A Computational Model of Auditory Entrainment

Shems Saleh (shems@cs.toronto.edu)
Department of Computer Science, University of Toronto
686 Bay St, Toronto, ON, M5G 0A4, Canada

Eric Di Tomasso (ericdito@cs.toronto.edu)
Department of Computer Science, University of Toronto
10 King's College Road, Toronto, ON, M5S 3G4, Canada

Introduction

Attention can be thought of as an information processing bottleneck. Given a number of stimuli, attention restricts which of them are allocated the most cognitive resources. Attention has been categorized into 4 types: Sustained, Selective, Divided and Alternating. Selective attention corresponds to the ability to consciously focus on a subset of presented stimuli, while ignoring others. In audition, selective attention corresponds to "tuning out" sounds or words (Woldorff et al., 1993).

Auditory neurons respond based on the following properties of auditory stimuli: volume, frequency, and rhythm. Volume corresponds to the magnitude of the input current to the neuron. Roughly speaking, higher volume translates into more action potentials. Frequency corresponds to the pitch of the stimulus, and rhythm corresponds to how often the stimulus is presented.

Neurons in primary auditory cortex have been shown to play an important role in auditory selective attention (Lakatos et al., 2013). Each neuron has an associated tuning curve, and the frequency to which the neuron is most sensitive is referred to as its preferred frequency (Wang, 2013). It has been suggested that entrainment of neuronal activity is the process by which our brains filter out unattended auditory stimuli (Lakatos et al., 2013). On a neuronal level, this corresponds to oscillating levels of excitability in neurons where the phase and frequency of the oscillation depends on the cell's preferred frequency, as well as properties of the stimuli. Informally, what this means is that in the presence of an attended rhythmic stimulus, cells whose preferred frequency is close to the stimulus frequency will oscillate in phase with the stimulus, and cells whose preferred frequency is far from the stimulus frequency will oscillate out of phase with the stimulus. The alignment of neuronal oscillation with stimulus rhythm is known as phase locking. The result of this process is a spectrotemporal filtering mechanism which amplifies responses to attended stimuli and dampens responses to unattended stimuli. Figure 1 illustrates the concept of entrainment.

Previous work on modeling auditory attention can be categorized into bottom up and top down algorithms (Kaya & Elhilali, 2017). Our entrainment model can be classified as a top-down single neuron model, as we assume that we know which stimuli are attended or unattended. In similar models

auditory receptive fields were modeled as spectrotemporal filters which adapted to tasks in order to optimize the distance between attended and unattended stimuli (Mesgarani, Fritz, & Shamma, 2010; Carlin & Elhilali, 2015) . Our model performs a similar function, but adds the constraint that all adaptation occurs in the form of entrainment.
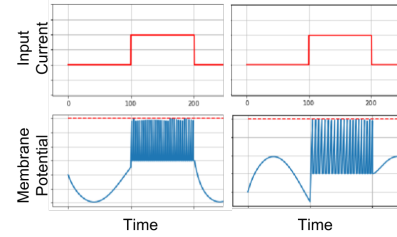

Figure 1: Illustration of auditory entrainment. The left column corresponds to the oscillation of a neuron entrained in phase with the presented stimulus. The right column corresponds to the oscillation of a neuron entrained out of phase with the presented stimulus.

Methods

Primary auditory neurons were modeled as phase-locked leaky integrate-and-fire (PL-LIF) cells. Each neuron had an associated preferred frequency and tuning curve. Preferred frequencies were evenly spaced over the interval [0.5kHz, 32kHz]. This interval corresponds to the hearing range of macaque monkeys (Lakatos et al., 2013). Tuning curves were modeled as normalized Gaussian functions centered at the preferred frequency of the corresponding neuron. All tuning curves had standard deviation 2.5kHz. To model PL-LIF cells, we modified the standard LIF equation (Gerstner & Kistler, 2002) as follows: $\frac{dv}{dt} = \frac{-(v - v_{rest}) + r*i(t) + a*sin(2\pi*f*t + b)}{\tau}$, where r is resistance and $i(t)$ is input current. Oscillation amplitude, a, was kept constant at 10mV for all experiments. Oscillation frequency, f, and phase, b, were determined by the entrainment function used in each experiment. Implementation of this neuron model was done using the Brian2 library for Python (Goodman & Brette, 2008). Auditory stimuli were modeled as step currents. Frequencies of all stimuli were in the interval [0.5kHz, 32kHz]. Currents were scaled for each cell by a factor sampled from the tuning curve of the cell at the frequency of the stimulus. Each stimulus was 100ms in length, with offset 100ms. Rhythm was set to 200ms for all stimuli.

We evaluated the following two entrainment functions:

Greedy Entrainment   Each neuron is entrained to the stimulus whose frequency was closest to the neuron's preferred frequency. The resulting oscillation of the neuron is phase locked to the rhythm of the closest stimulus and has no dependence on competing stimuli. If the closest stimulus is attended, the phase of the neuron's oscillation is equal to that of the stimulus rhythm. If the closest stimulus is not attended, the phase of the neuron's oscillation is exactly opposing the phase of the stimulus rhythm.

Attended Entrainment   Each neuron is entrained to the attended stimulus by shifting its phase by a value proportional to the difference between its preferred frequency and the attended stimulus frequency. Notably, this method assumed the existence of only a single attended stimulus, and so differs substantially from Greedy Entrainment in which an arbitrary number of stimuli can be attended. Formally, the Attended Entrainment function for a single neuron is as follows: $f(stim) = \frac{|pf - stim.freq|}{31.5kHz} * (2 * stim.rhythm)^{-1}$ where *pf* is the preferred frequency of the neuron, *stim* is the single attended stimulus, *stim.rhythm* is the rhythm of the stimulus in Hz, and 31.5kHz is a normalization factor corresponding to the size of the frequency range for stimuli.

Entrainment Function Comparison   In order to compare our entrainment methods against each other and against a baseline (no oscillation) condition, we implemented a Softmax Regression model and learned its parameters through gradient descent (Abadi et al., 2015). The input is a length 10 vector corresponding to the number of spikes a single neuron fired in response to the set of stimuli. The two outputs are the attended and unattended stimulus frequencies.

Results

Table 1 shows the number of spikes (NS), phase shift (PS) in ms, and distance (d) between a neuron's preferred frequency and stimulus frequency, in response to an attended and unattended stimulus.

Table 1: Spiking Behaviour

Entrainment f_n	Attended			Unattended		
	NS	PS	d	NS	PS	d
Attended	36	4.21	1.33	30	59.77	1.43
Greedy	62	0	0.39	42	100	1

Entrainment Function Comparison   Our predictive accuracy percentages, based on Softmax Regression, are 36%, 42% and 54% for the baseline, the attended entrainment function and the greedy entrainment function respectively.

Discussion

Greedy Entrainment   Since greedy entrainment is binary based whether the closest stimulus is attended, it works best when the stimuli are far apart and is weak when attended and unattended stimuli are close to one another.

Attended Entrainment   This function works best in the case that all stimuli are presented at the same time and with the same rhythm. Its main limitation is the restriction to a single attended stimulus.

Applications   Though the inspiration for this research has a neuroscientific basis, the implementation need not be restrained by physiological limitations. Abstractions that simplify the model, such as estimating tuning curves with gaussians, increase the ease of implementation and experimentation, while preserving the underlying filtering mechanism. As it has been shown that human speech has conserved frequency content and rhythmic properties (Gross et al., 2013), entrainment models could be used to improve speech recognition algorithms through more complex entrainment functions. In addition, models of this type may contribute to future noise-filtering methods (Kaya & Elhilali, 2017).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., . . . Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems.* Retrieved from http://tensorflow.org/ (Software available from tensorflow.org)

Carlin, M. A., & Elhilali, M. (2015). Modeling attention-driven plasticity in auditory cortical receptive fields. *Frontiers in computational neuroscience*, *9*.

Gerstner, W., & Kistler, W. M. (2002). *Spiking neuron models: Single neurons, populations, plasticity.* Cambridge university press.

Goodman, D., & Brette, R. (2008). Brian: a simulator for spiking neural networks in python.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol*, *11*(12), e1001752.

Kaya, E. M., & Elhilali, M. (2017). Modelling auditory attention. *Phil. Trans. R. Soc. B*, *372*(1714), 20160101.

Lakatos, P., Musacchia, G., OConnel, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron*, *77*(4), 750–761.

Mesgarani, N., Fritz, J., & Shamma, S. (2010). A computational model of rapid task-related plasticity of auditory cortical receptive fields. *Journal of computational neuroscience*, *28*(1), 19–27.

Wang, X. (2013). The harmonic organization of auditory cortex. *Frontiers in systems neuroscience*, *7*, 114.

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., & Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences*, *90*(18), 8722–8726.