

# Temporal dynamics of pitch strength in regular-interval noises: Effect of listening region and an auditory model

L. Wiegrebe<sup>a)</sup>

Zoologisches Institut der Universität München, Luisenstrasse 14, 80333 München, Germany

H. S. Hirsch

Centre for the Neural Basis of Hearing, Physiology Department, Cambridge University, Cambridge, United Kingdom and Lehrstuhl für Mensch-Maschine Kommunikation, Technische Universität München, Germany

R. D. Patterson

Centre for the Neural Basis of Hearing, Physiology Department, Cambridge University, Cambridge, United Kingdom

H. Fastl

Lehrstuhl für Mensch-Maschine Kommunikation, Technische Universität München, Germany

(Received 22 October 1999; revised 3 January 2000; accepted 8 March 2000)

Recently, it was demonstrated that the pitch strength of a stimulus denoted “AABB” differed from rippled noise (RN) despite the fact that their long-term spectra and autocorrelation functions are identical (Wiegrebe *et al.*, 1998). Rippled noise is generated by adding a delayed copy of Gaussian noise back to itself; AABB is generated by concatenating equal-duration, Gaussian-noise segments where every segment is repeated once. It was shown that a simple model based on a two-stage integration process separated by a nonlinear transformation explains the pitch-strength differences quantitatively. Here, we investigate how the spectral listening region influences pitch-strength differences between RN and AABB. Bandpass filtering the two stimuli with a constant bandwidth of 1 kHz revealed a systematic effect of center frequency. For relatively high pitches (corresponding to delays,  $d$ , of 4 or 5.6 ms, pitch strength differences between AABB and RN were absent when the pass band was between 0 and 1 kHz. When the pass band was between 3.5 and 4.5 kHz, pitch-strength differences were substantial. For lower pitches ( $d$  equal to or longer than 8 ms), AABB had a substantially greater pitch strength independent of the filter center frequency. The model presented in Wiegrebe *et al.* (1998) cannot capture these effects of center frequency. Here, it is demonstrated that it is possible to simulate the RN-AABB pitch-strength differences, and the effect of listening region, with a computer model of the auditory periphery. It is shown that, in an auditory model, pitch-strength differences are introduced by the nonlinear transformation possibly associated with half-wave rectification or mechano-electrical transduction. In this experimental context, however, the nonlinearity has perceptual relevance only when the differences in short-term fluctuations of AABB and RN are preserved in auditory-filter outputs. The current experiments relate the purely functional model introduced in the preceding paper to basic properties of the peripheral auditory system. The implication for neural time constants of pitch processing is discussed. © 2000 Acoustical Society of America. [S0001-4966(00)03406-8]

PACS numbers: 43.66.Ba, 43.66.Hg, 43.66.Jh, 43.66.Mk [SPB]

## INTRODUCTION

Both rippled noise (RN) and “AABB” noise are tonal stimuli derived from random noise by introducing temporal regularity into the waveform. Thus they both belong to the category of “regular-interval noises” (Yost, 1998). RN is produced by delaying a sample of random noise by “ $d$ ” ms and adding it back to itself; iterated rippled noise (IRN) is generated by iterating the delay-and-add process. Both have been used extensively to study mechanisms of pitch perception (Yost, 1996a, 1996b; Yost *et al.*, 1996; Patterson *et al.*, 1996; Wiegrebe and Patterson, 1999). Recently, Wiegrebe *et al.* (1998) used an AABB stimulus to investigate the time

constants associated with the pitches of regular-interval noises. AABB is generated by concatenating many independent noise segments of the same duration,  $d$ , with the constraint that every noise segment is repeated once before the next independent segment is concatenated. If we denote independent noise segments with letters A,B,C,..., the general form for the sound is AABBCC.... For convenience, this stimulus is called AABB. Wiegrebe *et al.* (1998) showed that AABB and RN have identical long-term spectra and autocorrelation functions when generated with the same  $d$ . Nevertheless, the pitch strengths of the two stimuli differ for pitches below about 250 Hz: AABB produces a stronger pitch than RN. To account for the pitch-strength differences, an autocorrelation model composed of a two-stage temporal integration mechanism separated by a nonlinear transforma-

<sup>a)</sup>Electronic mail: wiegrebe@zi.biologie.uni-muenchen.de

tion was introduced. The first stage was a running autocorrelation performed on the stimulus waveform. In this stage, temporal integration is imposed by an exponentially decaying integration window shifted over the waveform. The second stage relates the height of the peak in the running autocorrelation function with pitch strength in an exponentially nonlinear fashion. The third stage (second integration stage) involved averaging over the entire stimulus duration to produce a single pitch-strength value. The time constant of the first integration stage, the running autocorrelation of the stimulus waveform, was used to fit pitch-strength matching data between AABB and an IRN matching stimulus. The best-fitting time constant was 1.5 ms.

Although the model provided a good fit to the experimental data, it is not clear how the different model stages relate to properties of the auditory system. Before any kind of regularity analysis like running autocorrelation can be implemented in the auditory system, the stimulus is transformed by the outer and middle ear, cochlear nonlinear narrow-band filtering and the mechano-electrical transduction of the inner hair cells. The current study presents experiments whose results are incompatible with the purely functional modeling approach presented in Wiegrefe *et al.* (1998). However, it is demonstrated that the current experimental results as well as those in Wiegrefe *et al.* (1998) can be simulated in a computer model of the auditory periphery similar to that presented by Meddis and Hewitt (1991a, 1991b), Yost *et al.* (1996) and Meddis and O'Mard (1997). It is shown how the principal processing stages of the functional model in Wiegrefe *et al.* (1998) relate to specific processing stages in the peripheral human auditory system.

## I. EXPERIMENT I: THE EFFECT OF SPECTRAL LISTENING REGION ON RN-AABB DISCRIMINATION

In Wiegrefe *et al.* (1998) it was shown that despite the identity of the long-term spectra and autocorrelation functions, listeners could reliably discriminate between AABB and RN when  $d$  was longer than about 4 ms. As broadband stimuli were used, it was not possible to determine whether discrimination was supported by information available over the whole spectral range or if discrimination may depend on the spectral listening region. Here, we asked listeners to discriminate AABB from RN with bandpass filtered stimuli having constant bandwidth but different center frequencies. When using bandpass filtered stimuli, audible distortions may be generated below the pass band which can potentially interfere with the required experimental task. Because distortion products differ between RN and AABB, a small control experiment was performed where a random noise with a spectrum level only 10 dB lower than the stimulus spectrum level was added below the stimulus pass band. Introducing this low-pass masker, however, did not significantly alter performance in experiment I, and thus we are confident that the experimental results are not influenced by audible distortion products.

## A. Method

### 1. Stimuli

RN and AABB stimuli were produced exactly as described in the Introduction. The pitch, which corresponds to  $1/d$ , ranged from 31.25 to 250 Hz in half-octave steps. Within each run,  $d$  was the same for RN and AABB. Stimuli were generated in the time domain and then, using the fast Fourier transform, translated into the frequency domain. By setting the magnitude spectrum outside the pass band to zero and then performing an inverse Fourier transform, the bandpass filtered stimuli were obtained. The spectrum level was 30 dB; the pass-band width was fixed at 1 kHz; the center frequency was 500 Hz, 1 kHz, 2 kHz or 4 kHz. The stimulus duration was 327.68 ms, including 20-ms, cosine-squared ramps. Stimuli were generated and played back digitally using the Tucker Davis System II and software delay lines. The sampling rate was 25 kHz. All sounds were presented diotically via AKG K240 DF headphones. Listeners were tested individually in a double-walled sound-attenuating booth.

### 2. Procedure

As in Wiegrefe *et al.* (1998), psychometric functions for the discrimination of AABB from RN were obtained using a two-alternative, three-interval, forced-choice task. The first interval was always a RN. Either the second or the third interval contained an AABB stimulus to be discriminated from RN, whereas the remaining interval contained another RN. The listeners were instructed to choose the interval that differed from the other two. The listeners were provided with visual feedback. Each point on the psychometric function represents the mean of 99 trials obtained in three runs. Within a run, the value of  $d$  for each trial was chosen randomly from the full range with the constraint that each value of  $1/d$  was presented 33 times. Within each run, the bandpass center frequency was fixed at one of the four values given above. The different center-frequency conditions were presented in a nonsystematic order.

### 3. Listeners

Three listeners aged between 26 and 33 years took part in all experiments; L1 was the first author, L2 was the second author. Listeners L2 and L3 were given training until stable performance was achieved. Listener L1 had taken part in the experiments in Wiegrefe *et al.* (1998) and needed no further training. All listeners had absolute thresholds at all audiometric frequencies within 10 dB of the 1969 ANSI standard.

## B. Results

The results were very similar across listeners, and thus the results were pooled. The psychometric functions for the AABB-RN discrimination are plotted as a function of  $1/d$  in Fig. 1. Error bars represent cross-listener standard errors. When  $1/d$  is below about 100 Hz, discrimination of AABB from RN is above 80%, irrespective of filter center frequency. When  $1/d$  is 125 Hz and above, discrimination performance deteriorates with increasing  $1/d$  and there is a systematic effect of center frequency. The lower the center

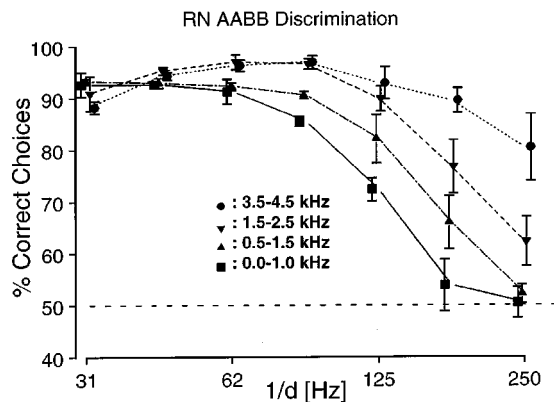


FIG. 1. Discrimination of AABB stimuli from RN stimuli as a function of the fundamental frequency ( $1/d$ ). Different line styles and symbols represent discrimination performance for bandpass filtered stimuli with a constant bandwidth of 1 kHz and a center frequency of 500 Hz (squares and solid line), 1 kHz (upward triangles and dash-dotted line), 2 kHz (downward triangles and dashed line) and 4 kHz (circles and dotted line). Error bars represent cross-listener standard errors.

frequency of the pass band, the worse the discrimination performance. When  $1/d$  is 250 Hz, discrimination is at chance levels for a center frequency of 500 Hz, but it is on average above 80% for a center frequency of 4 kHz.

## II. EXPERIMENT II: PITCH-STRENGTH MATCHING FILTERED AABB WITH EQUALLY FILTERED IRN STIMULI

Experiments II and III of Wiegrobe *et al.* (1998) showed that the basis of the RN-AABB discrimination with broadband stimuli is a difference in pitch strength. To establish if pitch-strength differences may also serve as a possible discrimination cue for the bandpass filtered stimuli, the pitch strength of the filtered AABB stimuli was matched with that of equally filtered IRN stimuli having the same  $d$ . Over trials, the actual value of  $d$  was randomized within  $\pm 3\%$  of the denoted  $d$  value to help to focus the listeners' attention to pitch strength, not pitch. IRN was produced in the add-same configuration (Yost, 1996a, 1996b) with 16 iterations. The pitch strength of RN and IRN stimuli can be continuously varied by changing the gain in the delay-and-add loop (Yost, 1996b; Wiegrobe *et al.*, 1998).

### A. Method

#### 1. Stimuli

Stimulus duration, presentation level and filter conditions were the same as in the previous experiment. AABB was generated as described above. IRN was generated with the same  $d$  and 16 iterations. Pitch-strength matches were obtained for  $1/d$ 's between 31.25 and 250 Hz in half-octave steps.

#### 2. Procedure

As in Wiegrobe *et al.* (1998), a two-alternative, forced-choice procedure was used to determine the pitch-strength matches. One randomly chosen interval contained a sample of filtered AABB; the other interval contained an equally filtered IRN. At the beginning of each run, the IRN gain was

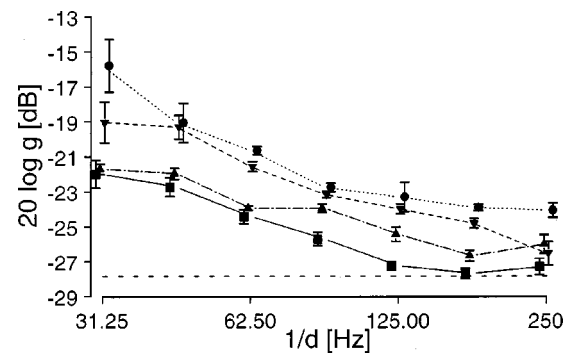


FIG. 2. Pitch-strength matches of AABB test stimuli with IRN matching stimuli. IRN was produced with 16 iterations and its gain ( $g$ ) was manipulated to match the IRN pitch strength to the AABB pitch strength. An increase in  $g$  corresponds to an increase in pitch strength. The different filter conditions are represented in the same format as in Fig. 1. The horizontal dashed line represents the pitch-strength matches as they would be obtained for RN test stimuli according to Yost (1996b). Error bars represent cross-listener standard errors. Generally, pitch-strength differences between AABB and RN decrease with increasing  $1/d$ . However, for a pass band from 3.5 to 4.5 kHz (dotted line and filled circles), there is still a pitch-strength difference when  $1/d$  equals 250 Hz where there is no pitch-strength difference when the stimuli are lowpass filtered at 1 kHz (solid line and filled squares).

set to unity so that the IRN pitch strength was initially higher than that of the AABB. Listeners were instructed to choose the stimulus with the greater pitch strength. Feedback was not provided. When they chose the IRN stimulus, the IRN gain was reduced for the next trial, when they chose the AABB stimulus, the IRN gain was increased for the next trial. When listeners switched from choosing the IRN to the AABB or vice versa, a reversal was registered. The step size for the IRN gain initially was 3 dB; it was reduced to 2 dB after the third reversal and to 1 dB after the fifth reversal. The average gain from reversals six to eleven was taken as the matching gain for that run. For each condition, four runs were obtained and their matching points were averaged. The listeners were the same as in the previous experiment. They were trained until stable performance was reached, which typically occurred after only a few runs for each value of  $d$ .

### B. Results and discussion

The IRN gain required to match the IRN pitch strength to that of equally filtered AABB with the same  $d$  is shown as a function of  $1/d$  in Fig. 2. Again performance across listeners was very consistent and so the data were pooled. Error bars represent cross-listener standard errors. Results for the different filter conditions are shown with the same symbols and line styles as in Fig. 1. The horizontal dashed line at an IRN gain of  $-28.8$  dB represents the pitch-strength match estimated for a RN test stimulus instead of an AABB test stimulus. The estimate is based on Yost's (1996b) finding that RN and IRN stimuli have the same pitch strength when the height of the first peak in the long-term autocorrelation function ( $h_1$ ) is the same. The pitch-strength difference between AABB and RN is represented as the difference between the IRN matching gain for AABB test stimuli and the horizontal dashed line. For all filtering conditions, there is a consistent trend toward lower matching gains with  $1/d$ . In



the lowest filter condition (0 to 1 kHz, solid line and filled squares), the pitch-strength difference was 6 dB when  $1/d$  was 31.25 Hz and reduces to zero for  $1/d$  of 125 Hz and above. In the highest filter condition (3.5 to 4.5 kHz, dotted line and filled circles), the pitch-strength difference was 12 dB at  $1/d$  equals 31.25 Hz, and it was still about 5 dB even for  $1/d$  equals 250 Hz. The results indicate that pitch-strength differences between RN and AABB may have served as a cue for the RN-AABB discrimination measured in experiment I. For broadband stimuli, this has been shown in Wiegrebe *et al.* (1998). For values of  $d$  and filter center frequency where pitch-strength differences are absent in experiment II, discrimination performance in experiment I is at chance levels. For two reasons, however, it must not be assumed that the pitch-strength matching curves of experiment II resemble the psychometric functions of experiment I: First, the psychometric functions for the discrimination are likely to be limited by ceiling effects, and second, the relation between pitch strength and IRN gain is monotonic but it is not linear (cf. Yost, 1996b). It is important to note that the pitch-strength matching paradigm only allows for comparisons of pitch strengths within one value of  $d$  and within one frequency band. As stated above, if a RN had been used as a test stimulus, matches would be fixed at about  $-28.8$  dB. This does not imply, however, that the pitch strength of RN is independent of  $1/d$ . It is well documented that pitch strength of RN stimuli increases with increasing  $1/d$  at least for the range of  $1/d$  tested here (Yost and Hill, 1978). Additionally, high-pass filtering complex harmonic stimuli decreases their pitch strength as shown, for example, by Houtsma and Smurzynski (1990). This is also supported by experiments on the dominance region of pitch (Plomp, 1967). Both these sets of experiments underline the perceptual dominance of the low harmonics or the low spectral peaks in regular-interval noises. The current experiments and the simulations in Sec. III only try to capture pitch-strength differences between RN and AABB within each of the experimental conditions of  $d$  and center frequency, not pitch strength across the different conditions.

### III. MODELING THE EFFECT OF LISTENING REGION ON THE PITCH-STRENGTH DIFFERENCE OF REGULAR-INTERVAL NOISES

#### A. The functional model and the effect of spectral listening region

Wiegrebe *et al.* (1998) used a functional model to account for RN-AABB pitch-strength differences; a block diagram of the model is shown in the left panel of Fig. 3. In this functional model, the waveform is subjected to a running normalized autocorrelation with a time constant of 1.5 ms. The correlation is calculated as a function of time, but only for one correlation lag which equals  $d$ , because only at this lag, a correlation peak ( $h1$ ) occurs in the long-term autocorrelation of RN and AABB. The result of this processing stage is called  $Rh1(t)$ ; it is subjected to an expansive non-linear transformation (exponentiation). The result is then averaged over the stimulus duration to obtain a scalar value for the pitch-strength estimate. In this section, this model is

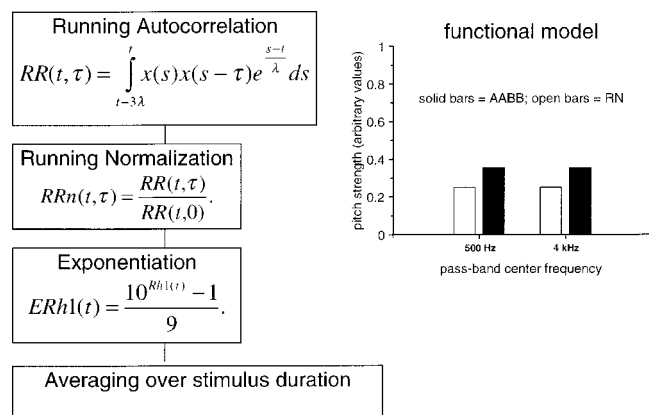


FIG. 3. Left panel: Block diagram of the functional model used to explain the AABB-RN pitch-strength differences in Wiegrebe *et al.* (1998). See text for a more detailed description of the model. The right panel shows simulated pitch-strength values for RN (open bars) and AABB (filled bars) for bandpass center frequencies of 500 Hz and 4 kHz. The simulation results are independent of center frequency and thus not suitable to simulate the results of experiments I and II.

tested to determine whether it can accommodate the effect of the listening region. To test the functional model, average pitch strength was calculated for 100 samples of RN and AABB;  $d$  was fixed at 4 ms and the stimuli were filtered with a center frequency of either 500 Hz or 4 kHz and a constant bandwidth of 1 kHz. The results of this simulation are shown in the right panel of Fig. 3. For both filter conditions, AABB with  $d$  equals 4 ms has a higher pitch strength than RN with the same  $d$ . Thus the model does not predict the effect of filter center frequency.

#### B. Toward an auditory model of pitch strength

In recent years, multichannel computer models of the human auditory periphery have been successfully used to simulate psychoacoustical experiments on pitch perception (Slaney and Lyon, 1990; Meddis and Hewitt, 1991a, 1991b, 1992; Patterson *et al.*, 1995; Yost *et al.*, 1996; Meddis and O'Mard, 1997; de Cheveigne, 1998). Common to these models is the desire to simulate the spectrotemporal distribution of auditory-nerve firing. The spiking activity or spike probability pattern is then subjected to an analysis of temporal regularity, typically implemented by channel-wise autocorrelation. The autocorrelation is typically implemented as a running autocorrelation with a time constant between 2.5 and 10 ms. We investigated the representation of RN and AABB stimuli in an auditory model whose peripheral preprocessing is identical to that of Meddis and O'Mard (1997).

Meddis and Hewitt (1991a, 1991b) and Meddis and O'Mard (1997) tested their model with a large variety of pitch experiments; some of which are related to pitch strength: the model successfully predicts the dominance region of harmonic complexes as determined by Plomp's (1967) paradigm, the sensitivity to mistuning of individual harmonics (Moore *et al.*, 1985), and the increase in threshold  $f0$ -differences with increasing listening region for band-pass filtered harmonic complexes (Houtsma and Smurzynski, 1990). The procedure of all these experiments involves the comparison of complex tones *within* the same frequency re-

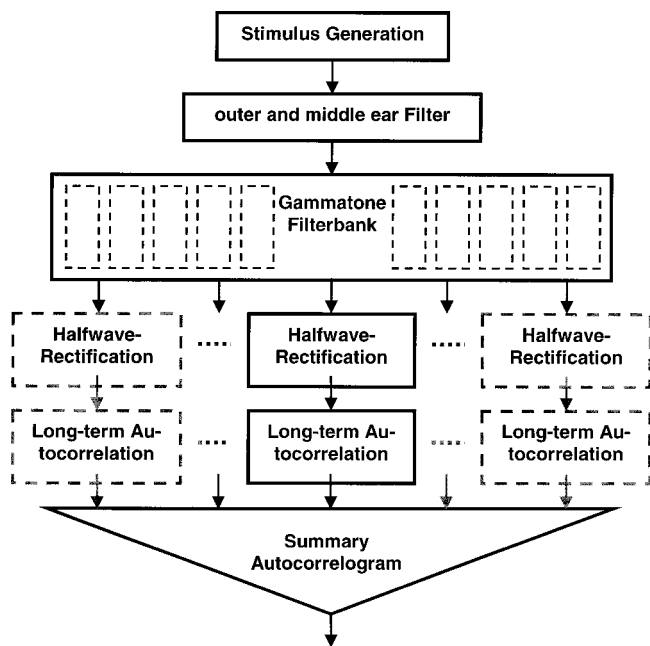


FIG. 4. Block diagram of the auditory model used to simulate pitch-strength differences between AABB and RN stimuli. The model is very similar to that of Meddis and O'Mard (1997) apart from the use of simple half-wave rectification instead of the Meddis hair cells and a long-term autocorrelation instead of a running autocorrelation with a 10-ms time constant.

gion and with similar  $f_0$ s. The same is true for the current experiments where pitch strength is compared for AABB and RN with the same  $d$  (i.e., with similar pitch) and in the same frequency region. In the next section, it will be shown the model can predict pitch-strength differences correctly, although the model is not able to predict absolute pitch strength and its changes across different  $d$ 's and frequency regions.

Outer and middle ear filtering is implemented by a second-order bandpass filter with cutoffs of 450 and 8500 Hz: spectral analysis is implemented by a bank of 60 gammatone filters equally spaced on a log frequency scale between 100 and 8000 Hz. Mechano-electrical transduction was very coarsely simulated by simple half-wave rectification. A bank of Meddis hair cells (Meddis, 1988) was shown to yield very similar results to half-wave rectification. Unlike Meddis and O'Mard (1997), autocorrelation of the rectifier output was implemented as a long-term autocorrelation, i.e., the width of the correlation window equaled the stimulus duration. The autocorrelograms were summed over frequency

channels to produce a summary autocorrelogram (SACG). A block diagram of the model is shown in Fig. 4.

### C. Simulation results

Yost (1996b) showed that for RN and IRN stimuli, the position of the first peak in the autocorrelation of the stimulus indicates the pitch of the stimulus and its height is monotonically related to pitch strength. The same has been suggested for the first peak in the SACG (Yost *et al.*, 1996; Patterson *et al.*, 1996); however, this hypothesis has not been quantitatively tested, so far. For the current simulations, it is assumed that pitch strength is monotonically related to the height of the first peak in the SACG. In the following, we compare the SACG peak heights for RN and AABB stimuli in the same stimulus conditions as used in experiment II. For each filter condition and each delay, 100 model responses were calculated and their peak heights averaged. In the left panel of Fig. 5, the SACG peak heights are plotted as a function of  $1/d$  for the four filter conditions, in the same format as for the experimental data. Simulation results for RN are plotted with solid lines; results for AABB stimuli are plotted with dashed lines. The right panel of Fig. 5 shows the SACG peak-height difference functions between RN and AABB.

The left and right panel of Fig. 5 represent the shortcomings and success of the model, respectively. The absolute SACG peak heights in the left panel are rather independent of  $d$ , suggesting a pitch strength independent of  $d$ . This is in contrast to the experimental results of Yost and Hill (1978), who found that when  $d$  increases from 5 ms, pitch strength decreases and is almost entirely absent for  $d > 20$  ms. Experiments on the dominance region and mistuning of individual harmonics showed that the strongest pitch is always mediated by the low harmonics (typically between two and five). The SACG peak heights in the left panel of Fig. 5, however, suggest that pitch strength is always lowest in the lowest frequency band and increases when the listening band is shifted to between 1.5 and 2.5 kHz. Thus the model in its current form is clearly not able to predict absolute pitch strength and its changes across different frequency regions and different  $d$ 's.

The general shapes of the peak-height difference functions in the right panel of Fig. 5, however, are consistent with the pitch-strength differences obtained in experiment II and shown in Fig. 2. For low values of  $1/d$ , pitch-strength

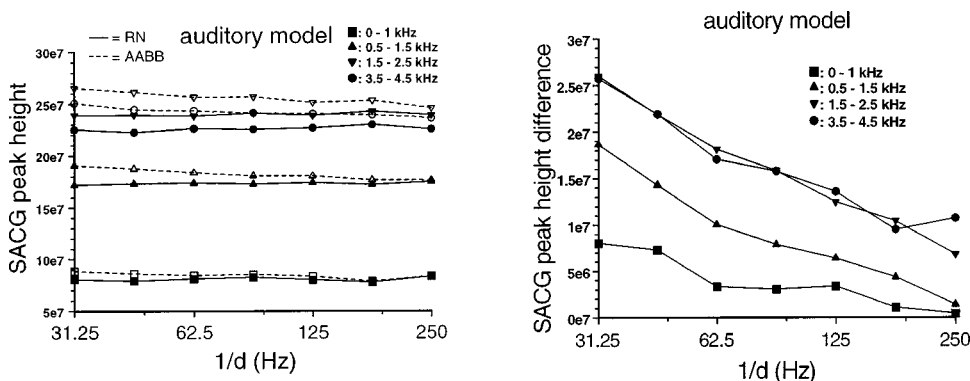


FIG. 5. Left panel: Height of the first peak in the summary autocorrelogram for AABB (dashed lines and open symbols) and RN stimuli (solid lines and filled symbols) as a function of  $1/d$ . The different symbols represent the different filter conditions. Peak-height differences between AABB and RN stimuli are plotted in the right panel. They are in good qualitative agreement with the pitch-strength differences obtained in experiment II (Fig. 2).

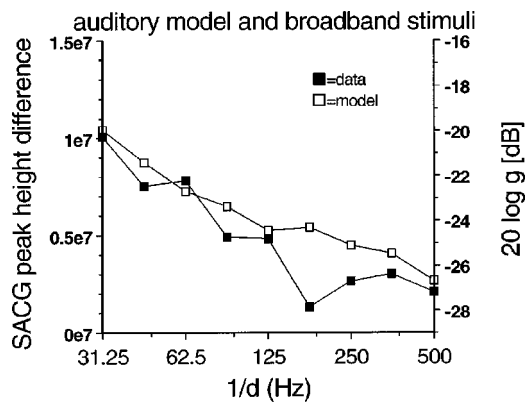


FIG. 6. SACG peak-height differences between broadband AABB and RN. The shape of the peak-height differences (open squares referring to the left Y-axis) are in reasonable *qualitative* agreement with the experimental data of experiment II in Wiegrebe *et al.* (1998) (filled squares referring to the right Y-axis). Note that the left and right Y-axes are not linearly related.

differences and SACG peak-height differences are large. With increasing  $1/d$ , pitch-strength differences and SACG peak-height differences decrease. When the pass band is from 0 to 1 kHz and  $1/d$  is 250 Hz, pitch-strength differences and peak-height differences become negligible. Pitch-strength differences are still substantial, however, when the pass band is from 3.5 to 4.5 kHz. To simulate the results of experiment II in Wiegrebe *et al.* (1998), the model was presented with broadband stimuli with  $1/d$  ranging from 31.25 to 500 Hz. The simulation results are shown in Fig. 6. Again, SACG peak-height differences are in good qualitative agreement with the experimental data of experiment II in Wiegrebe *et al.* (1998).

#### D. Crucial processing stages of the auditory model and their differential effect on RN and AABB stimuli

In this section, the manipulations imposed on the waveforms at the different processing stages of the auditory model are reviewed and the processing stages crucial to the differential representation of AABB and RN are highlighted. It will be shown that in the model, the cochlear filtering, together with the nonlinear transformation implemented by half-wave rectification, are responsible for the differences in the coding of AABB and RN periodicity.

##### 1. The effect of cochlear filtering

Cochlear filtering is approximated by a gammatone filterbank in the auditory model. The equivalent rectangular bandwidth (ERB) of a gammatone filter is

$$\text{ERB} = 24.7 + 0.108 * F_c.$$

The filter bandwidth grows with increasing filter center frequency  $F_c$  (in Hz). As it does, the filter impulse response becomes shorter, reducing the effective integration time of the filter. The effect of this frequency-dependent integration on the temporal structure in RN and AABB stimuli can be observed by comparing the running autocorrelation of the stimuli with a very short time constant, 1.5 ms, before filtering and after filtering with a gamma tone filter. As shown in Wiegrebe *et al.* (1998), the peak in the normalized running

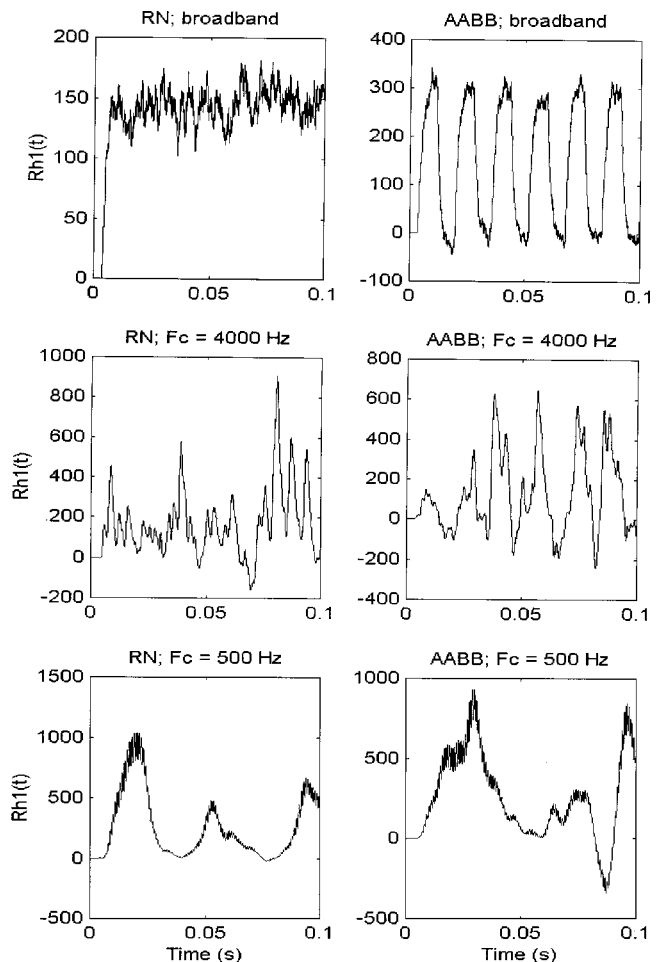


FIG. 7.  $Rh1(t)$  as a function of time for RN (left column) and AABB stimuli (right column). In the upper row,  $Rh1(t)$  is calculated on the raw waveform; in the middle row, stimuli were first filtered with a gammatone filter with  $F_c$  equals 4 kHz; in the lowest row,  $F_c$  equals 500 Hz. After the 4-kHz filtering, the  $Rh1(t)$  oscillations of AABB are still preserved whereas after the 500-Hz filtering where the filter bandwidth is only about one-sixth of the 4-kHz filter, the  $Rh1(t)$  oscillations are strong but random like those of RN. Note that, unlike in Wiegrebe *et al.* (1998),  $Rh1(t)$  is not normalized.

autocorrelation,  $Rh1(t)$ , fluctuates randomly around its average for RN stimuli whereas, for AABB stimuli, it oscillates with a period equal to two times  $d$ .  $Rh1(t)$  for RN stimuli with  $d$  equals 8 ms is shown in the left column of Fig. 7;  $Rh1(t)$  for AABB with  $d$  equals 8 ms is shown in the right column.  $Rh1(t)$  is shown for unfiltered stimuli in the upper row, for stimuli filtered with a 4-kHz gammatone filter in the middle row and for stimuli filtered with a 500-Hz gammatone filter in the lower row. The  $Rh1(t)$  oscillations for AABB are clearly visible for the unfiltered stimuli (upper right panel) and are still reasonably well preserved after the 4-kHz gammatone filter (middle right panel). However, for the AABB stimulus filtered with the 500-Hz gammatone filter (lower right panel), the period of the oscillations is smeared by the long integration time of the 500-Hz filter. It is important to point out that the degree of smearing does not depend on the absolute filter bandwidth but on the relationship between filter bandwidth and  $d$ . Thus although the oscillations in the autocorrelation of AABB are smeared out in the 500-Hz gammatone filter when  $d$  is 8 ms, the oscillations



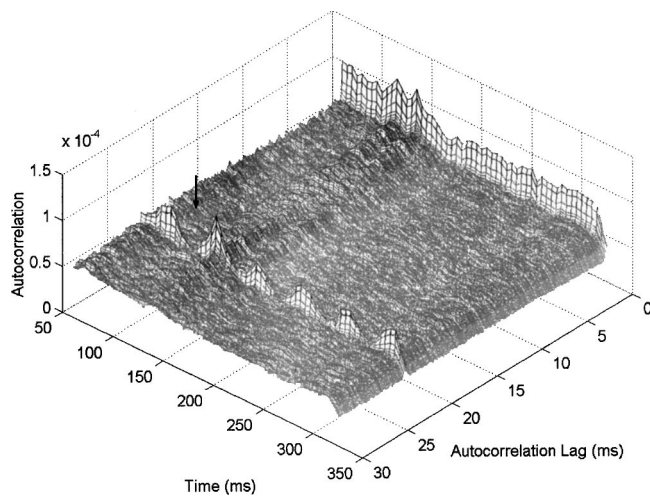


FIG. 8. Summary autocorrelograms (SACG) as a function of time for an ABB stimulus with  $d$  equal to 22.4 ms. The first peak in the SACG (arrow) fluctuates strongly and periodically over time. However, these fluctuations are not perceived as shown in Wiegrebe *et al.* (1998).

would still be preserved in the same filter when  $d$  is say 16 or 32 ms. In the next section, we show that the preservation of  $Rh1(t)$  fluctuations interacts with the nonlinear transformation (the half-wave rectification) following the gammatone filters.

## 2. The effect of half-wave rectification

As noted in the Introduction, the long-term autocorrelation functions of RN and ABB are identical. However, when the stimuli are half-wave rectified, this is no longer the case. Following half-wave rectification, the height of the first peak in the normalized stimulus autocorrelation function ( $h1$ ) is higher for ABB than for RN. The mathematical evidence for this interaction of half-wave rectification and temporal regularity is not trivial but the effect is found very consistently. When averaged over 10 stimuli,  $h1$  for ABB was 0.67, whereas  $h1$  for RN was only 0.60; the standard deviations were 0.0053 and 0.0119, respectively. The differential effect of half-wave rectification disappears, however, when the temporal oscillations of ABB are smeared out which, as shown above, happens when the oscillation period of the stimulus is in the same range as or shorter than the auditory-filter impulse response.

Thus the cumulative effect of frequency-dependent filter width and nonlinearity like half-wave rectification can account for the frequency-dependent pitch-strength differences found experimentally.

## 3. Running autocorrelation versus long-term autocorrelation

The third stage of the auditory model involves long-term autocorrelation instead of running autocorrelation with either a 10-ms time constant (Meddis and O'Mard, 1997) or a 2.5-ms time constant (Meddis and Hewitt, 1991a, 1991b). This was done for the following reason: Even with a 10-ms time constant, the running autocorrelation for ABB with a long  $d$  fluctuates strongly and periodically over time. This is illustrated in Fig. 8, where the summary autocorrelogram for

an ABB stimulus with  $d$  equals 22.4 ms is plotted as a function of time. Note, however, that these fluctuations are not perceived. This was demonstrated in experiment III of Wiegrebe *et al.* (1998). When  $d$  was 22.4 ms and the IRN pitch-strength was matched to that of ABB, the two stimuli were no longer discriminable although the autocorrelation fluctuations were periodic and large for ABB versus random and small for RN. Meddis and O'Mard (1997) and Meddis and Hewitt (1991a, 1991b) did not actually calculate *running* autocorrelation; rather they calculated the short-term autocorrelation of a stimulus snapshot at the end of the stimulus. Figure 8 shows that this approach is not applicable to ABB stimuli because, depending on the phase at the end of the stimulus, a pitch-strength estimate could be well above or well below the pitch strength that listeners hear. Thus a considerably longer integration time constant is required if running autocorrelation is used. The value of this time constant is not known and so, for the time being, we used long-term autocorrelation where the integration time is only limited by the stimulus duration.

## IV. SUMMARY AND CONCLUSIONS

The current experiments like those in Wiegrebe *et al.* (1998) reveal pitch-strength differences between ABB and RN stimuli, which represent a challenge for models of pitch perception because the long-term characteristics of ABB and RN are identical. The current experiments show that pitch-strength differences of ABB and RN stimuli depend systematically on the center frequency of the bandpass filter. When  $d$  is 4 ms, ABB-RN discrimination is poor and pitch-strength differences are small when the stimuli are lowpass filtered at 1 kHz, whereas discrimination is good and pitch-strength differences are large when the stimuli are filtered between 3.5 and 4.5 kHz. This effect cannot be explained with a purely waveform-based model as in Wiegrebe *et al.* (1998). Simulations with a model of the auditory periphery similar to Meddis and O'Mard (1997) show that auditory filtering and half-wave rectification affect the representation of RN and ABB differentially. The long-term autocorrelation of the ABB and RN stimuli processed in this way and summed up over frequency channels yields estimates of pitch-strength differences in good agreement with the current data. However, it must be pointed out that the model in its current form is not able to predict changes of pitch strength across different  $d$ 's and frequency regions. The current experiments, together with those of Wiegrebe *et al.* (1998), question the value of running autocorrelation with a fixed time constant in the range of 10 ms as suggested by Meddis and O'Mard (1997). With such a short time constant, the model would predict that ABB pitch-strength for higher  $d$ 's (lower pitches) would fluctuate periodically, which it does not (cf. Wiegrebe *et al.*, 1998, experiment III).

Based on the current experiments and simulations, it is possible to relate the processing stages of the functional model in Wiegrebe *et al.* (1998) to properties of the peripheral auditory system. The first integration stage implemented by the integration window of the running autocorrelation corresponds to the frequency-dependent temporal integration of the cochlear filters. The nonlinear transformation in the

functional model corresponds to the half-wave rectification in auditory mechano-electrical transduction. The second integration stage is implemented by averaging over the stimulus duration in both the functional and the auditory model. Despite the fact that the temporal properties of the second integration stage still remain obscure, the current research shows that pitch strength is substantially influenced not only by the filtering properties but also by the nonlinear transformations of the peripheral auditory system.

## ACKNOWLEDGMENTS

This study was supported by the Medical Research Council (G9703469), U.K. by a research grant from the "Deutsche Forschungsgemeinschaft" to the first author, and by a grant from the "Prof. Dr.-Ing. Erich Müller Stiftung" to the second author.

de Cheveigné, A. (1998). "Cancellation model of pitch perception," *J. Acoust. Soc. Am.* **103**, 1261–1271.

Houtsma, A. J. M., and Smurzynski, J. (1990). "Pitch identification and discrimination for complex tones with many harmonics," *J. Acoust. Soc. Am.* **87**(1), 304–310.

Meddis, R. (1988). "Simulation of auditory-neural transduction: Further studies," *J. Acoust. Soc. Am.* **83**, 1056–1063.

Meddis, R., and Hewitt, J. (1991a). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification," *J. Acoust. Soc. Am.* **89**(6), 2866–2882.

Meddis, R., and Hewitt, J. (1991b). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity," *J. Acoust. Soc. Am.* **89**(6), 2883–2894.

Meddis, R., and Hewitt, J. (1992). "Modeling the identification of concurrent vowels with different fundamental frequencies," *J. Acoust. Soc. Am.* **91**(1), 233–245.

Meddis, R., and O'Mard, L. (1997). "A unitary model of pitch perception," *J. Acoust. Soc. Am.* **102**(3), 1811–1820.

Moore, B. J. C., 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.

Patterson, R. D., Allerhand, M., and Giguere, C. (1995). "Time-domain modeling of peripheral 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**, 5.

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

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, pp. 357–360.

Wiegrefe, L., Patterson, R. D., Demany, L., and Carlyon, R. P. (1998). "Temporal dynamics of pitch strength in regular interval noises," *J. Acoust. Soc. Am.* **104**(4), 2307–2313.

Wiegrefe, L., and Patterson, R. D. (1999). "The role of envelope modulation in spectrally unresolved iterated rippled noise," *Hear. Res.* **132**, 94–108.

Yost, W. A. (1996a). "Pitch of iterated rippled noise," *J. Acoust. Soc. Am.* **100**(1), 511–518.

Yost, W. A. (1996b). "Pitch strength of iterated rippled noise," *J. Acoust. Soc. Am.* **100**(5), 3329–3335.

Yost, W. A., and Hill, R. (1978). "Strength of the pitches associated with ripple noise," *J. Acoust. Soc. Am.* **64**, 485–492.

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.

Yost, W. A. (1998). "Auditory processing of sounds with temporal regularity: Auditory processing of regular interval stimuli," in *Psychophysical and Physiological Advances in Hearing, Proceedings 11th International Symposium on Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London).