ECHO SUPPRESSION IN A COMPUTATIONAL MODEL OF THE PRECEDENCE EFFECT

Keith D. Martin

Machine Listening Group
MIT Media Lab, Rm. E15-401B
20 Ames Street, Cambridge, MA 02139
kdm@media.mit.edu

ABSTRACT

Recent neurophysiological evidence suggests that the so-called precedence effect is a composite of multiple phenomena, in particular echo suppression and "active" mechanisms that build up and release suppression. We propose a simple functional model of echo suppression in a population of low-frequency ITD[1]-sensitive neurons in the inferior colliculus. Our model is based on Zurek's 1987 proposal, and we show that it is consistent with Zurek's 1980 psychophysical data by presenting the results of two experiments. The current model is extensible to other localization cues represented by rate-place codes, and we suggest that a model such as this is a necessary component of computational models of spatial hearing.

1. INTRODUCTION

The precedence effect has often been viewed as an attempt by the auditory system to suppress echoes. To aid the listener in determining the physical position of a sound source, the auditory system focuses on the direct sound wave (the first wave-front) and largely ignores the information contained in reflective (later-arriving) wave-fronts for purposes of localization. This suppression is by no means complete; indeed, reflective energy contributes strongly to timbre, perceived spatiousness, and other sound qualities [1].

Our interest in building a computational model of the precedence effect is motivated by the limited success of applying computational spatial hearing models to real-world sounds. One case in point is our previous model [2], which used interaural differences to estimate source position (both azimuth and elevation). The model achieved a high degree of localization accuracy for single anechoic sources but failed completely in the presence of reverberation. We attribute this limitation to a poor model of onset salience and the precedence effect.

1.1. The precedence effect: echo suppression

In his influential book *Vision* [3], David Marr describes three levels at which any information-processing machine must be understood: the computational theory, the representation and algorithm, and the hardware implementation. At the computational theory level, we consider the *goal* of the computation. In the case of an auditory localization system, the primary goal is to determine the positions of sound sources in an (often reverberant) environment. Part of the strategy used to achieve this goal in the mammalian auditory system is a form of onset-mediated echo suppression. The suppression is seen in psychophysical data as a temporary increase in the just noticeable difference (JND) of localization cues (e.g., interaural time differences) immediately following an abrupt rise in signal energy

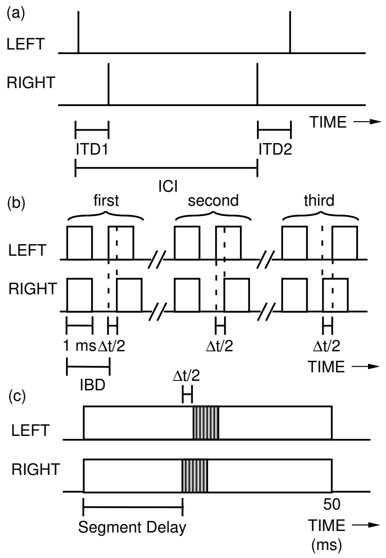

Figure 1: Three different stimulus paradigms for psychophysical investigation of the precedence effect: (a) binaural click pair (after [1]), (b) Zurek's 3AFC binaural noise burst pairs (after [4]), and (c) Zurek's "sub-burst" paradigm (after [4]). See text for details.

[4]. This suppression been shown to operate independently in different frequency bands [5], and it seems to be one of the primary components of the precedence effect, though it does not account for the "active" processes found in more recent psychophysical data (c.f., [5, 6]).

In building a computational model, we are most interested in Marr's second level, where we seek an understanding of the representation and algorithm employed to solve the information-processing problem (in this case echo suppression). From this viewpoint, it is instructive to examine auditory neurophysiology, but in doing so we must be careful to separate artifacts arising from the particular neural implementation from those arising from the underlying algorithm.

1.2. Neurophysiological correlates

Recent neurophysiological evidence reveals suppression of the neural response to the lagging click in a binaural click pair (see Figure 1a) in the auditory brain-stem. Fitzpatrick et al. [8] reported an increase in suppression from the level of the superior olivary complex (SOC; the first site of binaural interaction) to the inferior colliculus (IC; a nucleus strongly implicated in localization and a stopping point

[1] Interaural time difference, one of the primary localization cues.

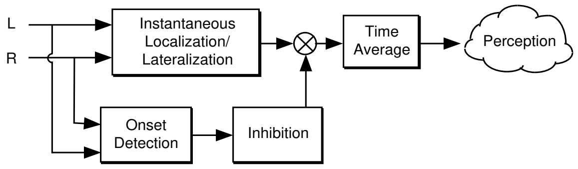

Figure 2: Zurek's proposed model to account for psychophysical precedence-effect data (after [1]).

between the SOC and the auditory cortex). In the IC, 50% second-click recovery times (measured as a ratio of click counts) ranged from 1–64 ms, with a median of 6 ms. Yin's measurements [7] ranged from 1–100 ms, with a median of 20 ms, but the increased suppression is probably due to the barbiturate anesthesia used.

These recovery times are consistent with much of the psychophysical precedence effect literature (c.f., [4]), but the measurements do not account for the "active release mechanism" or the buildup of suppression over time suggested by the results of [5] and [6] (indeed, Yin explicitly reports an absence of suppression buildup over repeated stimulus presentations). It seems likely that these effects are mediated by levels higher than the inferior colliculus, and they may be tied strongly to the processes of event formation and auditory scene analysis. In this report, we consider only echo suppression as it is seen at the level of the inferior colliculus.

It seems likely that suppression observed in the IC is mediated by inhibitory connections from both the ipsi- and contralateral SOC (either directly and/or via the lateral lemniscus). This results in a dependence of the degree of suppression on the ITD of the leading sound. The ITD-dependence varies from neuron to neuron, perhaps due to different combinations of inhibitory input [8]. The functional significance of this ITD-dependence is unknown, and since it does not vary predictably across the neural population (and it seems likely to be an artifact of the neural implementation), we have chosen not to model this aspect of the neural response.

2. THE MODEL

Zurek's proposed precedence effect model ([1], see Figure 2) also does not account for the "active" components of the precedence effect. It does, however, appear likely to serve as a good model of echo suppression. Put simply, Zurek proposes onset-mediated suppression with a time course suggested by the JND elevation data measured in psychophysical experiments.

In this section, we propose a computational model for echo suppression in a population of low-frequency ITD-sensitive neurons in the inferior colliculus. We have chosen this population because neurons of this type have been investigated in some depth with precedence effect stimuli, whereas other types of neurons (e.g., those sensitive to interaural intensity differences) have not. However, the type of model we propose is applicable to any localization cue that is represented by a rate-place code in the brain, so it is likely to be more general than it initially appears. The model is shown in schematic form in Figure 3. In the following sections, we describe its components in some detail.

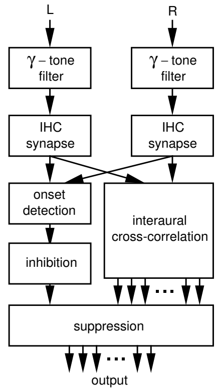

Figure 3: Block diagram depicting a model of onset-mediated suppression of an ITD-sensitive neural population.

2.1. Auditory periphery

We have used a simplified model of the auditory periphery. In the experiments described here, we are interested in ITD estimation within a single critical band, so we employ a gamma-tone filter [9] (in this paper we have used a center frequency of 500 Hz) to simulate the frequency selectivity of a small region on the basilar membrane. At this level, the processing of the signals from the two ears is kept separate, and each is passed through the Meddis-Hewitt inner hair-cell synapse (IHC) model [10] to provide the half-wave rectification and onset enhancement found there in the neurophysiology. We retain the firing probability of the IHC model output rather than a spike series, making the assumption that information from multiple hair cells is integrated at higher levels.

2.2. ITD estimation

Yin and Chan have shown [11] that ITD-sensitive neurons in the SOC perform a function very similar to a running cross-correlation on their inputs from the ipsi- and contralateral peripheries (the auditory nerve via the cochlear nuclei). Neurons in the IC exhibit similar ITD sensitivity, suggesting either a similar mechanism or a transfer of information from the SOC.

We have employed an explicit short-time cross-correlation, given by:

$$\mathrm{ITD}(t,\tau) = \int_{-\infty}^{\infty} L(t' - \frac{\tau}{2}) R(t' + \frac{\tau}{2}) w(t' - t) dt'. \quad (1)$$

where $L(t)$ and $R(t)$ are the signals arriving from the periphery. We choose $w(t)$ to be of the form:

$$w(t) = Ate^{-t/\tau}, \quad (2)$$

with $\tau = 1.5$ ms, and A set to yield unity D.C. gain.

2.3. Suppression by inhibitory input

The inhibition after a sharp onset begins approximately 1 ms after the onset, with strong suppression lasting for a few milliseconds, and gradual recovery on the time scale of approximately 10 ms [4]. To approximate this, we begin by passing the output of the IHC model through the IHC model again (to further enhance onsets), and then through a low-pass filter with a 1 ms group delay to extract

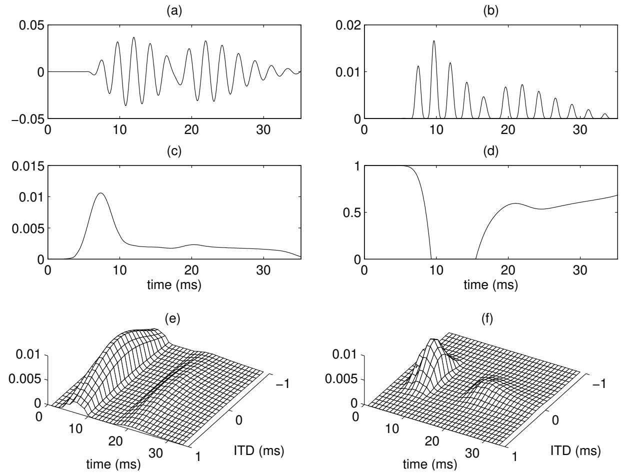

Figure 4: Output of the model at various stages for a diotic input signal (two 200 μs clicks separated by 10 ms): (a) basilar membrane filter output, (b) IHC output, (c) excitation envelope, (d) inhibition signal, (e) model output, and (f) "sharpened" output ($k = 10$).

an *excitation envelope*. The result is passed through a filter whose impulse response has the form given in Equation (2) with $\tau = 1.5$ ms, and A set to yield unity D.C. gain.

The inhibition signal is then transformed by:

$$y = 2 - e^{75x}, \tag{3}$$

and half-wave rectified. This is an ad hoc transformation that results in inhibition curves that are intuitively satisfying for a wide range of input stimuli.

Finally, the output of the interaural cross-correlation is multiplied by the inhibitory signal, as suggested by Zurek's model (Figure 2, [1]).

2.4. Model output

The output of the model as shown in Figure 4e and f is intended to represent the firing probability of a population of ITD-sensitive cells in the IC with center frequencies near 500 Hz. Figure 4e shows the unmodified output as a function of ITD and time. In the current implementation the ITD axis is sampled at one sample intervals (at a 22 kHz sampling rate, approximately 0.05 ms). To make the output more clear to human eyes, we "sharpen" the output by raising the output at each time slice to some power ($k = 5$ or 10 are typical values) and then normalize by the height of original peak, thereby maintaining the scale of the peak across the output as a function of time. Figure 4f shows the result for $k = 10$. Similar sharpening might be achieved in the neurophysiology by lateral inhibition, but we do not claim anything beyond the possibility of functional similarity here.

For most of the test signals employed in the next section, a human listener would hear a single, fused sound object, with a single source position defined by its lateral position within the head. For comparison with human psychophysical data, we average the unsharpened model output over the entire duration of the signal and report the centroid of the result, in ms, as the "perceived lateralization".

3. EXPERIMENTAL RESULTS

To test the proposed model, we simulated two of the experiments reported by Zurek [4], using binaural noise-burst pairs and the "sub-burst" paradigm. Both of Zurek's experiments employed a three-interval three-alternative forced-choice (3AFC) paradigm.

3.1. Binaural noise-burst pairs

The stimuli in the first experiment consisted of combinations of short (1 ms) noise bursts. In each stimulus interval, two binaural bursts were presented, the first of which was always diotic. The left- and right-channel signals of the second burst were identical except for a small interaural delay (labeled $\Delta t/2$ in Figure 1b). Zurek performed two separate experiments with this paradigm: one in which the second burst was identical to the first (the same random samples were used, modulo the ITD), and one in which the two bursts were independent noise samples. With our model (as with Zurek's measurements), the two experiments yielded similar results; only the results with independent noise bursts are reported here. Each of the three stimulus intervals employed identical stimuli, except that in one of the intervals, the left and right channels were swapped. The subject's task was to report which interval was different from the other two. The inter-burst delay (IBD) and the ITD of the trailing burst were varied, and the responses were used to estimate the JND of the trailing-burst ITD as a function of IBD.

By using a three-interval paradigm, Zurek avoided questions of directionality. This may be an issue at short inter-burst delays, where the two bursts might interact, leading to an "anomalous" perception (e.g., at a 1 ms inter-burst delay, the perceived direction is opposite the one expected for the particular second-burst ITD). The model exhibits a similar result (which is due to the width of the cross-correlation window), so we consider only the absolute value of the deflection without regard for its direction.

Figure 5 shows the results from one of Zurek's subjects (PZ) overlaid with results from the current model. Zurek's data points (open circles) are the estimated trailing-burst ITD resulting in 76% correct responses at each IBD. The model's results (filled squares) are estimates of the second-burst ITD needed to produce an overall deflection equivalent to a 40 μs ITD (each data point is based on a linear regression of data from 25 trials at each of 8 second-burst ITDs. The fit is somewhat awkward, but the general trends are correct. Notably, suppression exhibited by the model is not nearly as strong as that in

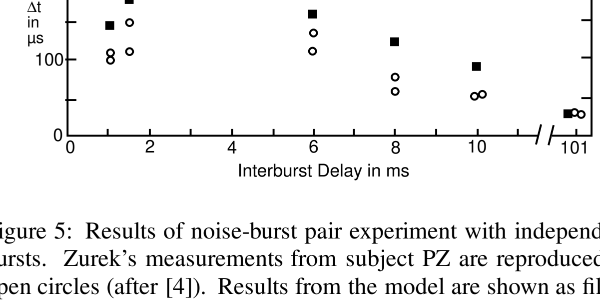

Figure 5: Results of noise-burst pair experiment with independent bursts. Zurek's measurements from subject PZ are reproduced in open circles (after [4]). Results from the model are shown as filled squares. See text for details and interpretation.

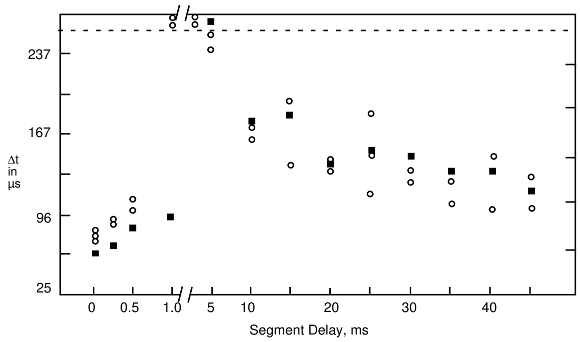

Figure 6: Results of sub-burst experiment. Zurek's measurements from subject PZ are reproduced in open circles (after [4]). Results from the model are shown as filled squares. See text for details and interpretation.

the psychophysical data in the 2-5 ms IBD range. We suspect that this is due to the ringing of the gamma-tone filter, which appears to be reduced in the physiology at levels above the auditory nerve.

3.2. Sub-burst paradigm

To test the hypothesis that the elevated JNDs were the result of suppression rather than heightened sensitivity at the onset, Zurek experimented with 50 ms diotic noise bursts in which a 5 ms sub-burst contained a small interaural difference. By varying the position of the sub-burst relative to the start of signal, he was able to map out the recovery function without introducing a second onset. Again, he employed a 3AFC paradigm, with a similar structure to that of the previous experiment, but with the stimulus in a particular interval taking the form shown in Figure 1c.

Figure 6 shows the results of the sub-burst experiment, with data from one of Zurek's subjects (PZ) overlaid with results from the model. Zurek's data points (open circles) are the estimated sub-burst ITD resulting in 76% correct responses at each sub-burst delay. The model's results (filled squares) are estimates of the sub-burst ITD needed to produce an overall deflection equivalent to an 80 μs ITD (as before, each data point arises from a linear regression of the results from 25 trials at 8 ITDs). The fit here is much closer. Since the model is stimulated throughout the stimulus duration, the ringing of the gamma-tone filters does not significantly affect the results.

4. CONCLUSIONS

We are encouraged by the experimental results reported above. It appears that this simple model of onset-mediated echo suppression confirms Zurek's expectations in that it produces brief elevation of JNDs shortly after an onset. Figures 5 and 6 show that the JND elevation seen in Zurek's psychophysical experiments is consistent with this simple form of echo suppression. The particular shape of the recovery curve is difficult to estimate from Zurek's data, and it appears to vary between subjects, so we have not attempted to make a close fit. Rather, we are satisfied that the general characteristics of the curves are similar.

Additionally, this simple model is consistent with recent neurophysiological results showing onset-mediated suppression at the level of the inferior colliculus. Our results are consistent with the interpretation that Zurek's data are a result of the suppression seen at the level of the IC. We claim that the "active" mechanisms that give rise to some precedence effect phenomena were not important factors in Zurek's data, and that echo suppression is therefore a sufficient explanatory mechanism.

This model is attractive both for it's simplicity and for its extensibility. It extends trivially to ITD detection in other low-frequency bands, and to high frequencies (above 1.5 kHz) with the intermediate step of envelope extraction after the IHC output. It may similarly be extended to any other localization cues that are represented by rate-place codes in the neurophysiology. This feature is attractive in that the time-scale of JND elevation for other localization cues (e.g., interaural level differences) is similar to that measured for ITDs [4].

5. ACKNOWLEDGMENTS

Special thanks are due to my area exam committee (Bill Peake, Pat Zurek, Ruth Litovsky, and Whitman Richards) for forcing me to take a close look at neurophysiological correlates of the precedence effect. Any misunderstandings, however, are mine alone. As always, thanks also to Dan Ellis and the Machine Listening Group for encouragement, support, and an excellent work environment.

References

1. Zurek, P.M., "The precedence effect", in Yost, W.A., and Gourevitch, G., eds., *Directional Hearing*, New York: Springer-Verlag, Chapter 4, pp. 85–105, 1987.
2. Martin, K.D., "Estimating azimuth and elevation from interaural differences", proceedings of the 1995 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, October, 1995.
3. Marr, D., *Vision*, New York: Freeman, 1982.
4. Zurek, P.M., "The precedence effect and its possible role in the avoidance of interaural ambiguities", J. Acoust. Soc. Am., Vol. 67 (3), pp. 952–964, 1980.
5. Hafter, E.R., Buell, T.N., and Richards, V.M., "Onset-coding in lateralization: Its form, site, and function", in Edelman, G.M., et al., eds., *Auditory Function: Neurobiological Bases of Hearing*, New York: Wiley, Chapter 22, pp. 647–676, 1988.
6. Clifton, R.K., "Breakdown of echo suppression in the precedence effect", J. Acoust. Soc. Am., Vol. 82 (5), pp. 1834-1835, 1987.
7. Yin, T.C.T., "Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat", J. Neurosci., Vol. 14 (9), pp. 5170–5186, 1994.
8. Fitzpatrick, D.C., Kuwada, S., Batra, R. and Trahiotis, C., "Neural responses to simple simulated echoes in the auditory brain stem of the unanesthetized rabbit", J. Neurophysiol., Vol. 74 (6), pp. 2469–2486, 1995.
9. Slaney, M., "An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank", Apple Computer Technical Report #35, 1993.
10. Meddis, R., Hewitt, M.J., and Shackleton, T.M., "Implementation details of a computation model of the inner hair-cell/auditory-nerve synapse", J. Acoust. Soc. Am., Vol. 87 (4), pp. 1813–1816, 1990.
11. Yin, T.C.T., and Chan, J.C.K., "Interaural time sensitivity in the medial superior olive of cat", J. Neurophysiol., Vol. 67 (3), pp. 952–964, 1992.