Deutsch et al. (1990), that the tritone paradox could also be the result of a developmental learning process derived from exposure to speech sounds. The present findings provide strong support for such a perceptual learning hypothesis with respect to the tritone paradox. They also lend indirect support to the hypothesis that certain other characteristics of pitch perception might also be based on perceptual learning, although these hypotheses await experimental verification.

Concerning the musical implications of these findings, we can conclude that under certain conditions at least, perception of music can be strongly influenced by the language spoken by the listener. The conditions under which this influence is manifest in natural musical situations remain to be determined. However, other work has shown that the tritone paradox can be produced by using a variety of tone complexes, provided that these contain some ambiguity of height (Deutsch, in press). In addition, related paradoxical effects have been shown to occur in the perception of certain two-part patterns (Deutsch, 1988; Deutsch et al., 1984, 1986): When such patterns are transposed from one key to another, the relative heights of the different pitch classes are preserved, so that there results a perceived interchange of voices. Further, when such patterns are presented in any one key, listeners differ strikingly in terms of which voice is heard as higher and which as lower, again reflecting differing orientations of the pitch-class circle with respect to height. It appears reasonable to conjecture that differences between listeners in perception of these patterns would also depend on linguistic subculture, in the same way as do differences in perception of the tritone paradox.

Another conclusion from the present findings, together with those obtained earlier on this class of paradoxes (Deutsch, 1986, 1987, 1988, 1989; Deutsch et al., 1984, 1986, 1987, 1990), is that, although absolute pitch is generally considered a rare faculty, the large majority of us exhibit a form of absolute pitch in making judgments of these patterns, in that we hear notes as higher or as lower depending essentially on their pitch classes. A related point has recently been made by Terhardt and Ward (1982) and Terhardt and Seewann (1983). These authors found that musicians were able to determine whether or not well-known passages were played in the correct key, even though most of their subjects did not have absolute pitch as traditionally defined.

In conclusion, the study reported here, coupled with the findings of Deutsch et al. (1990), provides, to the author's knowledge, the first demonstration of an influence of language on music perception. This influence appears to account for differences between listeners in how certain aspects of music are perceived, and we may therefore assume that such differences are cultural rather than innate in origin. In contrast, the handedness correlates that have been obtained with perception of other musical patterns [i.e., the octave and scale illusions (Deutsch, 1974, 1975a, 1975b, 1983)] indicate that differences in music perception can also be based on innate differences at the neurological level.

The finding that two different classes of musical pattern are associated with clear perceptual disagreement leads us to speculate that other such differences might also exist in music perception that have not yet been uncovered. Musical discourse is not precise or accurate enough for such perceptual differences to become apparent through normal communication, and it is only in the laboratory that we can develop a clear idea of what the listener really perceives. The possibility of basic disagreement at the perceptual level therefore should be considered in evaluating the issue of communication between composer, performer, and listener.[2]

References

Deutsch, D. An auditory illusion. *Nature,* 1974, ***251,*** 307-309.

Deutsch, D. Two-channel listening to musical scales. *Journal of the Acoustical Society of America,* 1975a, ***157,*** 1156-1160.

Deutsch, D. Musical illusions. *Scientific American,* 1975b, ***233,*** 92-104.

Deutsch, D. The octave illusion in relation to handedness and familial handedness background. *Neuropsychologia,* 1983, ***21,*** 289-293.

Deutsch, D. A musical paradox. *Music Perception,* 1986, ***3,*** 275-280.

Deutsch, D. The tritone paradox: Effects of spectral variables. *Perception and Psychophysics,* 1987, ***41***, 563-575.

Deutsch, D. The semitone paradox. *Music Perception,* 1988, ***6***, 115-132.

Deutsch, D. Pitch class and octave similarity. *Proceedings of the First International Conference on Music Perception and Cognition.* Kyoto: Japanese Society for Music Perception and Cognition, 1989, pp. 107-112.

Deutsch, D. A link between music perception and speech production. *Journal of the Acoustical Society of America,* 1990, ***88***, Suppl. 1, 139.

2. This work was first reported in abstract form in Deutsch (1990). The work was supported by grants from the Digital Equipment Corporation and the UCSD Biomedical Research Support Group. The author is grateful to Deborah Faast for help in collecting and tabulating the data.

Deutsch, D. The tritone paradox: Implications for the representation and communication of pitch structures. In M. R. Jones & S. Holleran (Eds.), *Cognitive bases of musical communication.* American Psychological Monograph Series. Washington, DC: American Psychological Association (in press).

Deutsch, D. The tritone paradox in relation to patterns of relative loudness for the components of the tones (in preparation).

Deutsch, D., Kuyper, W. L., & Fisher, Y. The tritone paradox: Its presence and form of distribution in a general population. *Music Perception,* 1987, ***5***, 79-92.

Deutsch, D., Moore, F. R., & Dolson, M. Pitch classes differ with respect to height. *Music Perception,* 1984, ***2,*** 265-271.

Deutsch, D., Moore, F. R., & Dolson, M. The perceived height of octave-related complexes. *Journal of the Acoustical Society of America,* 1986, ***80***, 1346-1353.

Deutsch, D., North, T., & Ray, L. The tritone paradox: Correlate with the listener's vocalrange for speech. *Music Perception,* 1990, ***7,*** 371-384.

Terhardt, E. Pitch, consonance, and harmony. *Journal of the Acoustical Society of America,* 1974, ***55,*** 1061-1069.

Terhardt, E. Music perception and sensory information acquisition: Relationships and low-level analogies. *Music Perception,* 1991, ***8,*** 217-40.

Terhardt, E., & Seewann, M. Aural key identification and its relationship to absolute pitch. *Music Perception,* 1983, ***1,*** 63-83.

Terhardt, E., & Ward, W. D. Recognition of musical key: exploratory study. *Journal of the Acoustical Society of America,* 1982, ***72,*** 26-33.

Whitfield, I. C. *The auditory pathway.* London: Arnold, 1967.

Whitfield, I. C. Central nervous processing in relation to spatio-temporal discrimination of auditory patterns. In R. Plomp & G. F. Smoorenberg (Eds.), *Frequency analysis and periodicity detection in hearing.* Leiden: Sijthoff, 1970, pp. 136-146.