Music Perception, 1990, *7,* 371-384.

The Tritone Paradox: Correlate with the Listener's Vocal Range for Speech

DIANA DEUTSCH, TOM NORTH, & LEE RAY
University of California, San Diego

In the tritone paradox, two tones are presented that are related by a half-octave. Each tone consists of a set of octave-related sinusoids whose amplitudes are scaled by a bell-shaped spectral envelope; thus the usual cues to height attribution are missing. When listeners judge whether such tone pairs form ascending or descending patterns, judgments are related in an orderly fashion 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 as lower. However, listeners differ strikingly in the orientation of the pitch-class circle with respect to height.

So far, the basis of the tritone paradox and the reasons for the individual differences in its manifestation have proved elusive. In the present study, a correlation is found between perception of the tritone paradox and the range of fundamental frequencies of the listener's speaking voice. To the authors' knowledge, this is the first demonstration of a close connection between the perception of a musical pattern on the one hand and the listener's speech characteristics on the other.

THERE is abundant evidence that we encode pitch information, both along a monotonic dimension of height and along a circular dimension of pitch class. Tones standing in octave relation are represented on the same point along the circular dimension, so that all tones in pitch class C are in a sense perceptually equivalent, as are all tones in pitch class C#, and so on (Babbitt, 1960, 1965; Bachem, 1948; Charbonneau & Risset, 1973; Demany & Armand, 1984; Deutsch, 1969, 1973, 1982; Meyer, 1904; Nakajima, Tsumura, Matsuura, Minami & Teranishi, 1988; Ohgushi, 1985; Revesz, 1913; Risset, 1971; Ruckmick, 1929; Shepard, 1964, 1965, 1982; Ueda & Ohgushi, 1987; Ward & Burns, 1982). It has recently been shown that interactions can occur between the monotonic and circular dimensions: In the case of tones that have a clearly defined pitch class but where the cues to height attribution are impoverished, an orderly relationship may be found between the perceived height of a tone and its position along the pitch-class

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

circle (Deutsch, 1986, 1987, 1988; Deutsch, Moore, & Dolson, 1984, 1986; Deutsch, Kuyper, & Fisher, 1987).

The simplest pattern in which this relationship is manifest is known as the tritone paradox (Deutsch, 1986, 1987, in preparation; Deutsch et al., 1987). Two tones are presented in succession 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 octave-related sinusoids, which are generated under a bell-shaped spectral envelope. A representation of one such tone pair is given in Figure 1. When listeners are asked to judge whether such tone pairs form ascending or descending patterns, in general their judgments show an orderly relation 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 as lower. This phenomenon is quite counterintuitive, because it violates the principle of equivalence under transposition, a principle that is generally considered to hold universally.

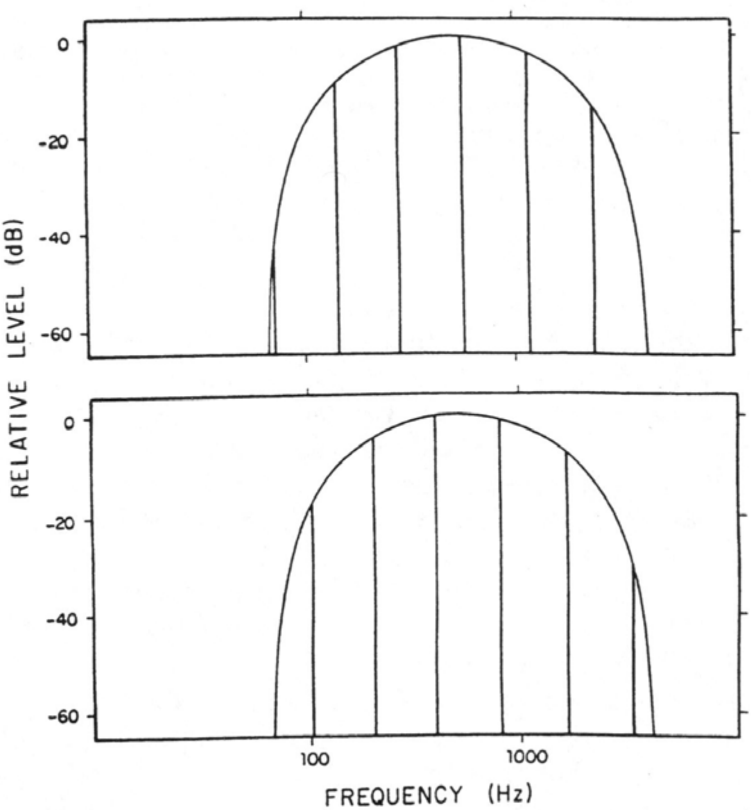

Fig. 1. Spectral composition of a tone pair producing the tritone paradox. Here the spectral envelope is centered at C_5. The upper graph represents a tone of pitch class D, and the lower graph represents a tone of pitch class G#.

Another remarkable aspect of this phenomenon is that the orientation of the pitch-class circle differs radically across listeners. For example, one listener may hear the C-F# pattern as ascending and the F#-C pattern as descending; reflecting the perception of F# as higher and C as lower. However, another listener may hear the C-F# pattern as *descending* and the F#-C pattern as *ascending;* reflecting the perception of C as higher and F# as lower instead.

The tritone paradox has been shown not to be attributable to patterns of relative salience for the different sinusoidal components of the tones. For many listeners, the function relating pitch class to perceived height is largely unaltered when the position of the spectral envelope is shifted over a three-octave range (Deutsch, 1987). Furthermore, these functions have been shown to be unrelated to patterns of relative loudness for the sinusoidal components when these are presented individually (Deutsch, in preparation). We can conclude, therefore, that the phenomenon cannot be explained in terms of low-level characteristics of the hearing mechanism.

In considering possible explanations, we may note that the tritone paradox has been found to occur in the large majority of listeners in a sizeable population, without any apparent relationship to musical training or experience (Deutsch et al., 1987). It appears reasonable to conjecture, therefore, that the phenomenon is extramusical in origin.

The present study was undertaken to test the hypothesis that the tritone paradox may be related to the processing of speech sounds. In particular, it was hypothesized that as a result of extensive exposure, the listener develops a long-term representation of the range of fundamental frequencies of his or her speaking voice. This representation includes a delimitation of the octave band in which the largest proportion of fundamental frequencies occurs, averaged over the long term. The pitch classes delimiting this speech band are then taken as defining the highest position along the pitch-class circle; this in turn determines the orientation of the pitch-class circle with respect to height.

For example, we can consider a subject for whom the upper limit of the octave band for speech is between C_4 and $C\#_4$. One would predict that, in this case, the orientation of her pitch-class circle would be such that the highest pitch classes would be close to C and C#. Similarly, we can take a subject for whom the upper limit of this speech band is between $D\#_3$ and E_3. One would predict that, for this subject, the orientation of the pitch-class circle would be such that the highest pitch classes would be close to D# and E.

In order to test this hypothesis, subjects were selected who showed clear and unambiguous relationships between pitch class and perceived height in judgments of the tritone pattern. In Experiment 1, this relationship was documented for each subject. In Experiment 2, a recording of speech was obtained from each subject, and from this recording the pitch classes delimiting the speech band were determined. Finally, for each subject, the limit of the speech band was compared with the highest position along the pitch-class circle as determined by judgments of the tritone pattern. In confirmation of the hypothesis, a significant correspondence between these two positions along the pitch-class circle was obtained.

Experiment 1

Experiment 1 was performed to determine, for each subject, the relationship between pitch class and perceived height, as reflected in judgments of the tritone paradox.

METHOD

Stimulus Patterns

All tones consisted of six sinusoidal components that were related by octaves and whose amplitudes were determined by a fixed, bell-shaped spectral envelope[1] (Figure 1). In order to control for possible effects based on the relative amplitudes or loudnesses of the components, tone pairs were created under envelopes that were placed at four different positions along the spectrum. These were centered at 262 Hz (C_4), 370 Hz ($F\#_4$), 523 Hz (C_5), and 740 Hz ($F\#_5$). All tones were 500 msec in duration, with no gaps between tones within a pair. The tones were of equal amplitude.

Twelve pairs of tones were generated under each of the four 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 thus 48 tone pairs altogether. These were presented in blocks of 12, with each block consisting of tones generated under one of the 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 orders. Sixteen blocks were thus created, with each of the four orderings occurring once for each of the four positions of the spectral envelope.

Procedure

Testing was performed 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 blocks were separated by

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, β = 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 γβ cycles above $f_{\min}$. Throughout, the values β = 2 and γ = 6 were used, so that the spectral envelope always spanned exactly six octaves, from $f_{\min}$ to $64f_{\min}$.

1-min pauses. Each subject served in two sessions, with all 16 blocks presented in each session. The results over the two sessions were averaged. A few practice trials were given at the beginning of each session.

Apparatus

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 delivered to subjects binaurally through headphones (Grason-Stadler TDH-49) at a level of approximately 72 dB SPL.

Subjects

Subjects were selected who had normal hearing as determined by audiometry and who showed pronounced and consistent functions relating pitch class to perceived height for the tritone pattern. Another criterion guiding the selection of subjects was that there should be a substantial variation across subjects in terms of the orientation of the pitch-class circle with respect to height. Ten subjects were originally selected, but one was dropped because he developed a respiratory infection and was unavailable for further scheduling. There remained five female subjects (AH, DM, DD, TT, and MD) and four male subjects (MC, MM, ES, and WB). Eight subjects were students at UCSD and were paid for their services; the first author (DD) also served as subject.

RESULTS

The percentages of judgments that a tone pair formed a descending pattern were tabulated as a function of the pitch class of the first tone of the pair. From these data were derived the percentages of trials on which each pitch class was heard as higher. Figures 2 and 3 display the results for each subject separately, averaged over all experimental conditions. It can be seen that all subjects showed highly orderly relationships between pitch class and perceived height; indeed, in each case the pitch-class circle could be bisected in such a way that all six pitch classes in one half of the circle were heard as higher, and all six pitch classes in the other half were heard as lower.

Large differences were, however, found across subjects in the form of the relationship between pitch class and perceived height (as in previous experiments on the tritone paradox). For subjects DM and WB, the pitch classes from C# to F# were heard as higher and those from G to C as lower, so that the highest position along the pitch-class circle for these subjects lay between D# and E (Figure 4). In contrast, for subject DD, the pitch classes from F to A# were heard as higher and those from B to E as lower, so that the highest position along the pitch-class circle for this subject lay between G and G# (Figure 4). For subjects ES and MC, the pitch classes from A# to D# were heard as higher and those from E to A as lower, so that the highest

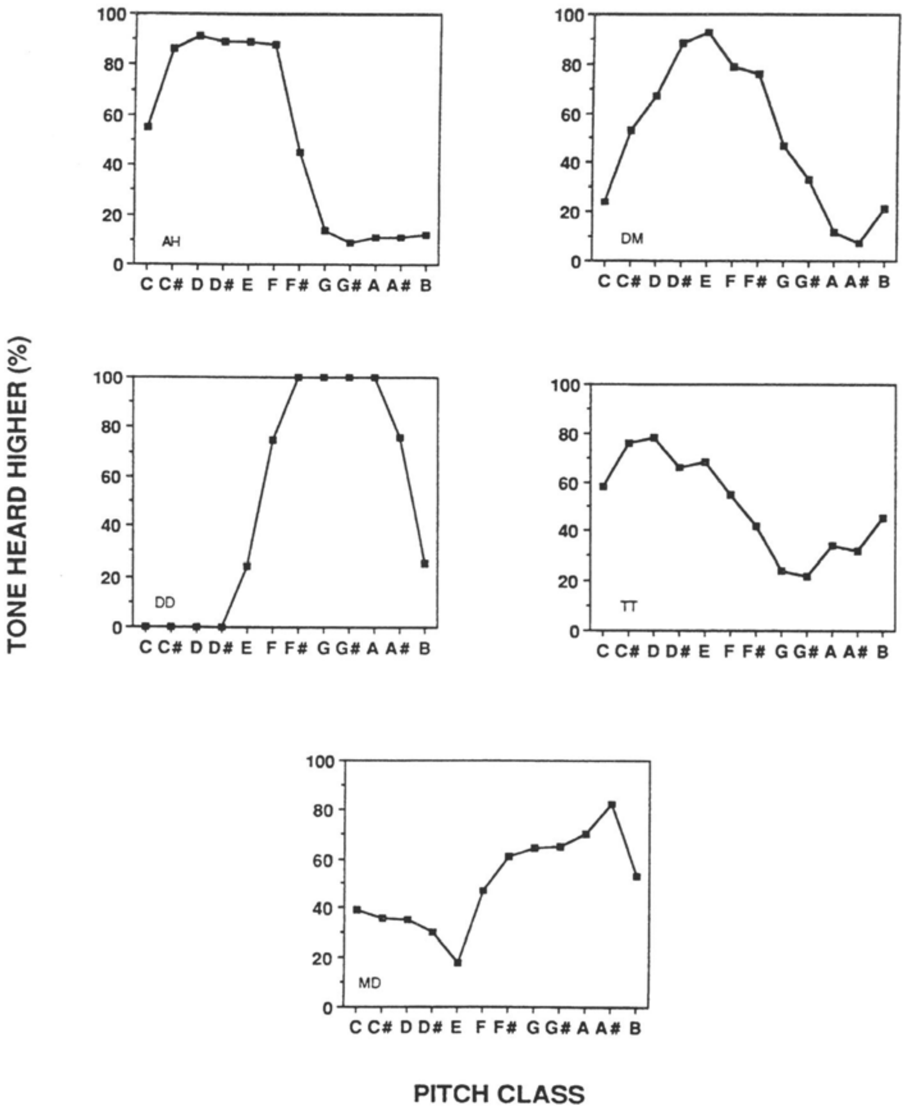

Fig. 2. Percentages of trials on which a tone was heard as the higher of a pair, plotted as a function of the pitch class of the tone. The results from the five female subjects are displayed here, averaged over all experimental conditions, and over two sessions.

position along the pitch-class circle lay between C and C#. For subjects AH, MM and TT, the pitch classes from C to F were heard as higher and those from F# to B as lower, so that the highest position along the pitch-class circle lay between D and D#. For subject MD, the pitch classes from F# to B were heard as higher and those from C to F as lower, so that the highest position along the pitch-class circle lay between G# and A.

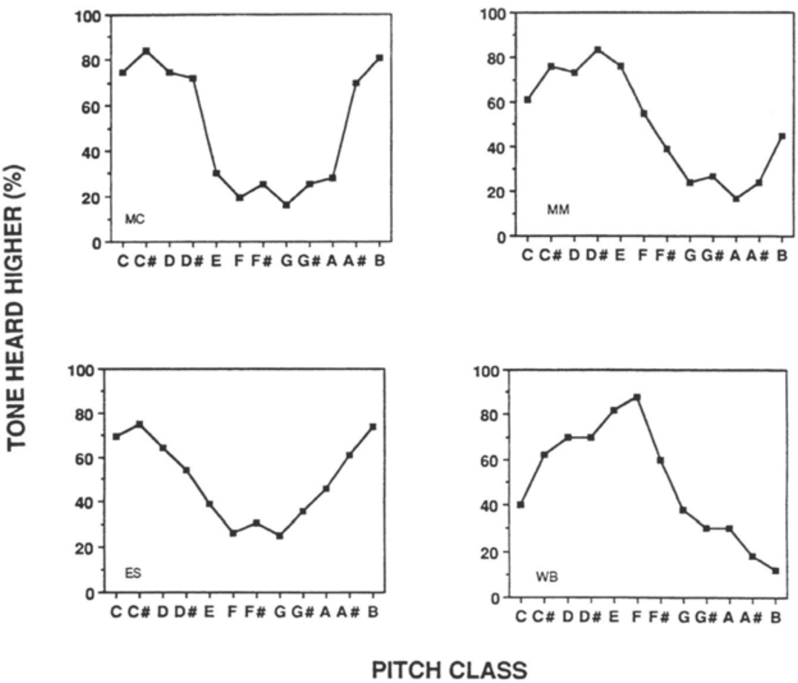

Fig. 3. Percentages of trials on which a tone was heard as the higher of a pair, plotted as a function of the pitch class of the tone. The results from the four male subjects are displayed here, averaged over all experimental conditions, and over two sessions.

Experiment 2

Experiment 2 was undertaken to determine, for each subject, the octave band containing the largest proportion of fundamental frequency (F_0) values in his or her spontaneous speech. The pitch classes delimiting this octave band were then compared with those defining the highest position along the pitch-class circle, as reflected in judgments of the tritone pattern.

METHOD

Procedure

The subjects were individually tested. They were seated before a microphone in a quiet room and instructed to speak freely on any topics of their choosing. Speech from each subject was recorded for roughly 15 min. The same experimenter interviewed all subjects, and spoke as little as possible during the interviews. Topics included the weather, traffic conditions on the way to campus, courses the subjects were currently taking, and their career plans. The subjects all performed remarkably well in this situation, and spoke with ease and fluency.

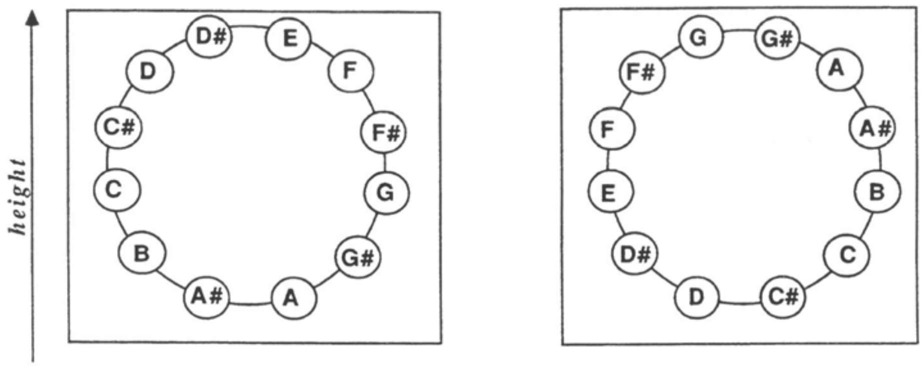

Fig. 4. Two orientations of the pitch-class circle with respect to height, derived from judgments of the tritone pattern (see Figures 2 and 3). The orientation for subjects DM and WB is shown on the left, and the orientation for subject DD on the right.

Apparatus

Speech was recorded via a Nakamichi microphone onto a Sony PCM-F1 digital audio processor. The speech samples were passed through a 16-bit analog-to-digital converter (DSC-200) and then stored and analyzed in a MicroVAX II computer.

Subjects

The same subjects participated as in Experiment 1, and they were paid for their services.

Anaysis Procedure

The digitized samples were recorded into computer memory at a sampling rate of 50 kHz.[2] The sound files were then converted to a sampling rate of 16 kHz, and band-pass filtered; the low and high cutoff frequencies were 50 Hz and 1300 Hz for the female subjects and 20 Hz and 650 Hz for the male subjects. F_0 estimates were then obtained at 256 estimates per second by using the parallel processing scheme of Rabiner and Shafer (1978).[3] The low and high boundaries for the F_0 estimates were set at 107 Hz and 639 Hz for the female subjects and at 53.5 Hz and 319.5 Hz for the male subjects. In addition, for each subject's recording, the time-varying energy level of the signal was obtained, and only those F_0 estimates that were associated with levels no lower than 25 dB below the peak level were saved for further analysis. This procedure was used in order to eliminate spurious F_0 estimates, such as obtained during pauses in the subject's speech.[4]

In the final stage of analysis, the F_0 estimates were allocated to semitone bins, with the center frequency of each bin determined by the equal-tempered scale. Thus the samples from each subject were placed in 31 semitone bins: For the female subjects, the center frequencies

2. The first 3 min of each subject's speech was treated as warmup and was not analyzed.

3. In this algorithim, six peak detectors operate in parallel, and a decision matrix is then used to determine the best estimate of the fundamental.

4. The cutoff frequencies for the band-pass filters were such as to ensure that there was no reduction in energy at the boundaries of the ranges set for the F_0 estimates.

of the bins ranged from A_2 (110 Hz) to D#$_5$ (622 Hz); for the male subjects the center frequencies ranged from A_1 (55 Hz) to D#$_4$ (311 Hz).

RESULTS

Figures 5 and 6 display the percentages of F_0 values obtained from each subject, plotted in semitone bins and so on a log frequency continuum. It

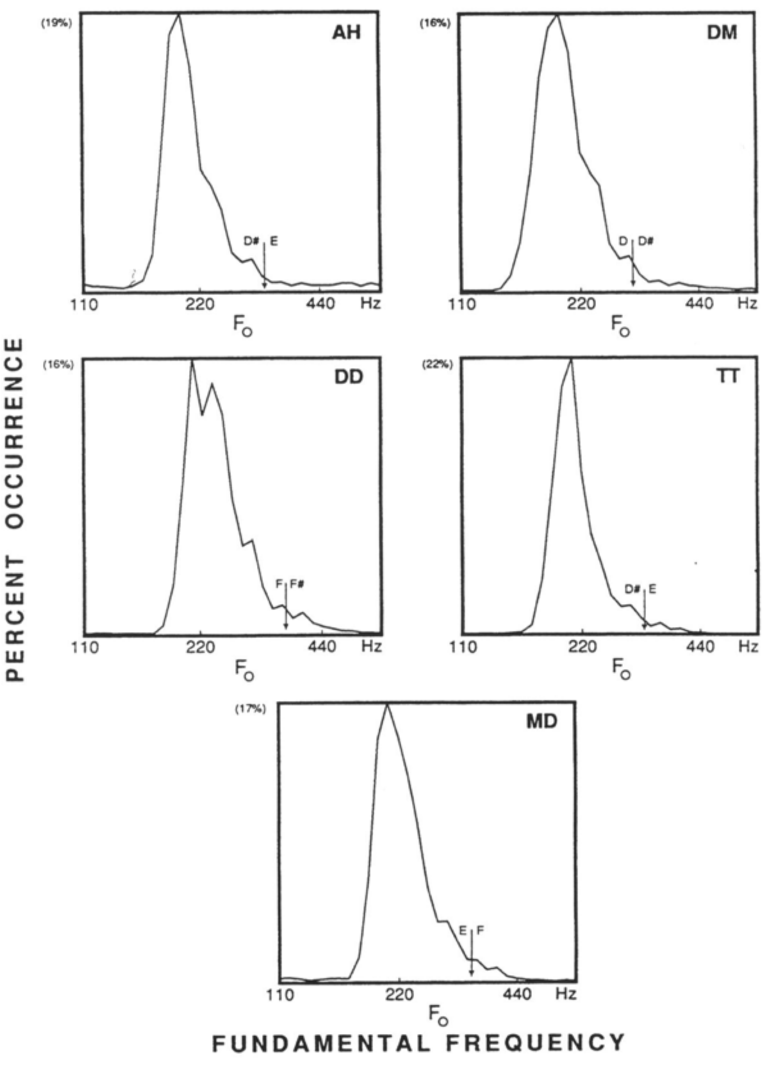

Fig. 5. Percentages of F_0 values obtained from the five female subjects, plotted in semitone bins and so on a log frequency continuum. Also displayed on each graph is the upper limit of the octave band containing the largest percentage of F_0 values, together with the pitch classes delimiting this band.

can be seen that in all cases, the number of F_0 values first increased sharply with an increase in frequency and then declined more gradually. The octave band containing the largest percentage of F_0 values was then determined for each subject. The upper limit of this band is also shown on each graph. Thus for subjects MM, WB, AH, and TT, the limit was bounded by pitch classes D# and E; for subject MC, it was bounded by A# and B; for subject ES, by C and C#; for subject DM, by D and D#; for subject DD, by F and F#; and for subject MD, by E and F.

Table 1 displays, for each subject, the pitch classes delimiting the octave band for speech, together with those defining the highest position along the pitch-class circle as determined by judgments of the tritone paradox. It can be seen that for eight of the nine subjects, the distance between these two positions was no greater than two semitones. As the largest possible unsigned distance between any two pitch classes is six semitones (a half-

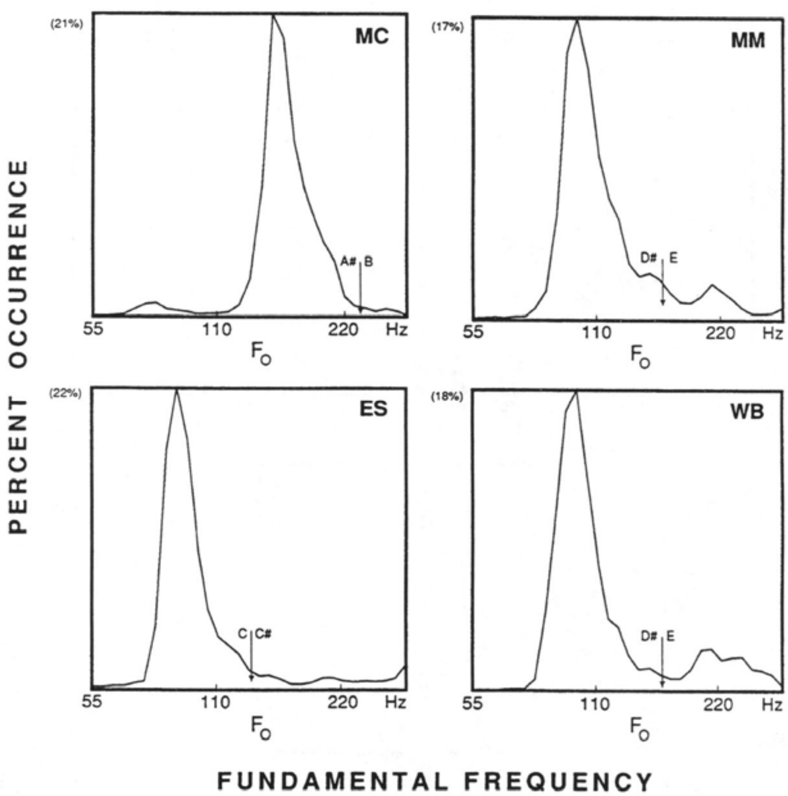

Fig. 6. Percentages of F_0 values obtained from the four male subjects, plotted in semitone bins and so on a log frequency continuum. Also displayed on each graph is the upper limit of the octave band containing the largest percentage of F_0 values, together with the pitch classes delimiting this band.

TABLE 1

Pitch Classes Delimiting Speech Band, Together with Those Defining Highest Position along the Pitch-Class Circle, Tabulated by Subject

Subject	Limit of Octave Band for Speech	Highest Position along Pitch-Class Circle	Distance in Semitones
AH	D#-E	D-D #	1
DM	D-D#	D#-E	1
DD	F-F#	G-G#	2
TT	D#-E	D-D#	1
MD	E-F	G#-A	4
MC	A#-B	C-C#	2
MM	D#-E	D-D#	1
ES	C-C#	C-C#	0
WB	D#-E	D#-E	0

octave), there were seven possible unsigned distances, ranging from 0 to 6 semitones. Since for eight of the nine subjects distances of less than three semitones were obtained, the correspondence between the two sets of values was statistically significant (p = .04, two-tailed, on a binomial test).

To obtain a further representation of the results, the orientation of the pitch-class circle was normalized across subjects,[5] and the normalized data were averaged. The limit of each subject's speech band was then plotted on the normalized curve in relation to the highest position along the pitch-class circle (Figure 7). It can be seen that the results are in accordance with the hypothesis proposed in the Introduction; namely that the orientation of the pitch-class circle for any given listener is related to the range of fundamental frequencies of his or her speaking voice.

Discussion

In this paper we have presented evidence that, at least under certain conditions, the perception of a musical pattern may vary in correlation with the listener's vocal range for speech. The most direct explanation of such a correlate is that it is acquired through learning: When presented with a series of pitch classes that are ambiguous with respect to height, we make reference to a circular template acquired through extensive exposure to speech sounds, in which the highest pitch classes are those at the upper end of our vocal range.

5. The normalization procedure was as follows. For each subject, the pitch-class circle was bisected so as to maximize the difference between the averaged scores within the two halves. The circle was then oriented so that the line of bisection was horizontal. The data were then retabulated, with the leftmost pitch class of the lower half of the circle taking the first position, its clockwise neighbor taking the second position, and so on.

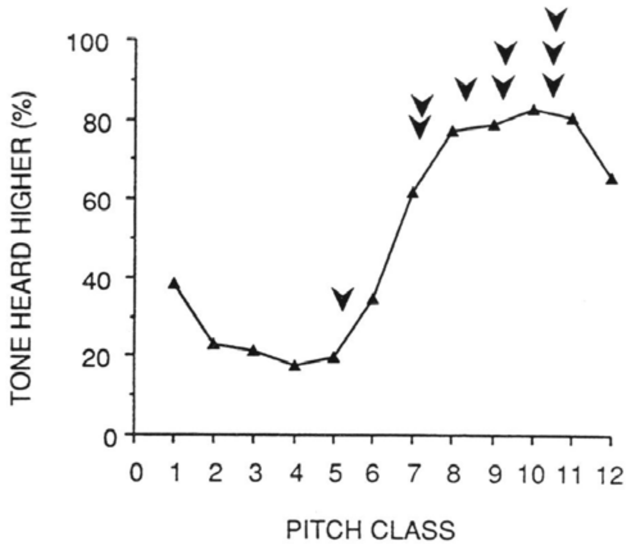

Fig. 7. Percentages of trials on which a tone was heard as the higher of a pair, with the orientation of the pitch-class circle normalized and averaged across subjects. Arrows show the limit of each subject's octave band for speech in relation to the highest position along the pitch-class circle.

The hypothesis that pitch perception may be based on a template that is acquired through experience is not new, having been suggested by several authors with reference to perception of the pitch of the fundamental from a series of harmonics. For example, Thurlow (1963) proposed that when presented with a complex tone the listener produces, either overtly or covertly, a comparison signal whose harmonics coincide with those of the complex tone presented. Whitfield (1967, 1970) made a related suggestion in proposing that patterns of neural firing resulting from combinations of harmonics are learned through extensive experience with complex sounds. When one such pattern occurs, we attribute the fundamental that we have learned is most probably associated with that pattern. Terhardt (1974) made a related argument with special reference to speech materials. He proposed that the attribution of a fundamental results from a learning process in which we associate the different components of a harmonic series, together with their fundamental (see also Terhardt, in press).

The present proposal with respect to pitch attribution for octave-related complexes is related to those just described in that it assumes that we make reference to a template acquired through extensive exposure to speech sounds. The proposal is more restricted in scope, however, because it is intended only to explain relative pitch perception for the case of tones that have the unusual property that pitch class is clearly defined but the usual cues for height attribution are missing. The question of whether such a template may be invoked under more general conditions remains to be addressed.[6]

References

Babbitt, M. Twelve-tone invariants as compositional determinants. *Musical Quarterly,* 1960, ***46***, 246-259.

Babbitt, M. The structure and function of music theory. *College Music Symposium,* 1965, ***5***, 10-21.

Bachem, A. Note on Neu's review of the literature on absolute pitch. *Psychological Bulletin,* 1948, ***45***, 161-162.

Charbonneau, G. & Risset, J. C. Circularite de jugements de hauteur sonore. *Comptes Rendus de L'Academie des Sciences, Ser. B.,* 1973, ***277***, 623.

Demany, L. & Armand, F. The perceptual reality of tone chroma in early infancy. *Journal of the Acoustical Society of America,* 1984, ***76***, 57-66.

Deutsch, D. Music recognition. *Psychological Review,* 1969, ***76***, 300-307.

Deutsch, D. Octave generalization of specific interference effects in memory for tonal pitch. *Perception and Psychophysics,* 1973, ***13***, 271-275.

Deutsch, D. The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music.* New York: Academic Press, 1982, pp. 271-316.

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. 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.

Meyer, M. On the attributes of the sensations. *Psychological Review,* 1904, ***11***, 83-103.

Nakajima, Y., Tsumura, T., Matsuura, S., Minami, H., & Teranishi, R. Dynamic pitch perception for complex tones derived from major triads. *Music Perception,* 1988, ***6***, 1-20.

Ohgushi, K. Circularity of the pitch of complex tones and its application. *Electronic Communications of Japan, Part 1,* 1985, ***68***, 1-10.

Rabiner, L. R. & Shafer, R. W. *Digital processing of speech signals.* Englewood Cliffs, NJ: Prentice Hall, 1978.

Revesz, *G. Zur grundlegun der Tonpsychologie.* Leipzig: Feit, 1913.

Risset, J. C. Paradoxes de hauteur: Le concept de hauteur sonore n'est pas le meme pour tout le monde. *Proceedings of the Seventh International Congress of Acoustics, Budapest,* 1971, p. 20, S10.

Ruckmick, C. A. A new classification of tonal qualities. *Psychological Review,* 1929, ***36***, 172-180.

6. This work was first reported in abstract form in D. Deutsch, T. North, and L. Ray (1989). Vocal range for speech determines perception of a musical pattern. *Journal of the Acoustical Society of America,* Suppl. 1 ***86***, S17.

This work was supported by grants from the Digital Equipment Corporation, the System Development Foundation, and from the UCSD Biomedical Research Support Group.

Shepard, R. N. Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America,* 1964, ***36***, 2345-2353.

Shepard, R. N. Approximation to uniform gradients of generalization by monotone transformations of scale. In D. I. Mostofsky (Ed.), *Stimulus generalization.* Stanford: Stanford University Press, 1965.

Shepard, R. N. Structural representations of musical pitch. In D. Deutsch (Ed.), *The psychology of music.* New York: Academic Press, 1982, pp. 344-390.

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* (in press).

Thurlow, W. R. Perception of low auditory pitch: A multicue mediation theory. *Psychological Review,* 1963, ***70***, 515-519.

Ueda, K. & Ohgushi, K. Perceptual components of pitch: Spatial representation using a multidimensional scaling technique. *Journal of the Acoustical Society of America,* 1987, ***82***, 1193-1200.

Ward, W. D., & Burns, E. M. Absolute pitch. In D. Deutsch (Ed.), *The psychology of music.* New York: Academic Press, 1982, pp. 431-452.

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.