Modeling temporal asymmetry in the auditory system

Roy D. Patterson
Centre for the Neural Basis of Hearing, Physiology Department, University of Cambridge, Downing Street, Cambridge, CB2 3EG, United Kingdom

Toshio Irino
ATR Human Information Processing Research Laboratories, 2-2 Hikaridai Seika-cho, Soraku-gun Kyoto, 619-0288, Japan

(Received 24 July 1997; accepted for publication 14 August 1998)

Sound sources in the environment produce waves that are almost invariably asymmetric in time, and human listeners are highly sensitive to temporal asymmetry. The spectral analysis and neural transduction processes in the cochlea enhance temporal asymmetry, as do time-domain models of cochlear processes, but it appears that the resulting asymmetry is not sufficient to explain the observed perceptual asymmetry. In the auditory image model (AIM) of hearing, the temporal asymmetry in the neural activity produced by the cochlea is further enhanced by the ''strobed'' temporal integration that converts the neural activity pattern into an auditory image, and the temporal asymmetry in the auditory image is sufficient to explain the perceptual asymmetry. Modern versions of the ''duplex model'' of pitch have time-domain cochlea simulations that produce neural activity with temporal asymmetry similar to that produced by AIM. In the final stage, however, they apply autocorrelation to the neural pattern and autocorrelation is a symmetric process in time. In this paper the effect of autocorrelation on temporal asymmetry is examined in a range of auditory models with varying forms of auditory filterbank, compression, and neural transduction. It is concluded that autocorrelation does not enhance temporal asymmetry and often reduces it, and that autocorrelogram models cannot explain the magnitude of the perceptual asymmetry in their current form. Then, the original version of strobed-temporal-integration is reviewed with regard to temporal asymmetry, and the delta-gamma theory of temporal asymmetry [Irino and Patterson, J. Acoust. Soc. Am. **99**, 2316–2331 (1996)] is used to develop a new version of strobed-temporal-integration that is more robust and physiologically more plausible. [S0001-4966(98)05711-7]

PACS numbers: 43.66.Ba, 43.66.Jh, 43.66.Mk [RVS]

INTRODUCTION

A. The perception of temporal asymmetry

The ASA set of ''auditory demonstrations'' (Houtsma *et al.*, 1987) includes a very compelling illustration of the effects of short-term temporal asymmetry on auditory perception (No. 29). A piece of music is played on the piano, and then it is repeated with the waves for the individual notes reversed in time. The melody and harmony are largely unaffected by the manipulation, but the instrument is perceived to change from a piano to a reed organ whose notes end in disruptive complex transients. The existence of the demonstration shows that the effect of time reversal has been known for some time. Nevertheless, there was little research on the topic until Patterson (1994a, 1994b) initiated a systematic study of auditory temporal asymmetry with ''damped'' and ''ramped'' sinusoids. The damped sinusoid has a repeating, exponentially decaying envelope, and is illustrated with a 4-ms half-life in Fig. 1(a); the ramped sinusoid has a repeating, exponentially rising envelope, shown with the same half-life in Fig. 1(b). The perceptions produced by these sounds have two components: a drumming component produced by the stream of abrupt transients at the start of each cycle of the envelope, and a continuous tonal component with the pitch and timbre of the carrier. As the half-life increases from 1 to 100 ms, the relative loudness of the drumming component decreases while that of the tonal component increases.

What makes the sounds interesting for auditory modellers is that pairs with the same half-life have identical power spectra when calculated over an integer number of periods, but they are, nevertheless, discriminable over a wide range of half-lives, envelope periods, and carrier frequencies. Thus, they pose a serious problem for traditional spectral models of auditory perception (Patterson, 1994a). Akeroyd and Patterson (1995) extended the phenomenon to noise carriers using the same discrimination technique. The carrier is heard as hiss rather than a tone, and listeners were asked to discriminate damped and ramped pairs with the same half-life on the basis of the relative loudness of the drumming and hiss components. The long-term power spectra of pairs with the same half-life are the same, and, in this case, the short-term power spectra are the same, except the level changes in the short-term spectra of the ramped sound come in the reverse order to those of the damped sound. In a spectral model, then, the fact that the ramped noise produces a relatively louder hiss component would have to be explained, *post hoc*, on the basis of the order of the short-term spectra. Fay *et al.* (1996) showed that goldfish, which have no basilar membrane, can nevertheless discriminate damped and ramped sinusoids, and that generalization from a ramped

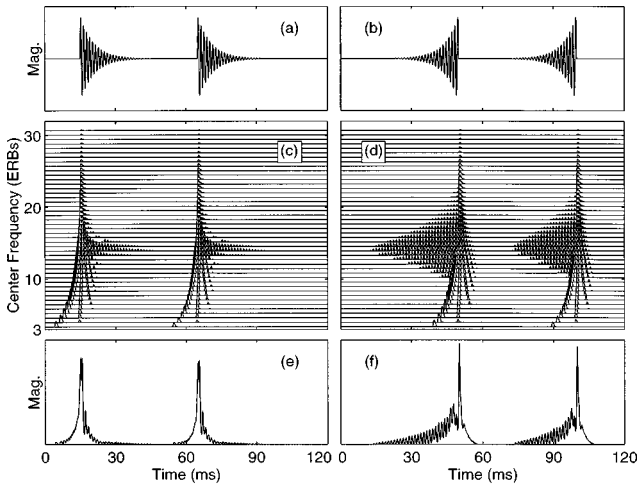

FIG. 1. Phase-compensated NAPs for damped (left) and ramped (right) sinusoids. The top row shows two cycles of the damped (a) and ramped (b) waves. The middle row shows the damped (c) and ramped (d) NAPs produced by the ***gtf/2dat*** model. The ordinate is channel center frequency on an ERB scale. The bottom row shows damped (d) and ramped (f) summary NAPs. The peak concentration is the activity in a 4-ms region about the peak divided by the activity in the complete cycle, and it is greater for the damped sound.

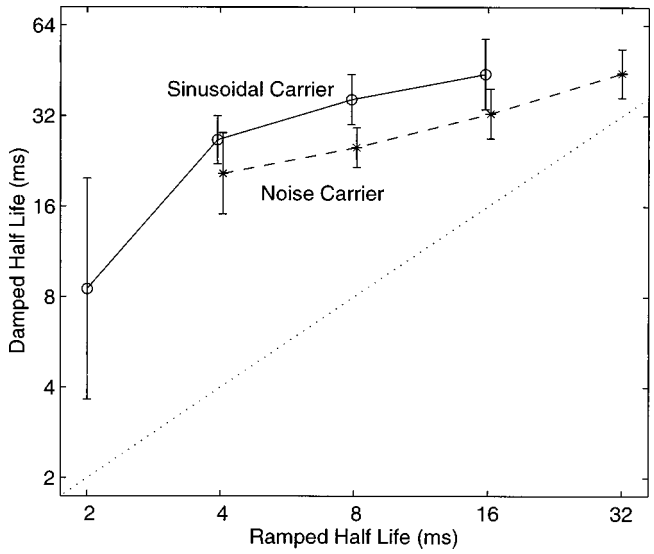

FIG. 2. Matching half-life values for damped and ramped sounds with sinusoidal carriers (open circles) and noise carriers (stars). Average data for four listeners. The matching half-life is the damped half-life required to produce the same perceptual click/carrier ratio as that produced by a given ramped sound. The equal half-life line is shown by the dotted diagonal; the average distance from the data to the diagonal is the measure of temporal asymmetry.

sinusoid to a flat-envelope sinusoid is stronger than generalization from a damped sinusoid to a flat-envelope sinusoid. Lorenzi *et al.* (1997) showed that cochlear implantees can discriminate damped and ramped sinusoids when the stimuli are presented on a single electrode, and that their performance was far superior to that of normals at longer half-lives. Since the implant bypasses the cochlea and stimulates the auditory nerve directly, it is difficult to see how a spectral model could explain this discrimination.

B. The measurement of temporal asymmetry

Irino and Patterson (1996) refined the damped/ramped discrimination experiment to provide a direct measure of auditory temporal asymmetry. The experiment is described in some detail because the measurement and quantification of the perceptual asymmetry are central to the modeling studies presented in the paper that follows. In a two-alternative, forced-choice experiment, listeners were presented a ramped sinusoid in one interval and a damped sinusoid having the same *or greater* half-life in the other interval. In one version of the experiment, the listeners were asked to choose the interval containing the sound with the louder tonal component; in another version with the same stimuli, they were asked to choose the interval with the louder drumming component. Between trials, the half-life of the damped sinusoid was varied to determine the ''matching point,'' that is, the half-life of the damped sinusoid required to equate the probability of choosing either the ramped or damped stimulus as the one with the louder tonal or drumming component. Then the experiments were repeated using noise carriers, and listeners were asked to choose either the interval with the louder hiss component or the interval with the louder drumming component, relative to the other component. The results showed that listener variability was exceptionally low; the type of response did not affect either the form *or the horizontal position* of the psychometric functions with either carrier, and so the data were averaged over listener and response type.

The average matching-point data for the sinusoidal and noise carriers are presented by open circles and asterisks, respectively, in Fig. 2 (Fig. 4 of Irino and Patterson, 1996). The abscissa is the half-life of the ramped sound, and the figure shows the half-life that the damped sound must have to produce a perception in which the two components have the same relative loudness. If there were no perceptual asymmetry, the data would lie along the dotted diagonal. All of the data lie well above this line, indicating that, relative to the drumming component, the carrier is substantially louder in the ramped sounds when the half-life is in the range 2–32 ms. The open circle above 4 ms shows the extreme case; the damped sinusoid has to have a half-life of about 25 ms to produce a tonal component with the same relative loudness as that produced by the 4-ms ramped sinusoid. In other words, to match the tonal component of the ramped sound in Fig. 1(b), the half-life of the damped sinusoid in Fig. 1(a) has to be extended to the point where the amplitude of the carrier at the end of the envelope cycle is still fully half the starting height! Comparison of the open circles and asterisks shows that the asymmetry with the sinusoidal carrier is greater than the asymmetry with the noise carrier. Irino and Patterson (1996) suggested that the main effects could be simply summarized for modeling purposes in terms of two ''asymmetry factors,'' one for the sinusoidal carrier and one for the noise carrier. The asymmetry factor was defined as the average distance in logarithmic units between the matching half-life and the equal half-life (dotted diagonal) for all of the ramped half-lives associated with one carrier. For the data in Fig. 2, the asymmetry factors for the sinusoidal and noise carriers

are 2.1 and 1.4, respectively.[1] That is, the matching half-lives of the damped sounds are, on average, about 4 and 2.5 times larger than those of the ramped sounds, respectively. The data in Fig. 2 and the asymmetry factors are the focus of the asymmetry modeling in this paper.

C. Modeling temporal asymmetry

In the damped/ramped papers mentioned above, the authors explain the perceptual asymmetries in terms of temporal asymmetries in the auditory image model (AIM) (Patterson *et al.*, 1992, 1995). The model consists of an auditory filterbank that produces a representation of basilar membrane motion (BMM), a multi-channel, neural transduction mechanism that produces a representation of the neural activity pattern (NAP) in the auditory nerve, and a bank of ''strobed'' temporal integration units that produce the model's representation of the auditory image that we hear in response to the sound. Irino and Patterson (1996) showed that all three stages of the model enhanced temporal asymmetry as the information passed through, and they argued that all three transformations were required to explain the magnitude of the perceptual asymmetry shown in Fig. 2. They summarize their studies in terms of a general ''delta-gamma'' theory of temporal asymmetry in the auditory system.

AIM has essentially the same architecture as the original autocorrelogram model of pitch perception proposed by Licklider (1951), computational versions of which have been developed by Lyon and colleagues (Lyon, 1982, 1984; Slaney and Lyon, 1990), Assmann and Summerfield (1989, 1990), Meddis and Hewitt (1991a, b, 1992), and Brown and Cooke (1994) among others. Brown and Cooke (1994) present a review of autocorrelation models and their uses. In each of these autocorrelogram (ACG) models there is an auditory filterbank that simulates BMM, a multi-channel transduction mechanism that simulates the NAP, and a bank of autocorrelators to produce the ACG, and so the primary difference between AIM and the traditional ACG model is in the final stage, where the former has strobed temporal integration and the latter has autocorrelation. AIM and the ACG model often produce very similar results. For example, both have been used to explain the pitch and pitch strength of high-pass filtered iterated rippled noise (IRN) (Yost *et al.*, 1996; Patterson *et al.*, 1996; Yost *et al.*, 1998), which appears to be beyond the capabilities of spectral model of hearing, even those based on short-term spectra. The delay-and-add process used to generate IRN, introduces a ripple into the power spectrum of the stimulus that can be used to explain the pitch in a spectral model when the ripple is resolved in the auditory system. However, the pitch persists when the stimulus is high-pass filtered to remove the frequency region where the ripple would be resolved.

AIM and the ACG model produce similar predictions for the pitch and the pitch strength of IRN (Yost *et al.*, 1996; Patterson *et al.*, 1996), which led to the question as to whether AIM and the ACG model could be distinguished by their ability to explain the perception of temporal asymmetry. The reason for doubting the ACG model was that autocorrelation is a symmetric process in time; given a periodic wave like a static vowel with an intraperiod waveform that is asymmetric in time, the autocorrelation of the sound is, nevertheless, symmetric within the autocorrelation cycle. For example, see the auditory image and the ACG of the vowel /ae/ presented in Figs. 2(c) and 3(c) of Patterson *et al.* (1995). Irino and Patterson (1996) argued that the strobed temporal integration mechanism in AIM accentuates the shape of ramped features in the NAP but not those of damped features (see Sec. III). As a result, it enhances the asymmetry of ramped and damped sounds in the auditory image to the level where it can explain the magnitude of the perceptual asymmetry shown in Fig. 2. If autocorrelation reduces temporal asymmetry rather than enhancing it, then it seems likely that ACG models will not produce sufficient temporal asymmetry to explain the magnitude of the observed differences in relative loudness.

D. Overview of the paper

In Sec. I, the temporal asymmetry produced by the default version of AIM (Release 7) is compared with that observed when AIM's strobed temporal integration is replaced by autocorrelation. Then, the temporal asymmetry observed with the autocorrelation model of Meddis and Hewitt (1991a, b) is compared with that observed when the autocorrelation module in that model is replaced by strobed temporal integration. In both cases, strobed temporal integration produces sufficient temporal asymmetry to explain the magnitude of the perceptual asymmetry, and, in both cases, autocorrelation does not produce sufficient temporal asymmetry to explain the perceptual asymmetry.

The first stage of both AIM and the Meddis and Hewitt model is a linear gammatone filterbank. Cochlear filtering, however, is nonlinear and arguably more asymmetric in time than gammatone filtering. Accordingly, in Sec. II, the gammatone filterbank of the Meddis and Hewitt model is replaced with the active cochlea simulation of Giguère and Woodland (1994) to determine whether this will lead to more asymmetry in the resulting ACGs. Once again the analysis reveals insufficient asymmetry to explain the magnitude of the perceptual asymmetry.

The form of compression varies from model to model and it was suggested that this might affect temporal asymmetry, and so a separate study was performed in which three forms of compression were crossed with three forms of neural transduction to determine the effects on asymmetry. It is concluded that compression has little effect on asymmetry and so the studies are presented as an Appendix.

The analysis of asymmetry in the auditory images produced by the default version of AIM (R7) (Sec. I E) reveals that the original strobe mechanism explains the form and magnitude of the perceptual asymmetry better than the more physiological ''delta-gamma'' strobe mechanism developed in Irino and Patterson (1996). In Sec. III, we first determine why this is so, and then use this information to develop a more robust version of the delta-gamma strobe that can explain the form and magnitude of the perceptual asymmetry.

Terminology: Over the course of this paper, upwards of a dozen different auditory models are reviewed with regard to their temporal asymmetry. To assist in distinguishing the models, they are referred to by abbreviations involving the

processes that they employ, and these labels are presented in bold, italic, lower-case symbols. The auditory representations produced by the models are distinguished by abbreviations presented in unbold, upper-case symbols. The Meddis and Hewitt (1991a) model consists of a gammatone auditory filterbank, ***gtf***, to simulate basilar membrane motion (BMM), a bank of Meddis (1986, 1988) haircells, ***med***, to simulate the neural activity pattern (NAP) in the auditory nerve, and a bank of autocorrelators, ***ac***, to produce the autocorrelogram (ACG). It is referred to as a ***gtf/med/ac*** model. AIM has the identical auditory filterbank for spectral analysis, it uses two-dimensional adaptive thresholding, ***2dat***, to simulate the NAP, and it uses strobed temporal integration, ***sti***, to produce the auditory image (AI) (Patterson *et al.*, 1995). So it is a ***gtf/2dat/sti*** model. The modules required to assemble the default version of AIM (R7) and the Meddis and Hewitt (1991a) model are available in the software package described by Patterson *et al.* (1995).[2]

I. ASYMMETRY IN AIM R7 AND MEDDIS AND HEWITT (1991a)

The filtering and transduction processes in the AIM and Meddis and Hewitt (1991a, b) are *all asymmetric in time*. It is a natural property of causal, physical systems. Thus, the question, as noted by Irino and Patterson (1996), is not so much 'Where does asymmetry arise in the auditory system?' but rather, 'Which asymmetry dominates and is primarily responsible for the perceptual asymmetry we hear?' To answer the question and evaluate the effect of ***ac*** on asymmetry, we follow the approach set out by Irino and Patterson. That is, we identify structures associated with the transient and carrier components of the perceptions in the model output, and develop a measure of the relative loudness of the components (the peak concentration, PC), a measure that is applicable to all of the representations of sounds produced by these models. The matching point for a given ramped sound is determined by calculating its PC and then finding the damped half-life required to produce the same PC value. In this way, the matching points for all conditions in the experiment can be calculated for a particular form of model output and converted to asymmetry factors to compare with those of the experimental data.

A. Asymmetry measures derived from auditory models

The auditory models described in this paper begin with 75-channel auditory filterbanks covering the frequency range 100 to 6000 Hz, and all of the NAPs, ACGs, and AIs have the same number of channels. The asymmetry information is distributed both in time and frequency in all these different representations. The decision statistic developed by Irino and Patterson to summarize the asymmetries and predict listeners' performance is illustrated in Fig. 1, which is adapted from Fig. 8 of Irino and Patterson (1996). The upper panels show two cycles of damped and ramped sinusoids with carrier frequencies of 800 Hz, half-lives of 4 ms, and envelope periods of 50 ms. The middle panels show phase-aligned NAPs produced by the AIM in response to the stimuli. Patterson (1994a) argued that the drumming component of the perception arises from the vertical structure in the NAP which is produced by the transient in each cycle of the stimulus, whereas the tonal component of the perception arises from the horizontal triangular structure associated with the carrier in each cycle; the triangular structure follows the vertical structure in the damped NAP and precedes it in the ramped NAP.

To provide a measure of the relative loudness of the transient and carrier components, Irino and Patterson (1996) averaged the NAPs across frequency to produce summary NAPs, as shown in Fig. 1(e) and (f). Information about the transient component is concentrated about the peak in the NAP while information about the carrier is contained in the region away from the peak. Information about the timbre of the components is contained in the fine structure of the summary NAP. For example, the time intervals in the ramped summary NAP are highly regular, like those in the NAP itself, indicating that the timbre of the carrier is a tone and the pitch of the tone is 1/1.25 ms, or 800 Hz. The summary NAP of the ramped noise has a similar envelope but the time intervals in the region away from the peak are highly irregular, revealing that the timbre of the carrier is a hiss, and there is no pitch (see Akeroyd and Patterson, 1995, Fig. 4). In the experiment of Irino and Patterson (1996), however, the listeners were not required to identify the carrier; rather they were instructed to focus on the relative loudness of the carrier component (tone or hiss) compared to the transient component (drumming sound). The information about relative loudness is contained in the areas occupied by the transient and carrier portions of the summary NAP. There is relatively more activity in the region of the peak in the summary NAP of the damped sinusoid, which produces the perception with the louder drumming component, and there is relatively more carrier activity away from the peak in the summary NAP of the ramped sinusoid, which produces the perception with the louder tonal component. This led them to suggest that the relative loudness of the drumming component of the perception might be characterized by the ''peak concentration'' in the NAP; that is, the ratio of ''the average activity level in the 4-ms segment around the peak'' to ''the average activity level in the remainder of the cycle.''

The decision statistic was defined to be the peak concentration of the damped sound over the peak concentration of the ramped sound, and it was designated the peak concentration ratio (PCR). To predict the matching point for a given ramped sound, the half life of the damped sinusoid was varied to find the value that produced a PCR of unity. That is, the summary NAP of a given ramped sound was generated and used to calculate its peak concentration, and then summary NAPs and peak concentration values were generated for damped sounds with varying half-lives to find the one that produced the same peak concentration value as that of the ramped sound. The process was repeated for all ramped half lives, separately with the tone and noise carriers, to generate a complete set of matching points to compare with those in Fig. 2; then the logarithms of the matching-point values were averaged to produce asymmetry-factor values for the tone and noise carriers. The details of the PCR cal-

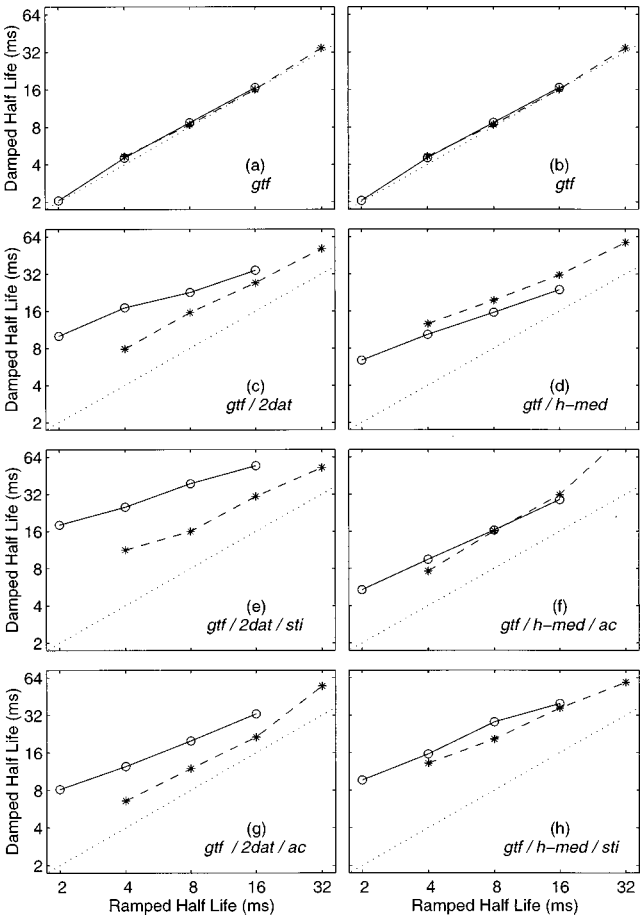

FIG. 3. Matching half-lives for sinusoidal (open circles) and noise (stars) carriers produced by the three stages of AIM (left column, ***gtf/2dat/sti***) and Meddis and Hewitt (1991) (right column, ***gtf/med/ac***). The gammatone filterbank [(a) and (b)] produces asymmetry only at 4 ms and the asymmetry is minimal. The cochlea simulation in AIM (c) produces more asymmetry than that of the Meddis and Hewitt model (d). The ***sti*** mechanism enhances asymmetry measured in the NAP (e) and (g); the ***ac*** module does not enhance asymmetry on average (f) and (h).

culation are presented in Appendix B of Irino and Patterson (1996).

B. Asymmetry in the auditory image model

The matching half-lives were determined for all of the conditions in the damped/ramped experiments at the output of each stage of the default version of AIM. The results are plotted in the upper three panels of the left-hand column of Fig. 3; each panel has the same format as the experimental data in Fig. 2. Then the ***sti*** module was replaced with an ***ac*** module and matching points were calculated from the resulting ACGs. The values are plotted in the bottom panel of the left-hand column of Fig. 3. Note that whenever there is asymmetry, it is invariably the case that the damped half-life is greater than the ramped half life. In this subsection, we compare the magnitude of the asymmetry at each level with that in the experimental data.

(i) ***gtf***: The measurement of asymmetry at the ***gtf*** stage was based on half-wave rectified versions of the individual filtered waves, which were phase aligned and averaged to produce a summary BMM in the same way as described for the summary NAP. The simulated matching-point data are presented in Fig. 3(a), which shows that there is virtually no measurable asymmetry for either carrier except at the 4-ms half-life where there is a small amount for both carriers. The impulse response of the gammatone filter has an exponential tail and its half-life is on the order of 4 ms in the region of 800 Hz. In this region, the response to the transient in the damped sound is slightly more concentrated than the response to the transient in the ramped sound. At shorter half-lives both ramped and damped sounds are like impulses to the gammatone filter; at longer half-lives the output just follows the stimulus with a short lag and the peak concentration is essentially the same for damped and ramped sounds. Thus, although it is true that gammatone filtering is asymmetric in time, and responses to ramped and damped sounds are different, the size of the difference is far too small for the PCR to explain the magnitude of the perceptual asymmetry observed in the experiments.

(ii) ***gtf/2dat***: The neural encoding stage in the AIM is a form of two-dimensional, adaptive thresholding ***2dat***, (Patterson and Holdsworth, 1996; Patterson *et al.*, 1995). It converts AIM's simulation of BMM into its simulation of the NAP. The adaptation in this module is highly asymmetric in time; the output of the module rises almost instantaneously with membrane amplitude, but the rate of decrease in the output after a peak is restricted to values on the order of 1 dB/ms. As a result, the asymmetry in adaptive thresholding interacts with that of the damped and ramped sounds over a larger range of half-lives than for gammatone filtering, increasing the negative slope of damped features and decreasing the positive slope of ramped features. The operation of adaptive thresholding and its role in asymmetry are discussed at length in Irino and Patterson (1996). The matching half-lives produced by sinusoidal and noise stimuli with ***gtf/2dat*** are presented in Fig. 3(c). The pattern of asymmetries is correct inasmuch as both carriers produce substantial asymmetry, and that for the sinusoidal carrier is greater than that for the noise carrier. The degree of asymmetry, however, is not as large as that observed in the experiments in the range 4–16 ms. The asymmetry factors for the experimental data (2.1 and 1.4 for the tone and noise carriers, respectively) are shown by the leftmost pair of bars in Fig. 4; they summarize the two sets of data presented in Fig. 2. The asymmetry factors for ***gtf/2dat*** [Fig. 3(c)] are presented by the second pair of columns in Fig. 4 over ''***2dat***,'' and they show in a more compact form that the asymmetry produced by ***gtf/2dat*** has the correct form but it is not large enough to explain the magnitude of the perceptual asymmetry.

(iii) ***gtf/2dat/sti***: The ***sti*** mechanism that converts NAPs into AIs in the original version of the AIM, employs an adaptive threshold somewhat like that in ***2dat***; it rises rapidly with level prior to a NAP peak and falls

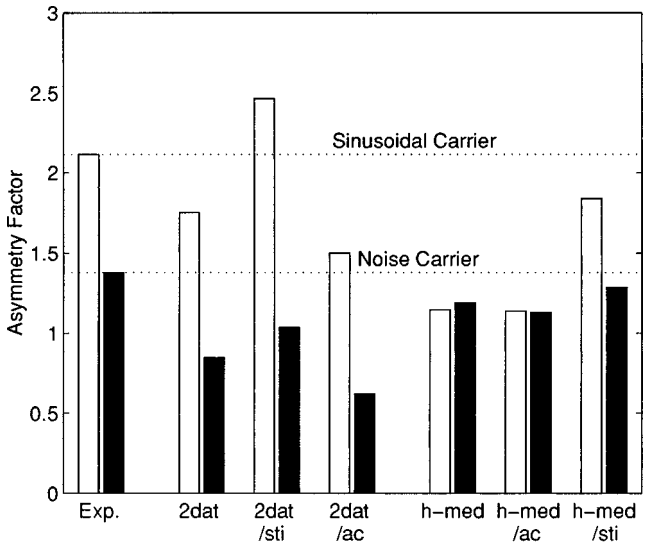

FIG. 4. Asymmetry factors calculated from the matching half-lives measured in the perceptual experiment (Exp), and from the representations underlying the predicted matching points in panels (c)–(h) of Fig. 3: on the left, the NAPs (***2dat***), AIs (***2dat/sti***), and ACGs from AIM (***2dat/ac***); on the right, the NAPs (***h-med***), ACGs (***h-med/ac***), and AIs (***h-med/sti***) from Meddis and Hewitt (1991a). Whereas ***sti*** enhances temporal asymmetry in the NAP, ***ac*** does not.

slowly after the NAP peak. In this version of ***sti***, however, the rate of decrease is 2%/ms, which is much slower than in ***2dat***. The strobe asymmetry enhances the damped/ramped asymmetry in the auditory image over that in the NAP, but in a different way; it has virtually no effect of damped features because they decay faster than the strobe threshold, but it does reduce the positive slope of ramped features because the ramped features induce multiple strobes per envelope period. The effect is described in detail in Sec. III B.

(iv) Auditory images are automatically phase aligned by the ***sti*** mechanism because it works asynchronously on individual channels, and the NAP peak that initiates temporal integration is mapped to the 0-ms interval in the auditory image (Patterson *et al.*, 1992). Thus, the PCR measure developed for NAPs is directly applicable to AIs. The matching-point data produced by damped and ramped sounds with sinusoidal and noise carriers in the auditory image, using the default version of the AIM, are presented in Fig. 3(e); the asymmetry factors are presented in the third pair of columns in Fig. 4. The pattern of asymmetries is correct, and, in this case, the average degree of asymmetry is close to that in the data. The matching points produced by sinusoidal carriers are a little greater than those observed in the experiments, and the matching points for noise carriers are a little less than those observed. Nevertheless, it is an excellent fit when compared with the problems of explaining temporal asymmetry with spectral models and leaky-integrator models (Patterson, 1994a; Akeroyd and Patterson, 1995; Patterson and Irino, 1996).

(v) ***gtf/2dat/ac***: To provide a direct comparison of the effect of ***ac***, a ***gtf/2dat/ac*** version of an autocorrelogram model was assembled using the first two stages of the AIM, that is, ACGs were generated from the AIM NAPs that produced the matching points in Fig. 3(c). The NAP asymmetry is relatively large and the asymmetry for the sinusoidal carrier is substantially greater than that for the noise carrier, so these NAPs might be expected to provide a sensitive test of the effects of ***ac***. The matching points from the ACGs are presented in Fig. 3(g) and the asymmetry factors are presented in the fourth pair of columns in Fig. 4. Comparison of Fig. 3(g) with Fig. 3(e) shows that, when applied to the same NAPs, ***ac*** produces less temporal asymmetry than ***sti***. Comparison of the second and fourth pairs of columns in Fig. 4 shows that ***ac*** actually reduces the temporal asymmetry of the NAPs presented as input to the ACG module.

C. Asymmetry in the model of Meddis and Hewitt (1991a, b)

AIM was not originally designed as an autocorrelogram model and so the analysis was repeated with the autocorrelogram model of Meddis and Hewitt (1991a, b) which is perhaps the most commonly referenced computational version of Licklider's (1951) autocorrelation model. Matching half-lives were determined for all of the conditions in the damped/ramped experiments at the output of each stage of the Meddis and Hewitt model (***gtf/med/ac***), and the results were plotted in the upper three panels of the right-hand column of Fig. 3. Then the ***ac*** module used to construct the ACGs was replaced with the ***sti*** module used to construct auditory images in the AIM, and matching points were calculated from the AIs to compare with those from the ACGs. The matching points are plotted in the bottom panel of the right-hand column of Fig. 3.

(i) ***gtf***: The Meddis and Hewitt model (***gtf/med/ac***.) employs the same gammatone auditory filterbank as the default version of the AIM, and so the matching points produced by this stage, and plotted in Fig. 3(b), are the same as those for the AIM, plotted in Fig. 3(a). The subfigure is included simply to maintain the parallel presentation of the two models.

(ii) ***gtf/med***: In the Meddis and Hewitt model the transduction is performed by a bank of haircell simulators (Meddis, 1986, 1988). Table I of Meddis (1988) presents two sets of parameter values for haircells that lead to medium- and high-spontaneous rate firing in primary fibers, and Table II of Meddis *et al.* (1990) presents two different sets of parameter values. The default parameter values for the Meddis module in the AIM software package are those associated with the medium spontaneous-rate fiber in Meddis *et al.* (1990) which will be referred to as ***m-med***. It has the greatest dynamic range and so is least restricted by the maximum firing rate of the haircell. Matching points were calculated with this version of the haircell and they are positive, indicating that the asymmetry associated with adaptation in the Meddis haircell does interact with the asymmetry of the stimuli. The size of

the asymmetry, however, was not nearly large enough to explain the perceptual asymmetry. The asymmetry factors for the sinusoidal and noise carriers were about 0.4 and 0.6, respectively.

(iii) The AIM package also includes parameter values for the high-rate fiber of Meddis *et al.* (1990) which will be referred to as ***h-med***. The high-rate fiber is more reactive than the medium-rate fiber, with a stronger onset response and stronger adaptation that together make ***h-med*** more asymmetric in time; the dynamic range of ***h-med***, however, is about 40 dB rather than the 60 dB of ***m-med***. Matching points produced by ***gtf/h-med*** are presented in Fig 3(d), the asymmetry factors are presented by the fifth pair of columns in Fig. 4, over ''***h-med***.'' They show that ***h-med*** produces much more asymmetry than ***m-med***, but it is still less than that required to explain the perceptual asymmetry. Comparison of Fig. 3(d) with Fig. 3(c) shows that, on average, the asymmetry produced by ***gtf/h-med*** is about the same as that produced by ***gtf/2dat***; however, with ***h-med***, the sinusoidal carrier produces *less* asymmetry than the noise carrier, which is incorrect.

(iv) ***gtf/h-med/ac***: The multi-channel NAPs from the Meddis haircell are converted into ACGs by applying ***ac*** separately to each channel of the NAP. The resulting ACGs are automatically phase aligned because ***ac*** is phase insensitive. So the PCR measure developed for NAPs is directly applicable to ACGs. The matching-point data produced by ***gtf/h-med/ac*** in response to sinusoidal and noise stimuli are presented in Fig. 3(f), and the asymmetry factors are presented in the sixth pair of columns in Fig. 4, over ''***h-med/ac***.'' Both figures show that the asymmetry in the ACG is essentially the same as that in the NAP from which it was derived [compare Fig. 3(f) with Fig. 3(d), and compare the fifth and sixth pairs of columns in Fig. 4]. Thus, the ***ac*** process in the Meddis and Hewitt model neither enhances nor attenuates the temporal asymmetry observed in the NAPs.

(v) ***gtf/h-med/sti***: The ***ac*** module was replaced by the ***sti*** module in the default version of the AIM and the matching points produced by the resultant ***gtf/h-med/sti*** model are presented in Fig. 3(h); the asymmetry factors appear in the rightmost pair of columns in Fig. 4. The pattern of asymmetries has the correct form and the average level of asymmetry is almost large enough to explain the magnitude of the perceptual asymmetry. The difference in asymmetry between sinusoidal and noise carriers is rather smaller than in the perceptual data.

(vi) ***Summary***: The fact that autocorrelation does not enhance temporal asymmetry was to be expected, inasmuch as autocorrelation is a symmetric process in time. Nevertheless, it seemed important to begin by demonstrating that the ACGs produced by the Meddis and Hewitt (1991a, b) model, or the AIM with ***ac*** in place of ***sti***, do not have sufficient temporal asymmetry to explain the magnitude of the perceptual contrast between damped and ramped sounds. A similar conclusion has recently been reported by de Cheveigné (1998). It seems likely that the ACG models of Assmann and Summerfield (1989, 1990) and Brown and Cooke (1994) cannot explain the magnitude of the perceptual asymmetry either, insofar as they employ very similar filterbanks for their spectral analyses and the same Meddis haircell models for neural transduction. It is possible that an extra stage could be added to the ACG model to enhance the temporal asymmetry of the ACG, but the prospects for this solution do not seem good inasmuch as the asymmetry with sinusoidal carriers is barely greater than that with noise carriers.

II. ASYMMETRY AND AUTOCORRELATION WITH A LEVEL-DEPENDENT AUDITORY FILTERBANK

Whereas the gammatone filterbank is linear, cochlear filtering is level dependent; the bandwidth of the filter increases and the duration of the impulse response decreases as the intensity of the stimulus increases (Evans, 1977). The dynamic range of damped and ramped sounds is large, so there are significant changes in the duration of the impulse response over the cycle of the sound. The nonlinear cochlear filter has a relatively broadband response to the peak amplitude in the transient and a relatively narrow-band response to the low-level portion of the carrier, and it is difficult to predict how the nonlinearities might interact with the temporal asymmetry of damped and ramped sounds. It is also the case that the phase response of the cochlea differs from that of the gammatone filterbank and the difference has measurable perceptual effects. Kohlrausch and Sander (1995) and Carlyon and Datta (1997) have measured masking period patterns for time-reversed, temporally asymmetric sounds (positive and negative Schroeder-phase waves) and shown that the peak factor in the internal representation of the two sounds is different. But again, it is not clear how this might interact with the temporal asymmetry of damped and ramped sounds.

A. Meddis and Hewitt (1991) with a level-dependent filterbank

There are several level-dependent filterbanks which are used to simulate cochlear filtering, and they can assist in understanding the interaction of level and temporal asymmetry (e.g., Lyon, 1982; Strube, 1985; Giguère and Woodland, 1994). The latter two of these filterbanks have separate, nonlinear, feedback circuits in the individual sections of the filterbank to simulate the effects of outer haircells. The compression in the feedback circuit is similar in form to that thought to exist in the cochlea (Giguère and Woodland, 1994), and so these filterbanks are arguably more appropriate as preprocessors for the Meddis haircell than the gammatone filterbank. The original Meddis (1986) haircell algorithm has an input compressor that is not properly part of the haircell simulation. It was included to control the dynamic range of the haircell input when the haircell is driven by a linear filterbank as in Meddis and Hewitt (1991). The physiological version of the AIM (Patterson *et al.*, 1995, Fig. 1, right column) is like a Meddis and Hewitt (1991) model with a some-

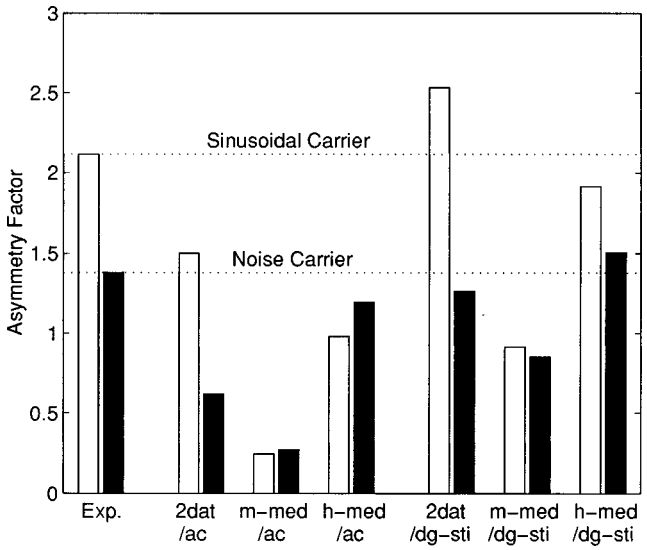

FIG. 5. Asymmetry factors calculated from the matching half-lives measured in the perceptual experiment (Exp), and from the matching half-lives predicted from six auditory models: On the left, three level-dependent ***(tlf)*** ACG models with different neural transduction stages (***2dat, m-med,*** and ***h-med***); they show that ***ac*** does not produce more asymmetry with a level-dependent filterbank ***(tlf)***. On the right, three level-independent (***gtf***) AI models (***dg-sti***) with the same neural transduction stages (***2dat, m-med,*** and ***h-med***); they show that ***dg-sti*** enhances temporal asymmetry in each case.

what more realistic cochlea simulation, that is, a level-dependent filterbank with compression. It is a combination of the Giguère and Woodland (1994) transmission-line filterbank (***tlf***), the Meddis *et al.* (1990) haircell model without the input compressor (***med***), and autocorrelation (***ac***). In this section, asymmetry factors for damped/ramped discrimination are calculated from the ACGs produced by this physiological version of the AIM (***tlf/med/ac***) using both the medium- and high-rate fibers from Meddis *et al.* (1990) to determine whether this would lead to greater asymmetry in the ACGs as suggested by Slaney (1994) (personal communication).

B. Phase-alignment in physiological NAPs

When calculating the PCR, phase alignment across channels is important for preserving the peak associated with the transient, as described in Sec. II A (Fig. 1). In the case of gammatone filtering, the appropriate phase compensation is a fixed temporal shift for a given channel, independent of level. This is not the case with transmission-line filtering, however, and we were not able to establish a reliable method of phase alignment and peak preservation for damped and ramped sounds. As a result, there are no asymmetry measures for the NAPs of the physiological model on their own. Nevertheless, ***ac*** produces phase alignment automatically, and so the temporal asymmetry of the model can be measured in the ACG.

C. Asymmetry in the level-dependent ACG

The asymmetry factors measured with the ***tlf/med/ac*** model are shown in the third and fourth pairs of columns in Fig. 5 for the medium- and high-rate fibers, respectively. They are essentially the same as the results produced with ***gtf*** in the original version of the Meddis and Hewitt model. There is not enough asymmetry to explain the magnitude of the perceptual asymmetry, and the asymmetry with sinusoidal carriers is not greater than that for noise carriers. The first pair of columns shows asymmetry factors measured in the experiment, as before. The second pair of columns shows the asymmetry factors produced by ***gtf/2dat/ac*** for comparison; these are the same values as in the fourth column of Fig. 4. Note that it is the combination of ***h-med*** with ***ac***, or ***m-med*** with ***ac***, that is the problem. In Sec. I C it was revealed that AIs produced with ***gtf/h-med/sti*** produce sufficient asymmetry, on average, to explain the magnitude of the perceptual asymmetry. And looking ahead briefly, in Sec. III B we show that AIs produced with ***tlf/h-med/dg-sti*** produce sufficient asymmetry to explain the perceptual asymmetry. So the problem is more with autocorrelation than with the Meddis haircell.

D. The effect of compression on asymmetry

The form of compression in the Meddis and Hewitt (1991) model is different from that in AIM, and the asymmetry measured at the output of the compressor stage in the AIM is a little greater than that with the filterbank on its own (Irino and Patterson, 1996, Fig. 5). It is also the case that the adaptation is different in the two models (***2dat*** vs. ***med***). As a result, we investigated the effect of compression on asymmetry using three forms of compression (none, square-root, and log), crossed with three forms of neural transduction (***2dat***, ***m-med***, and ***h-med***), to determine the contribution of compression to asymmetry. The analysis of these nine auditory models, and the argument as to why they represent a sufficient investigation of the compression issue, is fairly lengthy. Moreover, the analysis reveals that the form of compression has surprisingly little effect on asymmetry. As a result, the discussion is deferred to the Appendix.

III. ASYMMETRY IN STROBED TEMPORAL INTEGRATION

In Secs. I and II it was revealed that it is actually rather difficult to explain the pattern of asymmetries in the matching point data from the damped/ramped experiments, and that the successful fits produced by the AIM with the original version of ***sti*** are, in retrospect, somewhat surprising [Figs. 3(e) and (h)]. These asymmetry studies, with the original ***sti*** were performed explicitly for the comparisons described in Sec. I and had not been reported before; the asymmetry studies reported in Irino and Patterson (1996) were performed with the ''delta-gamma'' version of ***sti***. In this section of the paper, we examine the original strobe mechanism to determine how it enhances temporal asymmetry. Then, the results of the analysis are used to develop a new version of ***dg-sti*** which is more robust and physiologically more plausible, and which produces a slightly better fit to the asymmetry data. Finally, this new version of ***dg-sti*** is applied to the NAPs produced by the Meddis and Hewitt (1991) model to determine whether there is inherently sufficient temporal asymmetry in these NAPs to explain the perceptual asymmetry.

A. Ramped sounds and the original *sti*

The purpose of the strobe mechanism in the original version of AIM (Patterson *et al.*, 1992) was to identify local maxima in the NAP which were then used to initiate temporal integration. The NAPs of periodic sounds typically have only one maximum per cycle, and if integration is initiated on this maximum, the resulting auditory image is a stabilized copy of the pattern of pulses in the NAP. The local maxima are identified with the aid of an adaptive threshold and there is an independent adaptive threshold for each channel of the NAP. The operation is as follows: Threshold rises rapidly with NAP level on the leading edge of each NAP pulse up to the pulse peak, and the time of the peak is marked as a candidate strobe point. After the peak passes, the adaptive threshold decays at a rate of about 2%/ms. Before initiating temporal integration, however, the mechanism waits for up to 5 ms (the strobe lag) to see if another larger peak appears. If one does, the peak of the larger pulse becomes the strobe candidate and the strobe lag is reset to 5 ms. Eventually, the strobe lag times out, temporal integration is initiated, and the mechanism is reset.[3]

The NAPs of most natural periodic sounds have damped asymmetry, and they are readily stabilized by the simple strobe algorithm. The same is not true, however, for ramped sounds. Consider the response to a ramped sound with a long envelope period, say 500 ms, and a long half-life, say 64 ms. The strobe unit in the carrier channel would continually find larger NAP peaks on the way up the ramp and would keep on resetting the strobe lag to 5 ms until the end of the ramp. Thus, the sound level would be above absolute threshold for as much as 500 ms without any activity reaching the auditory image, which in the AIM means without the listener hearing anything. This is clearly incorrect; we would hear the onset of a slowly rising tone long before the local maximum at the offset of the tone. To solve this problem an extra condition was added to the strobe criterion; namely, ''Once a NAP pulse has been encountered, limit the duration of the search for a local maximum to a total of 10 ms.'' This version of ***sti*** is surprisingly successful in producing clean auditory images of ramped sinusoids; that is, images which preserve the carrier intervals of the horizontal triangular structure in the NAP *and* the filter-ringing intervals of the vertical structure in the NAP, even when the two occur adjacent to each other in the same channel. Patterson and Irino (1998) have shown that when carrier and filter-ringing intervals occur together in one channel, ***ac*** averages them and the ACG contains a range of time intervals that do not appear in the NAP.

This revised ***sti*** does not, however, produce an exact copy of repeating ramped NAP structures in the auditory image. In channels near the carrier frequency, the triangular structure is elongated horizontally and elevated in level because the mechanism strobes once or twice on the way up the ramp as well as at the end of the ramp. The time intervals in these channels are all carrier periods, both before and after the end of the ramp, and so the asynchronous strobing does not distort the fine structure in these channels. In channels farther away from the carrier region, the mechanism does not alter either the size or the form of the vertical transient structure. As a result, the peak concentration, which is the ratio of transient activity to carrier activity, is reduced for ramped sounds. This, in turn, means that the half-life of the matching damped sound has to be extended to reduce its peak concentration and restore the PCR to unity. Thus, the strobe mechanism enhances the asymmetry of damped/ramped sounds over that observed in the NAP.

The perception of the carrier component of a ramped sinusoid is pure, which in AIM means that there is the minimum of distortion of carrier intervals in the auditory image. Accordingly, the three strobe parameters, threshold decay rate, strobe lag, and strobe-lag limit, were varied to determine the values that would minimize time-interval distortion. The values of both strobe lag and strobe-lag limit were found to have little effect on image distortion over a fairly wide range and so they were left fixed at the default values of 5 and 10 ms, respectively. Threshold decay rate was found to have a substantial effect on strobe rate, and the value in the default version of the AIM (R7), 5%/ms, was found to produce too much strobing. Distortion in the image was reduced by decreasing the decay rate to 2%/ms, and this is the rate that was used to produce the matching points and asymmetry factors in Figs. 3 and 4. No parameter sets were found that produced one strobe per envelope period for the full range of ramped sounds. It seems that enhancement of temporal asymmetry is an unavoidable property of ***sti***—a property that emerges from the principles of image stabilization in the AIM.

B. Modification of the delta-gamma strobe

The original version of ***sti*** is based on logic; it is a simple algorithm for finding local maxima in NAP functions without regard to its physical or physiological plausibility. The ''delta-gamma'' version of ***sti*** developed by Irino and Patterson (1996) represents an attempt to move forward to a physical model in hopes of discovering general physical principles of auditory processing—principles that might eventually lead to a full physiological model of ***sti***. The delta gamma version of ***sti*** also has a better dynamic response than the original ***sti***; it overshoots less at onset and adapts to level changes faster. In this subsection we describe a revised version of ***dg-sti*** that produces the best asymmetry results to date. The process is illustrated with an extended example of the response of ***dg-sti*** to the NAPs of damped and ramped sounds with 16-ms half-lives and 50-ms periods, shown by the upper traces in Fig. 6(a) and (b), respectively. These are individual NAP functions from a channel about 200 Hz above the carrier frequency, 800 Hz. The perception of the damped sinusoid has the stronger click-to-tone ratio. In the AIM, this means that the mechanism issues strobe pulses less frequently for the damped sinusoid than for the ramped sinusoid. The bottom traces in Fig. 6(a) and (b) mark the strobe points issued by the modified version of ***dg-sti*** for

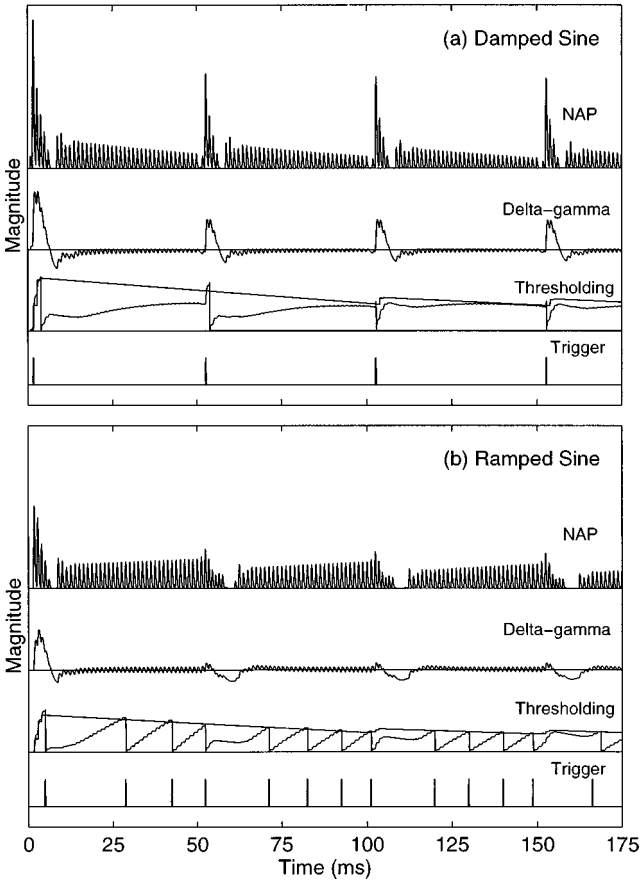

FIG. 6. Response of the delta-gamma strobe mechanism to damped (a) and ramped (b) sinusoids in the channel centred on 1.0 kHz. Delta-gamma (row 2) is the smoothed derivative of the NAP (row 1). It controls the rate at which activity accumulates for comparison with an adaptive threshold (row 3). When threshold is exceeded, a strobe pulse is issued (row 4). After the initial strobe pulse, delta-gamma causes activity to accumulate faster in the auditory image of the ramped sinusoid, thus enhancing temporal asymmetry.

these NAPs, and they show that strobing asymmetry occurs in ***dg-sti*** as in the original ***sti***. In this subsection we explain the operation of ***dg-sti*** and the matching-point data produced by this version of the AIM.

1. The architecture of the delta-gamma strobe

In AIM, there is a strobe unit for each channel of the NAP and they operate independently. The architecture of the ***dg-sti*** mechanism is presented in Fig. 7. The multi-channel NAP is represented schematically by the left-hand column of the figure. All of the modules to the right of the NAP column pertain to ***dg-sti*** for one NAP channel; namely, the one marked by the central, bold arrow of the set entering the summation. Delta-gamma is defined as the derivative of the smoothed envelope of the NAP. The envelope extractor is represented in Fig. 7 by the column of leaky integrators (LI) and the summation sign just to the right of the NAP column. The envelope is the weighted average of the smoothed NAPs from channels in a 3-ERB band about the given channel. Tc-short is 3 ms and the weighting is a Hamming function with unit area. The inclusion of the frequency dimension in the envelope calculation reduces the variability of the envelope estimate for a given time constant. The envelope is fed to the delta-gamma operator shown in the bottom panel of the central column. The envelope is also fed to an adaptive threshold (top panel) and an accumulator (middle panel) which between them determine precisely when a strobe should be issued. The delta-gamma operator controls the rate at which activity from the NAP accumulates in the decision process. When the level of activity in the accumulator exceeds the level of the adaptive threshold, a strobe pulse is issued and temporal integration is initiated.

Delta-gamma is basically the derivative of the envelope of the NAP. In this implementation, the derivative operator is preceded and followed by leaky integrators with a short Tc, 3 ms, to smooth the input and output. To limit the influence of extreme values, the delta-gamma value is passed through a sigmoid function with floor and ceiling values of 0 and 1, respectively. The slope of the sigmoid near its midpoint is a parameter of the model and in the current fits it is 2.

The output of the delta-gamma sigmoid controls the proportion of the NAP envelope that enters the accumulator which is a simple LI with a long Tc, 30 ms. The output of the accumulator is compared with the level of an adaptive

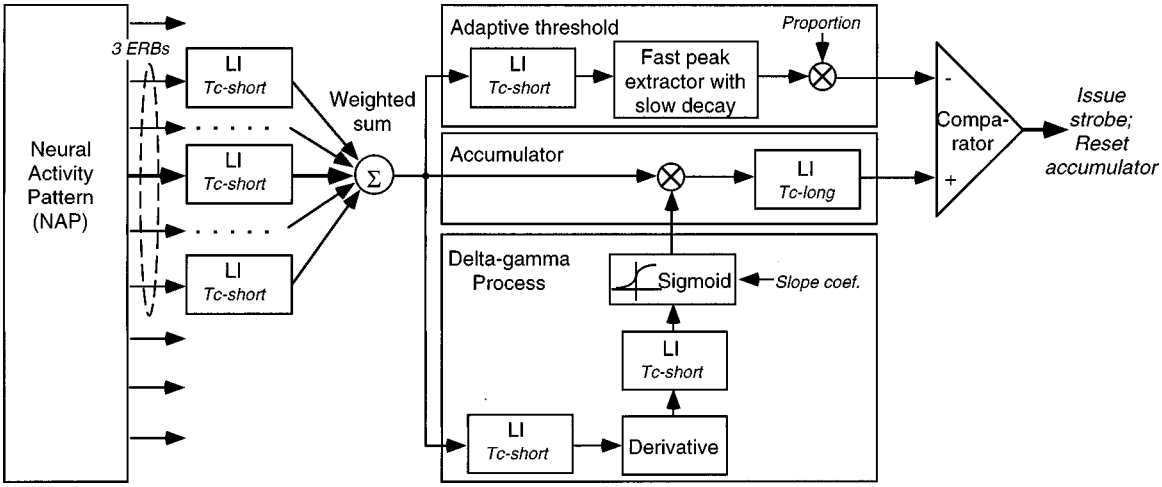

FIG. 7. Architecture of the delta-gamma strobe mechanism: The envelope of the NAP (col. 1) is extracted (col. 2) and fed to the delta-gamma process (col. 3) which determines the rate at which activity accumulates in the comparator (col. 4). When the activity level exceeds the adaptive threshold level, a strobe pulse is issued and it is reset.

threshold. The purpose of the adaptive threshold is to maintain the comparison value in roughly the same range as the level in the NAP channel. In order to strobe promptly in response to abrupt onsets, the mechanism must estimate the NAP level rapidly, and so the onset time-constant for the adaptive threshold is short (3 ms). In order to hold the estimated level for comparison over a reasonable length of time, the mechanism has a relatively slow decay (0.2%/ms). With these onset and offset characteristics, the adaptive threshold value tends to be set by onset responses where the neural transduction mechanisms are inclined to overshoot. As a result, the peak value is scaled down by a fixed proportion before being used for comparison; in the current simulations the proportion is about 0.15.

2. The operation of the delta-gamma strobe

The delta-gamma functions produced in response to the damped and ramped NAPs at the top of Fig. 6(a) and 6(b) are shown by the traces directly under those NAPs. The adaptive thresholds and accumulation functions produced in response to the damped and ramped NAPs are shown by the pair of traces directly under the delta-gamma traces. The slowly decaying trace is the adaptive threshold; the sawtooth function is the accumulator output. Every time the accumulator value exceeds the adaptive threshold, a strobe pulse is issued, as shown in the bottom trace of each figure, and then the accumulator is rest to zero.

Delta-gamma rises rapidly at the onset of both NAPs but the positive peak of the delta-gamma is greater for the damped sound and so the adaptive threshold for the damped NAP in Fig. 6(a) rises faster and to a higher level than that for the ramped NAP in Fig. 6(b). However, the accumulation rate is very high for both NAPs and so the accumulation value exceeds adaptive threshold shortly after onset in both cases and strobe pulses are issued. Shortly thereafter, delta gamma turns negative. The value is more negative for the damped sound because the recovery from overshoot is stronger in the damped sound. Moreover, the average value remains negative for the damped sound longer than for the ramped sound because the slope of the envelope of the NAP is negative for the damped sound, whereas it is positive for the ramped sound. As a result, the accumulation of NAP activity is relatively slow for the damped sound, and, since the adaptive threshold is relatively high, the accumulator does not exceed threshold until the start of the next cycle. The rising slope of the ramped sound leads to greater output from the delta-gamma operator, and, so, activity from the ramped NAP accumulates relatively quickly. The adaptive threshold is lower for the ramped NAP and so the level in the accumulator soon exceeds the adaptive threshold. The result is that ***dg-sti*** strobes two or three times during the rising portion of the cycle of the ramped sinusoid.

3. Asymmetry factors with the modified dg-sti

The asymmetry factors produced by the modified ***dg-sti*** operating on NAPs from the default version of AIM (R7) are presented by the fifth pair of columns in Fig. 5. The corresponding values with the same NAPs and the original version of ***sti*** were presented in the third pair of columns in Fig. 4. The comparison reveals that the modified ***dg-sti*** increases the noise asymmetry factor, bringing it closer to the observed values. The matching-point values produced by the modified ***dg-sti*** are a little smaller than the observed values for the shorter half-lives, and a little larger than the observed values for the longer half-lives; in general, however, they are closer to the observed values than those from the original version of ***sti***.

The final two pairs of columns show the asymmetry factors produced by ***dg-sti*** for NAPs from the level-dependent Meddis model, ***tlf/med/dg-sti***. With medium-rate fibers, ***dg-sti*** produces more asymmetry than ***ac*** (compare the sixth and third pairs of columns), but it is not sufficient to explain the magnitude of the observed asymmetry, and there is no carrier difference. With the high-rate fibers, ***dg-sti*** produces sufficient asymmetry to explain the observed asymmetry (seventh pair of columns), and there is a carrier difference that is in the correct direction. It is not as great as the observed carrier difference, but it is large enough to suggest that with a little tuning, ***tlf/h-med/dg-sti*** could explain the data as well.

IV. SUMMARY AND CONCLUSIONS

The autocorrelogram (ACG) model of hearing was developed to explain the pitch of complex sounds and it does this exceptionally well, including the pitch of high-passed iterated rippled noise which is beyond the scope of spectral models of pitch. This paper attempted to determine whether the ACG model could be extended to explain the timbre differences between spectrally matched pairs of damped and ramped sounds. Patterson and colleagues had demonstrated that the auditory image model (AIM) could explain the perception of damped and ramped sounds, and the initial stages of AIM are quite similar to those of ACG models. Accordingly, damped and ramped sounds were presented to two general forms of ACG model–the well-known Meddis and Hewitt (1991a, b) model and a version of AIM in which the strobed-temporal-integration (***sti***) module was replaced with an autocorrelation (***ac***) module. The results showed that neither form of ACG model could explain the magnitude of the perceived temporal asymmetry; that is, the fact that listeners require half-lives in damped sounds to be three to four times greater than those in ramped sounds to equate the relative loudness of the transient and carrier components of the perception. When the ***ac*** module of the Meddis and Hewitt model is replaced with an ***sti*** module, it produces sufficient asymmetry to explain the perceptual asymmetry provided the parameters of the Meddis haircell are set to simulate a high-spontaneous-rate fiber.

A series of studies was then performed with the autocorrelation model to determine whether damped/ramped asymmetry in the autocorrelogram, ACG, could be enhanced using a level-dependent auditory filterbank (Giguère and Woodland, 1994) and/or different forms of compression prior to the transduction stage in these models. The manipulations led to minor changes in the size and form of the temporal asymmetry, but there was never sufficient asymmetry in the ACG to explain the magnitude of the perceptual asymmetry. This suggests that, while autocorrelation models

are very successful in explaining pitch perception, they will need modification, and probably the addition of an extra stage, to explain the magnitude of the timbre discriminations associated with temporal asymmetry.

An analysis of the temporal integration of damped and ramped sounds with the original version of ***sti*** revealed that it is particularly difficult to stabilize the patterns of phase-locking information produced by periodic sounds where a rising tonal component ends in a transient (e.g., a ramped sinusoid). The analysis led to the development of a new version of delta-gamma strobe (***dg-sti***) that can stabilize ramped sounds accurately and explain the magnitude of the perceptual asymmetry quantitatively. The new form of ***dg-sti*** is also more plausible physiologically.

ACKNOWLEDGMENTS

Much of the modeling work and the first version of the paper were produced while the first author was a visiting researcher at NTT Basic Research Laboratories, Atsugi, Japan in the autumn of 1996. The authors would like to thank Malcolm Slaney for his suggestions with regard to reorganizing and simplifying the paper.

APPENDIX: COMPRESSION AND ASYMMETRY

1. The form of compression in AIM and Meddis and Hewitt (1991)

In the default version of AIM (R7) (Patterson *et al.*, 1995), there is logarithmic compression at the input to the neural transduction module, ***2dat***. The logarithmic function was chosen because the tail of the impulse response of the gammatone filter is a decaying exponential; log compression linearizes the decay function which greatly simplifies the adaptive thresholding algorithm. Log compression was also justified on the grounds that the pitch and timbre of complex sounds is largely invariant with level and log compression best captures this property; that is, the patterns that appear in the NAP vary least with level when the compression is logarithmic. In the Meddis and Hewitt (1991a, b) model, there is quasi-log compression at the input to the haircell module, ***med***, and there is a limit on the output firing rate that operates as a compressor at higher levels. It is this which limits the dynamic range of ***m-med*** to about 60 dB, and that of ***h-med*** to about 40 dB. In this Appendix, we investigate the effect of compression on asymmetry using three forms of compression (none, square-root, and log), crossed with three forms of neural transduction (***2dat***, ***m-med***, and ***h-med***).

We installed a variable compressor unit in the AIM software package after the filterbank and before the neural transduction modules (***2dat*** and ***med***). The unit was capable of applying no compression, log compression, or power-function compression. Several authors have recently reported that a power-law compressor with an exponent in the region of 0.5 (applied to amplitude) is a reasonable approximation to the compression applied by outer haircells over a wide range of levels in the normal cochlea (e.g., Allen, 1996; Plack, 1996; Oxenham, 1996). Consequently, we used power-function compression with an exponent of 0.5, that is, square-root compression as applied to amplitude.

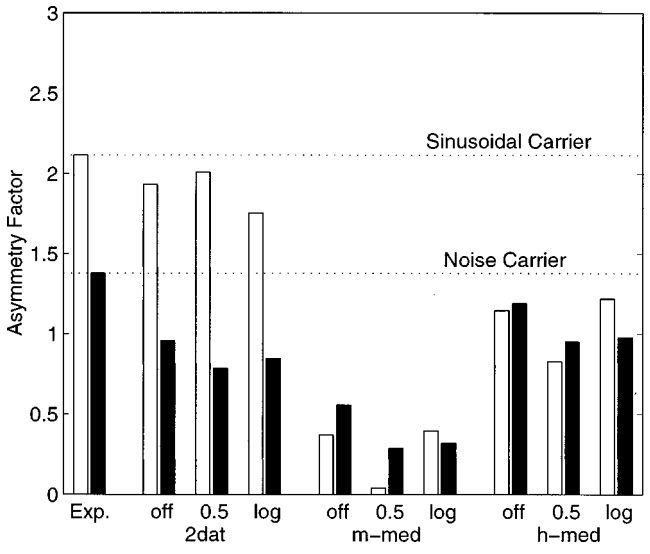

FIG. A1. Asymmetry factors calculated from the matching half-lives measured in the perceptual experiment (Exp), and from the NAPs produced by three auditory models, each with three different forms of compression, none (off), square root (0.5), and logarithmic (log). The NAPs were produced by AIM (***gtf/2dat***), and the Meddis and Hewitt model with medium-rate fibers (***gtf/m-med***) and high-rate fibers (***gtf/h-med***) fibers. The figure shows that the asymmetry factor is largely independent of compression.

2. Compression Effects in AIM and Meddis and Hewitt (1991)

The results are presented in terms of asymmetry factors in Fig. A1. The analysis was restricted to auditory models with the linear filterbank ***gtf***, since the analysis of models with the level-dependent filterbank, ***tlf***, revealed that the resultant ACGs had essentially the same temporal asymmetry. The linear filterbank has one further advantage with regard to the space of auditory models that need to be investigated; the PCR measure can be applied directly to the NAPs produced with ***gtf***, which is not the case for ***tlf***, as noted above in Sec. III B. This means that the effect of compression on asymmetry can be assessed independent of the effects of ***ac*** or ***sti***, which, in turn, reduces the number of auditory models that need to be assessed considerably.

The asymmetry factors produced with AIM (***gtf/2dat***) are presented over the ''***2dat***'' section of the abscissa. They show that the asymmetry factor is largely independent of compression. The ''log'' values are the asymmetry factors for the default version of AIM as shown in Fig. 3(c). Thus, for all three types of compression, there is sufficient asymmetry to explain the perceptual asymmetry when ***gtf/2dat*** is accompanied by ***sti*** in either the original form or the delta-gamma form. Moreover, the asymmetry has the correct form; the asymmetry for the sinusoidal carrier is greater than that for the noise carrier. This indicates that it is the asymmetric adaptation in ***2dat***, rather than compression, that dominates in the production of asymmetry at the NAP level in AIM.

The asymmetry factors for the Meddis and Hewitt model with medium-rate fibers and with high-rate fibers are presented over the ''***m-med***'' and ''***h-med***'' sections of the abscissa, respectively. In both cases, the asymmetry factor is largely independent of compression. The medium-rate model, ***m-med***, does not produce sufficient asymmetry to

explain the experimental data. Moreover, the asymmetry factors for the sinusoidal carriers are less than those for the noise carriers when there is no compression or power-function compression. The high-rate model, ***h-med***, produces more asymmetry and, if accompanied by ***sti***, it would probably be sufficient to explain the average magnitude of the perceptual asymmetry. However, there is insufficient difference between the asymmetry factors for sinusoidal and noise carriers to explain the perceptual data.

3. Summary

The results indicate that compression does not have a large effect on temporal asymmetry as measured by the asymmetry factor. With hindsight, perhaps this is not surprising. The PCR measure and the asymmetry factor are both relative measures. Compression is instantaneous and monotonic, and so when ratios are computed the effect of the compression is minimized. Nevertheless, it is useful to know that compression is not a complicating factor when modeling temporal asymmetry in the auditory system.

[1]The asymmetry factor reported by Irino and Patterson (1996) for sinusoidal carriers, 2.3, is incorrect. It should be *2.1* as stated in this paper rather than 2.3. This means that, in round numbers, the matching half-life for a damped sinusoid is about *four* times the ramped half-life rather than five times the ramped half life as reported in Irino and Patterson (1996).

[2]The package is available by ftp from ftp.mrc-cbu.cam.ac.uk. Alternately, see the web page for AIM on the internet at http://www.mrc-cbu.cam.ac.uk/aim.

[3]The strobe-criterion details are presented with demonstrations in the AIM documentation file '/docs/aimStrobeCriterion'.

Akeroyd, M. A., and Patterson, R. D. (**1995**). ''Discrimination of wideband noises modulated by a temporally asymmetric function,'' J. Acoust. Soc. Am. **98**, 2466–2474.

Allen, J. B. (**1996**). ''A review of active and passive basilar membrane cochlear mechanics,'' J. Acoust. Soc. Am. **99**, 2582.

Assman, P. F., and Summerfield, A. Q. (**1989**). ''Modelling the perception of concurrent vowels: Vowels with the same fundamental frequency,'' J. Acoust. Soc. Am. **85**, 327–338.

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

Brown, G. J., and Cooke, M. (**1994**). ''Computational auditory scene analysis,'' Comput. Speech Lang. **8**, 297–336.

Carlyon, R. P., and Datta, A. J. (**1997**). ''Masking period patterns of Schroeder-phase complexes: Effects of level, number of components, and phase of flanking components,'' J. Acoust. Soc. Am. **101**, 3648–3657.

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

Evans, E. F. (**1977**). ''Frequency selectivity at high signal levels of single units in cochlear nerve and nucleus,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 185–192.

Fay, R. R., Patterson, R. D., and Chronopoulos, M. (**1996**). ''The sound of a sinusoid: Perception and neural representations in the goldfish,'' Aud. Neurosci. **2**, 377–392.

Giguère, C., and Woodland, P. C. (**1994**). ''A computational model of the auditory periphery for speech and hearing research. I. Ascending path,'' J. Acoust. Soc. Am. **95**, 331–342.

Houtsma, A. J. M., Rossing, T. D., and Wagenaars, W. M. (**1987**). ''Auditory Demonstrations,'' CD (IPO, Acoust. Soc. Am., Eindhoven, The Netherlands).

Irino, T., and Patterson, R. D. (**1996**). ''Temporal asymmetry in the auditory system,'' J. Acoust. Soc. Am. **99**, 2316–2331.

Kohlrausch, A., and Sander, A. (**1995**). ''Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets,'' J. Acoust. Soc. Am. **97**, 1817–1829.

Licklider, J. C. R. (**1951**). ''A duplex theory of pitch perception,'' Experientia **7**, 128–133. Reprinted in E. D. Schubert (ed.), *Psychological Acoustics* (Dowden, Hutchinson and Ross, Stroudsburg, PA, 1979).

Lorenzi, C., Gallego, S., and Patterson, R. D. (**1997**). ''Discrimination of temporal asymmetry in cochlear implantees,'' J. Acoust. Soc. Am. **102**, 482–485.

Lyon, R. F. (**1982**). ''A computational model of filtering, detection, and compression in the cochlea,'' in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, Paris, France.

Lyon, R. F. (**1984**). ''Computational models of neural auditory processing,'' in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, San Diego, CA.

Meddis, R. (**1986**). ''Simulation of mechanical to neural transduction in the auditory receptor,'' J. Acoust. Soc. Am. **79**, 702–711.

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

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

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

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

Meddis, R., Hewitt, M., and Shackleton, T. (**1990**). ''Implementation details of a computational model of the inner-haircell/auditory-nerve synapse,'' J. Acoust. Soc. Am. **87**, 1813–1816.

Oxenham, A. J. (**1996**). ''Peripheral origins of the upward spread of masking,'' J. Acoust. Soc. Am. **99**, 2542.

Patterson, R. D. (**1994a**). ''The sound of a sinusoid: Spectral models,'' J. Acoust. Soc. Am. **96**, 1409–1418.

Patterson, R. D. (**1994b**). ''The sound of a sinusoid: Time-interval models,'' J. Acoust. Soc. Am. **96**, 1419–1428.

Patterson, R. D., and Holdsworth, J. (**1996**). ''A functional model of neural activity patterns and auditory images,'' in *Advances in Speech, Hearing and Language Processing*, edited by W. A. Ainsworth, Vol. 3, Part B (JAI, London), pp. 547–563.

Patterson, R. D., and Irino, T. (**1998**). ''Auditory Temporal Asymmetry and Autocorrelation,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. Palmer, A. Rees, Q. Summerfield, and R. Meddis (Whurr, London), pp. 554–562.

Patterson, R. D., Allerhand, M., and Giguère, C. (**1995**). ''Time-domain modelling 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**, 3286–3294.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand M. (**1992**). ''Complex sounds and auditory images,'' in *Auditory Physiology and Perception*, edited by Y Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 429–446.

Plack, C. J. (**1996**). ''Basilar membrane non-linearity and the growth of masking,'' J. Acoust. Soc. Am. **99**, 2543.

Slaney, M. (**1994**). Personal communication.

Slaney, M., and Lyon, R. F. (**1990**). ''A perceptual pitch detector,'' in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, Albuquerque, NM.

Strube, H. W. (**1985**). ''A computationally efficient basilar-membrane model,'' Acustica **58**, 207–214.

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., Patterson, R. D., and Sheft, S. (**1998**). ''The role of the envelope in processing of iterated rippled noise,'' J. Acoust. Soc. Am. **104**, 2349–2361.