

A model of the perceptual asymmetry between peaks and troughs of frequency modulation

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 15 January 1998; revised 3 March 1999; accepted 14 January 2000)

Pitch discrimination at peaks of frequency modulation is better than at troughs [L. Demany and K. I. McAnally, *J. Acoust. Soc. Am.* **96**, 706–715 (1989)]. A similar asymmetry emerges within a time-domain pitch perception model based on autocorrelation. The model requires the following assumptions: (a) The neural discharge patterns must be temporally sharpened to a single narrow pulse per period (possibly by neural convergence within the cochlear nucleus). (b) Autocorrelation must be implemented as a cross correlation between the neural pulse train and a delayed pulse train convolved with a short kernel function. This kernel function must be asymmetric in time. (c) Pitch discrimination must rely on higher-order modes of the autocorrelation function. This particular implementation of the autocorrelation model produces modes that are sharper for peaks than for troughs, and thus accounts for the pitch discrimination asymmetry observed experimentally. As a by-product it can account for “hyperacute” discrimination observed at peaks of triangular modulation. © 2000 Acoustical Society of America.

[S0001-4966(00)05104-3]

PACS numbers: 43.66.Ba, 43.66.Hg, 43.64.Bt [JWH]

INTRODUCTION

Demany and McAnally (1994) discovered an interesting perceptual asymmetry between peaks (local maxima) and troughs (local minima) of frequency modulation: the *perceptual salience* of peaks is greater than that of troughs. In a first experiment, subjects were presented with a frequency-modulated pure tone with a modulating waveform that had as many peaks as troughs. When requested to report the sequence of notes that they heard, subjects consistently reported maxima more often than minima. The asymmetry could not be ascribed to a mere advantage of high frequencies over low: in some conditions certain peaks were *lower* in frequency than certain troughs, but nevertheless more salient.

In a second experiment it was found that increased salience of peaks over troughs is reflected also by better *frequency discrimination*. Demany and his colleagues subsequently carried out an extensive set of experiments that showed that the peak/trough asymmetry is quite general. The many conditions that they explored allowed a wide range of hypotheses to be tested and, for many, rejected. These data are reviewed in detail in the next section. The ubiquity and strength of the peak/trough asymmetry (threshold ratios averaged over subjects as large as 5 in some experiments) suggest that it might reflect some important property of the pitch perception system. The later sections of this paper propose a model of this effect.

I. REVIEW OF THE EXPERIMENTAL DATA

In their second experiment, Demany and McAnally (1994) presented subjects with tones modulated by a wave-

form shaped either as a single cycle of a cosine (“trough”), or as the opposite of the same (“peak”). The modulator, defined on a logarithmic scale, had an amplitude of $\frac{1}{2}$ oct. For peaks, the initial frequency was 707 Hz, and for troughs 1414 Hz, but the vertex frequency, at the stimulus center, was about 1000 Hz in both cases [Fig. 1(a)]. In the present paper, “vertex” is used to designate the extremum of frequency modulation, either maximum or minimum, upon which the pitch discrimination task is focused. The subjects were requested to discriminate these stimuli from stimuli that started and stopped at the same frequency, but had a deeper frequency modulation, and thus a slightly different frequency at the vertex. Discrimination was clearly better for peaks than for troughs: thresholds for troughs were about twice those of peaks.

Demany and his colleagues have shown the phenomenon to be quite general. The asymmetry has been observed for modulated pure tones with vertex frequencies of 250, 500, and 1000 Hz (Demany and Clément, 1995a, 1998). It also exists at 4000 Hz, but the trough/peak threshold ratio is smaller (Demany and Clément, 1995a). It exists for harmonic complex tones with vertex frequencies of 200 or 500 Hz (Demany and Clément, 1995b, 1998), and in particular for Shepard tones that consist of an infinite series of octave-spaced partials shaped by a fixed band-limited spectral window (Shepard, 1964). The case of Shepard tones is interesting because it allows an ambiguity in the interpretation of the asymmetry to be resolved. Peak and trough stimuli have similar frequencies at the vertex, but the frequency at the beginning of the stimulus is an octave higher for the trough than for the peak (for a modulation amplitude of 0.5 oct), and the same is of course true at the end. Perceptual asymmetry

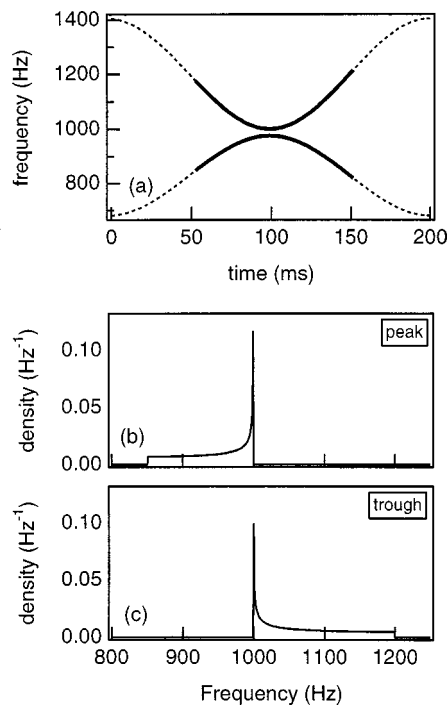


FIG. 1. (a) Top: modulation waveform of a trough. Bottom: modulation waveform of a peak. (b) Instantaneous frequency distribution of a peak over the portion marked as a full line in (a). (c) Same, for a trough.

could conceivably result from that aspect of the stimuli. Shepard tones spaced an octave apart are identical, so this interpretation can be ruled out in their case (and probably in general). The asymmetry should be attributed to other cues such as the shape of the vertex or the direction of the FM glides that flank it.

The asymmetry has been observed for modulation amplitudes of 0.5, 1, or 1.5 oct (in the latter case for unidirectional glides, Demany and Clément, 1997). When the vertex frequency was roved within a 0.5-oct range the trough/peak threshold ratio was only slightly reduced at 250 or 1000 Hz, but practically abolished at 4000 Hz. The asymmetry is not affected by training despite a general decrease in thresholds (Demany and Clément, 1995a), but it seems to differ somewhat between subjects. It has been investigated mainly with stimuli 200 or 400 ms in duration, but it exists for stimuli as long as 800 ms or as short as 50 ms but not 25 ms (Demany, 1997). The effect persisted when an amplitude modulation was superimposed on the FM, even when the AM was designed to favor troughs over peaks (Demany and McAnally, 1994; Demany and Clément, 1998).

The effect has been studied principally with cosine-shaped modulation waveforms [Fig. 1(a)], but it exists also with modulation waveforms that have extrema that are either sharper (almost triangular) or duller (with flat portions as long as 100 ms). The trough/peak ratio tends, however, to be greater for sharp than for blunt extrema (Demany and Clément, 1997). The effect was also observed for ramps consisting of half a cosine modulation waveform, either the first half or the second half (Demany and Clément, 1997).

This impressive sum of data suggests that the perceptual asymmetry between peaks and troughs might reflect fundamental characteristics of the pitch perception process, even if

it has so far resisted explanation in such terms. Several explanations that seem plausible *a priori* can be rejected based on the experimental results. One might, for example, imagine that the general improvement of discrimination with frequency up to 2 kHz, where relative thresholds are smallest (Moore, 1973), might somehow favor peaks over troughs. By showing that the asymmetry was not reversed at 4000 Hz, Demany and Clément (1995a) ruled out this explanation. Demany and McAnally (1994) had already shown that the effect could not be due to a mere change of amplitude or loudness with frequency (FM-induced AM).

The effect is not likely to be due to the asymmetry of excitation patterns on the basilar membrane. These should, if anything, favor troughs over peaks, as there is an upward spread of masking for pure tones at 70 dB (Demany and McAnally, 1994). The argument was strengthened by Demany and Clément (1995a) when they replicated the experiment at 35 dB, at which level the asymmetry of the excitation pattern should be opposite of that at 70 dB. It was further strengthened by the observation of the same asymmetry for complex harmonic tones (in particular tones shaped by a fixed spectral window). The excitation pattern of a complex tone is more complex and spread out than that of a pure tone, and less likely to differ between peaks and troughs in a way that could favor the former over the latter.

The fact that the asymmetry is observed for amplitude-modulated high-pass noise, that lacks spectral cues, suggests that it stems from a process operating in the time domain (Demany and Clément, 1995b). This is congruent with the fact that it is weaker at 4000 Hz, or even nonexistent for roving stimuli at that frequency (Demany and Clément, 1995a). On the other hand, Demany and Clément (1997) have also argued *against* a time-domain process, on the basis of the proactive and retroactive effects that they observed with ramps. For those authors a retroactive effect suggests a central mechanism, whereas high-resolution time-domain information is not usually observed beyond the periphery and neural relays below the inferior colliculus.

The fact that the asymmetry was observed with Shepard tones rules out any role of the initial (or final) frequency difference between peak and trough stimuli, as Shepard tones differing by an octave are identical. It also rules out a simple effect of the temporal pattern of extreme frequencies (low-high-low versus high-low-high). The asymmetry must therefore depend on dynamic frequency changes, for example, the frequency ramps preceding or following the temporally centered vertex. Demany and Clément (1995b, 1997) argued that discrimination cannot be based on differences in FM slope between standard and target, at least in the case of peaks: the differences are too small compared to FM slope thresholds measured by Dooley and Moore (1988). Overall, it is unlikely that the effect depends on a perceptual asymmetry of the ramps by themselves, but rather on some interaction between ramp and peak.

A priori, one can imagine either a proactive effect of the ramp preceding the vertex, or a retroactive effect of the ramp following it, or both. By cutting the cosine-modulated stimulus in two (either the first half or the second), Demany and Clément (1997) were able to demonstrate that the effect of a

ramp on the discrimination of a vertex can be *both* proactive and retroactive. Is the effect beneficial to peaks, or detrimental to valleys, or both? Demany and Clément (1995b) suggested a detrimental effect on valleys, at least for amplitude-modulated noise. However, Demany and Clément (1997) found, for triangular-shaped modulation, “hyperacute” perception at the peak. This latter effect is surprising: shift detection thresholds of practically durationless peaks were about 10 cents or 0.6%. This is of the same order as the threshold of a 6.25-ms steady-state pure tone at 1 kHz (Moore, 1973). Frequency would vary by about 18 cents within a 6-ms window centered on the peak (and by the double for a window situated on either side), and it is difficult to understand how “moving targets” of the sort can be discriminated with such acuity.

Clément (1996) proposed a model to explain the asymmetry, based on a formal neural network. The input layer received a tonotopic representation of the instantaneous frequency of the stimulus (derived by an infinite-resolution frequency extraction process that was not described). The cells in this layer were interconnected laterally by asymmetric inhibitory connections with temporal integration properties. This layer favored the representation of frequencies preceded by lower frequencies, at the expense of those preceded by higher frequencies. The second layer served to integrate temporally this modified frequency representation. Asymmetric lateral inhibition within the first layer accounts for the perceptual advantage of peaks over troughs, while the temporal integration of the second layer accounts for the fact that effects can be both “proactive” and “retroactive.”

Clément’s explanation of the asymmetry is plausible, but not entirely satisfying. In a sense, it is a direct translation of the phrase “the auditory system favors peaks over troughs” by which one might summarize the experimental results. It offers a plausible account of *how* such an asymmetry might be implemented, but not *why* things should occur in that way. An argument for the functional advantage of the asymmetric lateral inhibition, or an explanation in terms of *emergent* properties of the pitch perception process, would be more satisfying (though not necessarily more correct). The purpose of the present paper is to propose an explanation of the latter sort.

II. THE MODEL

A. Structure

The model is designed to produce a quantity (“pitch estimate”) that allows the pitch of two stimuli to be compared, for example, between two intervals of a discrimination task. Pitch asymmetries can be explained if the pitch estimate is less accurate for valleys than for peaks. The overall structure of the model is illustrated in Fig. 2. It is similar to the autocorrelation model of Licklider (1956) or Meddis and Hewitt (1991a, b), but differs in three important ways. First, the discharge probability at the output of the basilar-membrane/haircell model is sharpened before further processing. Second, the standard autocorrelation function (ACF) of Licklider’s model is replaced by an “asymmetric autocorrelation function” (AACF, defined presently). Third, the

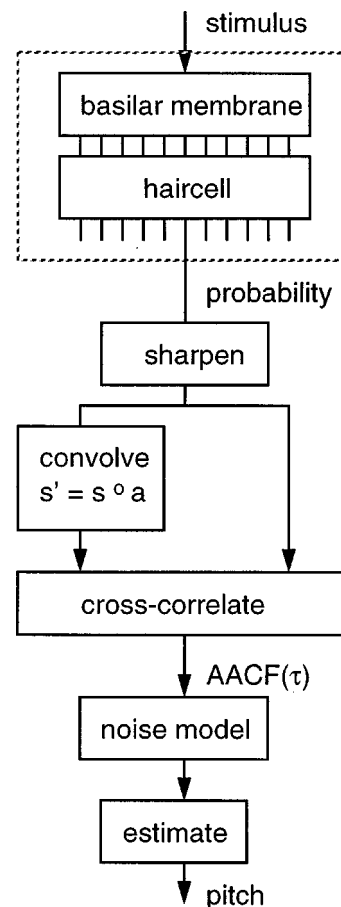


FIG. 2. Overall structure of the model. The part within dotted lines is implemented simply as a half-wave rectifier. Its output is first sharpened (by raising it to a power) and then cross correlated with a version of itself convolved with an asymmetric function, $a(t)$, to obtain an “asymmetric autocorrelation function” (AACF). After application of a noise model, the period is estimated from a high-order mode of the AACF.

pitch estimate involves higher-order modes of that function, rather than just the first-order “pitch peak.” These assumptions are detailed in the next paragraphs, after the presentation of a “lemma” that embodies the principle on which the model is based. All assumptions are examined critically in Sec. IV.

B. Lemma: Convolution of asymmetric peaked functions

The purpose of this section is to point out that the sharpness of the *convolution* of two functions that are sharp and asymmetric depends on whether the asymmetries agree or not. The “lemma” is the root of the model. No formal proof is attempted; the lemma is simply illustrated for two windowed exponential functions (Fig. 3). On the left, the asymmetries of the functions to be convolved agree. Their convolution has a peak that is somewhat blunt. On the right, one function has been time reversed, and their asymmetries are thus opposed. Their convolution has a sharp peak, related to the existence and orientation of the discontinuities in the functions. If the position of the maximum had to be determined precisely, for example, in the presence of a small amount of noise, it is clear that accuracy would be greater on the right than on the left.

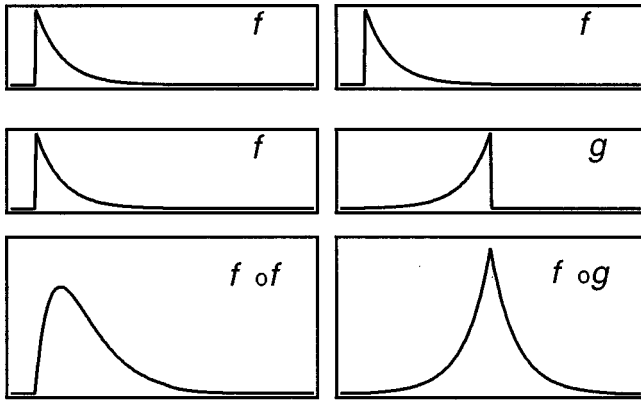


FIG. 3. Top and middle left: two asymmetric peaked functions with similar orientations. Bottom left: their convolution. Top and middle right: two asymmetric peaked functions with opposite orientations. Bottom right: their convolution. Note the difference in “peakiness” between the two convolutions.

The strategy followed by this paper relies on this lemma: (a) identify appropriate asymmetric functions within the auditory process, (b) show that the asymmetry of one of these functions is reversed for peaks and valleys of modulation, and (c) argue that pitch depends upon their convolution. One asymmetric function can be found in the distribution of instantaneous period (inverse of instantaneous frequency) in the region of the vertex [Fig. 1(b) and (c)]. The asymmetry of this distribution is opposite for peaks and valleys. The second asymmetric function will be introduced by assumption 2, within the autocorrelation calculation. Assumptions 1 and 3 are of less fundamental importance, but they make the model easier to illustrate, and allow it to handle a wider range of conditions.

C. Assumptions

Assumption 1: Temporal sharpening of neural responses. The discharge probability of an auditory nerve fiber generally resembles the half-wave rectified motion of the basilar membrane. In response to a pure tone, the probability function consists of peaks that occupy a large proportion of the period, and are therefore quite wide (Ruggero, 1992) [Fig. 4(a)]. Peaks of the autocorrelation function (ACF) are also wide [Fig. 4(b)], and the auditory system might have difficulty deriving an accurate pitch estimate on this basis. Probability functions may, however, be sharpened, for example, within onset cells in the cochlear nucleus (Palmer and Winter, 1992), as illustrated in Fig. 4(c). As a result of this sharpening, peaks of the ACF are also much sharper [Fig. 4(d)]. *Assumption 1* is that a sharpening stage is involved in the pitch perception process. A full discussion of the necessity and plausibility of this assumption is deferred to Sec. IV A.

Assumption 2: Asymmetric implementation of autocorrelation. The autocorrelation function (ACF) is, by definition, the cross correlation of a signal with itself. The product of two terms, the signal $s(t)$ and the delayed version $s(t - \tau)$, is summed over a window $w(t)$:

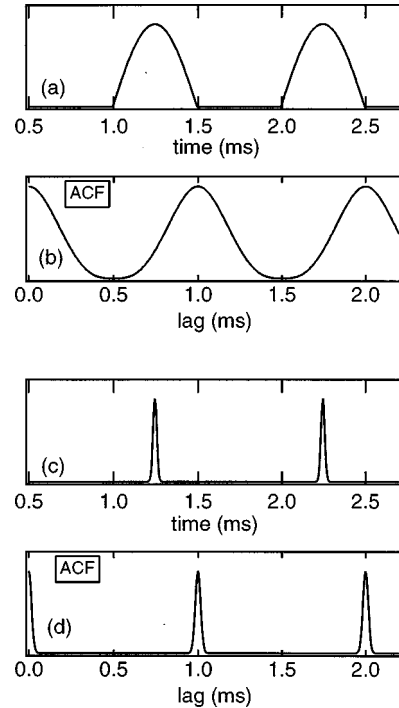


FIG. 4. (a) Discharge probability in response to a 1-kHz pure tone. (b) Autocorrelation function (ACF) of discharge probability. (c) Sharpened discharge probability function, obtained by raising the waveform of (a) to the power 200. (d) ACF of sharpened discharge probability function.

$$\text{ACF}_t(\tau) = \int_{-\infty}^t w(\theta - t) s(\theta) s(\theta - \tau) d\theta, \quad (1)$$

where θ is a temporal integration variable. This is actually the definition of a running autocorrelation function indexed by t . Suppose now that the delayed function $s(t - \tau)$ is convolved with a kernel function $a(t)$ before multiplication:

$$s' = s \circ a. \quad (2)$$

The kernel function $a(t)$ is supposed to be short and asymmetric: skewed with a tail towards large t . Let us call “asymmetric autocorrelation function” (AACF) the running cross-correlation function between s and s' :

$$\text{AACF}_t(\tau) = \int_{-\infty}^t w(\theta - t) s(\theta) s'(\theta - \tau) d\theta. \quad (3)$$

Figure 5 illustrates this process. In (a) is plotted the discharge probability function, sharpened as argued in the previous paragraph. In (b) is shown its convolution with a short kernel function shaped as a decaying exponential of time constant t_a . In (c) is shown the cross correlation of waveforms in (a) and (b), called AACF. Its modes are asymmetric and skewed in the direction of shorter lags. Other than that, the AACF is not too different from the ACF, hence the name “asymmetric autocorrelation function.” *Assumption 2* is that the pitch perception process involves such an asymmetric autocorrelation function, rather than the standard symmetric autocorrelation that is usually postulated by autocorrelation-based models of pitch (Licklider, 1956; Meddis and Hewitt, 1991a, b). This assumption is justified in Sec. IV A.

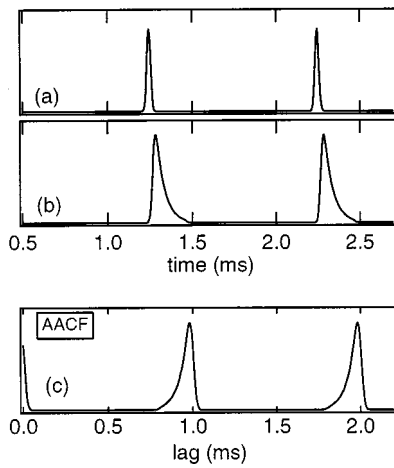


FIG. 5. (a) Sharpened discharge probability function. (b) Same, after convolution with a window shaped as a decaying exponential of time constant 0.25 ms. (c) “Asymmetric autocorrelation function” (AACF) of the sharpened probability function [cross correlation of (a) and (b)].

Assumption 3: Pitch is derived from higher-order modes. In the autocorrelation model of pitch, it is usually assumed that pitch is derived from the first mode (“period peak”) of the ACF. However, it has been argued that the auditory system might take into account higher-order modes (Srulovicz and Goldstein, 1983; de Cheveigné, 1989; Slaney, 1990). Higher-order modes have a smaller *relative* width, and this helps account for the small size of discrimination thresholds. The maximum order is limited by the stimulus duration, so the use of higher-order modes might explain the approximate inverse dependency on stimulus duration observed by Moore (1973) below 1–2 kHz. *Assumption 3* is that higher-order modes of the ACF (or AACF) are exploited in the task of discriminating the vertex frequency of peaks or troughs.

D. Modeling conventions

The following conventions are made in the interest of clarity. Basilar membrane filtering is ignored and the properties of hair-cell transduction are reduced to a simple half-wave rectification. The integration window of the running ACF and AACF is chosen to be square (rather than exponential as in the models of Licklider, 1956, or Meddis and Hewitt, 1991a, b):

$$\text{AACF}_t(\tau) = \int_{t-D}^t s(\theta) s'(\theta - \tau) d\theta, \quad (4)$$

where D is the duration of the window. The square shape allows the position of the window to be defined more precisely, so the effects of window size and position can be better understood. Note that, for nonzero values of τ , the calculation includes samples that lie *outside* the window. In signal processing the signal is often set to zero outside the integration window, to avoid the contribution of such samples. That custom is not followed here.

In the simulations, the duration D is chosen long relative to the period in order to avoid rapid fluctuations of the period estimate, rather than on the basis of physiological or psycho-

physical estimates. This is necessary for the model to play its role. Note that it implies some circularity, as the period is not known before estimation. This consideration determines the *lower* limit of the window size. The upper limit is treated as a free parameter, as is the position of the window within the stimulus.

The convolution kernel function $a(t)$ used in the AACF is taken to be shaped like a decaying exponential. Its time constant t_a must be small relative to the period, but otherwise it is treated as a free parameter.

Pitch is supposed to depend on the position of the *maximum* of one or more modes of the ACF (or AACF), rather than on, say, their center of gravity or positions of maximum slope. The accuracy with which this position can be determined is assumed to depend on the “sharpness” of the mode (inverse of the width at some proportion of the height). This assumption is reasonable if noise is superimposed on the function.

III. BEHAVIOR OF THE MODEL

To set the stage, the simple case of an unmodulated tone is considered first. Then the model is applied to the frequency-modulated tones used by Demany and colleagues. In every case, two versions of the model are illustrated: standard autocorrelation (ACF) and asymmetric autocorrelation (AACF).

A. Unmodulated tone

Figure 6(a) represents the discharge probability evoked by a 1-kHz pure tone of duration 10 ms. The representation is idealized (no transitory effects of filtering or haircell transduction). Figure 6(b) shows the ACF of the probability function in (a). The integration window size was 5 ms, and it was placed to cover the end of the stimulus. Had it been chosen as large as (or larger than) the stimulus duration, the envelope of the ACF would have been triangular. Had it been chosen shorter, for example, equal to 1 ms, the envelope would have been flatter (had it been chosen shorter still, ACF_t would have fluctuated strongly with t). Whatever the window size, the ACF is zero for lags greater than the stimulus duration.

Figure 6(c) shows the AACF in response to the same tone. The time constant of the convolution kernel function was $t_a = 0.25$ ms. The integration window was the same as for the ACF. The AACF has the same envelope as the ACF, and can be calculated for the same range of lags. However, each peak is skewed, as in Fig. 5(c).

The purpose of this example was to provide a simple reference to help understand the more complex patterns obtained with modulated stimuli. The stimulus was chosen relatively short to illustrate the fact that the running autocorrelation is limited to lags shorter than the stimulus duration. It also shows how the envelope of the ACF or AACF depends on the shapes of the stimulus envelope and integration window.

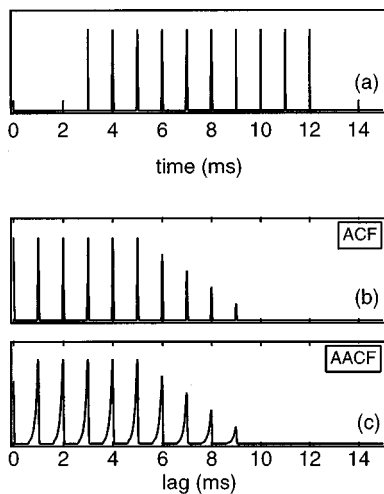


FIG. 6. (a) Idealized discharge probability evoked by a 10-ms pure tone of frequency 1 kHz. Filtering and transitory effects are ignored in this simulation. (b) The ACF of the discharge, for positive lags. The integration window covered the last 5 ms of the stimulus. If it had covered the entire stimulus, the ACF would have had a triangular envelope. (c) The AACF for the same stimulus, with the same integration window. The time constant of the convolution function in the AACF was 0.25 ms.

B. Cosine-shaped peaks and troughs of modulation

The stimulus is a 200-ms frequency-modulated pure-tone, such as used by Demany and McAnally (1994). The modulation waveform is shaped as one period of a cosine function (valley) or its opposite (peak) [Fig. 1(a)]. Its amplitude is 0.5 oct and the frequency at the vertex is 1 kHz.

Figure 7(a) shows the standard ACF for a peak. It was calculated with a window of duration $D=40$ ms placed before the modulation vertex (i.e., covering the last 40 ms of the ramp preceding the vertex). The effect of window position is discussed further on. The modes are widened as a result of modulation, and their width increases with mode order. They are also highly skewed, with a sharp peak and a tail towards longer lags. This reflects the nonuniform instantaneous frequency distribution a consequence of the rounded shape of the peak of the modulation function.

Figure 7(b) shows the standard ACF for a valley. The integration window was placed as in (a). The modes are skewed as in (a) but in the opposite direction. They are, however, just as sharp: if pitch were derived from the position of the modes of the ACF, there would be no reason for discrimination to be worse for valleys than for peaks of modulation.

Figure 7(c) shows the AACF for a peak. It was calculated using the same integration window as for Fig. 7(a). The time constant of the convolution kernel function was $t_a=0.25$ ms. The modes are wider at their base than those of the ACF (an expected effect of the convolution), but they remain quite sharp at their tips. Low-order modes are skewed to the left, and high-order modes to the right, while intermediate modes are more or less symmetrical: this reflects the interplay of the fixed width of the convolution kernel function with the increasing width of the ACF modes due to modulation.

Figure 7(d) shows the AACF for a valley. The modes

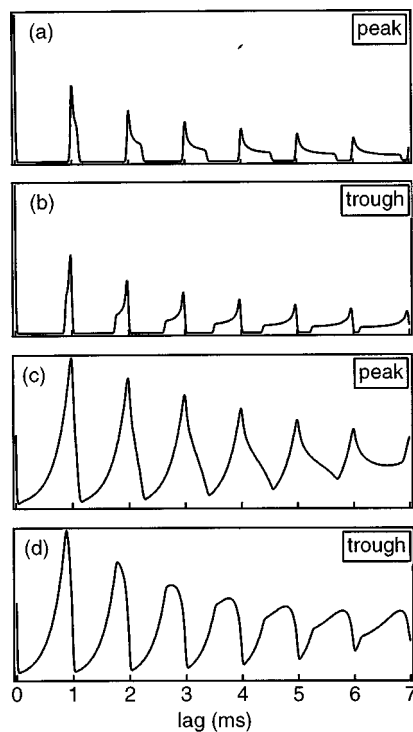


FIG. 7. (a) The ACF for a cosine-shaped modulation peak, calculated over a 40-ms window placed before the vertex. (b) The ACF for a cosine-shaped trough, calculated with a similarly placed window. Peaks are skewed in opposite directions for peaks and valleys, but their sharpness is the same. (c) The AACF for a cosine-shaped peak. (d) The AACF for a cosine-shaped trough. Peaks are rounder for valleys than for peaks, particularly peaks of higher order.

are less sharp than for the peak, especially higher-order modes. This difference can be understood from the ‘lemma’ of Sec. II B. The two functions that, convolved together, produce the mode shape are the *frequency distribution*, on one hand (skewed in different directions for peaks and troughs) and the *exponential-shaped kernel function* of the AACF, on the other. This difference explains, according to the present model, the asymmetry found by Demany and colleagues for cosine-shaped stimuli. As the explanation involves only samples preceding the vertex, it applies also to the asymmetries that they found between end-of-rise and end-of-fall of half cosine-shaped ramps (Demany and Clément, 1997).

1. The effect of window position

The shape of the AACF (and of the ACF) is quite sensitive to window position. Figure 8 (thin line) shows the AACF for a 40-ms window placed *after* a modulation peak (i.e., covering the first 40 ms of the decreasing ramp). The thick line represents the envelope of the AACF for the previous window position (before the apex).

Higher-order modes are sharper and more prominent when the window is placed after the peak rather than before. The difference may seem paradoxical, as windows placed before and after the apex cover similar portions of the waveform. The explanation is simple: the ACF and AACF calculations include samples that *precede* the window. For a window placed before the vertex, these samples belong to the

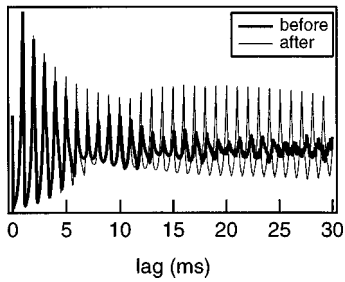


FIG. 8. Thick line: AACF for a cosine-shaped peak, calculated over a 40-ms window placed before the vertex. Thin line: same for a window placed after the vertex.

same slope of the modulation waveform as the window itself, whereas for a window placed after the vertex they belong to the opposite slope. A mode of order k can be understood as a histogram of the durations of the population of interpulse intervals of order k . For a window placed after the vertex, many of these intervals span the vertex, that is, the region where the sweep rate is smallest. In addition, one end of a k th order interval follows a slope of increasing frequency (decreasing period) while the other follows the opposite slope. For both reasons, the spread of k th-order interval durations is reduced when the window follows the vertex. This effect is greatest for large k . Aside from this difference, one expects an asymmetry similar to that observed with a window placed before the apex, because modes in response to a trough (not shown) are less sharp than for a modulation peak.

2. Noise model

The difference in peak shapes between Fig. 7(c) and (d) accounts qualitatively for the peak/trough asymmetry. To derive quantitative predictions of discrimination thresholds, the AACF can be assumed to be corrupted by addition of Gaussian noise with a standard deviation proportional, for each sample of the AACF, to the square root of the value of that sample. This noise model approximates the random accumulation of counts in a histogram bin, such as might occur if the AACF were calculated from neural spike trains within a coincidence network:

$$\text{AACF}'(\tau) = \text{AACF}(\tau) + N_\sigma \text{AACF}(\tau)^{0.5}, \quad (5)$$

where AACF' is the noisy pattern, AACF is the noiseless pattern [normalized by division by $\text{AACF}(0)$], and N_σ is a Gaussian noise of standard deviation σ and mean 0. It is further assumed that pitch is derived from the *position of the maximum* of the noisy pattern near its mode of order k . Figure 9 (top) shows the simulated standard-deviation-to-mean ratio of this statistic as a function of σ , for $k=3$.

Variability is greater for troughs than for peaks, especially at small values of σ for which it is determined mainly by the shape of the modes near their peaks. For larger values of σ , gross aspects of the mode shape affect the variability which becomes more similar for peaks and troughs. The ratio of variabilities between troughs and peaks therefore decreases as a function of σ , as plotted in Fig. 9 (bottom). It

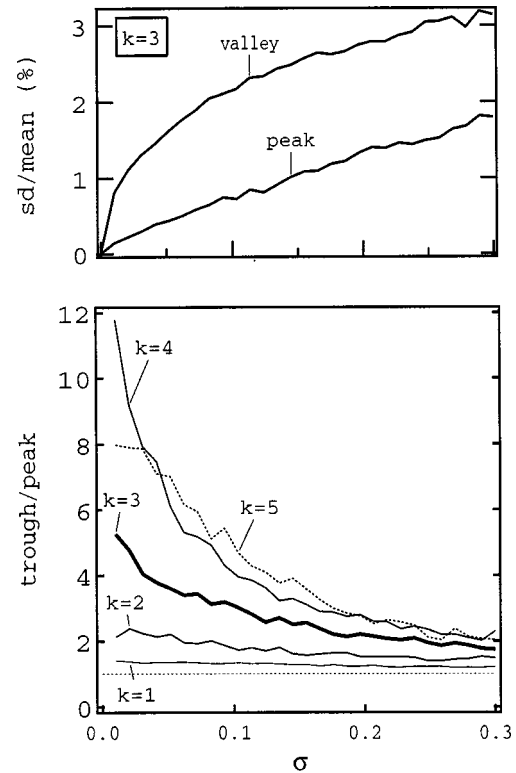


FIG. 9. Top: ratio of standard deviation to mean of pitch estimate as a function of noise standard deviation (σ) for peak and trough. The pitch estimate was derived from the position of the mode of order 3 of the AACF. Bottom: peak/trough ratio of pitch discrimination thresholds for various values of the mode order k .

also depends on the order of the mode: it increases up to $k=4$ or 5, and decreases slightly beyond that value (not shown).

The simulation used a sampling rate of 100 kHz, and variability was estimated over 5000 repetitions. The vertex frequency was 0.5 kHz. Probability functions were sharpened by raising the output of the haircell model (scaled to 0–1) to the 1000th power. The AACF was calculated with an exponentially shaped kernel of time constant $t_a=0.4$ ms, over a square integration window of duration $D=40$ ms placed before the vertex.

3. Comparison with experimental data

The model has many parameters (degree of sharpening, kernel time constant t_a , window size and position, mode order k , noise magnitude σ , etc.) that affect thresholds. To obtain specific figures, let us arbitrarily choose $\sigma=0.1$ and $k=3$. The variability of the estimate of the mode position can be read from Fig. 9 (top). Supposing that this factor governs pitch discrimination accuracy, relative thresholds are expected to be larger by a factor of $\sqrt{2}$ (Hartmann, 1997). For $\sigma=0.1$, thresholds for peaks should be about 1% (17 cents), and for troughs 3% (50 cents), with a trough/peak threshold ratio of 3. These figures are in rough agreement with those of Demany and McAnally (1994) or Demany and Clément (1995a) for cosine-modulated pure tones.

However, those authors also observed a wide range of interindividual differences. To some extent they can be ac-

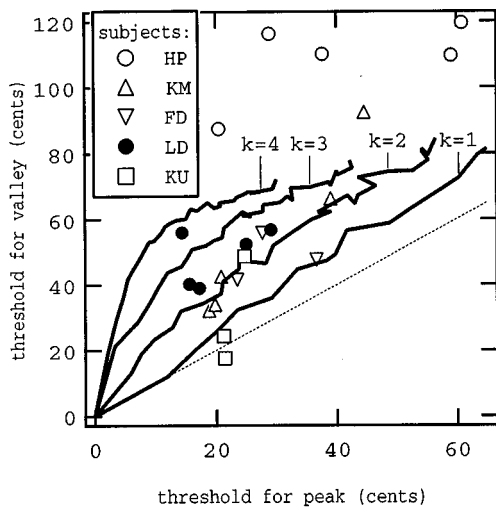


FIG. 10. Points: plot of thresholds for valleys (ordinate) versus peaks (abscissa), for different subjects and sessions of Demany and McAnally (1994). Lines: thresholds predicted by the model for various values of the mode order k .

counted for by assuming different parameter values. Figure 10 shows a scatter plot of experimental data points obtained by Demany and McAnally (1994), together with thresholds predicted by the model for several values of k , and σ ranging between 0 and 0.3. Most data points can be approximated by model predictions. However, a few data points (subject HP) lie beyond the range of model predictions. They could possibly be accounted for by assuming an additional source of variability, but asymmetry is likely to be reduced if this variability affects peaks and troughs equally. It is not clear how this issue can be resolved.

To summarize, the model produces pitch thresholds and peak/trough ratios that are in approximate agreement with experimental data obtained for cosine-modulated pure tones. In principle the explanation is also valid for complex tones, although this was not verified formally. The model would produce similar predictions at other frequencies if one supposed a uniform scaling of its temporal parameters (time constant, window size, etc.). As there is no compelling reason to expect perfectly uniform scaling, the differences across frequency noted by Demany and Clément (1995a) are not unexpected. The next section examines whether the model can be generalized to two other forms of modulation used by Demany and colleagues: triangular-shaped modulation and half-cosine ramps.

C. Triangular peaks and troughs of modulation

So far the reasoning relied on the fact that the frequency distribution is not uniform near the vertex of cosine-shaped modulation. This is no longer true if the modulation is triangle shaped, because the linear sweeps on either side of a sharp vertex imply an almost uniform distribution. The model should fail in that case, yet we know that the perceptual asymmetry is, if anything, *greater* for sharp than for flattened vertices (Demany and Clément, 1997). How to ac-

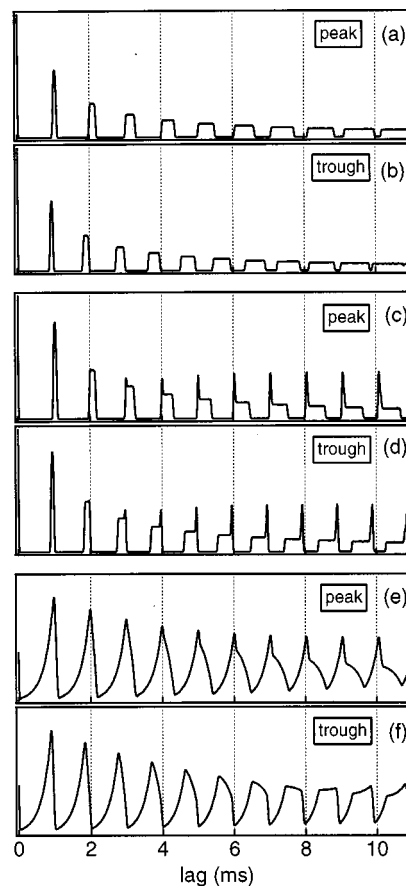


FIG. 11. (a) The ACF for a triangle-shaped peak, calculated over a 40-ms window placed before the vertex. (b) Same for a trough. (c) The ACF for a triangle-shaped peak, calculated over a 40-ms window placed after the vertex. (d) Same for a trough. (e) The AACF for a triangle-shaped window placed after the vertex. (f) Same for a trough.

count for the asymmetry, and also for the “hyperacute” frequency discrimination that Demany and Clément observed for peaks of triangular modulation?

If the integration window is placed *before* the vertex, indeed no asymmetry is observed. Figure 11(a) and (b) show the ACFs for a peak and a trough, respectively. Modes are blunt and low, all the more so as their order is high, and it does not seem that they would support accurate discrimination. If the window is placed before the vertex, one can account for neither perceptual asymmetry nor hyperacute discrimination.

If instead the integration window is placed *after* the vertex, each mode consists of two parts: a rectangular pedestal and a sharp “spire,” as shown in Fig. 11(c) for a peak of modulation. The spire represents intervals that span the peak, while the pedestal represents intervals that do not. The spire can be attributed to the effect noted in Sec. III B: if the pulse that ends an interval is displaced along one slope of the vertex, the pulse that begins it moves along the other. Any FM-induced change of period at one end is compensated by the opposite change at the other. The compensation is yet more effective than for cosine modulation, because the slopes on either side are more nearly constant and equal. For a valley of modulation the modes are similarly skewed, but in the opposite direction [Fig. 11(d)]. The spires are equally sharp

for peaks and valleys, so one cannot expect an asymmetry if discrimination is based on the ACF.

For the AACF, on the other hand, modes are sharper for a peak than for a trough [Fig. 11(e) and (f)]. This difference explains the asymmetry that Demany and Clément (1997) observed for *triangular* modulation, according to the present model. The explanation is the same as for cosine modulation, although the origin of the nonuniform interval distribution is different. *En passant*, the model accounts for the “hyperacute” discrimination observed for triangular peaks, a consequence of the mechanism that produces sharp spires in the higher-order modes. A similar hyperacuity would hold for troughs if it were not for the deleterious effect of the mechanism that produces the asymmetry.

D. Ramps

1. Cosine-shaped ramps

As pointed out earlier, the reasoning of Sec. III B for windows placed before the vertex applies directly to *ends* of ramps shaped like half-cosine periods. The model predicts an asymmetry in favor of rising ramps, as was observed by Demany and Clément (1997). For *beginnings* of ramps, one can apply the results of Sec. III B for windows placed after the vertex, with the important difference that samples to the left of the window now fall outside the signal (rather than on the opposite slope of a vertex). As a result, high-order modes are less well defined than in the case of full cosines, and discrimination should be less accurate, as Demany and Clément observed. One nevertheless still expects a peak/trough asymmetry which was indeed found (not shown).

2. Linear-shaped ramps

Contrary to all the conditions considered so far, the model predicts no asymmetry in the case of *linear* ramps. The frequency distribution is uniform, and the “compensation” mechanism that produced sharp spires is ineffective as it depends on the existence of ramps on both side of a vertex. The perceptual asymmetry should thus not be observed for linear ramps. Hyperacuity of pitch discrimination also should not be observed. Discrimination might thus actually be *worse* for the end of a ramp than for a peak flanked by linear ramps on both sides, despite the absence of the potentially interfering portion following the peak. Unfortunately linear ramps were not included in the stimuli investigated by Demany and Clément (1997), so we cannot test this prediction against their data.

IV. DISCUSSION

This paper offered an account for the perceptual asymmetry found by Demany and colleagues. It requires several strong assumptions that are reviewed here.

A. Importance and plausibility of the assumptions

Assumption 1. Assumption 1 (narrowing of the pulses of the spike probability function s) allows the shape of pulses of the convolution $s' = s \circ a$ to be determined mainly by the shape of the asymmetric kernel function $a(t)$. Such narrowing is not critical to the principle of the model, but if the

pulses of s were wider, their shape would dominate the convolution and the asymmetry of the pulses of s' would be reduced, and it is uncertain whether the model could explain peak/trough ratios as large as those observed. Certainly, the model could accommodate *several* sharp, well-spaced pulses per period instead of one.

Sharp pitch-following responses have been found in the cochlear nucleus. Temporal sharpening might be the result of the convergence of fibers on coincidence counting neurons (Kim *et al.*, 1986, 1988; Palmer and Winter, 1996), inhibition (Møller, 1983), refractory effects, etc. Standard deviations of first spike latency as small as 100 μ s have been reported by Rhode and Greenberg (1992), and there is evidence that time constants of spike integration can indeed be short, and possibly also tunable by varying the resting potential of neuron membranes (Banks and Smith, 1992). It has been proposed that such sharp responses might be a good substrate for pitch, but the proposition has also been questioned, because sharp pulses tend to be quite phase sensitive, and disappear for stimuli that lack a deeply modulated temporal envelope (Palmer and Winter, 1992). Other arguments against sharpening were put forward by Shofner (1999) and de Cheveigné (1999).

Sharpening is not a common assumption in pitch models based on autocorrelation (Licklider, 1956; Meddis and Hewitt, 1991a, b) or cancellation (de Cheveigné, 1998), but those models could certainly accommodate it. The strobed temporal integration (STI) of Patterson *et al.* (1992) is equivalent to a cross correlation between the discharge probability function and a strobe function with a single pulse per period. That strobe function is similar to the sharpened probability function assumed here, although the roles they play in the respective models are quite different.

Assumption 2. Assumption 2 (asymmetric autocorrelation) introduces the “peaked asymmetric function” that, convolved with the asymmetric interval distribution, produces a difference in accuracy between the frequency representations of peaks and troughs. This assumption is essential. It should be stressed that the opposite Kernel shape would produce the *opposite* perceptual asymmetry. It is therefore essential to justify the orientation of the asymmetry of the convolution kernel function involved in the AACF.

One can imagine at least two physiological accounts for smoothing of the delayed probability function (but not the undelayed function) by convolution with a short, rightward-skewed kernel function. (a) A first hypothetical account is that a jitter arises in the neural delay line, for example by reduction of conduction velocity following a spike. Spikes that follow closely after another spike would be delayed, adding a “tail” to the distribution of spike times on the side of greater delays. (b) A second hypothetical account is that the “coincidence counting neuron” of Licklider’s neural network is actually a *gating neuron*. The transmission of undelayed spikes is controlled by the density of delayed spikes, and the measure of this density naturally involves temporal smoothing. Each delayed spike contributes to the state of the neuron by an amount that decays exponentially, hence the exponential shape of the smoothing kernel function. Supposing that the gate controls the probability of

transmission of undelayed spikes without affecting their latency, the *undelayed* probability function remains unsmoothed. Spike probability at the output of the gate is thus the product of the undelayed input probability by the delayed probability convolved by an exponentially shaped smoothing kernel function. The spike count, across an array of such neurons indexed by delay, should resemble the AACF.

The second interpretation requires an assumption (gating neuron) that may seem unnecessary in the context of autocorrelation. However, it is congruent with the notion of “neural cancellation filter” that was used for pitch estimation in a model very similar to autocorrelation (de Cheveigné, 1998). The nuance between “gating neurons” and “coincidence counters” is significant if one imagines that the neural filters are one stage within a time-domain neural signal-processing network (Delgutte, 1984). A gating neuron can feed further processing, whereas a simple coincidence counter is a “dead end” as far as further time-domain processing is concerned. Note that if the first interpretation were true (jitter in the delay line), the convolution kernel function might be temporally wider for longer delays, rather than fixed as was assumed here.

Incidentally, the asymmetry of the AACF makes it sensitive to time reversal, and this might constitute yet another explanation of the ramp/damp asymmetries observed by Patterson (1994a, b; de Cheveigné, 1998).

Assumption 3. Assumption 3 (use of higher-order modes as pitch cues) is not essential for producing a peak/trough asymmetry for cosine modulation, although it makes things easier by “magnifying” the skew. It is required for producing an asymmetry for triangular modulation (the skew arises mainly for modes of order >1). It is also useful for explaining the hyperacuity of frequency discrimination at a triangular peak.

As pointed out earlier, higher-order mode cues are a useful hypothesis if one is to explain very accurate pitch discrimination, or the inverse dependency of thresholds on duration. They are indispensable in any model based on interval histograms derived from *single fibers*, as the first mode of both ISI and AC histograms disappears between 500 Hz and 1 kHz due to refractory effects. However, this constraint does not apply to interval statistics within *groups* of fibers, which are anyway a more reasonable assumption as they make much better use of the available temporal information (de Cheveigné, 1993, 1998).

Yost (1996) found that iterated-rippled-noise stimuli that differed by (waveform) autocorrelation modes other than the first had similar pitch strengths and were not readily discriminable. It is not clear whether that result can be generalized to imply that higher-order modes of autocorrelation histograms cannot be used in *any* task, in particular frequency discrimination with the stimuli of Demany and colleagues.

An obstacle to the use of higher-order modes is that they require long delay lines. The lack of physiological support for long delay lines is a weakness of time-domain models such as autocorrelation, and the higher-mode assumption rather makes it worse. High-order modes are also problematic in that they are multiple, and therefore ambiguous, and the auditory system would need a mechanism to keep track

of the order of the particular mode being used for a pitch match. The “narrowed autocorrelation function” (Brown and Puckett, 1989) has been proposed as a synthetic representation of all modes up to a certain order. It combines the accuracy of the higher-order modes with the lack of ambiguity of the lower-order modes (de Cheveigné, 1989; Slaney, 1990). Alternatively one can imagine an algorithmic process by which modes are registered in succession, starting at the lowest-order mode.

Other assumptions. Integration windows were chosen square for clarity, and it is not expected that other shapes would lead to significantly different results. Exponential-shaped windows are a natural choice, although estimates of temporal integration windows in other contexts (Plack and Moore, 1990) have lead to more rounded shapes.

Integration windows were chosen relatively large (40 ms) to produce legible effects in the plots, but asymmetries also occur for shorter windows. As argued above, windows should not be smaller than the stimulus period. The value of 2.5 ms suggested by Licklider (1956) and retained by Meddis and Hewitt (1991a, b) would be adequate at 1 kHz, but much too short at lower frequencies. It certainly should not be taken as a norm. One can argue that the auditory system has some latitude in choosing window sizes, and can within limits choose the best size for the task at hand (Hartmann and Klein, 1980).

The behavior of the model is evidently dependent on the range of mode *orders* that are allowed to enter the pitch estimate, and the rule by which evidence from modes of different orders is integrated. The range illustrated here was chosen arbitrarily. The asymmetry is also critically dependent on the use of the *position of the maximum* of a mode as the statistic. The matched filter of Sruлович and Goldstein (1983) would probably not show the same asymmetry. A statistic based on the position of maximum *slope* might show the opposite asymmetry, etc.

B. Significance of the model

The perceptual asymmetry found by Demany and colleagues appears to be robust and general over a wide range of stimuli. This paper offered an account on the basis of emergent properties of a sensory mechanism. One cannot exclude that the asymmetry is due instead to a cognitive process that somehow “favors peaks,” or a mechanism such as that proposed by Clément (1996) at some intermediate stage. At this point, it is probably unwise to cling dogmatically to one hypothesis over the others. The contribution of the present model is twofold: to demonstrate that a sensory-based explanation is possible, but also a *contrario* to reveal its cost in terms of assumptions.

This model should not be seen as a general-purpose model of pitch perception, or as an “improvement” over previous models. It is rather a description of the assumptions that must be added to a standard model, autocorrelation, possibly reflecting detailed properties of that model’s implementation, in order to account for this particular phenomenon. It is not worth burdening the general-purpose model with assumptions needed only to explain a restricted class of phe-

nomena. It is, however, worth keeping track of any evidence, behavioral or physiological, in favor or against them.

The number and strength of the assumptions “raise the stakes:” either the model is accepted and the assumptions constitute a set of strong constraints on auditory processes, or else it is rejected precisely for that reason, as it requires the conjunction of hypotheses that are too many and too strong. This choice depends on the plausibility of the hypotheses, which has been discussed. It also depends on the existence and plausibility of alternative explanations for both asymmetry and hyperacuity.

En passant, the model gave an account of hyperacute frequency discrimination at peaks of triangular modulation. This depends only on assumption 3, and not on the other assumptions, and so this model of hyperacute discrimination might be retained even if the complete model of pitch asymmetry is not.

Alternative explanations of the asymmetry might be built following the same principles but differing in particulars. For example, the “lemma” of Sec. II B might be applied to another representation of the frequency distribution, for example, spectral. Convolution with a kernel function of appropriate shape (skewed with a tail towards higher frequencies) might produce the appropriate form of asymmetry. Interestingly, such a shape is found in pure-tone excitation patterns at high amplitudes: convolution with the frequency distribution should yield a slightly sharper excitation pattern at peaks than at troughs. However, it was seen in the Introduction that there are several reasons to doubt an explanation based on the asymmetry of excitation patterns, and to prefer an account based on temporal cues, of which this paper is an attempt. Whether the model is a success or not, the phenomenon that was revealed by Demany and colleagues is sufficiently ubiquitous to justify serious efforts to elucidate it.

V. CONCLUSION

A model was proposed to explain the perceptual asymmetry between peaks and troughs of frequency modulation. It is based on a particular implementation of the autocorrelation model of pitch perception, in which the delayed signal is “smeared” before multiplication by the undelayed signal. The interaction between the asymmetric function that describes the smearing, and the asymmetric distribution of interspike intervals, produces cues that are more accurate for peaks than for troughs. To summarize, we have the following.

- (1) The perceptual asymmetry between peaks and troughs of frequency modulation can be explained within the context of a time-domain model of pitch perception based on a modified form of autocorrelation.
- (2) The delayed term that enters the calculation of the autocorrelation function is assumed to be convolved with a short kernel function before multiplication with the undelayed term. This kernel function is asymmetrical and skewed towards positive values of time.
- (3) The difference in pitch estimation accuracy between peaks and troughs arises from an interaction between the asymmetric convolution kernel function and the nonuni-

form distribution of interpulse intervals near the vertex of modulation. This interaction produces cues that are sharp in the case of peaks, and blunt in the case of troughs.

- (4) The nonuniform interval distribution results from two factors. The first is the rounded shape of the instantaneous frequency distribution near the peak or trough (for cosine-shaped modulation). The second results from the existence of two populations of interpulse intervals, one comprising intervals that span the vertex, the other intervals between pulses on the same slope of the vertex. For linear modulation (triangular shaped) the second factor contributes alone to the nonuniform distribution of intervals.
- (5) The relative invariance of interpulse intervals that span the vertex accounts also for “hyperacute” discrimination of triangular-shaped modulation peaks. Discrimination of troughs benefits less because of the smoothing mechanism that causes the perceptual asymmetry.
- (6) The model accounts for perceptual asymmetry between peaks and troughs of modulation, whether cosine shaped or triangular. It also accounts for perceptual asymmetry of half-cosine-shaped ramps. However, for *linear* ramps, neither perceptual asymmetry nor hyperacute discrimination are to be expected. No experimental evidence is yet available to test this prediction.
- (7) The model requires several strong assumptions in addition to the asymmetric form of the autocorrelation. Discharge probability functions originating in the periphery must be “sharpened” to narrow pulses before pitch estimation. Discrimination must rely on the high-order modes of the (asymmetric) autocorrelation histogram, rather than just the first mode.

ACKNOWLEDGMENTS

Part of this work was carried out at ATR Human Information Processing Research Laboratories, within a research agreement with the Centre National de la Recherche Scientifique and the University of Paris 7. The author thanks ATR for its kind hospitality, and the CNRS for leave of absence. Thanks to Laurent Demany for many discussions on the subject of this perceptual asymmetry, and to him, Malcolm Slaney, and one anonymous reviewer for comments on earlier versions of the manuscript.

- Banks, M. I., and Smith, P. H. (1992). “Intracellular recordings from neurobiotin-labeled cells in brain slices of the rat medial nucleus of the trapezoid body,” *J. Neurosci.* **12**, 2819–2837.
- Brown, J. C., and Puckette, M. S. (1989). “Calculation of a ‘narrowed’ autocorrelation function,” *J. Acoust. Soc. Am.* **85**, 1595–1601.
- Clément, S. (1996). “Modélisation de la perception auditive des modulations de fréquence,” Master’s thesis in Cognitive Science, Université Bordeaux 2, France.
- de Cheveigné, A. (1989). “Pitch and the narrowed autocoincidence histogram,” *Proc. ICMPAC*, Kyoto, pp. 67–70.
- de Cheveigné, A. (1993). “Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing,” *J. Acoust. Soc. Am.* **93**, 3271–3290.
- de Cheveigné, A. (1998). “Cancellation model of pitch perception,” *J. Acoust. Soc. Am.* **103**, 1261–1271.
- de Cheveigné, A. (1999). “Pitch shifts of mistuned partials: a time-domain model,” *J. Acoust. Soc. Am.* **106**, 887–897.

- Delgutte, B. (1984). "Speech coding in the auditory nerve: II. Processing schemes for vowel-like sounds," *J. Acoust. Soc. Am.* **75**, 879–886.
- Demany, L. (1997). Personal communication.
- Demany, L., and McAnally, K. I. (1994). "The perception of frequency peaks and troughs in wide frequency modulations," *J. Acoust. Soc. Am.* **96**, 706–715.
- Demany, L., and Clément, S. (1995a). "The perception of frequency peaks and troughs in wide frequency modulations. II. Effects of frequency register, stimulus uncertainty, and intensity," *J. Acoust. Soc. Am.* **97**, 2454–2459.
- Demany, L., and Clément, S. (1995b). "The perception of frequency peaks and troughs in wide frequency modulations. III. Complex carriers," *J. Acoust. Soc. Am.* **98**, 2515–2523.
- Demany, L., and Clément, S. (1997). "The perception of frequency peaks and troughs in wide frequency modulations. IV. Effects of modulation waveform," *J. Acoust. Soc. Am.* **102**, 2935–2944.
- Demany, L., and Clément, S. (1998). "The perceptual asymmetry of frequency modulation," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London).
- Dooley, G. L., and Moore, B. C. J. (1988). "Duration discrimination of steady and gliding tones: a new method for estimating sensitivity to rate of change," *J. Acoust. Soc. Am.* **84**, 1332–1337.
- Hartmann, W. H. (1997). *Signals, Sound and Sensation* (AIP, Woodbury, NY).
- Hartmann, W. M., and Klein, M. A. (1980). "Theory of frequency modulation detection for low modulation frequencies," *J. Acoust. Soc. Am.* **67**, 935–946.
- Kim, D. O., Rhode, W. S., and Greenberg, S. R. (1986). "Responses of cochlear nucleus neurons to speech signals: neural encoding of pitch, intensity and other parameters," in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson (Plenum, New York), pp. 281–282.
- Kim, D. O., and Leonard, G. (1988). "Pitch-period following response of cat cochlear nucleus neurons to speech sounds," in *Basic Issues in Hearing*, edited by H. Duifhuis, J. W. Horst, and H. P. Wit (Academic, London), pp. 252–260.
- Licklider, J. C. R. (1956). "Auditory frequency analysis," in *Information Theory*, edited by C. Cherry (Butterworth, London), pp. 253–268.
- Meddis, R., and Hewitt, M. J. (1991a). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification," *J. Acoust. Soc. Am.* **89**, 2866–2882.
- Meddis, R., and Hewitt, M. J. (1991b). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II. Phase sensitivity," *J. Acoust. Soc. Am.* **89**, 2883–2894.
- Møller, A. R. (1983). *Auditory Physiology* (Academic, New York).
- Moore, B. C. J. (1973). "Frequency difference limens for short-duration tones," *J. Acoust. Soc. Am.* **54**, 610–619.
- Palmer, A. R., and Winter, I. M. (1992). "Cochlear nerve and cochlear nucleus responses to the fundamental frequency of voiced speech sounds and harmonic complex tones," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 231–239.
- Palmer, A. R., and Winter, I. M. (1996). "The temporal window of two-tone facilitation in onset units of the ventral cochlear nucleus," *Audiol. Neuro-Otol.* **1**, 12–30.
- Patterson, R. D. (1994a). "The sound of a sinusoid: time-domain models," *J. Acoust. Soc. Am.* **96**, 1419–1428.
- Patterson, R. D. (1994b). "The sound of a sinusoid: Spectral models," *J. Acoust. Soc. Am.* **96**, 1409–1418.
- Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (1992). "Complex sounds and auditory images," in *Auditory Physiology and Perception*, edited by Y. Cazals, K. Horner, and L. Demany (Pergamon, Oxford), pp. 429–446.
- Plack, C. J., and Moore, B. C. J. (1990). "Temporal window shape as a function of frequency and level," *J. Acoust. Soc. Am.* **87**, 2178–2187.
- Rhode, W. S., and Greenberg, S. (1992). "Physiology of the cochlear nuclei," in *The Mammalian Auditory Pathway: Physiology*, edited by A. N. Popper and R. R. Fay (Springer Verlag, New York), pp. 94–152.
- Ruggero, M. A. (1992). "Physiology of the auditory nerve," in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer Verlag, New York), pp. 34–93.
- Shepard, R. N. (1964). "Circularity in judgments of relative pitch," *J. Acoust. Soc. Am.* **36**, 2346–2353.
- Shofner, W. P. (1999). "Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: analysis of neural autocorrelograms," *J. Neurophysiol.* **81**, 2662–2674.
- Slaney, M. (1990). "A perceptual pitch detector," *Proc. ICASSP-90*, pp. 357–360.
- Srulovicz, P., and Goldstein, J. L. (1983). "A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum," *J. Acoust. Soc. Am.* **73**, 1266–1276.
- Yost, W. A. (1996). "Pitch strength of iterated rippled noise," *J. Acoust. Soc. Am.* **100**, 3329–3335.