Harmonic fusion and pitch shifts of mistuned partials[a)]

Alain de Cheveigné
Laboratoire de Linguistique Formelle, CNRS/Université Paris 7, 2 place Jussieu, case 7003, 75251, Paris, France and ATR Human Information Processing Research Laboratories, 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan

(Received 30 May 1996; accepted for publication 14 April 1997)

A model is proposed to explain pitch shifts of components mistuned from a harmonic series [W. M. Hartmann and S. L. Doty, ''On the pitches of the components of a complex tone,'' J. Acoust. Soc. Am. **99**, 567–578 (1996)]. An internal random variable depends on the frequency of the partial and determines both its pitch and the probability that it will fuse within the complex. The model accounts for the sign of the pitch shifts (positive for positive mistunings, negative for negative mistunings), the fact that they saturate and become smaller for large mistunings, and, given reasonable assumptions on the underlying distributions, their overall shape and magnitude. No assumptions are required concerning the pitch extraction mechanism (place, time, etc.) other than that a common source of variability affects cues to both pitch and harmonic fusion. [S0001-4966(97)07607-8]

PACS numbers: 43.66.Ba, 43.66.Hg [JWH]

INTRODUCTION

Major aspects of pitch perception are predicted equally well by different pitch perception models. For this reason, minor effects (such as pitch shifts) take on great theoretical importance, as they alone can reveal the presence of particular underlying mechanisms. Hartmann *et al.* (1986, 1990) asked subjects to match the pitch of a mistuned component of a harmonic complex to that of an isolated tone of similar frequency. The task was relatively easy for components of low rank and large mistunings, and more difficult for higher ranks and smaller mistunings. An interesting aspect of their data, elaborated by Hartmann and Doty (1996), was that mistunings were systematically overestimated by the subjects. The pitch of the mistuned component was shifted by up to several percent, in a direction that matched that of the mistuning. The shift tended to peak at about 4% mistuning and decrease beyond.

On the basis of these shifts, Hartmann and Doty (1996) argued against the model of Terhardt (1979) that predicts shifts in one direction whatever the mistuning. They proposed instead two versions of a time-domain model based on peaks of interspike interval (ISI) histograms. In the first version, the histograms were supposed to be gathered from fibers of the same characteristic frequency as the mistuned component. The predicted shifts did not match those observed. In a second version, the histograms were derived from a higher-frequency region (because of the upward spread of masking). The second version successfully accounted for major aspects of the shifts (sign, magnitude) of all components of rank greater than 1. However it could not account for shifts observed at the fundamental, nor could it account for the saturation and decrease of the pitch shift beyond 4% mistuning.

Another possible explanation is that the shifts are a side effect of harmonic fusion. Partials that match the harmonic series of a complex tone tend to fuse with it, whereas mistuned partials segregate and are easier to hear (Moore *et al.*, 1985, 1986; Hartmann, 1988; Hartmann *et al.*, 1986, 1990; Martens, 1981). When the partials of a concurrent vowel pair follow the same harmonic series, the pair is heard as s single source more often than when there is a ΔF_0, or when either vowel is inharmonic (de Cheveigné *et al.*, 1997). If, in a hypothetical variant of Hartmann and Doty's experiment, the partial's frequency had been randomly distributed, successful matches would have been less likely on trials for which the partial fell closer to the harmonic series, with the result that the distribution of pitch matches would be biased in a way similar to that observed. In Hartmann and Doty's experiment frequencies were fixed and not random, but the same reasoning may be applied to an internal representation used by both pitch and harmonic fusion and randomly distributed from trial to trial. This hypothesis is explored in the present paper.

I. MODEL

We assume that a sine-wave component gives rise to an internal correlate randomly distributed from trial to trial. This correlate determines both the probability of fusion, and the pitch perceived on trials for which the partial is not fused. It varies on a scale x about a mean x_0 that is proportional to the frequency f of the component, expressed as percentage mistuning from the harmonic series. ''Trial'' may be interpreted as meaning a pitch match, some of which were successful, others not (Hartmann and Doty, 1996). It might also mean an individual presentation of the stimulus, each of which gave rise to a slightly different pitch. Or else a ''trial'' might be an elementary frequency estimate that contributed to the pitch with a probability (or weight) depending on how close it fell to the harmonic series.

Let us denote $a(x)$ as the probability density of x, distributed around its mean x_0 [Fig. 1(A)], and $b(x)$ as the probability of harmonic fusion as a function of x [Fig. 1(B)].

[a)]Parts of this work were presented at the Spring 1997 meeting of the Acoustical Society of Japan.

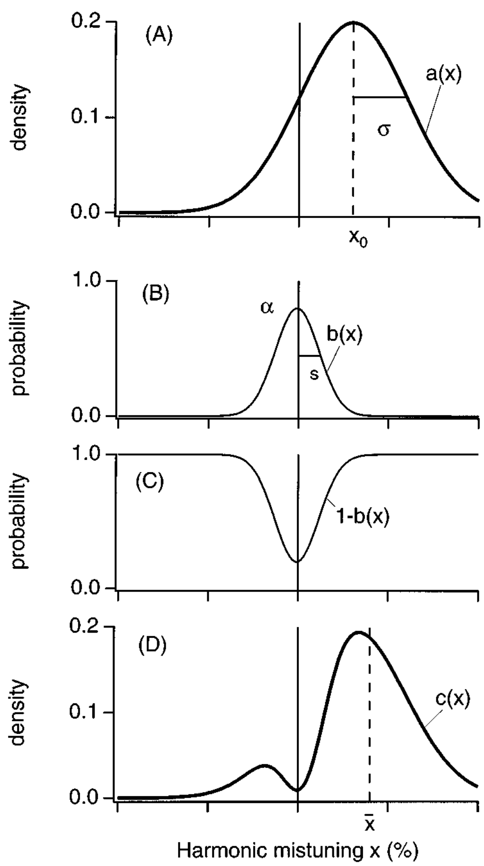

FIG. 1. (A) Distribution of x for all trials. The vertical dotted line marks the center of gravity of the distribution, x_0. (B) Probability of harmonic fusion. (C) Probability of a successful pitch match. (D) Distribution of x for successful matches. The vertical dotted line marks the center of gravity $\overline{x}$ of this distribution. The distance from the dotted line in (A) represents the pitch shift.

Because of fusion, pitch matches are less likely to be successful when x falls near zero [Fig. 1(C)]. The distribution of x for successful pitch matches [denoted as $c(x)$] is therefore distorted relative to the original distribution $a(x)$ [Fig. 1(D)]. Its center of gravity is shifted in the same direction as the mistuning, hence the pitch shift in that direction. With large mistunings the distribution of x is less affected by harmonic fusion, and pitch shifts should decrease. Both aspects were evident in the data of Hartmann and Doty (1996).

With a reasonable choice of the distribution of x and of the probability of fusion the shift may be calculated. Let us assume that x is normally distributed about its mean x_0:

$$a(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-x_0)^2/2\sigma^2}. \tag{1}$$

We suppose also that the probability of fusion conditional on x is shaped like a Gaussian centered at zero mistuning:

$$b(x)=\alpha e^{-x^2/2s^2}. \tag{2}$$

The parameter s controls the ‘‘width’’ of this slot of the harmonic sieve (Duifhuis *et al.*, 1982) and α, with a value between 0 and 1, its ‘‘permeability.’’ Together with σ, the model thus has three parameters. The distribution of successful matches is

$$c(x)=Aa(x)(1-b(x)), \tag{3}$$

where the normalization factor A ensures that the distribution sums to 1. This may be rewritten as the weighted sum of two normal distributions:

$$c(x)=A\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-e(x-x_0)^2/2\sigma^2} - B\frac{1}{D\sqrt{2\pi}}e^{-(x-C)^2/2D^2}\right)$$

with

$$B=\frac{\alpha s}{\sqrt{s^2+\sigma^2}}e^{-x_0^2/2(s^2+\sigma^2)},$$

$$C=\frac{fs^2}{s^2+\sigma^2}, \quad D=\frac{s\sigma}{\sqrt{s^2+\sigma^2}}.$$

Noting that a normal distribution has an integral of 1, we have

$$A=1/(1-B).$$

The mean $\overline{x}$ of the distribution of x for successful matches is

$$\overline{x}=\int_{-\infty}^{\infty}xc(x)dx=A(x_0-BC). \tag{4}$$

The pitch shift is the difference between the mean $\overline{x}$ of the distorted distribution and the mean x_0 of the original distribution. This can be calculated as

$$\overline{x}-x_0=x_0\frac{\sigma^2}{s^2+\sigma^2}\Bigg/\left[\frac{\sqrt{s^2+\sigma^2}}{\alpha s}e^{x_0^2/2(s^2+\sigma^2)}-1\right]. \tag{5}$$

Figure 2 illustrates the range of shift patterns that are obtained with various values of the three parameters α (‘‘permeability’’ of the harmonic sieve), s (width of harmonic sieve), and σ (standard deviation of the internal variable). In Fig. 2(A) parameter α was fixed at 1, and parameters s and σ were varied together from 1% to 10%. All curves have the same slope at the origin: When $s=\sigma$ this slope depends only on α. The dependency on α is illustrated in Fig. 2(B) for $s=\sigma=3\%$. Finally, in Fig. 2(C), α was equal to 1, and the geometric mean of s and σ was kept constant and equal to 3%, while the ratio s/σ was varied from 0.3 ($s=1\%$, $\sigma=9\%$) to 3 ($s=9\%$, $\sigma=1\%$) in 10 logarithmically spaced steps. When s/σ is small, shifts tend to increase gradually with mistuning, whereas when it is large they peak at a small mistuning and then decrease beyond. The model can thus produce a wide range of shift patterns that are all symmetrical about the origin.

II. COMPARISON WITH EXPERIMENTAL DATA

Symbols in Fig. 3(A) represent shifts observed by Hartmann and Doty (1996) for the fifth harmonic at a low level (28 dB per harmonic). The line represents shifts produced by the model for $\alpha=0.8$ and $s=\sigma=3\%$. These parameter values were chosen to produce a good fit ‘‘by eye.’’ The shifts have the same sign as the mistuning and peak at about 4% mistuning and decrease thereafter, as observed experimentally.

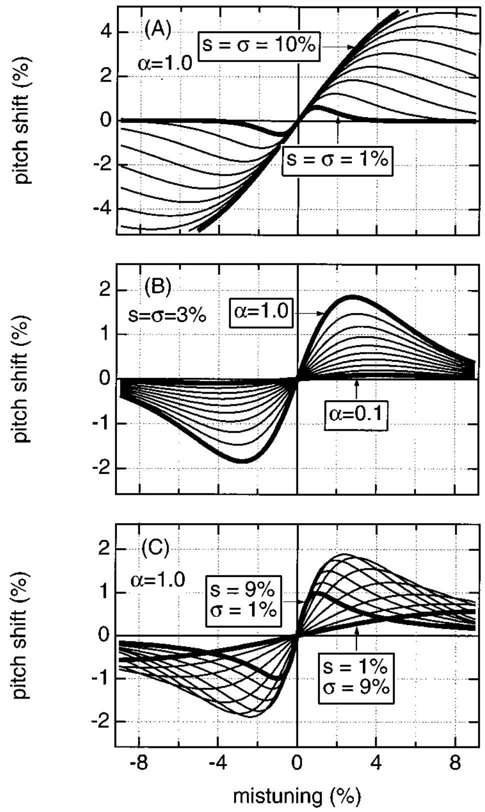

FIG. 2. (A) Pitch shifts produced by the model as a function of mistuning, for $\alpha=1.0$ and for values of $s=\sigma$ varying between 1% and 10%. (B) Same, for $s=\sigma=3\%$ and α varying between 0.1 and 1.0. (C) Same, for s/σ varying in logarithmically spaced steps from 0.3 to 3, while $\alpha=1.0$ and the geometric mean of s and σ remains constant and equal to 3%.

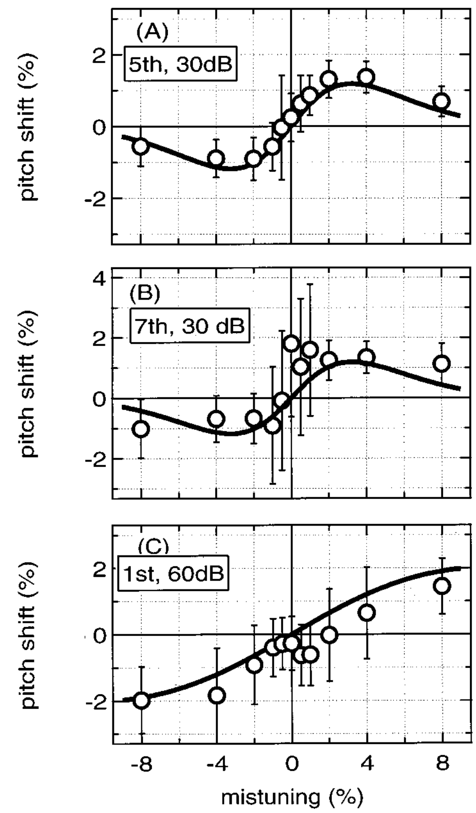

FIG. 3. (A) Symbols: pitch shifts observed for the fifth harmonic at a level of 28 dB per harmonic (Hartmann and Doty, 1996). Line: shifts produced by the model assuming $s=\sigma=3\%$ and $\alpha=0.8$. (B) As in (A), for the seventh harmonic at a level of 28 dB per harmonic. (C) Symbols: pitch shifts observed for the fundamental at a level of 58 dB per harmonic. Line: shifts produced by the model assuming $s=\sigma=8\%$ and $\alpha=0.6$.

Symbols in Fig. 3(B) represent shifts observed for the seventh harmonic at 28 dB per harmonic. The line represents shifts produced by the model using the same set of parameters as in Fig. 3(A). The ''fit'' is less good, but the model-produced shifts nevertheless fall within experimental error bars. The same can be said of most other conditions: harmonics 2, 3, and 4 at 28 dB per harmonic, and harmonics 3, 4, 5, and 11 at 58 dB per harmonic. Other conditions require a different choice of parameters. The somewhat larger shifts observed by Hartmann and Doty at harmonics 9 and 11 at 28 dB per harmonic, and harmonic 7 at 58 dB per harmonic (not shown) can be accommodated by assuming $s=\sigma=4\%$, and $\alpha=0.8$ as before. The monotonically increasing shifts at the fundamental (58 dB per harmonic) can be accounted for by assuming $s=\sigma=8\%$ and $\alpha=0.6$ [Fig. 3(C)]. Finally, shifts at the second harmonic at 58 dB per harmonic (not shown) can be modeled by assuming $s=\sigma=6\%$ and $\alpha=0.6$. Thus to account for the entire set of Hartmann and Doty's data one must assume that the parameters vary somewhat between harmonics and level, which is perhaps not surprising. Given experimental error and the tradeoffs between parameters, it is however difficult to make precise estimates of each parameter.

In all three examples shown in Fig. 3 there was a tendency for experimental data points to mostly fall above or below the curve predicted by the model. Systematic shifts of up to 2% were observed in a mistuned harmonic experiment by Lin and Hartmann (1996), who found considerable variation between conditions and subjects. The model makes predictions that are symmetric relative to the origin, and thus we cannot explain such systematic shifts. They might be explained by a response bias such as the tendency to make pitch adjustments slightly sharp as observed by Jarosewski (1992), or by shifts caused by neighboring partials as predicted by Terhardt (1979). Lin and Hartmann argue against this last interpretation, as the shifts do not seem to require the presence of either neighboring partial. A ''sideways'' discrepancy might be explained within the context of our model by assuming that the harmonic sieve is slightly mistuned. Apart from such systematic shifts, the model accounts for major aspects of Hartmann and Doty's data. Pitch shifts have the same sign as the mistuning, and with a reasonable choice of parameters they peak at about 4% mistuning and decrease beyond. This last aspect, highly reproducible according to Hartmann and Doty, was not explained by their model. Shifts at the fundamental, also not accounted for by their model, are explained here in the same fashion as for other harmonics.

III. SHIFTS OF THE LOW PITCH OF THE COMPLEX

Darwin *et al.* (1994) presented listeners with a complex tone that had a mistuned fourth harmonic. Instead of matching the pitch of the mistuned partial, as in Hartmann and

Doty's experiments, the subjects matched the low pitch of the entire complex. That pitch varied with the frequency of the partial up to about 3% mistuning, at which point it had shifted by about 0.5%. With further mistuning the low pitch shifted back, and by about 8% mistuning it had resumed its initial value, in agreement with the original observations of Moore *et al.* (1985). The pitch could be accurately modeled as

$$F_0 = a + k\Delta f e^{-\Delta f^2/2s^2}, \tag{6}$$

where F_0 is the frequency of the tone that matches the low pitch, Δf is the mistuning, and a is a constant close to the nominal pitch of the complex. The constant k is the rate of change in low pitch relative to changes in frequency of the partial, and s determines the range over which mistunings affect the low pitch. For mistunings expressed in percentage, their values were about 0.2% and 3%, respectively. When plotted, the complex pitch shift (Darwin *et al.*, 1994, Fig. 1) looks very similar to the partial pitch shift [Fig. 3(A)]. One must keep in mind their different nature: a shift of the low pitch of the complex relative to the fundamental of the fixed harmonic series in the former case, a shift of the pitch of the partial relative to its *own* mistuned frequency in the later. The low pitch is as it were ''pulled'' by the mistuned partial, whereas the pitch of the partial is ''pushed'' away from the harmonic series.

It is tempting to view the pitch shift of the complex as the outcome of the same probabilistic process that produces pitch shifts of the partials. According to our model (Sec. I) the partial's pitch was available on trials for which there was *not* fusion. Suppose we assume that on trials for which there *was* fusion, the partial contributed to the low pitch instead. In that case, the decrease in complex pitch shift for larger mistunings [Eq. (6)] would simply reflect the fact that fusion becomes less probable, as described by Eq. (2) and illustrated in Fig. 1(B). The parameter s would be the same in both equations, and this would explain why the value found by Darwin *et al.* (1994) (3%) works well in our model. This hypothesis is attractive because it allows the principle of disjoint allocation (Bregman, 1990) to apply at the microscopic level of the ''trial,'' even if it seems to fail at the macroscopic level of pitch matches (Moore, 1987; Darwin and Carlyon, 1995). However, we know of no direct experimental evidence in favor of this hypothesis. As it is unnecessarily strong in the context of our model, we shall explore it no further.

Our model assumes that pitch shifts are the consequence of harmonic fusion rather than interactions between neighboring partials. In support of this assumption, Lin and Hartmann (1996) found shifts at harmonic positions within a harmonic complex, even when neighboring components were missing. For example, the pitch of a 200 Hz ±8% partial showed shifts in the presence of a harmonic template formed by harmonics between 1000 and 3200 Hz. Similarly, pitch shifts at the third harmonic were significant in presence of a harmonic complex that lacked either the second or the fourth harmonic, or both. In general, however, the shifts tended to be larger if all background harmonics were present, in particular the harmonic of next-lowest rank relative to the mistuned partial.

Similar shifts have been observed for the pitch of tones preceded by a tone of similar frequency (Hartmann and Kanistanaux, 1979). Those authors proposed a model in which the excitation pattern evoked by a tone was distorted by adaptation, thus causing a pitch shift. If the quantity $a(x)$ in Eq. (3) is viewed as a deterministic excitation pattern, and $[1-b(x)]$ as distortion caused by the ''harmonic sieve,'' our model can be interpreted as formally similar to theirs, the effects of adaptation being replaced by those of the harmonic sieve. Kashino and Nishida (1996, 1997) found similarly shaped shifts in the localization of sources preceded by a source of similar ITD (interaural time delay). They interpreted them as resulting from a recalibration of the ITD-extraction mechanism, to enhance the representation of ITDs near that of the adaptor. This rescaling of perceptual dimensions would play a functional role similar to that proposed for visual aftereffects by Barlow (1990).

IV. CONCLUSION

Pitch shifts of mistuned partials were explained by postulating a noisy internal variable that determined both the pitch of the partial and the probability of fusion within the harmonic complex. Aspects accounted for were: (a) the sign of the pitch shifts (same as the mistuning); (b) their magnitude; (c) the fact that they peak at about 4% and decrease beyond; and (d) the fact that they affect individually audible harmonics of all ranks, including the fundamental. The explanation is congruent with the notion of harmonic fusion and segregation observed in many situations. No assumption was required concerning a particular model of pitch perception (place, time, etc.). This weakens the argument made by Hartmann and Doty (1996) in favor of ISI-based rather than other temporal or place mechanisms, on the grounds that they allow the shifts to be explained. Instead, a different but interesting conclusion may be drawn: Pitch and harmonic fusion must share a common physiological substrate subject to random variations from trial to trial. The alternative hypothesis, that pitch and fusion depend on separate mechanisms, each with its own source of variability, is contradicted by Hartmann and Doty's data.

ACKNOWLEDGMENTS

This work was carried out within a research agreement between ATR and CNRS. Thanks to Bill Hartmann for providing access to unpublished data and manuscripts, and for detailed comments on several versions of this manuscript, that also benefited from comments of one anonymous reviewer.

Barlow, H. B. (**1990**). ''A theory about the functional role and synaptic mechanism of visual after-effects,'' in *Vision: Coding and Efficiency*, edited by C. Blakemore (Cambridge U. P., Cambridge, England), pp. 363–375.

Bregman, A. S. (**1990**). *Auditory scene analysis* (MIT, Cambridge, MA).

Darwin, C. J., and Carlyon, R. P. (**1985**). ''Auditory grouping,'' in *Handbook of Perception and Cognition: Hearing*, edited by B. C. J. Moore (Academic, New York), pp. 387–424.

Darwin, C. J., Ciocca, V., and Sandell, G. J. (**1994**). ''Effects of frequency and amplitude modulation on the pitch of a complex tone with a mistuned harmonic,'' J. Acoust. Soc. Am. **95,** 2631–2636.

de Cheveigné, A., McAdams, S., and Marin, C. (**1997**). ''Concurrent vowel segregation. II: Effects of phase, harmonicity, and task,'' J. Acoust. Soc. Am. **101,** 2848–2856.

Duifhuis, H., Willems, L. F., and Sluyter, R. J. (**1982**). ''Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception,'' J. Acoust. Soc. Am. **71,** 1568–1580.

Hartmann, W. M., and Kanistanaux, D. C. (**1979**). ''The effect of a prior tone on the pitch of a short tone,'' in *Proceedings of the Research Symposium on the Psychology and Acoustics of Music,* edited by W. V. May (University of Kansas, Lawrence, KS), pp. 199–215.

Hartmann, W. M. (**1988**). ''Pitch perception and the segregation and integration of auditory entities,'' in *Auditory Function—Neurological Bases of Hearing,* edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 623–645.

Hartmann, W. M., McAdams, S., and Smith, B. K. (**1986**). ''Matching the pitch of a mistuned harmonic in a complex tone,'' IRCAM annual report.

Hartmann, W. M., McAdams, S., and Smith, B. K. (**1990**). ''Hearing a mistuned harmonic in an otherwise periodic complex tone,'' J. Acoust. Soc. Am. **88,** 1712–1724.

Hartmann, W. M., and Doty, S. L. (**1996**). ''On the pitches of the components of a complex tone,'' J. Acoust. Soc. Am. **99,** 567–578.

Jarosewski, A. (**1992**). ''A study of constant (systematic) errors in pitch discrimination of short tone pulses,'' Acustica **77,** 106–110.

Kashino, M., and Nishida, S. (**1996**). ''Auditory localization aftereffects,'' J. Acoust. Soc. Am. **99,** 2596.

Kashino, M., and Nishida, S. (**1997**). ''Adaptation in the processing of interaural time differences revealed by the auditory localization aftereffects,'' J. Acoust. Soc. Am. (submitted).

Lin, J. L., and Hartmann, W. M. (**1996**). ''The pitch of mistuned harmonics: evidence for a template model,'' J. Acoust. Soc. Am. (submitted).

Martens, J. P. (**1981**). ''Comment on 'Algorithm for extraction of pitch and pitch salience from complex tonal signals' [J. Acoust. Soc. Am. **71**, 679–688 (1982)],'' J. Acoust. Soc. Am. **75,** 626–628.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1985**). ''Thresholds for the detection of inharmonicity in complex tones,'' J. Acoust. Soc. Am. **77,** 1861–1867.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1986**). ''Thresholds for hearing mistuned partials as separate tones in harmonic complexes,'' J. Acoust. Soc. Am. **80,** 479–483.

Moore, B. C. J. (**1987**). ''The perception of inharmonic complex tones,'' in *Auditory Processing of Complex Sounds,* edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 180–189.

Terhardt, E. (**1979**). ''Calculating virtual pitch,'' Hearing Res. **1,** 155–182.