Auditory Temporal Asymmetry and Autocorrelation

Roy D. Patterson* and Toshio Irino**

* *Center for the Neural Basis of Hearing, Physiology Department, Cambridge University, Downing Site, Cambridge CB2 3EG, U.K.*

** *ATR Human Information Processing Research Laboratories, 2-2 Hikaridai Seika-cho Soraku-gun Kyoto, 619-02, Japan*

1. Introduction

Vowels and musical notes produce complex repeating structures in the neural activity pattern (NAP) flowing from the cochlea. A number of groups have demonstrated that the pitch of these sounds can be extracted by autocorrelating the activity in the individual channels and constructing a multi-channel autocorrelogram (ACG) (e.g. Meddis and Hewitt, 1991; Slaney and Lyon, 1990). Recently, Cariani and Delgutte (1996) showed that the physiological equivalent of the ACG, a multi-channel, all-order interval histogram, is also an excellent predictor of pitch. Several authors have gone farther and argued that the autocorrelograms (ACGs) of speech and musical sounds could also explain vowel quality and musical timbre (e.g. Meddis and Hewitt, 1992). There is a problem, however; the structures produced by natural sounds in the NAP are highly asymmetric. Autocorrelation is symmetric in time and it converts *a*symmetric NAP structures into symmetric structures in the ACG. Patterson (1994b) and Akeroyd and Patterson (1995) have shown that we are highly sensitive to temporal asymmetry and they have argued that, for timbre analysis at least, autocorrelation (AC) should be replaced with a form of 'strobed' temporal integration (STI) which produces a similar representation but which preserves temporal asymmetry. Section 2 of this paper compares the asymmetry processing of AC to STI. Section 3, introduces a new form of STI that is more like what we might expect to find in the auditory system. It is based on the delta-gamma operator of Irino and Patterson (1996).

2 Autocorrelation verses Strobed Temporal Integration

In response to periodic sounds, the cochlea produces an elaborate, multi-channel, pattern of phase-locking information that repeats once per cycle of the wave. The auditory images that we hear in response to periodic sounds are static rather than oscillating, indicating that some form of temporal integration is applied to the NAP prior to our initial perception of the sound. The rich timbres of musical notes suggest that at least some of the phase-locking information in the NAP is preserved in the auditory image. The traditional model of temporal integration removes phase-locking information from the internal representation of the sound. Strobed temporal integration was introduced to produce stable auditory images while preserving the phase-locking information produced by periodic sounds (Patterson et al., 1992). It is not a difficult problem if you know the moment in time at which the pattern in the NAP will repeat. Consider the example of temporally asymmetric, 'damped' and 'ramped' sinusoids from Patterson (1994a) presented in row *a* of Figure 1 (columns 1 and 2). In both stimuli, the frequency of the sinusoidal carrier is 800 Hz, the

Patterson, R.D. and Irino, T. (1998). "Auditory Temporal Asymmetry and Autocorrelation," In: Psychophysical and physiological advances in hearing: Proceedings of the 11th International Symposium on Hearing, Eds. A. Palmer, A. Rees, Q. Summerfield and R. Meddis. Whurr, London, 554-562.

envelope is exponential, and the envelope period is 25 ms; the only difference is that the envelope of the ramped sinusoid is reversed in time. The second row shows the NAPs produced by the auditory image model (AIM) in response to these sounds in the 800-Hz channel. In essence, AIM applies gammatone auditory filtering, log compression and two-dimensional adaptation to the wave to produce its simulation of the neural response (Patterson et al., 1995).

The stabilised representation of the NAP is produced by setting up an image buffer for each channel, and at the start of each cycle, adding a copy of the NAP function to the image function, point-by-point. In the image buffer, activity does not move from right to

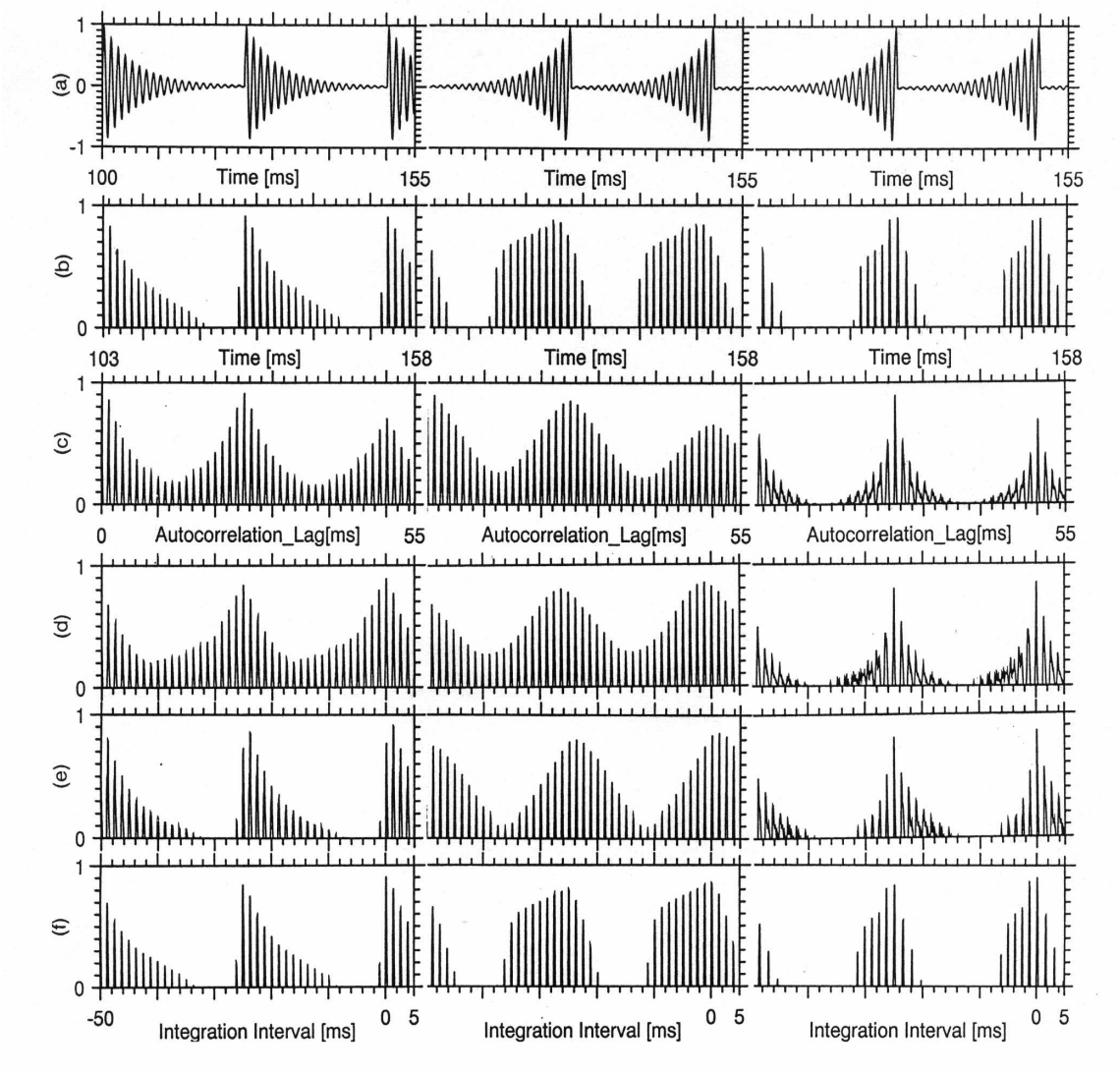

Figure 1. Autocorrelograms (c) and auditory images (d-f) of the neural activity patterns (b) produced by the cochlea in response to damped and ramped sinusoids (a). The left and centre columns show activity in the carrier channel, 800 Hz; the right column shows activity in the 640-Hz channel. Autocorrelation forces symmetry on the representation and it introduces distortions in off-frequency channels. Strobed temporal integration with a local-max criterion preserves the original asymmetry in the auditory image.

left, it simply decays into the floor over time, and provided it does so slowly with respect to the rate of cycles, periodic sounds will produce stabilised images (Patterson et al., 1992).

Patterson, R.D. and Irino, T. (1998). "Auditory Temporal Asymmetry and Autocorrelation," In: Psychophysical and physiological advances in hearing: Proceedings of the 11th International Symposium on Hearing, Eds. A. Palmer, A. Rees, Q. Summerfield and R. Meddis. Whurr, London, 554-562.

The auditory images produced by the default version of AIM are shown in the bottom row of the figure where it can be seen that temporal asymmetry is preserved in the stabilised auditory image. The corresponding ACGs are presented in row *c* ; they are symmetric and only vaguely reminiscent of the NAPs that they represent. The peak of the ramped ACG is more rounded and it has a smaller peak-to-trough ratio, but the shape information is largely lost. The preservation of the NAP pattern by STI, and our sensitivity to temporal asymmetry, are the primary bases for preferring STI over AC for timbre analysis. With regard to pitch perception, AC and STI produce largely the same results (e.g. Yost et al. 1996). The question, then, is how does STI preserve temporal asymmetry?

2.1 Strobe Criterion and Temporal Asymmetry

Auditory filters are relatively narrow and, as a result, the NAPs of tonal sounds typically have one local maximum per cycle. Thus, the problem of producing a stabilised auditory image from an oscillating NAP reduces to one of finding local maxima in the NAP. In AIM R7 (Patterson et al., 1995), the local maxima are identified by a nested sequence of strobe criteria. In this Sub-section, we show that the least restrictive criterion produces a symmetric image like AC, and that progressive restrictions designed to isolate local maxima also restore temporal asymmetry to the auditory image, step by step. With regard to the all-order interval histogram, the progressive restrictions remove higher order intervals from the histogram and preserve 'time intervals measured from the next local maximum in the NAP'. The initial criterion for strobing is simply 'strobe temporal integration on every non-zero point in the NAP' (SC=1). The next criterion restricts strobing to the point at the peak of each NAP pulse (SC=2). The former is much slower but the results of the two are quite similar (Allerhand and Patterson, 1992). Auditory images with SC=2 are shown in row *d* of Figure 1. The ramped image is very similar to the ramped ACG; the damped image is not quite as symmetric as the damped ACG; both images have reduced peak to trough ratios relative to their NAPs. Auditory images are plotted with the short intervals on the right so that asymmetric structures in the image will have the same temporal orientation in the image as they have in the NAP.

The next criterion, SC=3, restricts strobing to the larger NAP peaks with the aid of an adaptive strobe threshold which is temporally asymmetric. That is, when the strobe unit encounters a pulse, strobe threshold rises with NAP level to the peak of the NAP pulse without delay, but after the pulse peak, strobe threshold is restricted to decaying no faster than about 5% per ms, and strobing only occurs if a NAP peak exceeds this slowly decaying threshold. This is referred to as the 'temporal shadow' criterion because it eliminates strobing on small peaks in the temporal shadow of large peaks. Row *e* of Figure 1 shows that this criterion solves the asymmetry problem for the damped sinusoid, and it restores much of the peak-to-trough ratio for the ramped sinusoid, but it does not restore ramped asymmetry. The problem is that every peak along the rising section of the ramped NAP produces a strobe because each peak is larger than its predecessor in this region. The solution in this case is to introduce a 'strobe lag'; that is to delay strobing by a few milliseconds after each NAP pulse that exceeds the adaptive threshold, to determine whether another, larger, NAP pulse is about to occur. If one does, it becomes the new strobe candidate and the 'strobe lag' is reset. Strobing occurs when the lag expires without encountering a pulse larger than the candidate. This is referred to as the 'local max' criterion (SC=4) and it produces the auditory images in row *f* of Figure 1. By eliminating strobing on the rising portion of ramped NAPs, it preserves asymmetry in the auditory images of ramped

Patterson, R.D. and Irino, T. (1998). "Auditory Temporal Asymmetry and Autocorrelation," In: Psychophysical and physiological advances in hearing: Proceedings of the 11th International Symposium on Hearing, Eds. A. Palmer, A. Rees, Q. Summerfield and R. Meddis. Whurr, London, 554-562.

sounds. It also restores the original peak-to-trough ratio, with the result that both auditory images are now essentially stabilised versions of the NAP pattern.

2.2 Autocorrelation Distortion in Off-Frequency Channels

The responses produced by ramped sinusoids in channels above and below the carrier frequency reveal two more problems with AC. Responses from the 640-Hz channel are presented in the right-hand column of Figure 1. Detailed examination of the NAP (row *b*) reveals that, whereas the time intervals just prior to the peak are all 1.25 ms -- the period of the carrier, those after the peak are 1.56 ms -- the period of the centre frequency of the channel. In short, the stimulus drives the channel at the stimulus frequency on the way up the ramp and rings at the centre frequency of the channel after the offset of the ramp (Patterson, 1994a). Autocorrelation of this single-channel, dual-frequency NAP mixes the two components so that neither is properly represented in the ACG (row *c*). The time intervals around the peak in the ACG are all the same, but they are neither the carrier period nor the channel period, they are a weighted sum of the two. Moreover, there are many, irregular peaks at lower levels in the ACG, but this autocorrelation 'noise' is not heard as noise by listeners. Similar problems appear in the auditory images with the least restrictive strobe criteria; the image for SC=2 is shown in row *d*. The shadow criterion (row *e*) restores the carrier frequency to the left of the main peak; but it requires the local-max criterion (row *f*) to restore the channel frequency to the right of the peak.

In correlation terms, STI with a local-max criterion is like cross-correlation between the NAP and a function composed of delta pulses at local maxima in the NAP. In interval histogram terms, it is like a dynamic interval histogram composed of time intervals measured from moments of peak neural activity.

3 Delta-Gamma Strobed Temporal Integration

Recently we have developed a version of STI in which the nested set of strobe criteria is replaced with a mechanism that is more like what we would expect to find in the auditory system. The architecture of this 'delta-gamma' STI is presented in Fig. 2. The multi-channel NAP produced by the cochlea is represented schematically in the left-hand column of the figure. All of the modules to the right pertain to the strobe unit for one NAP channel, namely, the central channel entering the summation. Delta-gamma is defined as the derivative of the smoothed envelope of the NAP; the *envelope extractor* is represented by the column of leaky integrators, LI, and the summation sign. The envelope is the weighted average of smoothed NAPs from channels in a 3-ERB band about the central channel. The time constant, Tc-short, is 3 ms. The inclusion of the frequency dimension in the envelope calculation makes it possible to estimate the envelope with shorter time constants. The envelope is fed to the three components of the *delta-gamma strobe* shown in the central column; the delta-gamma process (bottom panel), an accumulator (middle panel) and an adaptive threshold mechanism (top panel). The delta-gamma operator

Patterson, R.D. and Irino, T. (1998). "Auditory Temporal Asymmetry and Autocorrelation," In: Psychophysical and physiological advances in hearing: Proceedings of the 11th International Symposium on Hearing, Eds. A. Palmer, A. Rees, Q. Summerfield and R. Meddis. Whurr, London, 554-562.

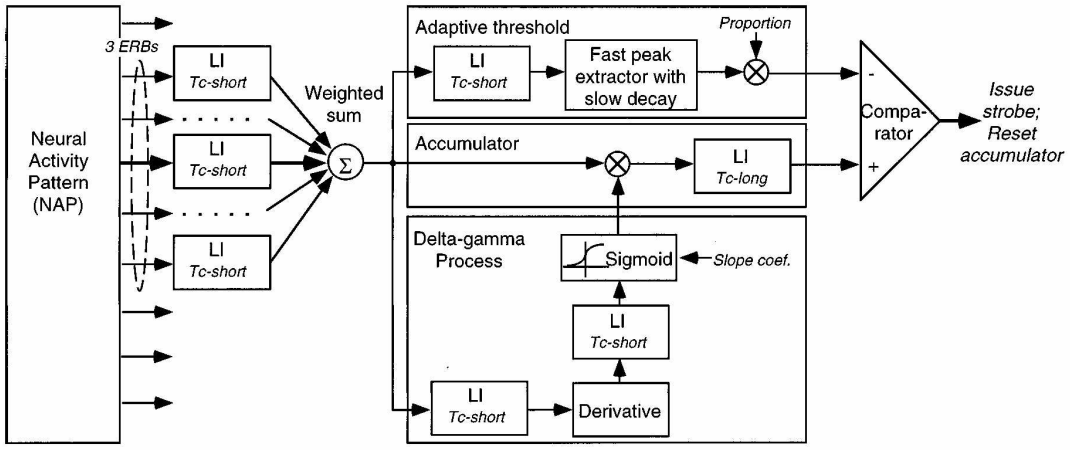

Figure 2. Architecture of the delta-gamma strobe mechanism: The envelope of the NAP (col. 1) is extracted (col. 2) and fed to the delta-gamma process (col. 3) which determines the rate at which activity accumulates in the comparator (col. 4). When the activity level exceeds the adaptive threshold level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

controls the rate at which activity from the NAP accumulates in the decision mechanism in the fourth column. When the level of activity in the accumulator exceeds the level of the adaptive threshold, a strobe pulse is issued and the accumulator is reset to zero.

In the delta-gamma process, 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 currently 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 leaky integrator with a long Tc, currently 30 ms. The output of the accumulator is compared with the level of an adaptive threshold whose purpose 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 (0.5 %/ms)**.**

The operation of delta-gamma STI is illustrated in Figure 3: The NAPs produced by AIM in response to damped and ramped sinusoids with 16 ms half lives and 50 ms envelope periods are presented in the top rows of Figures 3a and 3b. The delta gamma functions produced in response to these NAPs are shown in the second row. The adaptive thresholds and accumulation functions are shown in the third row: the slowly decaying trace is the adaptive threshold; the sawtooth function is the accumulator output. Every time the

Patterson, R.D. and Irino, T. (1998). "Auditory Temporal Asymmetry and Autocorrelation," In: Psychophysical and physiological advances in hearing: Proceedings of the 11th International Symposium on Hearing, Eds. A. Palmer, A. Rees, Q. Summerfield and R. Meddis. Whurr, London, 554-562.

accumulator value exceeds the adaptive threshold, a strobe pulse is issued, as shown in the bottom row of each figure, and then the accumulator is reset 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 sinusoid and so the adaptive threshold for the damped NAP in Figure 3a rises faster and to a higher level than that for the ramped NAP in Figure 3b. 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 and the value is more negative for the damped sinusoid because the recovery from overshoot is stronger in the damped sinusoid. As a result, the accumulation of NAP activity is relatively slow for the damped sinusoid, 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 sinusoid 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 the mechanism strobes three times during the rising portion of the cycle of the ramped sinusoid.

Irino and Patterson (1996) measured the perceptual asymmetry of the tonal and drumming components of damped and ramped sinusoids in a discrimination matching experiment and showed that the damped half life has to be about four times the ramped half life to produce a perceptual match. In AIM, this arises because of the ramped sinusoid induces more strobing and the half life of the damped sinusoid has to be increased to restore balance. Delta-gamma STI was installed in AIM (R8) and both it and AC were used to simulate the full discrimination matching experiment. The results showed that AIM with delta-gamma STI produces sufficient temporal asymmetry to explain the asymmetry in the matching half lives whereas AIM with autocorrelation does not (Irino and Patterson, 1997).

4. Conclusions

It seems unlikely that ACG models of pitch perception can be extended to explain vowel quality or musical timbre because the highly asymmetrical neural patterns produced by natural sounds are rendered symmetric in the ACG. This conclusion applies equally to the all-order interval histogram.

The ACG model of pitch will have to be modified if it is to explain why the distorted pitch values produced in off-frequency channels with dual-frequency NAPs do not affect the pitch of ramped sinusoids.

The strobed temporal integration mechanism designed to stabilise the patterns of phase-locking information produced by periodic sounds preserves and enhances temporal asymmetry provided strobing is somehow restricted to local maxima in the NAP.

The derivative of the envelope of the NAP, delta-gamma, enhances temporal asymmetry and illustrates how STI might be implemented in the auditory system.

Patterson, R.D. and Irino, T. (1998). "Auditory Temporal Asymmetry and Autocorrelation," In: Psychophysical and physiological advances in hearing: Proceedings of the 11th International Symposium on Hearing, Eds. A. Palmer, A. Rees, Q. Summerfield and R. Meddis. Whurr, London, 554-562.

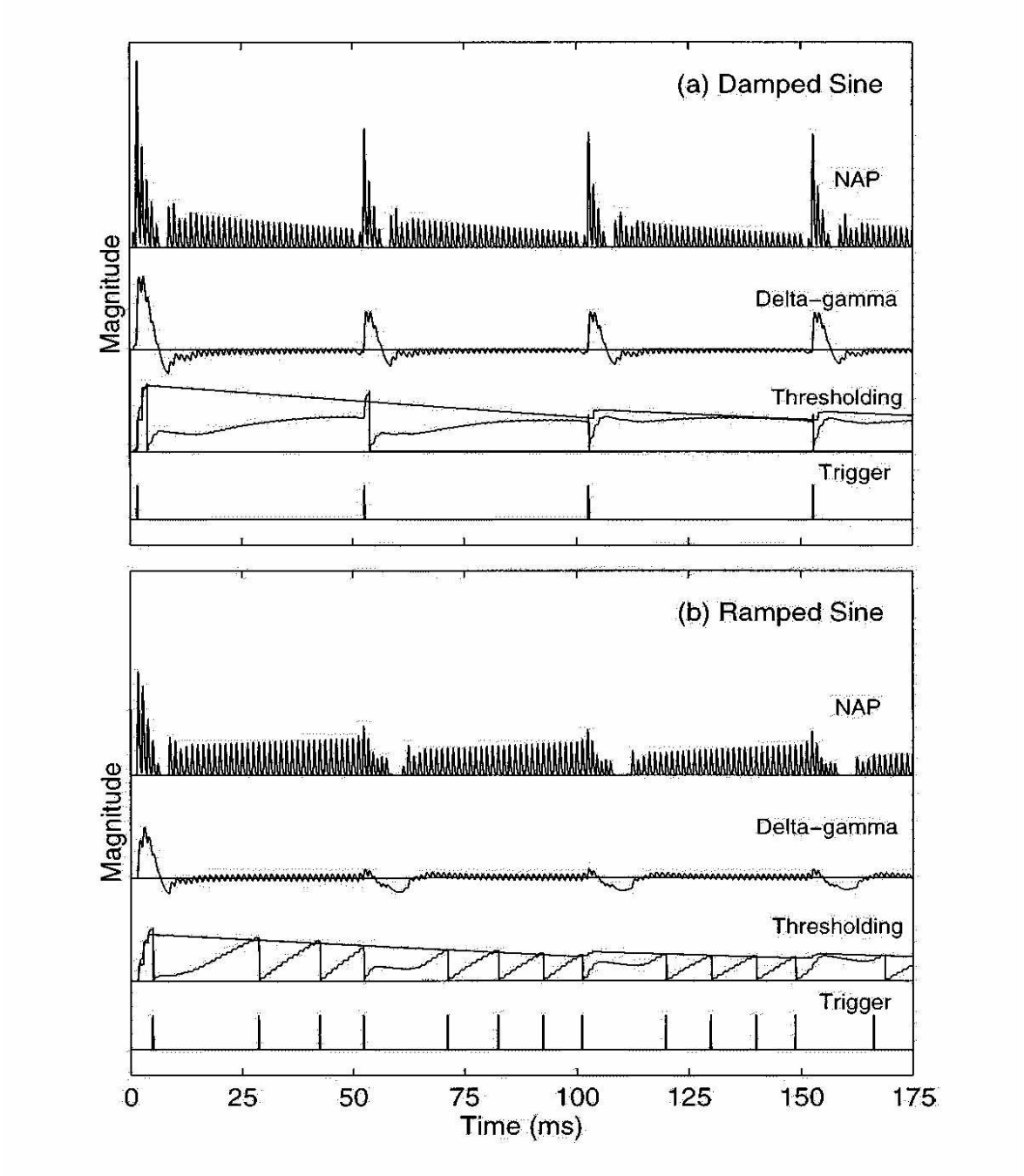

Figure 3. Response of the delta-gamma strobe mechanism to damped (a) and ramped (b) 800-Hz sinusoids in the channel centred on 1.0 kHz. Delta-gamma (row 2) is the smoothed derivative of the NAP (row 1). It controls the rate at which activity accumulates for comparison with an adaptive threshold (row 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.

Patterson, R.D. and Irino, T. (1998). "Auditory Temporal Asymmetry and Autocorrelation," In: Psychophysical and physiological advances in hearing: Proceedings of the 11th International Symposium on Hearing, Eds. A. Palmer, A. Rees, Q. Summerfield and R. Meddis. Whurr, London, 554-562.

4.1 Acknowledgements

Delta-gamma strobe as it appears in Irino and Patterson (1996) was developed while the second author was seconded to the Applied Psychology Unit, Cambridge, U.K.; the version presented here was developed while the first author was a visiting researcher, and the second author was a full-time researcher, at NTT Basic Research Laboratories, Atsugi, Japan. The authors would like to thank Dr T. Hirahara of NTT BRL for his continuing support and for making both of these collaborations possible.

4.2 References

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.

Allerhand, M. and Patterson, R. (1992). Correlograms and auditory images. In Proceedings of the Institute of Acoustics, Vol. 14, Part 6, 281-288.

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

Irino, T. and Patterson, R.D. (1996). "Temporal asymmetry in the auditory system," J. Acoust. Soc. Am. 99, 2316-2331.

Irino, T. and Patterson, R.D. (1997). "Explaining perceptual temporal asymmetry with autocorrelation versus strobed temporal integration," Kyoto meeting of the Acoustical Soc. of Japan, I, 455-456.

Meddis, R. and M. J. Hewitt (1991). “Virtual pitch and phase sensitivity of a computer model of the auditory periphery: I pitch identification,” J. Acoust. Soc. Am. 89, 2866-82.

Meddis, R. & Hewitt, M.J. (1992) "Modelling the identification of concurrent vowels with different fundamental frequencies," J. Acoust. Soc. Am. 91, 233-245.

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, Y Cazals, L. Demany, K. Horner (eds), Pergamon, Oxford, 429-446.

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., Allerhand, M., and Giguere, C., (1995). "Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. 98, 1890-1894.

Slaney, M. and Lyon, R.F. (1990). “A perceptual pitch detector,” in Proc. IEEE Int. Conf. Acoust. Speech Signal Processing, Albuquerque, New Mexico.

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.

This is post-print constructed from the text and scanned versions of the original figures to provide an electronic version of the paper.

Patterson, R.D. and Irino, T. (1998). "Auditory Temporal Asymmetry and Autocorrelation," In: Psychophysical and physiological advances in hearing: Proceedings of the 11th International Symposium on Hearing, Eds. A. Palmer, A. Rees, Q. Summerfield and R. Meddis. Whurr, London, 554-562.