A DUPLEX THEORY OF SPIKE CODING IN THE EARLY STAGES OF THE AUDITORY SYSTEM

Ismail Uysal, Harsha Sathyendra, John G. Harris

Computational NeuroEngineering Laboratory
University of Florida
Gainesville, FL
{ ismail, hsathyen, harris }@cnel.ufl.edu

ABSTRACT

We propose a duplex theory of spike coding in the early stages of the auditory system based on the intensity and noise levels of the acoustic stimuli. According to this concept, at low intensity levels, where auditory nerve firings cannot generate a high enough synchrony among neuron ensembles, rate coding is more likely favored against phase-locking via synchrony coding. To the contrary, at conversational intensity levels, phase synchrony coding is preferred due to its superior and highly noise robust performance. The theory is supported by both evidence from biology, as well as from experimental simulations using biologically plausible models of the entire processing chain from spike generation to recognition.

Index Terms— Spike coding, phase synchrony, speech perception, audition, psychoacoustics

1. INTRODUCTION

One of the ultimate goals of this research is to develop possible process descriptions to explain simple recognition mechanisms in the less complex levels of the auditory system. In order to accomplish this goal, one must decide on many different system parameters regarding the tools and the architecture. Our previous work introduced a biologically plausible algorithm for vowel recognition exclusively utilizing spikes in both the feature extraction and recognition stages [1]. The hypothesis suggested that one of many reasons why humans are exceptionally robust to noise for acoustic recognition tasks, lies within the redundancy of action potentials from many nerve fibers in the form of phase synchrony. We compared this biologically plausible algorithm to a typical speech recognizer with well-known Mel Frequency Cepstrum Coefficients (MFCC) [2]. The results supported this belief by demonstrating the superior and highly noise robust performance of synchrony coding combined with a spike-based rank order classifier.

This paper explores the information potential and robustness of different spike coding schemes, along with phase synchrony coding, for the early stages of the auditory pathway.

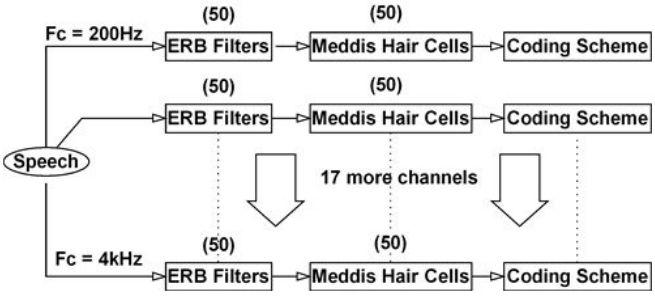

Fig. 1. The front-end, speech-to-spike conversion block

Recent research has introduced different spike based classifiers such as liquid state machines (LSMs), temporal integrators with transient synchrony and rank order coding (ROC) [3], [4], [5]. Different from the ROC classifier of the initial algorithm, an LSM with supervised learning is used for better generalization and to see the degree of correlation between the performances of coding schemes and particular system architectures.

2. SYSTEM ARCHITECTURE

The front-end of the algorithm, shown in Figure 1, converts acoustic stimuli to spike trains in the auditory nerve fiber. First, the speech is filtered through a series of gamma-tone equivalent rectangular bandwidth filters and the spiking probabilities are obtained by passing the filter bank output to the Meddis Hair Cell Model (MHCM) [6], [7]. This is followed by the feature extraction where different spike coding schemes are employed to obtain the temporal and spectral features of the generated spike trains. Finally, these features are passed on to the spike-based classifier.

2.1. The Meddis Hair Cell Model

The MHCM physiologically formulates the transduction of acoustic stimuli to neural signals observed in the auditory nerve fiber. According to the model, the sound pressure waves

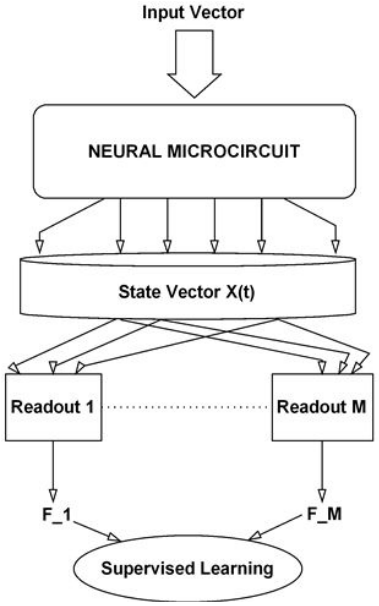

Fig. 2. The basic flow diagram for a generic LSM with supervised learning algorithm

are first converted into mechanical motion at the basilar membrane to which hair cells are attached. Hair cells are deflected by the basilar membrane motion which results in changes in the permeability of the cell membranes. This is followed by the release of neurotransmitters into the synaptic cleft which starts the typical process of action potential generation. The model associates the amount of neurotransmitters in the synaptic cleft to the probability of spike generation with a function derived from real experimental data.

The particular model used in this paper also accounts for the non-linear responses typically observed in the auditory system such as adaptation and the temporal properties of the human auditory pathway [7].

2.2. Liquid State Machine with Supervised Learning

The concept of using the inherent transients of high dimensional dynamic structures to perform computational tasks, was first introduced independently by Maass and Jaeger in the form of liquid state machines (LSM) and echo state networks respectively (ESN) [3], [8]. Where LSM uses randomly connected spiking neural circuits to carry the input to a higher dimension, ESN makes use of the multi-layer neural network architecture with much larger hidden layers and denser recurrent connections. This paper chooses to implement LSM for its spike-based classification, however please note that one can also use ESN in a similar way by converting spike train inputs to continuous signals via low-pass filtering or exponential kernels.

For the thorough analysis and formulation of LSM dynamics the reader is referred to the original paper by Maass [3]. Figure 2 provides the flow diagram for a generic LSM with supervised learning block. The input vector, *u(t)* , which might be a train of spikes, a rate code, or a degree of synchrony map for the application in this paper, is passed onto the neural microcircuit with random recurrent connections via randomly distributed dynamic spiking synapses. The basic idea relies on the assumption that at any given time *t*, the state of the liquid (neural microcircuit) holds all the necessary information about the current and past inputs. This state information is mapped by the memoryless readout functions to desired outputs with a supervised learning algorithm. More formally,

$$\vec{X}(t) = NMC(\vec{u}(t))$$

where NMC is the neural microcircuit or the liquid filter which operates on the input vector. The state vector is then mapped to a desired output vector via memoryless readout functions,

$$\vec{y}_{desired}(t) = F_{1:M}(\vec{X}(t))$$

which are trained using a supervised learning algorithm. This paper implements a backward-propagation algorithm to train the readout functions. Section 4 discusses the implementation of LSM in the overall model in more detail.

3. SPIKE CODING SCHEMES

Three common spike coding schemes are investigated: rate coding, explicit time coding and phase synchrony coding. The goal is to determine the information potential and noise robustness of these schemes for the early stages of the auditory pathway, given a simple acoustic classification task.

3.1. Rate Coding

A rate code basically implies that the frequency of spike occurrence is a means to carry information. This has been proven to some extent via physical experiments [9]. In order to find the rate, the spike count can be averaged over time (rate as spike count), several experimental trials (rate as spike density) or populations of neurons (rate as population activity).

Experimental evidence shows that rate coding is not reliable at normal conversational sound pressure levels (SPL) (60dB) as most nerve fibers are saturated (firing as fast as they can), which eliminates the possibility to differentiate between two different acoustic stimuli [10]. This paper raises the question whether it is still a viable technique for low intensity acoustic inputs with an SPL around 10dB.

3.2. Explicit Time Coding

Explicit time coding is a member of the temporal coding class which include many schemes ranging from time-to-first-spike

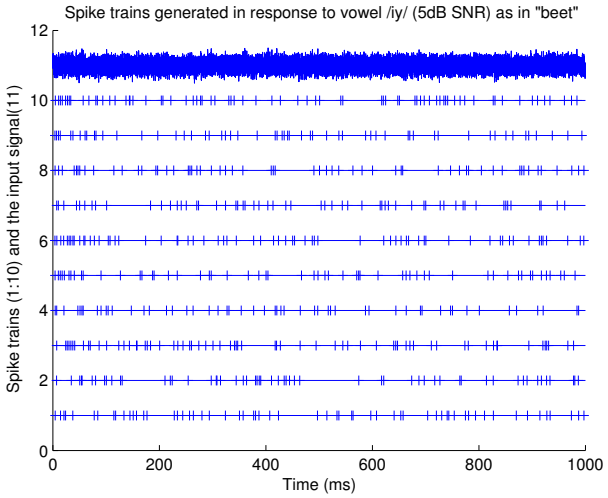

Fig. 3. Spike train outputs for a set of 10 hair cells

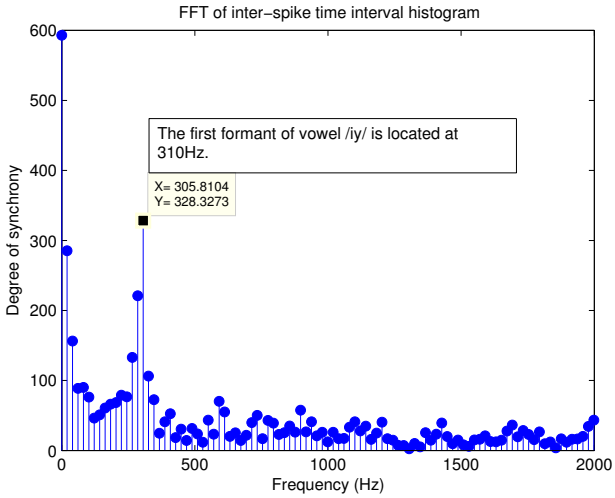

Fig. 4. Magnitude spectrum of the inter-spike time interval histogram

coding to phase coding. Explicit time coding implies the direct use of spike firing times without enforcing any kind of coding scheme. This corresponds to using all of the timing information by inputting the spike train outputs of all the 20 channels of the cochlea to the neural microcircuit via randomly distributed synapses.

3.3. Synchrony Coding

In most basic terms, synchrony coding groups neurons with similar firing times. This type of coding has been studied extensively in literature and it is common consensus that synchrony plays a significant role in the group communications of neurons [11].

To explain synchrony coding, let us start with a simple example where the vowel /iy/ as in "beet" is presented as the acoustic input stimuli to a set of hair cells with characteristic frequencies around 300Hz. Figure 3 shows the output spike trains for such a set of 10 neurons. Even though there are many different definitions of synchrony, the one used in this paper is the "*spectrum of inter-spike time interval histogram*". We further define the magnitude of this spectrum as the "*degree of synchrony*" which is shown in Figure 4. As the figure indicates, even with such a noisy input signal (5dB SNR) the neurons are still able to phase lock to the first formant frequency of the input vowel at 310Hz.

4. TEST SETTINGS AND RESULTS

To measure the robustness and information potential of each coding scheme a simulation experiment is performed to try and identify the vowel class of an utterance from 5 possible classes /iy/, /ae/, /aa/, /ao/ and /uh/ as in "beet", "bat", "hot", "bought", "foot". The results are also compared to a baseline classifier which employs MFCC features and a nearest neighbor classifier (***MFCC /w 1NN***) to perform the same task. Training (200) and testing (200) vowels for each class are chosen from the commonly used TIMIT database such that they are multi-speaker and multi-gender.

For the overall system model, the 20 output channels of the MHCM are supplied to the LSM via random dynamic synapses as vectors of firing rate (***RC as in rate code, analog***), degree of synchrony (***DoS, analog***) and exact spike timings (***EST, digital***) respectively. The analog inputs are connected to the neural circuit via analog synapses modeling the membrane potential, whereas the digital input for EST is connected via spiking synapses. The neural microcircuit is chosen to be a randomly connected neural circuit with 150 excitatory (80%) and inhibitory (20%) leaky integrate-and-fire neurons. The state of the neural circuit is found by low-pass filtering the spiking outputs of all 150 neurons (F_c=200Hz). For 5 vowel classes, the readout function was chosen to be a single hidden layer feed-forward neural net, trained with backward propagation algorithm to output an analog value between zero and five depending on the input class. Finally, the results also include the performance of phase synchrony coding when used with a rank order decoder (***DoS /w ROC***) [1].

The algorithm is tested with both pink and white noise for 3 noise levels ranging from 25dB to 5dB SNR. Finally the tests are performed at both 10dB (low intensity) and 60dB (typical conversation) sound intensity levels.

The results are provided in tables 1 and 2 for pink noise (white noise performances are very similar and thus omitted). Table 1 shows the results for a low input intensity value (10dB SPL). Even though all three schemes perform

Table 1. Percentage of vowels correctly classified at 10dB SPL

Scheme \ SNR in dB	25	10	5
RC	***77.9%***	***74.2%***	***63.0%***
DoS	76.2%	71.6%	58.4%
EST	77.8%	72.0%	59.8%

Table 2. Percentage of vowels correctly classified at 60dB SPL

Scheme \ SNR in dB	25	10	5
RC	36.2%	35.4%	35.2%
EST	80.2%	72.5%	64.9%
DoS	79.8%	77.0%	***76.6%***
DoS /w ROC	80.6%	80.2%	***79.9%***
MFCC /w 1NN	80.3%	71.1%	***69.8%***

similarly at high SNR levels, the performances of EST and DoS degrade faster than RC with increasing noise. There are two explanations for this phenomenon. First of all, both EST and DoS require as many spikes generated by the real acoustic input as possible to code information effectively which is not the case for 10dB SPL. Moreover, at low intensity values, phase synchrony is easier to occur at low frequency channels rather than at high frequency channels. Hence, when noise is introduced to some of the low frequency channels where there is no real spectral content from the actual acoustic input, the randomly generated synchrony might negate the contribution of real synchrony happening at higher frequencies. This is not true for rate coding, because at 10dB SPL, the firing rates for every channel change almost linearly with changes in acoustic input intensity. Moreover, inherent averaging over time also contributes to its higher performance under noise.

Nevertheless, as indicated in Table 2, for typical conversational intensity levels around 60dB SPL, DoS significantly outperforms other coding schemes as well as the baseline classifier, especially at low SNR values. The close-to-chance performance of RC is expected as most nerve fibers are saturated at 60dB SPL. A system highly depending on the exact spike timings, EST, is likely to fail as well at low SNR values. As the last three rows signify, **regardless of the system architecture**, DoS drops only slightly (2%-3%) in performance with a change of 20dB in the SNR value whereas the performance of the baseline classifier drops by 11%.

Future goals include a computational complexity vs. performance comparison for engineering purposes. However, preliminary analysis for the proposed DoS feature extraction technique indicates roughly the same number of computations when compared to MFCC extraction: same number of filterbanks, synchrony detection via histogram FFT vs. decorrelation via discrete cosine transform, etc.

5. CONCLUSIONS

The information potentials of different spike coding schemes are investigated for early levels of auditory processing. For high intensity signals synchrony coding displays a superior and noise robust performance when compared to other coding schemes. This is expected, regarding the fact that most nerve fibers are saturated during normal conversational levels. For low intensity signals, rate coding starts to regain its information potential and manages to outperform synchrony coding, however by a small margin. This suggests a duplex theory of spike coding, correlated with the intensity and noise levels of the input stimuli, and with more emphasis on phase synchrony because of its robustness and high performance at typical conversational intensity levels.

6. REFERENCES

[1] I. Uysal, H. Sathyendra, and J. G. Harris, "A biologically plausible system approach for noise robust vowel recognition," in *IEEE Proc. of the Midwest Symposium on Circuits and Systems*. 2006, CDROM Proceedings.

[2] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 28, pp. 357–366, 1980.

[3] W. Maass, T. Natschlager, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," *Neural Computation*, vol. 14, no. 11, pp. 2531–2560, 2002.

[4] J. J. Hopfield and C. D. Brody, "What is a moment? transient synchrony as a collective mechanism for spatiotemporal integration," *Proc. Natl. Acad. Sci. USA*, vol. 98, no. 3, pp. 1282–1287, 2001.

[5] S. J. Thorpe and J. Gautrais, "Rank order coding," in *Computational Neuroscience: Trends in Research*, J. Bower, Ed., pp. 113–119. New York: Plenum Press, 1998.

[6] R. Meddis, "Simulation of mechanical to neural transduction in the auditory receptor," *J. Acoust. Soc. Am.*, vol. 79, pp. 702–711, 1986.

[7] C. J. Sumner, E. A. Lopez-Poveda, L. P. O'Mard, and R. Meddis, "Adaptation in a revised inner-hair cell model," *J. Acoust. Soc. Am.*, vol. 113, no. 2, pp. 893–901, 2003.

[8] H. Jaeger, "The "echo state" approach to analysing and training recurrent neural networks," Tech. Rep. GMD Report 148, German National Research Center for Information Technology, 2001.

[9] F. Rieke, D. Warland, R. de Ruyter can Steveninck, and W. Bialek, *Spikes - Exploring the Neural Code*, MIT Press, Cambridge, MA, 1996.

[10] M. B. Sachs, "Neural coding of complex sounds: speech," *Annual Review of Physiology.*, vol. 46, pp. 261–273, 1984.

[11] D. Terman and D. Wang, "Global competition and local cooperation in a network of neural oscillators," *Physica D.*, vol. 81, pp. 148–176, 1995.