Using depressing synapses for phase locked auditory onset detection

Leslie S. Smith

Centre for Cognitive and Computational Neuroscience, Department of Computing Science and Mathematics, University of Stirling, Stirling FK9 4LA, Scotland
lss@cs.stir.ac.uk

Abstract. Auditory onsets are robust features of sounds: because the direct path from sound source to ear is the shortest path, the onset is unaffected by reverberation. Many cells in the cochlear nucleus (in the auditory brainstem) are very sensitive to onsets. We propose a neurobiologically inspired spiking onset detector which spikes in phase with the incoming signal, and which can, as a result, be used to measure relatively small inter-aural time differences, permitting sound source direction estimation.

1 Introduction

The auditory system is particularly sensitive to onsets, and this is clear right from the signal on the auditory nerve all the way to cognitive perceptual experience. Hewitt and Meddis[6] suggested a role for what amount to depressing synapses causing onset sensitivity of the auditory nerve. Very strong onset responses have been detected in the bushy and octopus cells of the cochlear nucleus (reviewed in [7]). In auditory signal processing, it has been suggested that onsets may be useful for sound segmentation[8], lip-synchronisation[10], for monaural streaming of signals [3, 4] and for providing interaural temporal differences to enable sound source direction finding which is robust against reverberation [9].

We present results from a novel spiking phase locked onset detector which has a low latency, and whose spikes occur in phase with the incoming signal. This allows it to be used to estimate interaural time differences (ITDs) of the order of 10's or 100's of μs even although actual onset durations are much larger. We present first a brief discussion of why phase locking onset detectors are useful, and of the depressing synapse which underlies this particular implementation, followed by results showing the estimation of shortest path ITDs.

1.1 Phase locked onset detection

Detecting onsets means responding when the input signal increases in intensity. Increases in intensity are characterised by their duration, rate and amount of increase. A useful onset detector must not be too sensitive (or noise will be classified as onsets), nor too insensitive (or onsets will be missed). Further, the onsets it should respond to will have a particular range of increase rate. For example, auditory onset detectors for use with speech will ignore sounds that increase too slowly. Too rapid increases will not normally be a problem in this application because of the way the sound is produced, and also because preprocessing includes bandpass filtering.

Initial onsets are of interest because they are unaltered by reverberation. For incoming sound, the shortest path is always the direct path, so that these onsets come from the direct path. This work is intended as a step towards using onsets for sound streaming [9] where (i) the sound is filtered in many bandpassed channels, (ii) initial onset times and onset based ITDs are computed in each channel and then (iii) those channels in which onsets occur within some small period of time are grouped together. Precise measurement of onset timing is crucial to this approach. Actual onset duration is upwards of 5ms, but ITDs are at maximum about 0.9ms.

To achieve this accuracy needs more than low latency onset detection (where the onset spike is emitted just after the onset). In [5] phase locked spikes are used for ITD estimation: the technique here provides for phase locking of the onset spike. Thus, even although the actual onset takes more than 10ms, the onset spike will always be emitted near a particular point in the phase of the input signal. As a result, we can achieve much better accuracy in computing onset time differences, and hence use onsets to compute interaural time differences relatively accurately. Thus we do not actually use onset time differences, but the phase difference between the signals at onset. The advantage over [5] is that the ITD is computed only for the direct path.

1.2 Modelling depressing synapses

Bertram[1] discusses two mechanisms of depressing synapse: vesicle depletion and G-protein inhibition. The mechanism we model is neurotransmitter (vesicle) depletion. The model we are using is a 3 reservoir model used in[6] in the context of inner hair cell to auditory nerve fibers, and later in[11] to model rat neocortex synapses. The model has three interconnected populations of neurotransmitter: M, the presynaptic neurotransmitter reservoir, C, the amount of neurotransmitter in the synaptic cleft, and R, the amount of neurotransmitter in the process of reuptake. These neurotransmitter reservoir levels are interconnected by first order differential equations as follows:

$$\frac{dM}{dt} = \beta R - gM \tag{1}$$

$$\frac{dC}{dt} = gM - \alpha C \tag{2}$$

$$\frac{dR}{dt} = \alpha C - \beta R \tag{3}$$

where α and β are rate constants, and g is positive during a spike, and zero otherwise. We do not model loss or manufacture of neurotransmitter. We take the amount of post-synaptic depolarisation to be directly proportional to C.

2 Experimental techniques and results

We first show that a simple depressing synapse allied with an integrate and fire neuron can act as a phase locked onset detector. The audio signal was cochlear filtered, half-wave rectified, then turned into a spike generation probability at each sampled time instant using a logistic function so that when spikes occur they do so at some point during the positive-going part of the rectified signal. This is a much simplified model of the behaviour of the auditory nerve fibres. We are not interested in spike length effects, so we set the spike length to be one sample interval. These spikes were then applied to a depressing synapse, described by equation 3. g was set to 30,000, (which, at a sampling rate of 44100 samples/second is equivalent to 680 (30,000/44.1) for a 1ms pulse), α was set to 3,000, and β to 50. Thus, the first input pulse results in a large EPSP, which rapidly decreases. The EPSP from this synapse was applied to the onset cell, an integrate and fire unit. This unit has a high dissipation (1500) (the cell is very leaky), and has a sufficiently high synaptic weight (20,000) so that a single isolated input pulse causes it to fire. The threshold is set to 1, and the unit has a refractory period of 2ms. The result is illustrated in figure 1 where the input

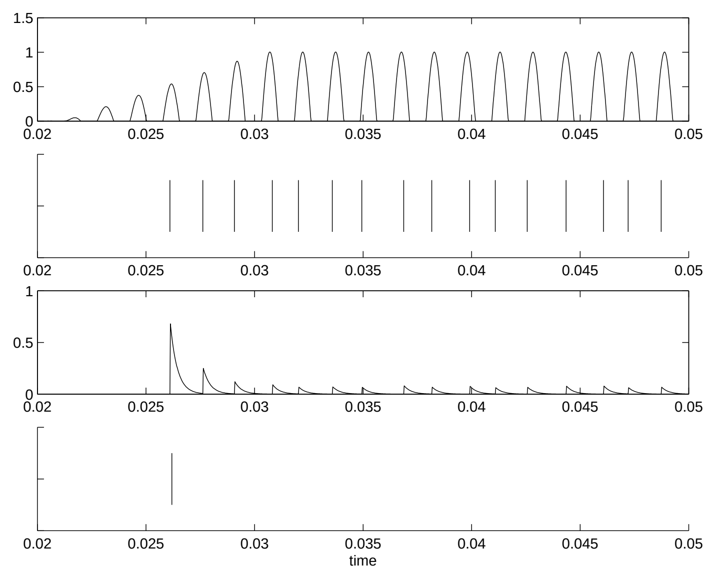

Fig. 1. Top shows 660Hz signal after cochlear filtering($F_c = 660Hz$) and rectification. Next line is the probabilistically generated spikes, followed by the depressing synapse EPSPs. Last line is the onset cell spike output. With above parameters, the onset cell fires on first input spike, but this is not crucial in this application.

signal is a simple sinusoid with a frequency of 660Hz, and a rise time of 10ms.

From figure 1 it is clear that the output spike occurs at onset and is in phase with the signal because the spike generation maintains phase, and both the depressing synapse and the integrate and fire neuron react quickly. Phase locking will be maintained up to a high frequency, as the auditory nerve signals will remain phase locked, and the there is no delay in either the depressing synapse EPSP or the onset cell itself. These last two factors are not biologically plausible.

To illustrate the use of the technique for ITD estimation, signals were generated digitally (16 bit, 44100 samples/second) using the CoolEdit 2000 package, and played using small amplifier and loudspeaker to the binaural recording system ("Gloria"). This consisted of two matched omni-directional microphones (AKG type C417), placed one in each auditory canal of a model head. This head consisted of a realistic model skull with a latex covering modelling real flesh and skin, and complete with latex rubber pinnae. The head was mounted on a simple model torso. Signals from the microphones were digitised at 44100 samples/second, and stored for later analysis. Signals were played to the binaural recording system from the loudspeaker, at the same height as the model head pinnae, at a distance of 1.5m. They were played at 10° intervals from +30° (right) to −70° (left). The signals used were short tone pulses at 220, 380, 660, 1000, 1500, 2250 and 3250Hz, with (linear) rise times of 10ms. We do not currently have equipment to measure SPL at the pinnae, but the sound level was about that of normal speech.

The stochastic nature of spike generation from the signal means that with a single onset cell for each side, the time difference between onsets for the left and right ears will consist of the estimate of the ITD (EITD) $\pm\frac{n}{f}$ where n is some integer, and f is the frequency of the signal. In addition, the stochastic nature of spike generation means that each onset can occur within one half-period, making the each EITD approximate. To improve the estimate, we model a number of auditory nerve signals and associated onset cells in each channel so that we have a number of onset spikes for each side. To achieve this, we run the simulation a number of times (N_{iter}). We then compute the mean onset firing times for each channel by averaging the near-coincident spike times (i.e. those which fall within $\frac{1}{2f}$ of each other). The ITDs are then calculated from these means. Results with $N_{\text{iter}} = 20$ are shown in figure 2.

3 Discussion and conclusions

The figure shows that the ITD can be estimated, and that best accuracy is obtained at small ITDs (low angles) using higher frequencies. Although we have not quantified it precisely, it is clear that an accuracy of better than ±5 deg can be obtained at 2250Hz and above for source angles between −10 deg and +10 deg. At higher angles, lower frequencies need to be used. The ITD error becomes larger, and accuracy decreases. We note that 0 deg does not result in an ITD of 0, and this is due to both slight asymmetry in the model head, and possibly differential delays in the audio equipment. We note that the ITD appears

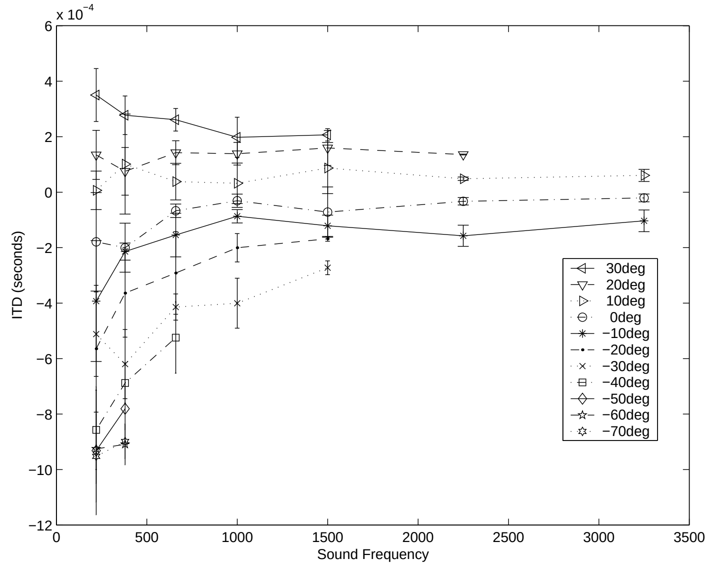

Fig. 2. Graph of ITD found for source angle between −70 deg (i.e. left) and +30 deg (right). Each line shows the ITDs found for stimuli of a particular frequency. Error bars are STDs of between 3 and 5 measurements. Results for low ITD (low angle) at lower frequencies have been omitted as they have a very high STD.

to increase as the signal frequency decreases, particularly for signals from the left, but we do not have any explanation for this.

In common with many spike-based algorithms, accuracy requires many parallel measurements: using a larger value for N_{iter} improves estimates. This onset detector system is able to overcome the problem in [8] of latency being affected by signal intensity. The time of occurrence of the onset spikes remains affected by the actual size of the onset, but because of the probabilistic nature of the system, they occur throughout the duration of the onset. Since the onset spikes are phase locked, signal intensity has little effect on the ITD estimate. This also helps when the left and right signals differ in intensity due to head shadow: ITDs can still be computed.

We have concentrated on the ITD at the onsets of these sounds because in an echoing environment, (such as the one the experiments were carried out in), the relative signal phase varies. We note that the relative phase difference between the left and right signals can vary considerably during the sound. This is due to the additive effect of multiple reflections.

The onset detector described has been made as simple as possible. It lacks dynamic range so that if a sound starts, then gets louder, this system will not detect the second onset. Where the second onset occurs soon after the first, this emulates human performance (the precedence effect [2]): however, here, the effect continues for longer. In the auditory system, transduction of the pressure wave to a neural signal is more than rectification: there is onset enhancement as well. The depressing synapse we simulate is logically situated at the onset cell: in the auditory system, there appear to be two depressing synapses in series, one at the inner hair cell/auditory nerve synapse, and (putatively) one at the auditory nerve/onset cell synapse.

The results here illustrate the technique: we note considerable improvement in human performance with broadband sounds [12], and this suggests we should use wideband sounds, thus permitting multiple concurrent ITD estimations. We note that the spikes from the onset detector code two different pieces of information, namely the signal onset and the signal phase. This suggests that single spikes can be quite rich sources of information, and this may be the source of some novel engineering solutions which emulate the neurobiological system's algorithms.

Acknowledgements

The author thanks David Sterratt, Bruce Graham and Amir Hussain for useful discussions, Frank Kelly, Val Shatwell and Caroline Wilkinson for helping construct Gloria, Nhamo Mtetwa and Frank Kelly for assistance in performing experiments, and the anonymous referees for suggesting improvements.

References

1. R. Bertram. Differential filtering of two presynaptic depression mechanisms. *Neural Computation*, 13:69–85, 2000.
2. J. Blauert. *Spatial Hearing*. MIT Press, revised edition, 1996.
3. A.S. Bregman. *Auditory scene analysis*. MIT Press, 1990.
4. M. Cooke. *Modelling Auditory Processing and Organisation*. Distinguished Dissertations in Computer Science. Cambridge University Press, 1993.
5. W. Gerstner, R. Kempter, J. Leo van Hemmen, and H. Wagner. A neuronal learning rule for sub-millisecond temporal coding. *Nature*, 383(6595):76–78, 1996.
6. M.J. Hewitt and R. Meddis. An evaluation of eight computer models of mammalian inner hair-cell function. *Journal of the Acoustical Society of America*, 90(2):904–917, 1991.
7. R. Romand and P. Avon. Anatomical and functional aspects of the cochlear nucleus. In G. Ehret and R. Romand, editors, *The Central Auditory System*. Oxford, 1997.
8. L.S. Smith. Onset-based sound segmentation. In D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, editors, *Advances in Neural Information Processing Systems 8*, pages 729–735. MIT Press, 1996.
9. L.S. Smith. Method and apparatus for processing sound. World Patent application number WO 00/01200, published 6 Jan 2000, January 2000.
10. C. Tait. *Wavelet analysis for onset detection*. PhD thesis, Department of Computing Science, University of Glasgow, 1997.
11. M.V. Tsodyks and H. Markram. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Nat Acad Sciences*, 94:719–723, 1997.
12. A. van Schaik, C. Jin, and S. Carlile. Human localisation of band-pass filtered noise. *International Journal of Neural Systems*, 5:441–446, 1999.