Psychophysical evidence against the autocorrelation theory of auditory temporal processing

Christian Kaernbach[a)]
Institut für Allgemeine Psychologie, Universität Leipzig, Seeburgstraße 14/20, D-04 103 Leipzig, Germany

Laurent Demany[b)]
Laboratoire de Neurophysiologie, UMR CNRS 5543, Université Bordeaux 2, 146 rue Léo-Saignat, F-33 076 Bordeaux, France

(Received 6 July 1997; revised 30 January 1998; accepted 22 June 1998)

Nowadays, it is widely believed that the temporal structure of the auditory nerve fibers' response to sound stimuli plays an important role in auditory perception. An influential hypothesis is that information is extracted from this temporal structure by neural operations akin to an autocorrelation algorithm. The goal of the present work was to test this hypothesis. The stimuli consisted of sequences of unipolar clicks that were high-pass filtered and mixed with low-pass noise so as to exclude spectral cues. In experiment 1, ''interfering'' clicks were inserted in an otherwise periodic (isochronous) click sequence. Each click belonging to the periodic sequence was followed, after a random portion of the period, by one interfering click. This disrupted the detection of temporal regularity, even when the interfering clicks were 5 dB less intense than the periodic clicks. Experiments 2–4 used click sequences that showed a single peak in their autocorrelation functions. For some sequences, this peak originated from ''first-order'' temporal regularities, that is from the temporal relations between consecutive clicks. For other sequences, the peak originated instead from ''second-order'' regularities, relative to nonconsecutive clicks. The detection of second-order regularities appeared to be much more difficult than the detection of comparable first-order regularities. Overall, these results do not tally with the current autocorrelation models of temporal processing. They suggest that the extraction of temporal information from a group of closely spaced spectral components makes no use of time intervals between nonconsecutive peaks of the amplitude envelope. © *1998 Acoustical Society of America.* [S0001-4966(98)00810-8]

PACS numbers: 43.66.Ba, 43.66.Hg, 43.66.Mk [RHD]

INTRODUCTION

A complex tone with a rich spectrum, such as a vowel, is normally perceived as *one* sound with *one* pitch. The heard pitch is very close to that of a pure tone with the same period, even if the corresponding pure tone is actually absent in the spectrum of the complex tone (the ''missing fundamental phenomenon''). It is much more difficult to perceive a complex tone as a sum of pure tones with various pitches. This is very remarkable, for two reasons. First, the cochlea behaves as a spectral analyzer and resolves the lower harmonics of a vowellike sound. Second, these lower harmonics appear to play a more important role than the higher harmonics, unresolved by the cochlea, in the process of pitch extraction (Plomp, 1967; Ritsma, 1967; Moore *et al.*, 1985).

In the 1970s, ''pattern-recognition'' theories of pitch perception were proposed to account for the missing fundamental phenomenon and the importance of spectral resolution in pitch extraction (e.g., Terhardt, 1972; Goldstein, 1973). Basically, these theories assume that the pitch of a complex tone is extracted by a centrally located processor of the frequency relationships between resolved spectral components. The pattern-recognition theories can explain, in addition to the missing fundamental phenomenon, the pitch percepts induced by inharmonic complex tones, or the pitch ambiguity of complex tones consisting of only few harmonics. However, a fundamental problem for these theories is that pitch percepts can be elicited by the periodicity of sounds consisting of completely unresolved harmonics, or by other stimuli providing no spectral pitch cues (Burns and Viemeister, 1976, 1981; Moore and Rosen, 1979; Houtsma and Smurzynski, 1990).

To account for the latter fact, it is necessary to admit that pitch can be extracted by a mechanism working exclusively in the temporal domain. One may think that this temporal mechanism is used only for the processing of *amplitude envelopes*, and coexists with a completely different central processor of spectral cues (Terhardt, 1972; Carlyon and Shackleton, 1994). However, it is well established that the frequency of a resolved harmonic has in itself a temporal representation in the spike trains conveyed by the auditory nerve fibers responding to this harmonic (Sachs and Young, 1980; Horst *et al.*, 1986). Thus, temporal information might be used to identify the frequency of individual harmonics as a basis of a pattern recognition process (Srulovicz and Goldstein, 1983).

Moore (1977, 1997) argued that most of the psychophysical data concerning the pitch of complex sounds can be understood on the basis of a simpler model. According to Moore, the pitch of a complex sound would simply correspond to the most frequent interspike interval (ISI) occurring

a)Electronic mail: chris@psychologie.uni-leipzig.de
b)Electronic mail: Laurent.Demany@psyac.u-bordeaux2.fr

in the responses of all the auditory nerve fibers excited by this sound. In a nerve fiber excited by a resolved spectral component with frequency f (Hz), consecutive spikes will typically be separated by ISIs corresponding to $1/f, 2/f, 3/f, \ldots, n/f$ s (Sachs and Young, 1980). In other nerve fibers excited by another resolved component, the ISIs will be partly different, but common ISIs will occur if the two components are harmonically related (i.e., if the sound is periodic). The smallest of the common ISIs will correspond to the period of the sound. As the corresponding ISI should also occur in fibers excited by the sum of several harmonics rather than by a single harmonic (Evans, 1978), this ISI should be overall the most frequent one. Note that although Moore's model posits that the pitch extraction process is the same for spectrally resolvable sounds and unresolvable sounds, it is possible in this conceptual framework to make sense of the fact that resolved harmonics provide more salient pitch cues than unresolved harmonics (see Moore, 1997).

Moore's model identifies a possible correlate of the pitch of complex sounds at the auditory nerve level, but does not specify how pitch is neurally represented at higher levels of the auditory system. Because it seems that fine-grain temporal information cannot be represented directly in the auditory cortex (e.g., de Ribaupierre *et al.*, 1972; Steinschneider *et al.*, 1980), any temporal correlate of pitch in the auditory nerve is likely to be recoded beyond into place information (Pantev *et al.*, 1989; Langner, 1992; Langner *et al.*, 1997). How could this be done?

More than two decades earlier, Licklider (1951) hypothesized that the auditory system is able to calculate the autocorrelation (AC) function of a neural spike train, and to transform in this way temporal regularities into a place code for pitch. The neural scenario imagined by Licklider is depicted in Fig. 1. Nowadays, this specific neural scenario is often judged unrealistic, but Licklider's basic proposal is still very influential (Lyon, 1984; Slaney and Lyon, 1990; Lazzaro and Mead, 1989; Assmann and Summerfield, 1990; Meddis and Hewitt, 1991; de Cheveigné, 1993, 1998; Hartmann, 1993; Patterson *et al.*, 1996; Yost *et al.*, 1996; Cariani and Delgutte, 1996a, 1996b). Remark, however, that the temporal regularities liable to be picked up in a spike train by an AC process are not identical to those considered as relevant for pitch by certain pitch theorists (Goldstein and Srulovicz, 1977; Srulovicz and Goldstein, 1983; Ohgushi, 1978; van Noorden, 1982). These authors posited that the relevant information is limited to *first-order* ISIs, that is to intervals between *consecutive* spikes. By contrast, an autocorrelator will not distinguish first-order ISIs from higher-order ISIs, that is, consecutive spikes from nonconsecutive spikes (see Hartmann, 1997, p. 355).

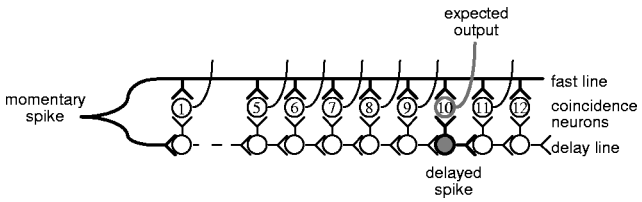

FIG. 1. A neural autocorrelator (after Licklider, 1951). A set of coincidence neurons is placed between a fast line and a delay line. The delay line is realized as a chain of neurons losing approximately 1 ms per synaptic transmission. The figure shows what happens when a spike enters into the system while another spike came in 10 ms earlier. The 10-ms coincidence neuron is about to fire. In Licklider's original model, each coincidence neuron is followed by another neuron doing some temporal integration.

The present study was intended to test the idea that the analysis of temporal regularity can be based on an AC process. To this end, we performed psychophysical experiments using sound stimuli that did not provide spectral cues and had the advantage of producing precisely predictable temporal patterns of neural activity. These stimuli were high-pass filtered nonperiodic click sequences, mixed with low-pass noise. They did not elicit truly *musical* pitch sensations, in so far as their pitches were weak and could not be used to build identifiable musical intervals. Nevertheless, it is reasonable to consider that they are able to provide information on a mechanism of pitch perception. Pitch can probably be derived from both spectral and temporal features of sound waves. In order to isolate the temporal mechanism of pitch perception, one has to eliminate spectral cues. By doing so one eventually reduces pitch strength up to a point where musical judgments can no longer be made. The remaining perceptual correlate of temporal regularity has been named ''rattle pitch'' by Plomp (1976, Chap. 7).

I. EXPERIMENT 1. THE POOR DETECTABILITY OF ''SECOND-ORDER PERIODICITY''

A. Preliminary observations

Consider a periodic click sequence in which consecutive clicks are separated by a constant interclick interval (ICI) of, e.g., 10 ms. Let us remove the resolvable spectral components of this stimulus by high-pass filtering it at, e.g., 6000 Hz. In addition, let us mix it with low-pass noise to ensure that its *internal* (i.e., auditory) power spectrum will not contain resolved components arising from cochlear nonlinearities (Plomp, 1976, Chap. 2). Under such conditions, one can hear a clear ''rattle pitch,'' which must be extracted from purely temporal information at the auditory nerve level. The AC function of the filtered click sequence shows a series of sharp peaks for delays of 10, 20, 30,... ms. It is reasonable to assume that, in the auditory nerve, each filtered click produces a short burst of activity, so that the ICIs are represented by ISIs of the same duration (Kiang *et al.*, 1965; Ruggero, 1992).

Suppose now that one ''interfering click'' is inserted at a random position within each 10-ms first-order ICI of the periodic click sequence. The top panel of Fig. 2 shows a segment of a stimulus obtained in doing so. There are no longer *first-order* temporal regularities in the click sequence: The ICIs of consecutive clicks have a flat statistical distribution. However, temporal regularities appear in the *second-order* ICI statistics, and the sequence can be said to have a ''second-order periodicity'' (SOP). Its AC function is displayed in the left part of Fig. 3. In spite of the interfering clicks, prominent peaks are still present for delays of 10, 20, 30,... ms. Yet, in informal listening tests, we found that the sequence does not sound regular. Instead, it is perceived as

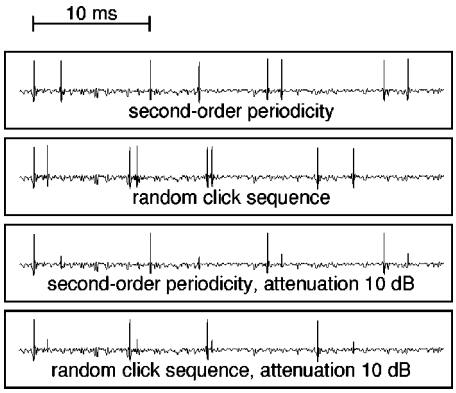

FIG. 2. High-pass filtered click sequences mixed with low-pass noise. The top panel shows a click train possessing a second-order periodicity (SOP) of 10 ms. It consists of an isochronous click train with one ''interfering'' click in each first-order ICI. It is not easy to realize visually that the first, third, fifth, and seventh clicks are equidistant. This click train looks basically similar to the random click train (without SOP) presented in the second panel. Here, each first-order ICI is randomly selected in the interval [0,10] ms. The SOP of the third train is much easier to see than the SOP of the first one. In this third train, the interfering clicks are three times smaller (10-dB attenuation) than the isochronous clicks. The bottom panel shows a comparable random click train with every even click attenuated by the same amount. While the SOP is easy to see with a 10-dB difference, it is difficult to hear.

similar to a sequence in which each first-order ICI is selected randomly, without any constraint, between 0 and 10 ms. We also noticed, however, that a perceptual discrimination between sequences with SOP and random sequences was possible on the basis of local (momentary) differences in click rate: In a random sequence, many short (or long) first-order ICIs sometimes occur in immediate succession; this cannot happen in a sequence with SOP. The corresponding discrimination cue disappears if, in the random sequences, the number of consecutive first-order ICIs falling above 5 ms, or below 5 ms, is prevented to exceed 2. Using this constraint, we observed that it was extremely hard to discriminate sequences with SOP from random sequences.

Experiment 1 was conducted to confirm this informal finding. In order to quantify the deleterious effect of the interfering clicks, we attenuated them by a variable amount, as illustrated in the third panel of Fig. 2. In the random sequence presented on the same trial as a given sequence with SOP, every even click was attenuated by the same amount (lowest panel of Fig. 2). For an attenuation of 20 dB, a sequence with SOP sounded perfectly regular. The goal of the experiment was to determine at which amplitude of the interfering clicks the detection of SOP would be disrupted.

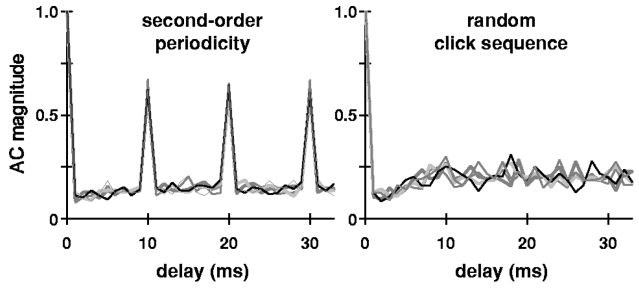

FIG. 3. Normalized AC functions of click trains used in experiment 1. Left part: AC functions of five click trains with a SOP of 10 ms. Right part: AC functions of five random click trains, excluding multiple repetitions of similar first-order ICIs (as explained in Sec. I A). The click trains analyzed here were composed of nonfiltered clicks with equal intensities. The AC functions were computed for delays of 0–33 ms, with a bin width of 1 ms.

B. Procedure

The click sequences were digitally generated at a sampling rate of 44 100 Hz. The unipolar clicks were high-pass filtered at 6000 Hz. The filter shape followed a logistic function with a 400-Hz transition region (from 10% to 90% of full amplitude). The low-pass noise mixed with the click sequences consisted of white noise filtered symmetrically, so that the entire stimulus had a flat spectral envelope (spectrum level: 35 dB) when the interfering clicks were unattenuated. The stimuli were presented diotically, via electrostatic earphones (Stax Lambda Pro), in a sound-proof booth.

On each trial, the subject was presented with two 1-s click sequences separated by a 333-ms pause. The low-pass filtered noise started 333 ms before the first sequence and ended 333 ms after the second one. The two sequences included a ''target'' sequence (with SOP) and a ''random'' sequence (excluding multiple repetitions of similar first-order ICIs, as explained in the previous section). The subject had to determine if the more regular sequence was the first or the second sequence. Feedback was provided following each trial. In a block of trials, the attenuation of the interfering clicks was initially set to 20 dB, and then varied according to the weighted up-down adaptive procedure (Kaernbach, 1991): Following a correct response, the attenuation decreased by 2 dB (before the first reversal) or 1 dB (after the first reversal); following each incorrect response, the attenuation increased by 3 dB. This continued until 60 trials were run. Three psychology students, without previous experience in psychoacoustic tasks, were tested each in two trial blocks.

C. Results

Since the results of the three subjects were very similar, they were pooled. (Author CK also produced similar results, which were not taken into account.) A psychometric function was fitted to the data using a maximum-likelihood procedure. Performance severely declined when the attenuation became smaller than about 10 dB. The 75% point of the psychometric function corresponded to an attenuation of 9.2 dB.

D. Discussion

This short experiment demonstrated that, for untrained subjects at least, the SOP of the target sequences was inaudible when the interfering clicks had the same intensity as the isochronous clicks, or when they were attenuated by as much as 5 dB. In the absence of any attenuation, the AC functions of the target sequences had prominent peaks at 10, 20, 30,... ms, as shown in the left part of Fig. 3. The prominence of the AC peaks was even larger when the interfering clicks were attenuated by 5 dB. Even then, the target sequences could not be discriminated from sequences with no

AC peak at all (right part of Fig. 3). This clearly casts doubts on the idea that the auditory system is able to compute AC functions.

The fact that the attenuation of the interfering clicks had to exceed as much as 5 dB before being effective is not so surprising if one considers data recently reported by Tsuzaki and Patterson (1998). These authors measured thresholds for the detection of amplitude jitter in high-pass filtered isochronous click trains. The obtained thresholds were remarkably high: For first-order ICIs of 10 ms (the shortest ICIs used by Tsuzaki *et al.*), they corresponded to interclick amplitude differences of no less than 7 dB.

It should be emphasized that the AC functions displayed in Fig. 3 are valid estimations of the AC functions of the activity produced by the click sequences in the auditory nerve. Consider, in this respect, a model of neural transduction assuming that the clicks are bandpass filtered in a number of frequency channels (with center frequencies exceeding 6 kHz), half-wave rectified, and finally low-pass filtered at about 1 kHz. In a given channel, the resulting signal will consist of smeared versions of the original clicks, spread over approximately 1 ms. Therefore, its AC function will be quite comparable to the AC function computed from the original click sequence with a bin width of 1 ms (the bin width we used). This will hold true as long as the clicks are all of the same amplitude and polarity. Meddis and Hewitt (1991) proposed that the neural AC functions should be averaged across channels. For high-pass filtered click sequences with a cutoff frequency as high as 6 kHz, the neural AC functions should be essentially identical across channels, so that averaging across the relevant channels will still produce results similar to those displayed in Fig. 3.

Our basic result is consistent with findings by Carlyon (1996) on the perception of mixtures of complex tones differing in fundamental frequency. In some of his experimental conditions, Carlyon mixed two spectrally unresolvable tones with identical spectral envelopes and amplitudes. He found that such a mixture is not heard as a sum of two tones differing in pitch but evokes instead a ''unitary noiselike or 'crackle' percept.'' This shows, like our own experiment, that the detection of a sound's periodicity can be dramatically disrupted by the simultaneous presentation of another sound which is not more intense. The disrupting sound was periodic in Carlyon's study, whereas it was not periodic in our experiment. In both cases, however, amplitude peaks of the disrupting sound occurred between consecutive amplitude peaks of the other (periodic) sound. Thus, both sets of data provide an indication that first-order time intervals are of paramount importance for the extraction of pitch in the temporal domain.

II. EXPERIMENTS 2 AND 3. THE PERCEPTUAL NONEQUIVALENCE OF FIRST-ORDER AND SECOND-ORDER TEMPORAL REGULARITIES

A. Purpose and general method

Experiment 1 demonstrated that second-order temporal regularities are difficult to hear in spectrally unresolvable click trains. It was the aim of experiments 2 and 3 to quantify the sensitivity of trained subjects to temporal regularities of different orders. The subject's task was again to discriminate ''target'' sequences with temporal regularities from irregular, ''random'' sequences. However, unlike the target sequences of experiment 1, those of experiments 2 and 3 had no ''periodicity'' of any kind. More precisely, there was only one peak in their AC functions. This peak originated from multiple occurrences of a fixed ICI, which was a first-order ICI for some targets (without second-order regularities), and a second-order ICI for other targets (without first-order regularities). In the targets with fixed first-order ICIs, the fixed ICI occurred more or less frequently, so that the AC peak was more or less prominent. The prominence of the AC peak was thus experimentally dissociated from the nature of the temporal regularities producing the peak.

In contrast to experiment 1, the component clicks of the sequences employed never differed from each other in intensity. They were always identical. Thus, discrimination performance was not assessed as a function of an intensity variable. In experiment 2 we manipulated instead of this the *length* of the sequences: For targets of various types, we measured how long a target sequence had to be, that is to say, how many fixed ICIs it had to contain, in order to be reliably discriminated from a random sequence of the same length.

B. Experiment 2: Procedure

In the target and random sequences presented on each trial of this experiment, the average click rate was the same and the first-order ICIs had an identical upper limit (of τ ms). Four types of targets were used. Their respective temporal characteristics are specified and illustrated in Fig. 4, as well as those of the comparison random sequences (constrained by a rule which was similar to that employed in experiment 1; see the caption of Fig. 4). Each target contained both fixed and random ICIs. The fixed ICIs were first-order ICIs for target types labeled **kxx**, **kxxx**, and **kxxxx**. In **abx** targets, by contrast, all the fixed ICIs were second-order ICIs; there was always one click at a random position between two clicks separated by the fixed ICI value. As shown in the bottom row of Fig. 4, the fixed ICIs of the various targets produced a single sharp peak in their otherwise noisy AC functions. (The singleness of this peak was due to the ''**x**'' ICIs.) The peak occurred at $\tau/2$ for the targets with first-order regularities, and at τ for the **abx** targets. Signal-to-noise ratios (S/N) were computed to quantify the prominence of each peak in the asymmetric noisy background surrounding it. In doing so, we selected the noise falling in a 6-ms region centered on the peak. Arranging the targets in order of decreasing S/N, one obtained: **kxx** (0.45), **abx** (0.34), **kxxx** (0.31), and **kxxxx** (0.24). There was no peak in the AC functions of the random sequences. Hence, under the assumption that the auditory system is able to perform operations akin to AC, one predicted that it would be easier to discriminate an **abx** target from a random sequence than to discriminate a **kxxx** or **kxxxx** target from a random sequence.

In the **kxx**, **kxxx**, and **kxxxx** targets, τ was always equal to 10 ms. In the **abx** targets, τ was set to 10 ms in some trial blocks, and to 5 ms in other trial blocks. This permitted a

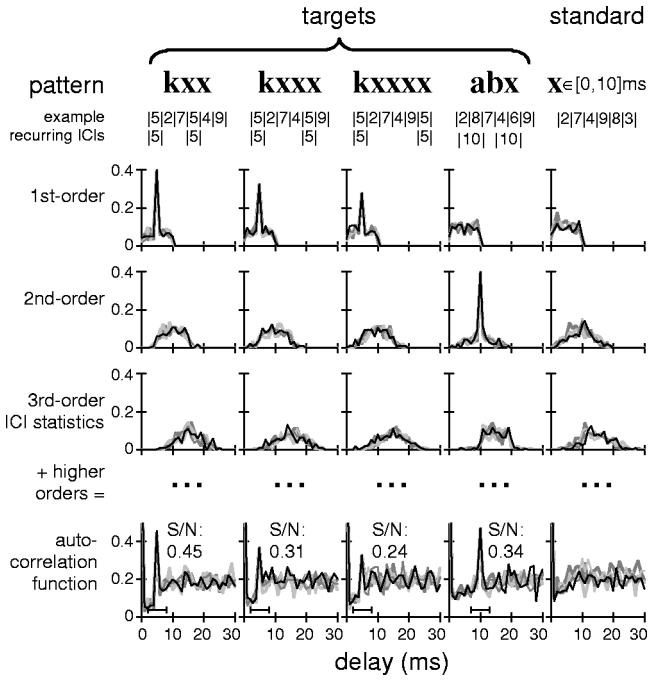

FIG. 4. Temporal characteristics of the sequences used in experiment 2. For each sequence type (patterns on the first row), we give on the second row a possible succession of first-order ICIs, in ms, τ being equal to 10 ms. The recurrence of a fixed ICI is emphasized on the third row. In a **kxx** sequence, a first-order ICI of $\tau/2$ (corresponding to **k**) is followed by two first-order ICIs (the **x**'s) which are randomly selected in the interval $[0,\tau]$; this pattern is then iterated. The **kxxx** and **kxxxx** sequences are constructed similarly, but with less frequent occurrences of the fixed ICI **k**. In an **abx** sequence, **a** is randomly selected in the interval $[0,\tau]$, **b** is such that $\mathbf{a}+\mathbf{b}=\tau$, and **x** is again taken randomly between 0 and τ. In a random (standard) sequence, all the first-order ICIs are taken randomly between 0 and τ, except that the number of consecutive first-order ICIs falling above $\tau/2$, or below $\tau/2$, cannot be larger than in the target sequence used on the same trial (e.g., larger than 3 for a **kxxx** target). Three rows of panels present normalized statistical distributions of the first-order, second-order, and third-order ICIs recorded in 1-s stimuli. The bin width is 1 ms. For the **kxx**, **kxxx**, and **kxxxx** sequences, the distribution of the first-order ICIs shows a peak and there is no peak in the higher-order distributions. In the **abx** case, by contrast, the peak occurs in the distribution of the second-order ICIs; the first-order ICIs are distributed exactly like the first-order ICIs of the random sequences. The sequences' AC functions, shown on the bottom row, are the sums of all their ICI statistics. The bar below each AC peak shows the region (−3 to +3 ms relative to the peak) that was integrated to compute S/N.

comparison between the detections of first-order and second-order temporal regularities for the same mean click rate (200 clicks per s) or for identical locations of the AC peak (a peak obtained for a 5-ms delay).

On each trial, the subject was presented with three successive click sequences, separated by 250-ms pauses. The clicks were high-pass filtered at 6000 Hz and mixed with low-pass noise as in experiment 1. The low-pass noise started 250 ms before the first sequence and ended 250 ms after the third one. The first sequence was a random sequence and served as a standard. The following two sequences included one target and one random sequence. The subject had to determine if the ''different'' sequence (the target) was the second or the third sequence. Feedback was provided immediately after the response. The stimuli were presented monaurally, via a Stax Lambda Pro earphone, at a 35-dB spectrum level.

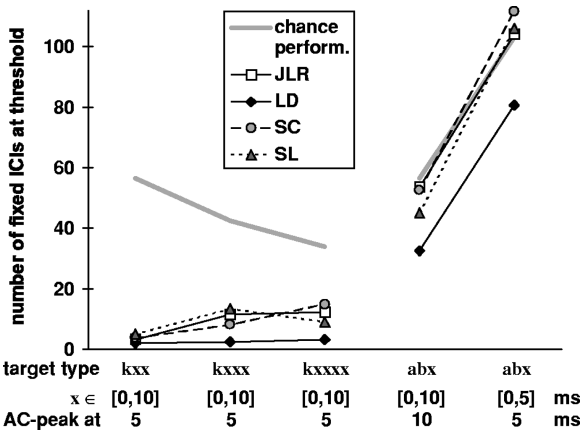

FIG. 5. Results of experiment 2. The ordinate shows the numbers of fixed ICIs needed by the four subjects to discriminate the various targets (abscissa) from random sequences. For the **abx** targets, the threshold estimates were strongly biased. A thick gray line indicates the results expected from a subject performing the task at chance level.

In each block of trials, the type of the targets and the value of τ were fixed. The duration of the sequences was initially set to 1 s and then varied following the same adaptive procedure as that used in experiment 1 (1 dB corresponding here to a duration change of approximately 26%). However, 1 s was the maximum possible duration. Due to the partial randomness of the ICIs, the number of fixed ICIs contained by a sequence of a certain duration could vary. Following each trial, we recorded the exact number of fixed ICIs which occurred on that trial. A block was finished after 100 trials. From the obtained data, we estimated the number of fixed ICIs for which the probability of a correct response was 0.75. This ''threshold'' was taken as the median of the numbers of fixed ICIs which had been presented on all the trials following the fourth reversal.

During a test session, one block of trials was run in each of the five experimental conditions. Each subject was tested for at least five training sessions before the formal experiment, run in five additional sessions. Four subjects were used: three students and author LD. Each subject had previously participated in other psychoacoustical experiments.

C. Experiment 2: Results

Figure 5 shows the mean threshold estimates obtained in the five final sessions. The thick gray line indicates the threshold estimates expected if the targets were actually not discriminated at all from random sequences. For a complete absence of discrimination, the adaptive procedure resulted in a random walk and simulations showed that the duration threshold would be estimated at 0.85 s (15% less than the maximum duration, 1 s). The number of fixed ICIs corresponding to this duration depended of course on the target type and on τ.

For the **abx** targets, the performance of three subjects did not differ significantly from the chance level indicated by the gray line; the measured thresholds were thus strongly biased by ceiling effects. By, contrast, unbiased thresholds could always be measured for the **kxx**, **kxxx**, and **kxxxx** targets. For each subject, the discrimination task was clearly

TABLE I. Percentages of correct responses obtained in the three conditions of experiment 3.

Target type		**kxxxx**	**abx**	**abx**
τ (ms)		10	10	5
Mean number of fixed ICIs		12	20	40
Subject	JLR	77.6	49.2	46.6
	LD	89.4	59.6	59.6
	SC	74.8	61.0	50.4
	SL	79.0	65.0	49.4
Mean		80.2	59.7	51.5

much more difficult for the **abx** targets than for all the targets with first-order regularities, even though S/N could be markedly smaller in the latter case.

D. Experiment 3

Experiment 3, a variant of experiment 2, was performed on the same subjects. In this new experiment, the duration of the three sequences presented on each trial was no longer varied adaptively but fixed at 300 ms. Thus, we did not measure discrimination thresholds but simply percentages of correct responses $[P(C)]$. This was done for three categories of targets: (1) **kxxxx** targets with $\tau=10$ ms; (2) **abx** targets with $\tau=10$ ms; and (3) **abx** targets with $\tau=5$ ms. In each of these three conditions, each subject was tested in five blocks of 100 trials. There was no need of preliminary training blocks as the subjects had been tested soon before in experiment 2.

Table I displays the $P(C)$ values obtained in the three experimental conditions. The third row of this table indicates the mean number of fixed ICIs contained by the various targets. Given that each subject performed 500 trials in each condition, a $P(C)$ larger than 55.3% exceeded the chance level with $p<0.01$. For the **abx** targets, $P(C)$ sometimes exceeded the chance level, but never markedly; two subjects (SC and SL) were more successful when τ was 10 ms than when τ was 5 ms, but this was not the case for the other two subjects (JLR and LD). For the **kxxxx** targets, $P(C)$ was always much higher, although the number of fixed ICIs contained by these targets was systematically smaller.

III. EXPERIMENT 4. SOME ADDITIONAL DATA ON THE DETECTABILITY OF FIRST-ORDER REGULARITIES

In experiment 4, we readopted the adaptive procedure used in experiment 2 to examine, in three trained listeners, the effect of τ on the discrimination of **kx** targets from random sequences. In a **kx** sequence, a fixed first-order ICI of $\tau/2$ ms alternates with a first-order ICI which is randomly chosen between 0 and τ ms. We employed such targets because these were the simplest sequences containing only first-order temporal regularities. The number of fixed ICIs necessary for their discrimination from random sequences (matched in average click rate, and with the same maximum first-order ICI) was measured using four values of τ: 5, 10, 20, and 40 ms. In ten blocks of 100 trials, ten threshold measurements were made for each subject and value of τ.

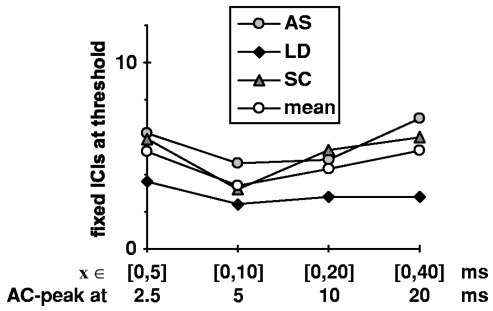

FIG. 6. Results of experiment 4. Numbers of fixed ICIs necessary to discriminate random sequences from **kx** targets with $\mathbf{k}=\tau/2$ and $\mathbf{x}\in[0,\tau]$, as a function of τ, for three subjects.

The average thresholds measured in the last five blocks of trials are displayed in Fig. 6. This figure shows that a reliable discrimination of the targets never required more than eight fixed ICIs. The best thresholds, corresponding to less than six fixed ICIs, were obtained for $\tau=10$ and 20 ms, that is, for the detection of pitches corresponding to 100 and 200 Hz. Remarkably, these are the typical pitches of male (100 Hz) and female (200 Hz) speech.

IV. GENERAL DISCUSSION

The present study was intended to test the idea that the auditory system acts as an autocorrelator in order to extract temporal information from a sound. A rudimentary version of that idea would be that the auditory system calculates the AC function of the sound waveform itself. To reject this model, it is sufficient to note that the AC function of a signal is independent of its phase spectrum but that the pitch of a complex tone consisting of unresolved harmonics does depend on the tone's phase spectrum (e.g., Moore, 1997; Houtsma and Smurzynski, 1990). More interesting is the idea that the AC is actually performed on sequences of neural spikes, after cochlear filtering. In recent years, this idea was supported by, e.g., Meddis and Hewitt (1991). They proposed a quantitative model of pitch perception according to which the auditory nerve response to a sound is processed by a bank of autocorrelators operating in different frequency-selective channels. The AC functions so obtained are supposed to be averaged across channels to generate a ''summary AC function,'' and the model assumes that the pitch of the sound corresponds to the highest point of this summary AC function. Other recent auditory models, especially the ''neural cancellation model'' of de Cheveigné (1993, 1998) and the ''auditory image model'' of Patterson (Patterson *et al.*, 1992, 1995), rest on very closely related assumptions.

Psychophysicists willing to test this family of models are clearly required to use stimuli that will have largely *predictable* temporal representations at the auditory nerve level. Such was the case of the high-pass filtered click sequences that we used. For a given auditory nerve fiber excited by these click sequences, one could reasonably consider that most of the ISIs would correspond to ICIs present in the stimuli (Kiang *et al.*, 1965; Ruggero, 1992). Moreover, it was reasonable to consider that the compound neural activity

of the fibers excited by the stimuli would be very similar to the stimuli themselves, each click being represented by many synchronous spikes (Cariani and Delgutte, 1996b). If this is admitted, then our results are not consistent with the models mentioned above. Especially, these models seem to be contradicted by our finding that it is much easier to detect temporal regularities in **kxxxx** sequences than in **abx** sequences (experiments 2 and 3). Our results apparently imply that very little or no information is conveyed by time intervals between nonconsecutive neural spikes.

From recent studies on the perception of ''iterated rippled noise'' (IRN), Yost (1996), Yost *et al.* (1996), and Patterson *et al.* (1996) concluded that the pitch of IRN is hard to explain in spectral terms but can be simply explained under the hypothesis that pitch extraction rests on the analysis of temporal regularities in the stimuli. More specifically, they suggested that the pitch salience of an IRN stimulus is determined only by the height of the first peak of its AC function. In support of this view, Yost *et al.* (1996) found that when two IRN stimuli have identical first AC peaks (for a delay d), they cannot be discriminated from each other, even when they differ with respect to the height of a second AC peak (at $2 \cdot d$). The representation of an IRN stimulus in a human auditory nerve is obviously more complex and less predictable than that of the click sequences used here. The first AC peak in the stimulus may be partly represented by first-order ISIs, but is probably also represented by higher-order ISIs. However, it is clear that a second AC peak will be represented by ISIs of an even higher order, on average. Thus, the finding that listeners are insensitive to the second AC peak is consistent with the idea that only first-order ISIs perceptually matter for pitch perception.[1]

We do not wish to conclude from our results that first-order ISIs at the auditory nerve level are a perfect predictor of pitch in any possible case. This conclusion would be at variance with some physiological data reported by Cariani and Delgutte (1996a). These authors conducted a large set of studies on the temporal correlates of pitch in the cat's auditory nerve. Their work shows that numerous pitch phenomena can be correctly predicted from the ISIs occurring in the auditory nerve. In one of their studies, the stimuli were periodic, vowellike complex tones. Such stimuli elicit a pitch corresponding to the period whatever the intensity. It was found that the highest point of the neural ''summary AC function'' (as defined by Meddis and Hewitt, 1991) did correspond to the period, and thus to the pitch heard, whatever the intensity. By contrast, when only the first-order ISIs were taken into account, the predicted pitch appeared to be somewhat dependent on intensity: It corresponded unambiguously to the period at 60 dB SPL, but not at 40 or 80 dB SPL.

It is plausible that the ''final'' temporal structure contributing to pitch sensations (either directly or after a conversion into a place code) does not occur in the auditory nerve but at a higher location in the auditory system. We believe that at this stage the ISIs that matter are first-order ISIs. However, the consecutive spikes bounding these ISIs may originate from nonconsecutive spikes at the auditory nerve level.[2]

It is important to keep in mind that our stimuli were spectrally unresolvable and that those having a detectable temporal regularity induced a percept of ''rattle'' pitch rather than ''musical'' pitch. We must acknowledge that the implications of our results may not be generalizable to spectrally resolvable sounds, which induce more salient and precise pitch sensations than those evoked by spectrally unresolvable sounds (Hoekstra, 1979; Houtsma and Smurzynski, 1990).[3] Carlyon and Shackleton (1994) provided experimental support for the idea that pitch extraction rests on different mechanisms for these two types of sounds. The task of their subjects was to detect differences in the periods of two simultaneous groups of harmonics, falling in separate frequency regions. Detection performance was good when listeners had to compare two resolvable groups, or two unresolvable groups, but poor when the comparison was between one resolvable group and one unresolvable group. It might be that the AC theory is in error for spectrally unresolvable sounds but is correct for resolvable sounds.[4]

Let us finally mention here some informal observations that we made using (high-pass filtered) click sequences which were not employed in the experiments described above. We constructed a sequence in which the first-order ICIs took two alternating values: 3, 5, 3, 5, 3, 5,... ms. This periodic sequence (''Period [3,5]'') sounds quite regular. It elicits a slightly ambiguous pitch, commonly identified as that of Period [5], but sometimes as that of Period [3] (very rarely as that of Period [8]). Period [3,5] sounds quite different from a sequence (''Random [3,5]'') in which the same two ICIs occur randomly, even when the randomness is limited by preventing an ICI to be repeated immediately more than once. This shows that the first-order ICI statistics are not sufficient to account for the perceptual effects of filtered click sequences. Period [3,5] and Random [3,5] are similar with respect to pitch, but Random [3,5] elicits a percept of temporal fluctuations that is not elicited by Period [3,5]. For more and more complex periodic sequences based on the same first-order ICIs, for instance, Period [3,3,5,5] or Period [3,3,5,3,5,5], there is an increasing perception of temporal fluctuations. However, these sequences still sound more regular than Random [3,5]. It does not seem reasonable to assume that the discrimination between Period [3,3,5,3,5,5] and Random [3,5] rests on the measurement of sixth-order ICIs (always equal to 24 ms in the periodic sequence, but variable in the random sequence). A neural model based on first-order ISI detection and fast synaptic plasticity (Kaernbach and Mohlberg, 1994) can account for such perceptual discriminations.

ACKNOWLEDGMENTS

We thank Peter A. Cariani, Robert P. Carlyon, Alain de Cheveigné, Bertrand Delgutte, Roy D. Patterson, and Lutz Wiegrebe for stimulating discussions. Special thanks are due to Christian Lorenzi for his useful computer simulations of the neural processing of our stimuli. We are also grateful to Ray Meddis, William A. Yost, and an anonymous reviewer for their helpful comments on a previous version of the manuscript. Part of the work was done while author CK was working at the Institut für Neuroinformatik, Ruhr-Universität Bochum, and supported by DFG Grant No. MA 697/4-2.

[1]Yost *et al.* (1996) submitted various IRN waveforms to a simple threshold device (including an absolute refractory period of 1 ms) and measured the first-order time intervals between successive threshold crossings. They found that the statistical distribution of these first-order intervals was a good predictor of the perceived pitch and pitch strength. This is consistent with our results. However, the waveforms used by Yost *et al.* were not high-pass filtered and, as acknowledged by these authors, their very simple processing of the waveforms did not provide a realistic picture of the neural spike trains induced by the corresponding sounds.

[2]Consider, for example, a small set of adjacent auditory nerve fibers responding to a vowellike complex tone (with a period of $p=1/f_0$ ms). If their characteristic frequency f_c is higher than about $4\cdot f_0$, they will not respond to a single harmonic. They will respond to a complex waveform resulting from the interaction of several, more or less attenuated, harmonics. The envelope of this complex waveform will have an amplitude peak every p ms. In a given fiber, many first-order ISIs may not be equal to p ms. They may often be close to $1/f_c$ and reflect time intervals between peaks of the waveform's fine structure. Across the set of fibers, however, peaks of the envelope will tend to elicit *synchronous* spikes, whereas peaks of the fine structure will fail to do so. One can imagine that at a higher stage of the auditory system, the synchronous spikes will produce excitation, thanks to some kind of summation effect, whereas many of the intervening nonsynchronous spikes will not survive. This could transform high-order ISIs at the auditory nerve level into first-order ISIs at the higher stage.

[3]The pitch of our target stimuli was weak because these stimuli were spectrally unresolvable but also because of their partial randomness.

[4]Moore (1997, Chap. 5) hypothesized that, in an auditory nerve fiber with characteristic frequency f_c, the temporal information on musical pitch is limited to ISIs smaller than about $15/f_c$. This would explain why a group of very high harmonics fails to evoke a sense of musical pitch corresponding to the fundamental.

Assmann, P. F., and Summerfield, Q. (**1990**). ''Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **88**, 680–697.

Burns, E. M., and Viemeister, N. F. (**1976**). ''Nonspectral pitch,'' J. Acoust. Soc. Am. **60**, 863–869.

Burns, E. M., and Viemeister, N. F. (**1981**). ''Played-again SAM: Further observations on the pitch of amplitude-modulated noise,'' J. Acoust. Soc. Am. **70**, 1655–1660.

Cariani, P. A., and Delgutte, B. (**1996a**). ''Neural correlates of the pitch of complex tones. I. Pitch and pitch salience,'' J. Neurophysiol. **76**, 1698–1716.

Cariani, P. A., and Delgutte, B. (**1996b**). ''Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, rate-pitch, and the dominance region of pitch,'' J. Neurophysiol. **76**, 1717–1734.

Carlyon, R. P. (**1996**). ''Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker,'' J. Acoust. Soc. Am. **99**, 517–524.

Carlyon, R. P., and Shackleton, T. M. (**1994**). ''Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?'' J. Acoust. Soc. Am. **95**, 3541–3554.

de Cheveigné, A. (**1993**). ''Separation of concurrent harmonics 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 Ribaupierre, F., Goldstein, Jr., M. H., and Yeni-Komshian, G. (**1972**). ''Cortical coding of repetitive acoustic pulses,'' Brain Res. **48**, 205–225.

Evans, E. F. (**1978**). ''Place and time coding of frequency in the peripheral auditory system: some physiological pros and cons,'' Audiology **17**, 369–420.

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

Goldstein, J. L., and Srulovicz, P. (**1977**). ''Auditory-nerve spike intervals as an adequate basis for aural frequency measurement,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).

Hartmann, W. M. (**1993**). ''On the origin of the enlarged melodic octave,'' J. Acoust. Soc. Am. **93**, 3400–3409.

Hartmann, W. M. (**1997**). *Signals, Sound, and Sensation* (AIP Press, Woodbury, NY).

Hoekstra, A. (**1979**). ''Frequency discrimination and frequency analysis in hearing,'' Doctoral dissertation, University of Groningen, The Netherlands.

Horst, J. W., Javel, E., and Farley, G. R. (**1986**). ''Coding of spectral fine structure in the auditory nerve. I. Fourier analysis of period and interspike interval histograms,'' J. Acoust. Soc. Am. **79**, 398–416.

Houtsma, A. J. M., and Smurzynski, J. (**1990**). ''Pitch identification and discrimination for complex tones with many harmonics,'' J. Acoust. Soc. Am. **87**, 304–310.

Kaernbach, C. (**1991**). ''Simple adaptive testing with the weighted up-down method,'' Percept. Psychophys. **49**, 227–229.

Kaernbach, C., and Mohlberg, H. (**1994**). ''A neural sequence-learning model to explain auditory periodicity analysis,'' in *Proceedings of the 1994 International Congress on Artificial Neural Networks*, Sorrento, Italy, edited by M. Marinaro and P. G. Morasso (Springer-Verlag, New York).

Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). ''Discharge Patterns of Single Fibers in the Cat's Auditory Nerve,'' Res. Monogr. 35 (MIT, Cambridge, MA).

Langner, G. (**1992**). ''Periodicity coding in the auditory system,'' Hearing Res. **60**, 115–142.

Langner, G., Sams, M., Heil, P., and Schulze, H. (**1997**). ''Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography,'' J. Comp. Physiol. A **181**, 665–676.

Lazzaro, J., and Mead, C. (**1989**). ''Silicon modeling of pitch perception,'' Proc. Natl. Acad. Sci. USA **86**, 9597–9601.

Licklider, J. C. R. (**1951**). ''A duplex theory of pitch perception,'' Experientia **7**, 128–134.

Lyon, R. F. (**1984**). ''Computational models of neural auditory processing,'' in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing 36.1* (IEEE, New York), pp. 1–4.

Meddis, R., and Hewitt, M. J. (**1991**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.

Moore, B. C. J. (**1977**). ''Effects of relative phase of the components on the pitch of three-component complex tones,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing* (Academic, London), 4th ed.

Moore, B. C. J., and Rosen, S. M. (**1979**). ''Tune recognition with reduced pitch and interval information,'' Q. J. Exp. Physiol. **31**, 229–240.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). ''Relative dominance of individual partials in determining the pitch of complex tones,'' J. Acoust. Soc. Am. **77**, 1853–1860.

Ohgushi, K. (**1978**). ''On the role of spatial and temporal cues in the perception of the pitch of complex tones,'' J. Acoust. Soc. Am. **64**, 764–771.

Pantev, C., Hoke, M., Lütkenhöner, B., and Lehnertz, K. (**1989**). ''Tonotopic organization of the auditory cortex: Pitch versus frequency representation,'' Science **246**, 486–488.

Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). ''Time-domain modeling of auditory processing: A modular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.

Patterson, R. D., Handel, S., Yost, W. A., and Datta, A. J. (**1996**). ''The relative strength of the tone and noise components in iterated rippled noise,'' J. Acoust. Soc. Am. **100**, 3286–3294.

Patterson, R. D., Robinson, K., Holdsworth, J. W., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). ''Complex sounds and auditory images,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford).

Plomp, R. (**1967**). ''Pitch of complex tones,'' J. Acoust. Soc. Am. **41**, 1526–1533.

Plomp, R. (**1976**). *Aspects of Tone Sensation* (Academic, London).

Ritsma, R. J. (**1967**). ''Frequencies dominant in the perception of the pitch of complex tones,'' J. Acoust. Soc. Am. **42**, 191–198.

Ruggero, M. A. (**1992**). ''Physiology and coding of sound in the auditory nerve,'' in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer-Verlag, New York).

Sachs, M. B., and Young, E. D. (**1980**). ''Effects of nonlinearities on speech encoding in the auditory nerve,'' J. Acoust. Soc. Am. **68**, 858–875.

Slaney, M., and Lyon, R. F. (**1990**). ''A perceptual pitch detector,'' in

Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Albuquerque, New Mexico (IEEE, New York), 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.

Steinschneider, M., Arezzo, J., and Vaughan, Jr., H. G. (**1980**). ''Phase-locked cortical responses to a human speech sound and low-frequency tones in the monkey,'' Brain Res. **198**, 75–84.

Terhardt, E. (**1972**). ''Zur Tonhöhenwahrnehmung von Klängen. II. Ein Funktionsschema,'' Acustica **26**, 187–199.

Tsuzaki, M., and Patterson, R. (**1998**). ''Jitter detection: a brief review and some new experiments,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 546–552.

van Noorden, L. (**1982**). ''Two channel pitch perception,'' in *Music, Mind and Brain*, edited by M. Clynes (Plenum, New York).

Yost, W. A. (**1996**). ''Pitch of iterated rippled noise,'' J. Acoust. Soc. Am. **100**, 511–518.

Yost, W. A., Patterson, R. D., and Sheft, S. (**1996**). ''A time domain description for the pitch strength of iterated rippled noise,'' J. Acoust. Soc. Am. **99**, 1066–1078.