Some new pitch paradoxes and their implications

DIANA DEUTSCH

Department of Psychology, University of California, San Diego, La Jolla, California 92093, U.S.A.

SUMMARY

This paper explores two new paradoxical sound patterns. The tones used to produce these patterns consist of six octave-related harmonics, whose amplitudes are scaled by a bell-shaped spectral envelope; these tones are clearly defined in terms of pitch class (C, C#, D, and so on) but are poorly defined in terms of height. One pattern consists of two tones that are separated by a half-octave. It is heard as ascending when played in one key, yet as descending when played in a different key. Further, when the pattern is played in any one key it is heard as ascending by some listeners but as descending by others (the tritone paradox). Another pattern that consists of simultaneous pairs of tones displays related properties (the semitone paradox). It is shown that the way the tritone paradox is perceived correlates with the speech characteristics of the listener, including his or her linguistic dialect. The findings suggest that the same, culturally acquired representation of pitch classes influences both speech production and also perception of this musical pattern.

1. INTRODUCTION: PITCH CLASS AND PITCH HEIGHT

Pitch, as defined by the American National Standards Institute (1973) is 'that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from high to low'. Indeed, most experimental research on the subject has assumed such a definition. However, it has long been recognized by musicians that pitch is not a unidimensional attribute; rather it varies along two dimensions: the monotonic dimension of pitch height and the circular dimension of pitch class (Westergaard 1975). Tones that are related by octaves; i.e. whose fundamental frequencies stand in the ratio of 2:1, are in some sense perceptually equivalent. Indeed, the system of notation for the traditional musical scale assumes such equivalence. The core of this scale consists of twelve tones, which are formed by the division of the octave into equal logarithmic steps, called semitones. Each tone is assigned a name (C, C#, D, and so on), and the entire scale is produced by repeatedly traversing the circle of note names across successive octaves.

Assuming, then, that pitch varies along two dimensions, the question arises as to whether these dimensions are orthogonal, or whether they interact in some fashion. Shepard (1964) performed an experiment which addressed this issue. He generated a series of tones, each of which consisted of ten harmonic components which were separated by octaves. The amplitudes of these components were scaled by a fixed, bell-shaped spectral envelope, and the pitch classes of the tones were varied by shifting the components up and down in log frequency, keeping the position and shape of the envelope constant. Subjects were presented with ordered pairs of such tones, and they reported in each case whether they heard an ascending or a descending pattern. When the tones within a pair were separated by one or two steps along the pitch class circle, judgments were almost entirely determined by proximity. For example, the pattern D-D# was always heard as ascending, and the pattern C#-C was always heard as descending. As the distance between the tones along the pitch class circle increased, the tendency for judgments to be determined by proximity lessened, and when the tones were separated by exactly a half-octave, ascending and descending judgments occurred equally often.

Shepard concluded from these findings that the dimensions of pitch class and pitch height are indeed orthogonal. However, this conclusion can be questioned on two grounds. First, since judgments in this study were largely determined by proximity, any influence of pitch class on perceived height could have been overwhelmed by this factor (Deutsch 1982). Second, the judgments were presented as averaged both over subjects and also over pitch classes, so that any influence of pitch class on perceived height would have been lost in the averaging process.

The foregoing considerations led the present author to embark on a series of experiments to determine whether an influence of pitch class on perceived height would be manifest for patterns in which proximity could not be used as a cue. The patterns used were composed of octave-related complexes similar to those employed by Shepard, and the data were analysed for each subject separately as a function of the pitch classes of the tones comprising each pattern. Using this procedure, the perceived heights of the tones were found to vary in an orderly fashion as a function of their positions along the pitch class circle, so that tones in one region of the circle were heard as higher and tones in the opposite region were heard as lower. Furthermore, the direction of this relationship varied substantially across subjects, so that any given pattern was heard by different listeners in strikingly different ways.

2. EQUIPMENT AND TONE COMPLEXES

Tones were generated on a VAX 11/780 computer, interfaced with a DSC-200 Audio Data Conversion System (16 bit, 48K sampling rate). They were recorded and played back on a Sony PCM-F1 Digital Audio Processor, and the output was passed through a Crown amplifier and delivered to subjects binaurally through headphones (Grason-Stadler TDH-49) at a level of approximately 72 dB SPL.

Each tone consisted of six sinusoids which were separated by octaves, the amplitudes of which were determined by a fixed, bell-shaped spectral envelope. The following is the general form of the equation describing the envelope:

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

$$A(f) = 0 \text{ elsewhere}$$

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 non-zero. The maximum frequency for which the amplitude is non-zero is thus $\gamma\beta$ cycles above $f_{\min}$. The values $\beta = 2$ and $\gamma = 6$ were used throughout, so that the spectral envelope always spanned exactly six octaves, from $f_{\min}$ to $64\,f_{\min}$.

3. THE TRITONE PARADOX

One-pattern that was explored consisted of an ordered pair of tones which were separated by a half-octave, or tritone (Deutsch 1986). Because the tones were diametrically opposed along the pitch class circle, proximity could not here be used as a cue in determining their relative heights. On each trial, subjects were presented with one such tone pair, and they judged whether it formed an ascending or a descending pattern; from these judgments it was inferred which tones were heard as higher and which as lower.

All tones were 500 ms in duration, and there were no gaps between tones within a pair. To control for possible effects based on the relative amplitudes of the harmonic components, and also to examine the effects of variations in overall height, the tone pairs were generated under envelopes that were placed at six different positions along the spectrum. The envelope peaks were spaced at half-octave intervals, and were at C_6 (1047 Hz), $F\#_5$ (740 Hz), C_5 (523 Hz), $F\#_4$ (370 Hz), C_4 (262 Hz), and $F\#_3$ (185 Hz). All twelve pitch-class pairings (C-F#, C#-G, D-G#, . . ., B-F) were presented equally often under each of the six positions of the spectral envelope.

Figure 1 shows the results from two subjects, in each

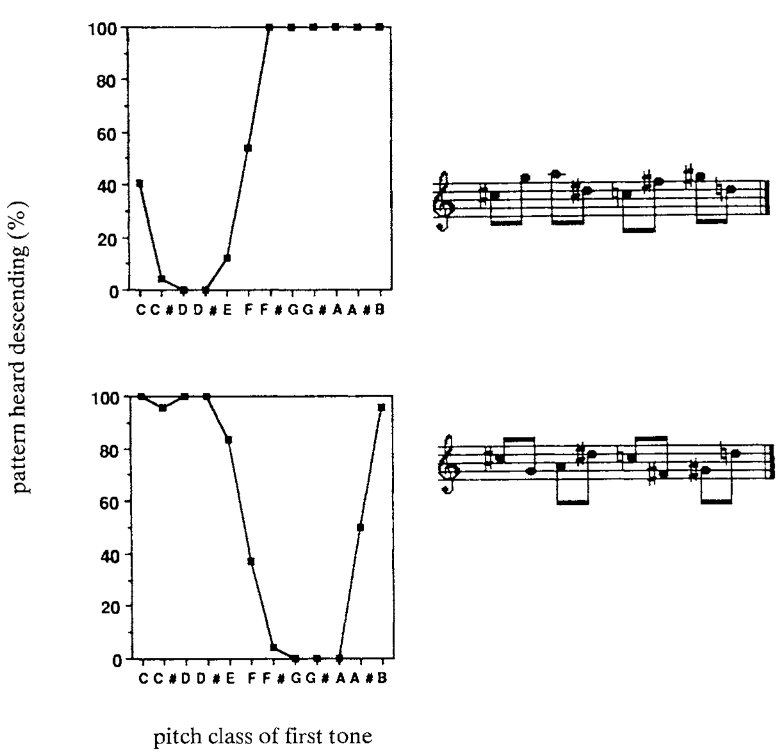

Figure 1. The tritone paradox. Graphs on left show percentages of judgments that a tone pair formed a descending pattern, plotted as a function of the pitch class of the first tone of the pair. The results from two subjects are here displayed, avaraged over all six positions of the spectral envelope, and over two experimental sessions. Musical notations on right show how these two subjects perceived the identical series of tone pairs C#-G, A-D#, C-F#, G#-D. (Data from Deutsch (1986).)

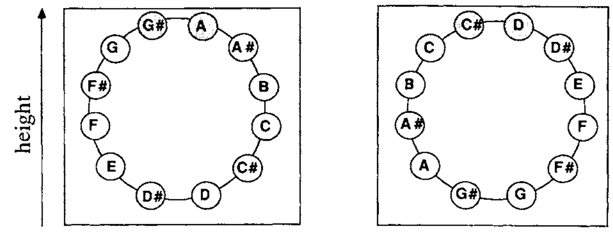

Figure 2. Perceptual orientations of the pitch class circle with respect to height, derived from the judgments of the two subjects shown in figure 1. Filled circles indicate the pitch classes that defined the highest position along the circle; these are termed peak pitch classes.

case averaged over two experimental sessions, and over all six positions of the spectral envelope. It can be seen that for both subjects, judgments depended in an orderly fashion on the positions of the tones along the pitch class circle. However, the direction of this dependence differed strikingly between the subjects: for the most part, when the first subject heard an ascending pattern the second subject heard a descending one, and vice versa. In consequence, extended passages composed of such tone pairs were heard by these two subjects as forming entirely different melodic contours. An example is given in the righthand part of figure 1.

Figure 2 depicts the two perceptual orientations of the pitch class circle with respect to height, which are derived from the judgments of two subjects shown in figure 1. For the first subject, the pitch classes that defined the highest position along the circle (hereafter referred to as peak pitch classes) were C# and D instead.

Deutsch (1987) investigated the effects of the spectral envelope position in detail. The data from four subjects were examined, and twelve envelope positions were employed: these varied in one-quarter octave steps, and so over a three octave range. It was found that in all cases judgments depended systematically on the positions of the tones along the pitch class circle, and also that the direction of this dependence varied from one subject to another. Although there were some interactions between pitch class and the relative amplitudes of the sinusoidal components, and also between pitch class and overall height, these interactions were not necessarily present. Further, even when interactions were found, their effects in absolute terms were generally quite small.

In a large-scale study (Deutsch *et al.* 1987), a group of subjects was selected on the only criteria that they were undergraduates at the University of California, San Diego, that they had normal hearing, and that they could judge reliably whether pairs of single sinusoids that were separated by a half-octave formed ascending or descending patterns. The subjects made judgments of the tritone paradox, and the results were analysed as follows. It was first determined for the scores of each subject separately whether the pitch class circle could be bisected in such a way that none of the scores in the upper half of the circle was lower than any of the scores in the lower half. This criterion was fulfilled by 22 of the 29 subjects. Next, the proportion of random permutations of the scores that could be so characterized was determined by computer simulation. Averged across subjects, this proportion was found to be 0.027 per subject. The probability of obtaining the combined result by chance was thus shown to be vanishingly small, and it can be concluded that the influence of pitch class on perceived height occurs to a highly significant extent in this general population.

In this study, the peak pitch classes were tabulated for each subject, and the distribution of peak pitch classes within the subject population was determined. An orderly bell-shaped distribution was obtained, so that for the most part pitch classes B, C, C#, D, and D# were heard as higher, and pitch classes F, F#, G, G#, and A were heard as lower. This finding will be further considered below.

4. THE SEMITONE PARADOX

We next consider what happens when more than one tone is presented at a time. To examine this, a pattern was created which consisted of two simultaneously presented pairs of tones; one pair formed a pattern that ascended by a semitone whereas the other formed a pattern that descended by a semitone (Deutsch 1988). The tone pairs were diametrically opposed along the pitch class circle, so that, again, proximity could not be used as a cue in making judgments of their relative heights. All tones were 500 ms in duration, and there were no gaps between tones within a pair. In preliminary work it was found that this pattern was generally perceived as two stepwise lines that moved in contrary motion. However, for any given instantiation of the pattern, some listeners heard the higher line as ascending and the lower line as descending, whereas other listeners heard the higher line as descending and the lower line as ascending. In the formal study, patterns were generated under envelopes which were placed at 12 different positions along the spectrum, which were spaced at one-quarter octave intervals. Under each position of the spectral envelope, each of the 12 pitch classes served equally often as the first tone of an ascending pair, and also as the first tone of a descending pair.

Four subjects were employed in this study, and it was found that, again, judgments reflected orderly relationships between the perceived heights of the tones and their positions along the pitch class circle. Also, again, the direction of the relationship between pitch class and perceived height varied considerably across subjects. This is exemplified in the judgements of two subjects shown in figure 3. It can be seen that as the pattern was transposed, the ascending and the descending lines appeared to interchange positions. However, for the most part, when the first subject heard the higher line as ascending the second subject heard it as descending, and vice versa. Thus extended passages composed of such patterns were perceived by these two subjects in quite different ways. This is exemplified in the passage shown on the righthand part of figure 3. Concerning the effects of envelope position, it was found that, as in the study of Deutsch (1987), there were some interactions between pitch class and the relative amplitudes of the sinusoidal

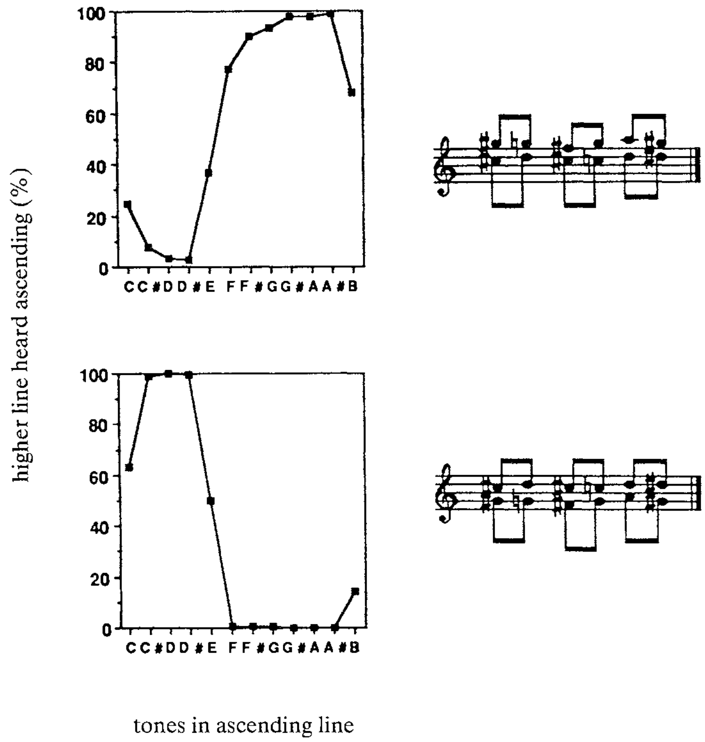

Figure 3. The semitone paradox. Graphs on left display the percentages of judgments that the higher line formed an ascending pattern, plotted as a function of the pitch classes of the tones in the ascending line. The results from two subjects are here displayed, averaged over twelve positions of the spectral envelope, and over nine experimental sessions. Musical notations on right show how these two subjects perceived the identical series of patterns G#-G/C#-D, F#-G/C#-C, A-G#/D-D#. (Data from Deutsch (1988).)

components, and there were also some interactions between pitch class and overall height. However, these interactions were not necessarily present, and even when they were, their effects in absolute terms were generally quite small.

5. BASIS OF THE TRITONE PARADOX

As described earlier, Deutsch *et al.* (1987) found that the tritone paradox occurred in the large majority of subjects in a sizeable population, showing that this phenomenon is not confined to a few selected individuals. Within this population, no correlate with musical training was found, either in terms of the probability of obtaining the phenomenon, or its size, or its direction. A number of other studies have also ruled out explanations in terms of low-level properties of the hearing mechanism. For many subjects, the profiles relating pitch class to perceived height were largely unchanged when the position of the spectral envelope was shifted over a three octave range (as, for example, in the study of Deutsch (1987)). In a recent experiment, the tritone paradox was produced when the odd-numbered components of each tone complex were presented to one ear and the even-numbered components were presented to the other ear. However, the effect was not produced when either the odd-numbered components or the even-numbered components were presented alone to both ears. Since the phenomenon can result from integration of information presented separately to the two ears, then it follows that it must be central in origin (Deutsch 1992*a*)†.

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 hypothesized that through long-term exposure to such sounds, the listener acquires a representation of his or her own 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 hypothesized that the pitch classes delimiting this octave band for speech are taken as defining the highest position along the pitch class circle (i.e. the position bounded by the peak pitch classes, as sown in figure 2), and that this in turn determines the perceived orientation of the pitch class circle with respect to height‡.

† Space does not permit a discussion of the reasons why the tritone paradox did not occur with tone complexes whose components were spaced at two-octave intervals. One possibility is that the subjects were here using spectral proximity in making their judgments. However, the explanation of this negative result is irrelevant to the point made here, that the tritone paradox can be produced by central combination of information presented to each ear separately.

‡ It may be noted that the spectral compositions of the tones used here to produce the tritone paradox differed substantially from those of speech sounds. However, the author recently performed a study using pairs of single tones, each of which comprised a full harmonic series, but in which the relative amplitudes of the odd and even harmonics were such as to produce ambiguities of perceived height. At least for some subjects, the tritone paradox was still obtained, and the judgments made were very similar to those made by the same subjects with the use of octave-related complexes (Deutsch 1992*b*). The conditions giving rise to the tritone paradox may therefore be rather general, provided that the presented sounds are such as to produce ambiguities of perceived height.

Table 1. *Pitch classes delimiting the octave band for speech, together with those defining highest position along the pitch class circle, tabulated by subject (from Deutsch et al. 1990)*

subject	limit of octave band for speech	highest position along pitch class circle	distance in semitones
A.H.	D#-E	D-D#	1
D.M.	D-D#	D#-E	1
D.D.	F-F#	G-G#	2
T.T.	D#-E	D-D#	1
M.D.	E-F	G#-A	4
M.C.	A#-B	C-C#	2
M.M.	D#-E	D-D#	1
E.S.	C-C#	C-C#	0
W.B.	D#-E	D#-E	0

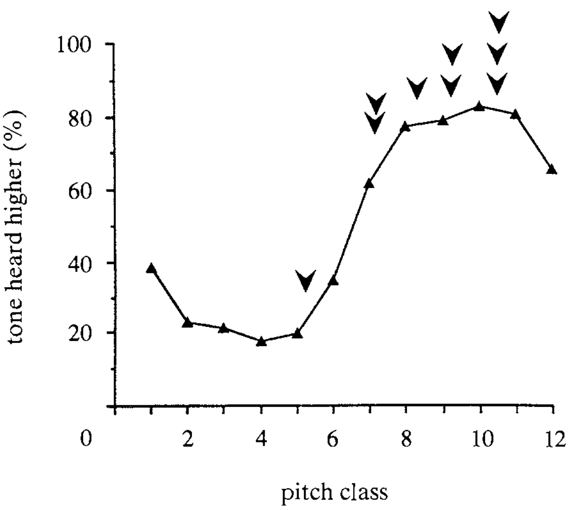

Figure 4. Percentages of trials in which a tone was heard as the higher of a pair in making judgments of the tritone paradox, with the orientation of the pitch class circle normalized and averaged across subjects. Arrows indicate the limit of each subject's octave band for speech, in relation to the highest position along the pitch class circle, as determined by judgments of the tritone paradox. (From Deutsch *et al.* (1990).)

Deutsch *et al.* (1990) performed an experiment to test this hypothesis. Nine subjects (four male and five female) were selected who showed clear and consistent 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 the speech samples were recorded into computer memory. F_0 estimates were then obtained from these speech samples at 4 ms intervals. These estimates were allocated to semitone bins, and histograms were derived of the percentage occurrence of F_0 estimates in each bin. From each histogram, the octave band containing the largest number of F_0 estimates was determined. Finally, for each subject, comparison was made 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. As shown in table I, for eight of the nine subjects, the values were separated by no more than two semitones, so that a significant correspondence between these values was established ($p=0.04$, two-tailed, on a binomial test). Figure 4 displays a further representation of these findings, in which the limit of each subject's octave band for speech is plotted on a curve showing the judgments of the tritone paradox, normalized and averaged across all subjects§.

The findings of Deutsch *et al.* (1990) are in accordance with the above hypothesis, which we may then consider in two versions. The first version makes no assumption that the listener's pitch range for speech is itself determined by a learned template. The second version assumes that such a template is acquired through exposure to speech produced by others, and that it is employed both to evaluate perceived speech and also to constrain the listener's own speech production. On this second hypothesis, the characteristics of such a template would be expected to be similar for people who speak in the same language or dialect, but to vary for people who speak in different languages or dialects.

The second hypothesis was tested in an experiment by Deutsch (1991). Two groups of subjects made judgements of the tritone paradox. The first group consisted of 24 subjects who had grown up in California, and the second consised of 12 subjects who had grown up in the South of England. None of the subjects in the Californian group had a parent who had grown up in England, and none of the subjects in the English group had a parent who had grown up in California.

For each subject, 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 distribution of peak pitch classes was then determined for the Californian and English groups separately. These distributions are shown in figure 5, and it can be seen that they are strikingly different: for the Californian group, B, C, C#, D, and D# occurred most often as peak pitch classes, but for the English group, F#, G, and G# occurred most often instead.

To evaluate the statistical significance of the difference between the two groups, the hypothesis was tested that the distribution produced by the Californian group would be similar to that obtained earlier in the study of Deutsch *et al.* (1987) on undergraduates at the University of California, but that the distribution produced by the English group would be a different one. Comparison was therefore made between the number of subjects in each group for whom the peak position along the pitch class circle lay in the half of the circle containing the larger number of peak positions in the study of Deutsch *et al.* (1987). It was found that 21 of the 24 Californian subjects fell

§ Because the correlate here was between pitch classes rather than frequency values, the statistical significance of this correlate did not depend on the hypothesis that the listener focuses on the upper limit of the octave band for speech in generating a pitch class template. However, it appears intuitively reasonable that the listener would indeed focus on the upper end rather than the lower one, as it is here that speech sounds are most salient. It should further be noted that the choice of an octave band follows logically from the findings on the tritone paradox, since this paradox is a pitch class phenomenon. The peak of the F_0 distribution in the speech histogram is not an appropriate measure here, as it lies somewhere in the middle of the distribution.

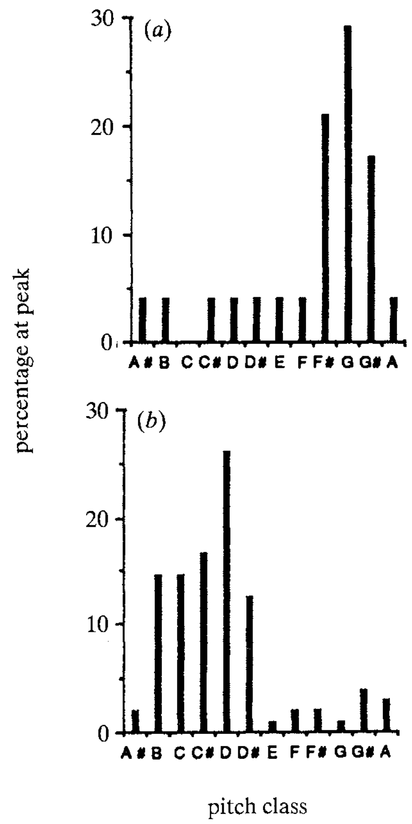

Figure 5. Distribution of peak pitch classes (see figure 2) as determined by judgments of the tritone paradox, in two groups of subjects. The first group (*a*) had grown up in the South of England, and the second (*b*) had grown up in California. (From Deutsch (1991).)

into this category; however, only three of the 12 English subjects did so. The difference between the two groups on this measure was therefore highly significant ($p < 0.001$ on a Fisher exact probability test). These findings are therefore in accordance with the hypothesis of a culturally acquired representation of the pitch class circle which is specific to the language or dialect of the listener.

What can be the evolutionary value of such a template? As one line of reasoning, it could be advantageous to determine the emotional state of a speaker through the pitch of his or her voice. A template such as hypothesized here could provide a framework, common to a particular dialect, within which the pitch of a speaker's voice may be evaluated, and this would in turn provide evidence concerning his or her emotional state. It may be observed that a template that is based on pitch class rather than pitch height has the useful feature that it can be invoked for both male and female speakers, even though their speech involves different pitch ranges. Such a template could also be invoked in communicating syntactic aspects of speech||.

|| It is interesting to note that, in retrospect, subjects who had shown large differences in the orientation of the pitch class circle with respect to height in making judgments of the tritone paradox had frequently come from different linguistic backgrounds. For example, the subject whose data are shown in the upper portion of figure 1 had grown up in London, and the subject whose data are shown in the lower portion of this figure had grown up in California.

The findings of Deutsch *et al.* (1990) and Deutsch (1991) taken together lead to the prediction that, for a large sample of individuals in a given linguistic group, the distribution of fundamental frequencies in spontaneous speech for males and females should stand in octave relation. It should be emphasized, however, that based on the data from Californians obtained by Deutsch (1991) the spread should be quite large (almost a half-octave). Further, the correlate between the peak pitch classes derived from judgments of the tritone paradox and the pitch classes delimiting the octave band for speech found by Deutsch *et al.* (1990) was obtained on subjects who were selected for showing clear and consistent functions in making judgments of the tritone paradox. On might expect, therefore, that in an unselected sample of subjects the correlate would be less clear.

6. FURTHER IMPLICATIONS

One further implication of these findings concerns theories of pitch perception. Two factors have previously been shown to contribute to the perceived pitch of a complex tone. First, when presented with a series of harmonics the listener perceives a pitch that corresponds to the frequency of the fundamental. Second, the listener assigns weightings to the harmonics on the basis of their relative amplitudes, and perceives a pitch in accordance with these weightings. The first factor is held to predominate for tones with fundamentals below roughly 900 Hz, and the second factor for tones with fundamentals above this value (Goldstein 1973; Scharf & Houtsma 1986; Terhardt *et al.* 1982). The present findings show that, at least under certain circumstances, a third factor operates also: the perceived height of a tone can be influenced by its position along the pitch class circle.

Another implication of the findings concerns musical transposition. It is generally taken as self-evident that when a passage is transposed from one key to another, the perceived relationships between the pitches are unchanged. The patterns explored here are striking cases in which this principle is violated: For the tritone paradox, transposition results in a perceived inversion; for the semitone paradox, transposition results in a perceived interchange of voices.

A further implication concerns the phenomenon of absolute pitch, which is characterized as the ability to name a note that occurs in isolation. In general, this faculty is considered very rare. However, the present paradoxes demonstrate that the large majority of individuals possess a form of absolute pitch, in that we perceive tones as higher or as lower depending solely on their pitch classes, or note names.

A final implication concerns relationships between speech and music. It has been acknowledged since antiquity that these two modes of communication have characteristics in common; however the bases for such commonalities have so far been undetermined. The findings reported here provide, to the author's knowledge, the first indication that one form of communication can exert a direct influence over the other, and this opens the door to the possible uncovering of other such influences.

REFERENCES

American National Standards Institute 1973 *American national psychoacoustical terminology*. S3.20. New York: Anonymous.

Deutsch, D. 1982 Grouping mechanisms in music. In *The psychology of music* (ed. D. Deutsch), (pp. 99–134). New York: Academic Press.

Deutsch, D. 1986 A musical paradox. *Mus. Percept.* **3**, 275–280.

Deutsch, D. 1987 The tritone paradox: effects of spectral variables. *Percept. Psychophys.* **41**, 563–575.

Deutsch, D. 1988 The semitone paradox. *Mus. Percept.* **6**, 115–132.

Deutsch, D. 1991 The tritone paradox: an influence of language on music perception. *Mus. Percept.* **8**, 335–347.

Deutsch, D. 1992*a* The tritone paradox and central integration. (In preparation.)

Deutsch, D. 1992*b* The tritone paradox produced by full harmonic series. (In preparation.)

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

Deutsch, D., North, T. & Ray, L. 1990 The tritone paradox: Correlate with the listener's vocal range for speech. *Mus. Percept.* **7**, 371–384.

Goldstein, J.L. 1973 An optimum processor theory for the central formation of the pitch of complex tones. *J. acoust. Soc. Am.* **54**, 1496–1516.

Scharf, B. & Houtsma, A.J.M. 1986 Audition II: Loudness, pitch, localization, aural distortion, pathology. In *Handbook of perception and human performance*, vol II, 15 (ed. K. Boff, L. Kaufmann & J. Thomas), pp. 1–60. New York: Wiley.

Shepard, R.N. 1964 Circularity in judgments of relative pitch. *J. acoust. Soc. Am.* **36**, 2345–2353.

Terhardt, E., Stoll, G. & Seewann, M. 1982 Pitch of complex signals according to virtual pitch theory: tests, examples, and predictions. *J. acoust. Soc. Am.* **71**, 671–678.

Westergaard, P. 1975 *An introduction to tonal theory*. New York: Norton.