3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

A THEORY OF ASYMMETRIC INTENSITY ENHANCEMENT AROUND ACOUSTIC TRANSIENTS

IRINO Toshio *and* *Roy D. PATTERSON*

NTT Basic Research Laboratories
3-1, Morinosato Wakamiya, Atsugi-shi,
Kanagawa, 243-01, JAPAN
e-mail: irino@nttlab.ntt.jp

MRC - Applied Psychology Unit
15 Chaucer Road, Cambridge,
CB2 2EF, UNITED KINGDOM
e-mail: roy.patterson@mrc-apu.cam.ac.uk

ABSTRACT

A theory of asymmetric intensity enhancement around acoustic transients is presented. Experiments with ramped and damped tones (Patterson 1994a,b) have shown that reversing an asymmetric envelope changes the timbre of the sound. A 'delta-gamma' filter with thresholding produces asymmetric intensity enhancement around transients that can explain the experimental results obtained with damped and ramped tones. If we assume that the onset of a sound is more important perceptually than what comes shortly thereafter, then this onset information can be enhanced by the delta-gamma process. Nonsimultaneous masking experiments were also performed but between subject variability precluded explanation by the delta-gamma theory.

I. INTRODUCTION

The sound quality of tones with decreasing (damped) and increasing (ramped) envelopes is quite different, although their Fourier spectra are identical (Patterson, 1994a). The damped tone sounds like a click with a weak tonal component; the ramped tone sounds like a click with a stronger continuous sinusoid. This implies that the envelope information affects the representation of the sound in the auditory system prior to our conscious awareness of the sound. The temporal aspects of sound are usually explained in terms of lowpass filtering, or leaky integration (Plomp and Bouman, 1959; see Viemeister and Plack, 1993 for a review). Although these models can explain phenomena like gap detection, they cannot explain important temporal phenomena like forward masking and the perception of damped and ramped tones. This paper describes a theory of auditory processing in which asymmetric intensity enhancement is applied around acoustic transients and a computational model of the theory is implemented to demonstrate the effect of the enhancement. The theory was tested using Patterson's (1994b) perceptual data. A new experiment on nonsimultaneous masking or "transient masking" (Duifhuis, 1973) was also performed to test the delta-gamma process.

II. COMPUTATIONAL THEORY

2.1 Detection of sound intensity change

Figure 1(a) shows the envelope of one cycle of Patterson's (1994a,b) damped and ramped tones. Intensity changes can be detected by looking at the sign of the derivative of the sound envelope. Figure 1(b) shows that the value of the derivative changes rapidly around the envelope peak, it is positive during increasing intensity and negative during decreasing intensity. The duration of the positive part of derivative is shorter in the damped sound than in the ramped sound. If, in the initial representation of the sound, activity is suppressed by thresholding when the derivative of the envelope is negative, there will be a difference in the internal representations that could explain the listener's choice of the ramped sound.

Although it is a very simple explanation, it is nevertheless interesting to consider how it might be performed because it is unlikely that the auditory system computes the differential of the sound envelope directly. First of all, the auditory system does not have a differential processor between the ear drum and the cochlea, and once in the cochlea, the envelope of the raw sound exists only as the envelopes of individual frequency components. It is also the case that the direct differential is not well defined at the rapid change point on the envelope, nor in the case of a noisy envelope (Poggio, Torre and Koch, 1985).

The cochlear vibration can be simulated by a gammatone auditory filter bank; whose envelope in time

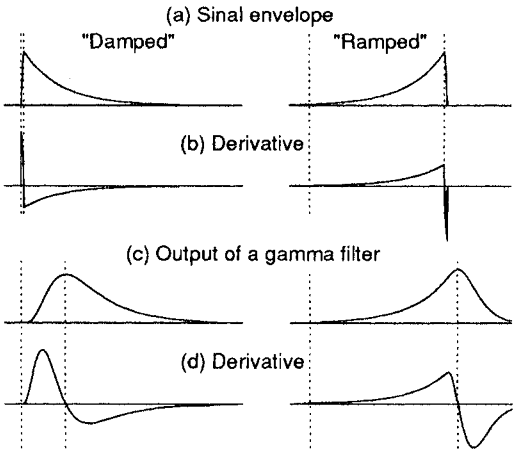


Fig. 1 The envelopes and their derivatives of damped and ramped tones

dimension is a gamma function. Figure 1(c) shows the output of one such auditory filter for the envelopes of the damped and ramped tones. Figure 1(d) shows their derivatives. The duration of the positive part of the derivative for the damped envelope is less than that for the ramped envelope. Thus, the derivative of the envelope of the filtered sound is consistent with the perceptual results, in the same way that the direct differential of the envelope was. Thus, the auditory system could determine sound intensity change using this information.

This operation is formulated and modified as follows:

$$\frac{\partial}{\partial t}\{G_m(f,t)*|s(f,t)|\} \equiv G_m(f,t)*\frac{\partial |s(f,t)|}{\partial t}$$
$$\equiv \frac{\partial G_m(f,t)}{\partial t}*|s(f,t)| \equiv \Delta G_m*|s(f,t)|$$

where $G_m(f,t)$ is the gamma function at frequency f, $|s(f,t)|$ is the sound envelope and * denotes the convolution operator. Thus, the following three operations are equivalent: 1) taking derivative of the output of a gamma filter, 2) filtering the derivative of the sound envelope with a gamma function and 3) filtering the sound envelope with the derivative of a gamma function – a 'delta-gamma' filter. Since the gamma function is smooth, the auditory system could perform a robust differential operation on the envelope of the incoming sound in this way. Increasing and decreasing the intensity of the sound correspond to the positive and negative outputs of the delta-gamma filter. Note that this formulation is similar to edge detection in vision where the second derivative of a Gaussian filter is used (Marr and Hildreth, 1980).

The description above is only for sound envelopes. Irino and Patterson (1994) have shown that the envelope operation can be performed in a cochlear filter bank even if the carrier frequency is fluctuating. The report shows that compression, adaptation and envelope estimation error do not affect the sign of the output of the delta-gamma filter, provided these processes respond relatively slowly to the envelope change.

2.2 Enhancement of activation level

If the filter output is suppressed by thresholding when the output of the delta-gamma filter is negative, then the auditory system would extract only the initial section of a sound with decreasing intensity. So, if negative values caused total suppression, sounds with slowly decreasing intensity would be inaudible. To avoid this problem, the threshold should be a "soft" function like a sigmoid, $S_g(x) = 1/(1+e^{-k(x-x_0)})$. The sigmoid function increases smoothly from zero to one as the variable goes from negative to positive; it is a reasonable function because sounds with constant or slowly decreasing intensity would be passed by the thresholding device, albeit at a reduced level. Delta-gamma filtering and thresholding are referred to as 'the delta-gamma process' in the following.

2.3 Algorithm

Figure 2 presents a diagram of the initial stage of the Auditory Image Model (AIM) (Patterson et al., 1992; also in this volume) and the delta-gamma process, for a single channel. The gammatone filter, rectifier and compressor (Tap A) are connected to a two-dimensional adaptation process (Tap B) and the delta-gamma process (Tap C). The Neural Activation Pattern (NAP) (Patterson et al., 1992) at each tap is used for calculating the measure used to evaluate each stage.

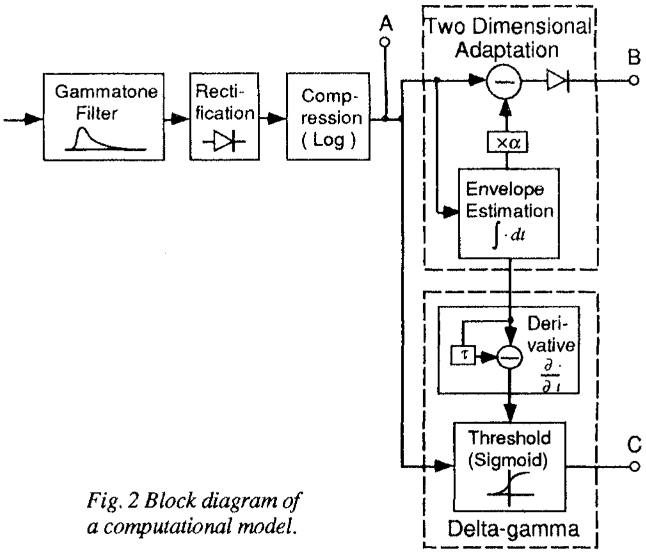


Fig. 2 Block diagram of a computational model.

III. TONALITY EXPERIMENT

The theory was tested using the experimental results obtained by Patterson (1994b) with damped and ramped tones (Fig. 3(a)). In that experiment, listeners were presented a pair of damped and ramped tones with the same half life and asked to choose the interval with the stronger sinusoidal sound quality in a two-interval two-alternative forced-choice task. The damped and ramped tones had one of five carrier frequencies (400, 800, 1600, 3200 and 4800 Hz). The duration of the envelope period was 40 times the duration of the carrier period.

3.1 Measure

Patterson (1994b) proposed a measure of sinusoidal quality to explain his experimental results. The measure is the 'carrier-period activity level' (CPAL) in his simulation of the internal representation of the sound, referred to as the auditory image. The number of carrier intervals that appear in the NAP is affected by the envelope of the NAP. If the non-zero part of the envelope is long, the number of carrier intervals is large and then the CPAL is large, and vice versa. This suggests that it might be possible to define a measure that is compatible with the CPAL measure but which works in terms of the envelope of the NAP.

The damped tone sounds like a stream of clicks with a weak tonal component; the ramped tone sounds like a stream of clicks with a stronger continuous sinusoid. It appeared that a measure of the 'click-to-background ratio' or 'peak concentration' might be compatible with listeners' judgements provided we assume that the auditory system could generate some measure of click strength relative to tone strength, and that listeners would base their judgements on the ratio of these relative click strength values when presented with a pair of damped and ramped tones.

The mean activity in the NAP was calculated by averaging across channels in the 2-octave range around the carrier frequency. We can define 'peak concentration' to be the average activity in the 4-ms segment around the peak, relative to the average activity level in one cycle of the NAP. We assume that the decision statistic is the 'peak concentration ratio'; that is, the ratio of the peak concentration for the damped tone divided by that for the ramped tone.

3.2 Results

Figure 3(a) shows the experimental results. Figures 3(b), (c) and (d) show the peak concentration ratios calculated from the mean NAPs after compression (Tap A), after two dimensional adaptation (Tap B), and after the delta-gamma process (Tap C). In Fig. 3(a), the half life of the maximum in the curve for 400 ms is 16 ms. The maximum tends to move toward lower half lives and lower ratios as the carrier frequency increases. In Fig. 3(b), the half life of the maximum for 400 Hz is between 4 ms and 8 ms and, for each carrier frequency, the maximum half life is less than that in Fig. 3(a). Thus, the pattern of predicted results is different from the experimental results at Tap A. In Fig. 3(c), the half lives of the maxima are almost the same as those in Fig. 3(a). Although the peak concentration ratios are large and different at each carrier frequency, these curves could be fitted to the experimental curves if the ramped response saturated above a certain peak concentration ratio. In Fig. 3(d), the half lives of the maxima are almost the same as those in Fig. 3(a) and the pattern of the predicted curves resembles that of the experimental curves. In summary, two dimensional adaptation and the delta-gamma process both produce the correct patterns of results, whereas the lowpass filter model does not.

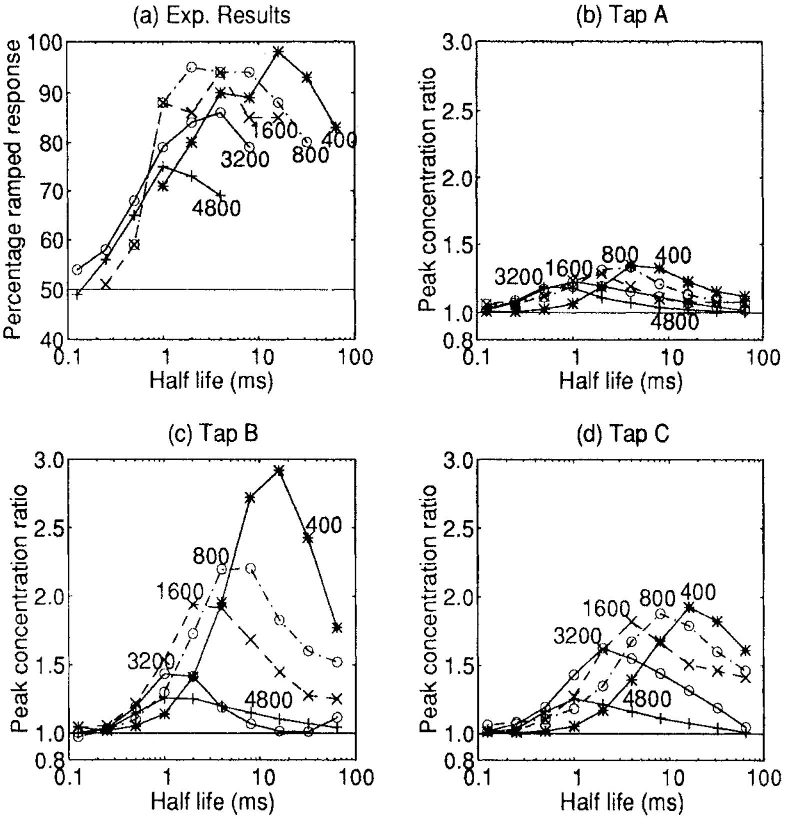


Fig. 3 (a) Experimental results (Patterson 1994b) : percentage of ramped tone selection. Peak concentration ratio for damped and ramped tones calculated from the mean NAPs at (b) Tap A, (c) Tap B and (d) Tap C in Fig. 2. The numbers in the panels give the carrier frequency in Hz.

3.3 The relationship between delta-gamma and 2D adaptation

Two dimensional adaptation (Fig. 2, upper right hand section) is based on the difference between the original activation pattern and the estimated envelope, multiplied by a fractional constant ($0<\alpha<1$). The envelope of the original activation pattern is slightly displaced from the estimated envelope because of the time lag produced by the leaky integrator. The subtraction of these envelopes produces the derivative of the envelope. Thus, the activation pattern after the subtraction is biased by the derivative of the envelope. Then the negative values are cut off by the rectifier. So, two dimensional adaptation is quite similar to the delta-gamma process which also uses the difference between the estimated envelope and the delayed envelope to control the threshold function. Thus, two dimensional adaptation is another algorithm of the delta-gamma theory.

IV. TRANSIENT MASKING EXPERIMENT

An experiment similar to the "transient masking" of Duifhuis (1973) was performed to determine threshold for 1-kHz damped and ramped sinusoids in an attempt to produce evidence of the delta-gamma process.

4.1 Method

Listeners were presented a pair of tone pips, one with and one without a probe signal, and asked to choose the interval with the 'louder' sound (a standard two-interval, two-alternative, forced-choice task). The stimuli were produced as follows: 1) The impulse response of a gammatone filter centered at 1 kHz is used to produce a damped tone. 2) Its time reversal is used as a ramped tone. 3) Both of the tone pips are filtered by a linear-phase bandpass filter with center frequency 1kHz, band width 1 ERB and duration 16 ms in order to reduce off-frequency splatter. 4) In the first condition, the filtered, damped tone pip is used as a masker and the ramped tone pip is used as a probe. In the second condition, the masker is a ramped tone pip and the signal is a damped tone pip. 5) Each sound is repeated 4 times with a period of 40 ms. 5) Fringe sounds are added before and after the masker to prevent subject's making a length judgement. 6) The time delay of the probe sound is taken to be the peak time relative to the peak of the masker sound. 7)

The intensity of the masker is 75 dB SPL at the peak.

4.2 Results

Figure 4 shows experimental and theoretical results for masking level as a function of the time delay between the peaks of the masker and probe. In Fig. 4(a), the thresholds of SH at time delays 0 and -5 ms are much less than those of the others. The threshold of TM at time delay 15 ms is about 15 dB above those of the others. In Fig. 4(b), the thresholds of SH and TM at time delay -10 ms are about 10 dB above that of IT1 and 20 dB above that of IT2. The threshold of SH at time delay 5 ms is about 10 dB above that of TM. These subject differences suggest that different subjects use different cues.

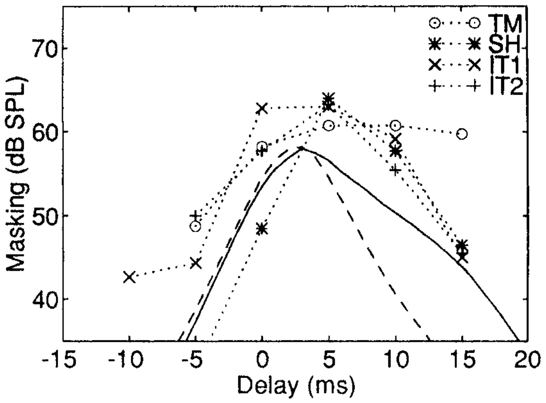
Theoretical masking thresholds can be calculated using the total activation level in the NAPs after compression (Tap A) and after the delta-gamma process (Tap C). We can calculate masking threshold from the difference between the total activation levels with and without the signal. The masking threshold can be defined as the intensity level of the signal which produces a just detectable increment in the internal activation level. The dashed and solid lines shown in Fig. 4 show the contour lines for summed NAPs for Tap A and Tap C when the difference limen (DL) was assumed to be 1.5 dB.

In Fig. 4(b), the solid line seems to fit the experimental results better than the dashed line except at time delay -10 ms where thresholds are quite high in subjects TM and SH. The solid and dashed lines in Fig. 4(a) are lower than the dotted lines showing the experimental results. If we assume that the DL for the damped masker is greater than that for the ramped masker, the solid line could be elevated by increasing DL. However, the theoretical curves cannot explain the variability between the subjects. It is difficult, then, to get definite support for the delta-gamma process in this experiment because it allows subjects to use different cues.

V. CONCLUSION

This paper has presented a theory of asymmetric intensity enhancement around acoustic transients. The asymmetric enhancement could be performed by thresholding the output of the filterbank in accordance with the output of the delta-gamma filter. The theory was shown to explain the experimental results of damped and ramped tones. Transient masking experiments did not provide clear confirmation because of large variability between subjects.

(a) Masker: "Damped", Probe: "Ramped"



(b) Masker: "Ramped", Probe: "Damped"

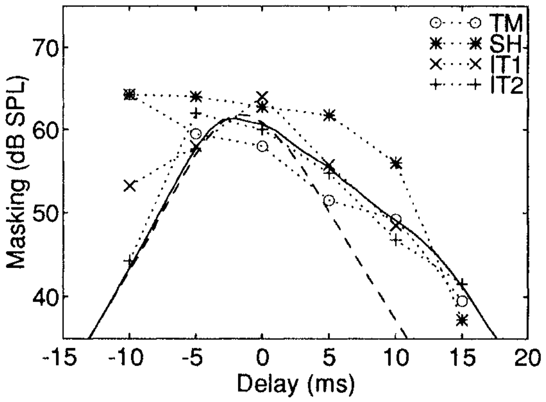


Fig. 4 Experimental results and theoretical results for transient masking. Subjects TM and SH: results of one run. Subject IT: results of two runs shown separately. The dashed lines show the theoretical results using summed NAPs after compression (Tap A in Fig.2). The solid lines show the theoretical results using summed NAPs after the delta-gamma process (Tap C in Fig.2).

REFERENCE

Irino,T. and Patterson, R.D. (1994) : "A computational theory of asymmetric intensity enhancement around acoustic transients," NTT Basic Research Laboratories, Research Report, ISRL-93-9.

Duifhuis, H. (1973): "Consequence of peripheral frequency selectivity for nonsimultaneous masking," J. Acoust. Soc. Amer., 54(6), pp.1471-1488.

Marr, D. and Hildreth, H. (1980) : "Theory of edge detection", Proc. R.Soc. Lond. B207, pp.187-217.

Patterson, R.D., Robinson, K., Holdsworth, J., McKeown,D., Zhang, C. and Allerhand, M. (1992): " Complex sounds and auditory images, " in Cazals.Y, Dermany.L, Horner, K. Eds, "AUDITORY PHYSIOLOGY AND PERCEPTION," Pergamon, Oxford.

Patterson, R.D. (1994a): " The sound of a sinusoid: spectral models," J. Acoust. Soc. Amer., (in press).

Patterson, R.D. (1994b): " The sound of a sinusoid: time-interval models," J. Acoust. Soc. Amer., (in press).

Plomp,R. and Bouman, M.A.(1959) :"Relation between hearing threshold and duration for tone pulses," J. Acoust. Soc. Amer., 31(6), pp.749-758.

Poggio,T., Torre, T. and Koch, C. (1985): "Computational vision and regularization theory," Nature, 317 (26), pp.314-319.

Viemeister, N.F. and Plack, C.J. (1993) : " Time analysis," in Yost ,W.Y., Popper,A.N. and Fay, R.R. Eds," HUMAN PSYCHOPHYSICS," Springer-Verlag, London.