Neurogranular Synthesis: Granular Synthesis Controlled by a Pulse-coupled Network of Spiking Neurons

Kevin McCracken, John Matthias, Eduardo Miranda

Faculty of Arts, University of Plymouth, UK
{kevin.mccracken, john.matthias, eduardo.miranda}@plymouth.ac.uk

Abstract. We introduce a method of generating grain parameters of a granular synthesiser in real-time by using a network of artificial spiking neurons, the behaviour of which is determined by user-control of a small number of network parameters; 'Neurogranular synthesis'. The artificial network can exhibit a wide variety of behaviour from loosely correlated to highly synchronised, which can produce interesting sonic results, particularly with regard to rhythmic textures.

Keywords: Spiking neurons, granular synthesis, interactive musical control systems.

1 Introduction

A recent development in the field of autonomous interactive musical control systems [1], which has received a great deal of attention [2] [3] [4], utilises the adaptive nature of an artificial recurrent network of nonlinear spiking neurons [5]. Several methodologies have been developed in which sonic events are triggered by neuronal firing in a network of nonlinear Integrate-and-Fire (IF) neurons which have been applied in various artistic contexts [6][7]. These firing events are attractive from a musical point of view for several reasons; the collective temporal dynamics of firing occur within a natural temporal scale for musical phrases, they have the potential to be controlled by external stimulation by a musician and can adapt and change according to relationships between external stimulation, internal 'noisy' currents and plasticity within the network of neurons.

In this paper, we focus on the neuronal control of a granular synthesis engine, in which grains of synthesised sound are triggered by neuronal firing events in a simple spiking network model [8]. We will introduce granular synthesis, develop a rationale for the control of a granular synthesis engine using artificial neural networks and introduce a prototypical model, which we intend to develop in further work.

2 Granular Synthesis

Granular synthesis is a method for generating sound using a series of audio 'grains', typically of a few tens of milliseconds in duration [9]. Each grain is usually an envelope-modulated waveform (as shown in fig. 1) and the grain duration, envelope function, amplitude, pan, waveform, frequency, etc. must be specified as control data. The grains may be temporally isolated or may overlap. The grain envelope removes clicks and glitches caused by discontinuities in the audio waveforms produced at the grain boundaries. The form of this grain envelope has a considerable effect upon the character of the resultant sound comprised of sequences or clouds of multiple grains.

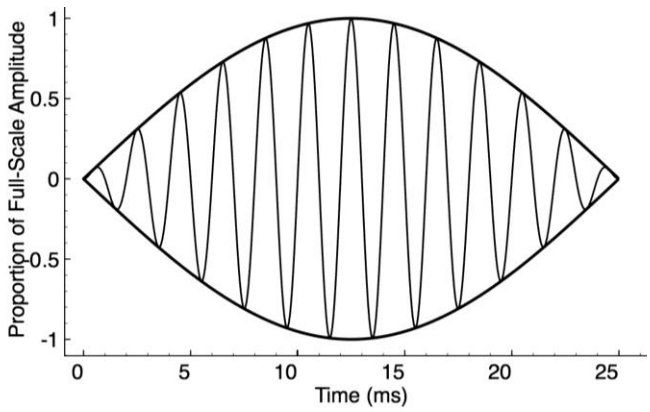

Fig. 1. A 'grain' is an envelope-modulated tone or 'chirp', typically a few tens of milliseconds in duration.

All of the parameters mentioned above constitute control data, which influence the perceived sound. Since each grain can have a dozen or more parameters associated with it and since grain durations can be as little as 10ms, the necessary control data rate to realise granular synthesis is quite large. It is not possible for a human operator to produce this amount of control data in real-time, so it must be generated, either by some sort of deterministic algorithm (set of instructions/equations) or a non-deterministic stochastic process.

Most current granular synthesis employs stochastic processes to generate grain control data [9]. However, granular synthesis with stochastically generated grain parameters will necessarily tend to produce 'noisy' synthesis and greatly limit the controllability of any system generating control information in this manner. A grain control generation system producing output data, which is, to some degree, correlated with previous data, is likely to produce more interesting and flexible results, particularly in terms of temporal information. To move forward with granular synthesis, novel methods of generating real-time grain control data are likely to prove fruitful.

3 Research Background

One of the main driving forces behind this work is that the level of synchronisation in nonlinear distributed systems is controlled by the strength of individual interactions. A desirable temporal output for granular synthesis in music would be a controllable signal, which lies at the border between randomness and synchrony. Two-dimensional reduced spiking neuronal models (reduced from the four-dimensional Hodgkin-Huxley model [10]) have been shown recently to self-organise at this boundary [11]. Recent artistic projects, The Fragmented Orchestra [7] and the Neurogranular Sampler [5][7] focused on the two-dimensional Izhikevich spiking network model [8] and trigger sound samples upon neuronal firing. In this work, we introduce a Neurogranular Synthesiser, which triggers the synthesised grains upon neuronal firing within the simple Izhikevich network.

3.1 Why Izhikevich Spiking Neurons?

Any one of a number of types of artificial neural network are possible candidates for real-time control of a granular synthesiser, such as McCulloch-Pitts, Hopfield or back-propagation networks, self-organising maps, Multi-Layer Perceptrons (MLP), Radial Basic Function (RBF) networks and, indeed, other spiking neuron models, such as IF neurons. The Izhikevich neuron was chosen for the following reasons;

1) behavioural richness; real neurons exhibit a diverse range of firing behaviour and the Izhikevich model's biological plausibility provides a similarly rich behavioural palette to be mapped to grain parameters.
2) dynamical richness; Spiking Neural Networks (SNNs) in particular manifest a rich dynamical behaviour. Transmission delays between neurons can give rise to separate functional groups of neurons (polychronisation) with the potential to be harnessed when mapping to grain parameters.
3) spiking; certain useful behavioural features require spiking behaviour, e.g. Spike Timing-Dependent Plasticity (STDP), discussed below.
4) computational efficiency; the Izhikevich model was considered an excellent trade-off between accuracy of biological triggering behaviour simulation and the amount of computer processing to implement it [12].
5) We know from our own 'experience' that networks of real spiking neurons can produce a rich and adaptive behaviour.

3.2 The Izhikevich Neuron Model

The Izhikevich neuron model is defined by the following pair of first-order differential equations [8];

$$\frac{dv}{dt} = 0.04v^2 + 5v + 140 - u + i \,. \tag{1}$$

$$\frac{du}{dt} = a(bv - u). \tag{2}$$

and the reset condition when v ≥ +30mV;

$$\begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases}. \tag{3}$$

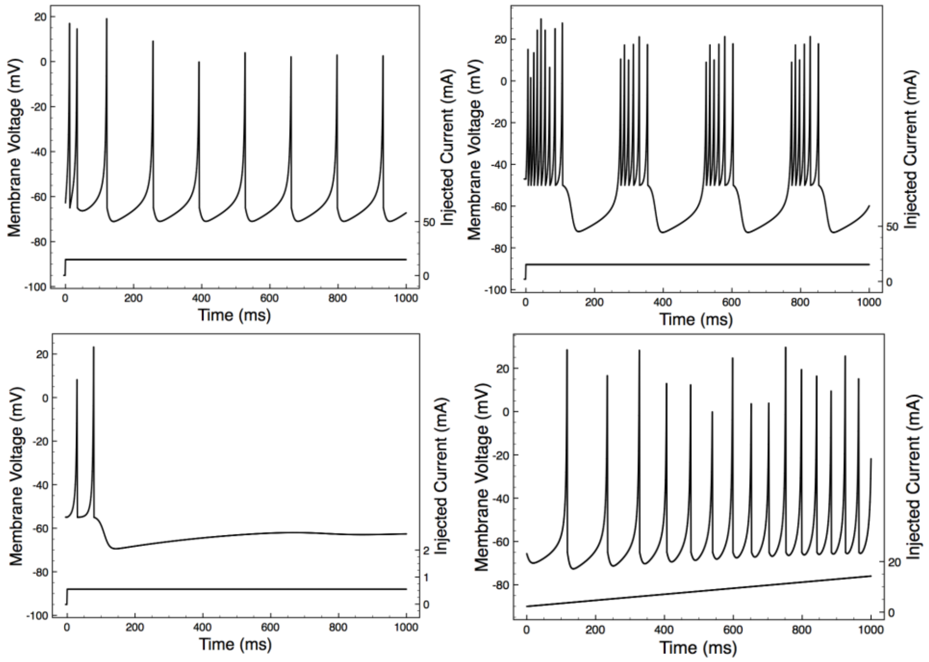

Fig 2. Some Izhikevich behaviours produced by a single neuronet neuron. Top left to right bottom; tonic spiking (TS), tonic bursting, phasic bursting & TS firing freq. vs. input stimulus.

The dependent variables are the recovery variable u and the cell membrane voltage v, respectively. The behaviour of an individual Izhikevich neuron is determined by the four neuron parameters (a, b, c and d) and the input stimulus, i. The mammalian cortex contains many different types of neuron whose spiking behaviour differs in response to an input stimulus. Appropriate selection of these four parameters models the full range of these behaviours, some of which are illustrated in fig. 2.

3.3 Networks of Izhikevich Neurons

The fact that the four Izhikevich neuron parameters in conjunction with the input stimulus can produce such a wide variety of spiking behaviour [8] and that spike and grain timescales are of the same order, makes the Izhikevich model appropriate for real-time control of a granular synthesiser.

When artificial neurons are interconnected in a network, with each receiving inputs from many others, their collective behaviour becomes more complex and interesting. Connections to other neurons (or synapses) vary in strength and type. There are two types of synapses: excitatory, which increase membrane potential in response to a positive synaptic message and inhibitory, which decrease it. The ratio of excitatory to inhibitory neurons in a mammalian cortex is approximately 4:1. Synapses are simulated in the network of N neurons discussed below via an N×N synaptic connection weight (SCW) matrix S with elements $s_{i,j}$, such that firing of the jth neuron instantaneously changes the membrane voltage of the ith neuron v_i by $s_{i,j}$, where;

$$s_{i,j} = S_e c_{i,j} : 1 \le i, j \le N, S_e = 0.5 . \tag{4}$$

for excitatory neurons and

$$s_{i,j} = S_i c_{i,j} : 1 \le i, j \le N, S_i = -1.0 . \tag{5}$$

for inhibitory neurons. The parameters S_e and S_i are weighting factors for excitatory and inhibitory synapses respectively and $c_{i,j}$ are random variables such that;

$$c_{i,j} \sim U(0,1) \tag{6}$$

For each neuron in the network, the SCW weights from all connected neurons which have fired at a given instant are summed and added to the input stimulus, i.e. at each instant of time t, the total input I_i (t) to the ith neuron in the network,

$$I_i(t) = i_i(t) + \sum_{j=fired}^{N} s_{i,j} . \tag{7}$$

where i_i (t) is the input stimulus to the ith neuron at time t.

3.4 Propagation Delays, STDP and Spike Coding Schema

Electrochemical signals in biological neural networks are transmitted between neurons at speeds of the order of a few metres per second, giving rise to time differences between sent presynaptic and received postsynaptic signals. These propagation delays range from fractions of a millisecond to a few tens of milliseconds and are responsible for polychronisation [13]. The research presented here employs propagation delay as a prospective tool in mapping behavioural and dynamical features to grain parameters in a granular synthesiser.

Research from a number of sources has shown that synaptic connection strengths are modulated by relative timing between pre-synaptic action potentials and post-synaptic firing. Within a time difference of around 10-50 ms, long-term synaptic connection strengths are increased when pre-synaptic action potentials precede post-synaptic firing and decreased when they follow post-synaptic firing [14]. Such a mechanism is termed Spike Timing-dependent Plasticity (STDP) and is a means of learning a response to a particular stimulus. STDP will be implemented in the final version of the neurogranular synthesiser described below.

The method by which spike trains are translated into meaningful grain control parameters can greatly influence the resultant sound of a granular synthesiser driven by them. Possible spike-train coding schemas are as follows;

1) temporal coding - a spike could simply trigger the production of a grain
2) rate coding – the spike count per unit time could be used to generate parameter values; i) for a single neuron, or ii) for a group.
3) spike coding – relative spike times of a group of neurons could be scaled to generate a grain parameter value, i) as time to first spike, or ii) phase relative to stimulus for a group.

The efficacy of possible spike-coding schema in mapping spiking behaviour to grain parameters (in terms of timbral malleability) will be evaluated in future work.

4 Research Objectives

The first neurogranular synthesiser was *Neurosynth* [15]; a real-time implementation of neurogranular synthesis using 5 networks of 10 identical Izhikevich neurons, each triggering a half sine-wave windowed sine oscillator, the frequency of which is determined via rate-coding. The 5 network output grain streams are summed to produce the final audio output. The four Izhikevich parameters for each network are available for user control, as are the grain amplitude, frequency, duration and density.

The research will address the limitations of *Neurosynth* and extend the scope of neural network control and modularity of the neurogranular approach, allowing the user to configure parameters such as the following;

- number of neurons in the network
- mappings between oscillators and neurons
- inhibitory neuron proportion/arrangement
- network constitution from homogeneous (with identical neurons) to heterogeneous (with neurons of varied behaviour)
- topology or interconnectivity of the neurons
- propagation delays
- Spike Timing Dependent Plasticity (STDP)
- multiple oscillator waveforms
- multiple grain envelope functions

5 Building a Neurogranular Synthesiser

The neurogranular synthesiser was coded in the C language as two interconnected real-time external objects for Cycling74's™ MAX/MSP audio development environment using Cycling74's™ Applications Programming Interface (API): the neural network neuronet and audio grain generator multigrain~. The neuronet SNN objects output passes scheduling and grain parameter data to multigrain~'s input.

5.1 The Grain Generator Object

The multigrain~ object accepts an argument specifying the required polyphony, P. Each GG produces a 'chirp' of specified waveform, frequency, amplitude, duration, pan and grain envelope shape when a properly formatted message is received at the input. At the time of writing, a basic one-to-one prototypical mapping of neurons to grains has been implemented. Currently, the grain oscillators have fixed frequencies derived from P equal subdivisions of a five-octave range and all other grain parameters are fixed. In future versions of the objects, much greater flexibility in the mapping from neurons to grains and codification of frequency and scheduling will be implemented. Each GG produces non-overlapping grains constituting 'a grain stream'. All streams are mixed to produce the output audio. The following grain parameters are provided for each grain by the triggering input message;

- waveform (sine, triangle, sawtooth, square or white noise)
- frequency
- grain amplitude (0.0 to 1.0)
- grain envelope (rectangular, triangular, trapezoidal, or Tukey)
- grain envelope fade-in slope (trapezoidal and Tukey waveforms only)
- grain duration (10-100ms)
- pan left to right (-50 to +50)

5.2 The Izhikevich Network Object

The neuronet object simulates a real-time N-neuron network of Izhikevich artificial neurons, N being accepted as an argument. Neuronet outputs data to control the multinet~ grain generator and has four inputs for real-time control of the Izhikevich parameters a, b, c and d from which individual neuron parameters are derived. On instantiating a neuronet object of N neurons, the Izhikevich parameters are initially randomised to produce a spread of neuron behaviours. The network can be made homogeneous, with all neuron parameters being identical, or heterogeneous, with individual neuronal parameters for the Izhikevich model being statistically distributed around the four user-supplied average network parameter values (a, b, c and d). Any user adjustment of the Izhikevich parameter magnitudes (a, b, c and d) input to the neuronet object, moves the statistical centre of those parameters through the phase-space for every neuron in the network, so modifying the spiking behaviour of the entire network.

The leftmost input also accepts other real-time messages to configure network parameters such as propagation delay modelling, network topology, constitution (homogeneity or heterogeneity) and input stimulus, etc. On instantiation, a matrix of propagation delay times is calculated from each neuron in the network to every other neuron. At the time of writing, the synaptic connection strengths stored in the SCW matrix remain static after initial randomisation, but future versions will implement STDP by allowing the SCW matrix elements to be varied in real-time.

6 Discussion

The first implementations of granular synthesis were 'offline' systems which did not operate in real time. Pioneering work by Iannis Xenaxis combined music concrète approaches and Denis Gabor's ideas on audio quanta to realise audio grains with many short sections of pre-recorded sound on magnetic tape. Curtis Roads and John Alexander's Cloud Generator program [16] and Barry Truax's GSX and GSAMX programs [17] were the first computer-based implementations of granular synthesis. Cloud Generator is an offline (non-real-time) system providing granular synthesis and sampling. All parameters are determined with either linear functions between user-specified start and end points, statistical algorithms with user-specified maximum and minimum bounds, or random processes and the audio output is rendered as a sound file. Barry Truax's GSX and GSAMX programs perform granular synthesis in real-time with a polyphony of 20 voices. Grain control parameters changed via single keyboard strokes are either fixed values, linear functions (or ramps), or statistically derived around a fixed mean value. More recent examples of real-time granular synthesis include Stochos [18] and Propellerheads'™ Maelström graintable synthesiser. Both Stochos and Maelström employ wavetable-based oscillators and numerous grain control parameters are generated in real-time by deterministic or statistical functions. In all cases, setting bounds for these functions for numerous grain parameters constitutes a control interface of some complexity.

For any synthesis methodology, one of the most important issues, when designing an instrument, is managing the complexity of the controlling parameters to provide an interface which responds in an intuitive manner to the composer. Granular synthesis in particular requires the manipulation of a very large number of grain parameters. The majority of approaches utilise combinations of deterministic and statistical functions to generate the necessary grain control parameters in real-time, but still require the user to specify many parameters.

Using neural networks to generate grain parameters has the potential to greatly simplify the user interface of a granular system. The Izhikevich model of artificial neuron employed has a behaviour which is determined by four parameters and the behaviour of every neuron in the neuronet object's is derived from these four parameters (in a homogeneous network, the Izhikevich parameters of each neuron are identical and, in a heterogeneous network, the parameters of each neuron are derived as a statistical variation around those four values). Thus, the behaviour of the whole network can be radically altered by changing just four parameters. In fact, this could be further simplified once musically useful behaviours are mapped to regions in the network parameter space so that, potentially, one controller might be varied in real-time to allow the user to move through all of the musically 'useful' network behaviours. The input stimulus to the network influences network firing rate and the nature of that stimulus influences how correlated network firing is. Thus, for example, a simple 2-dimensional controller could provide effective adjustment of grain density and character with the y axis controlling amplitude and the x axis varying the nature of the input continuously from constant to random.

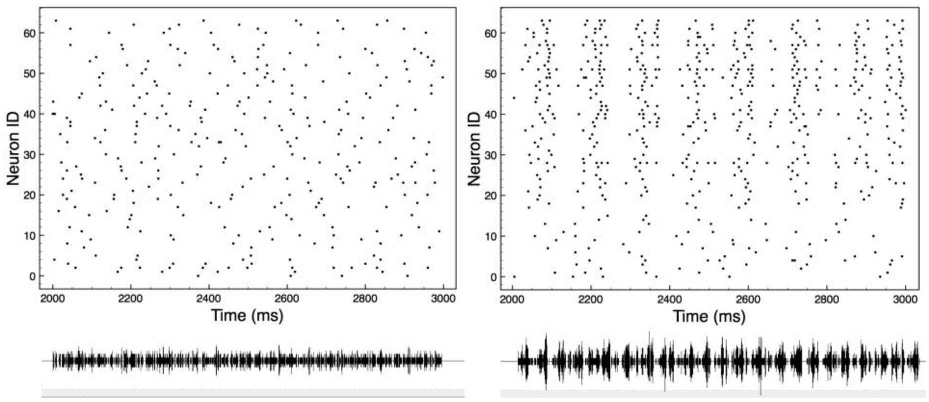

Fig. 3. Raster plots and output waveforms from neuronet driving multigrain~ for a heterogeneous 64-neuron network with input noise without (left) and with (right) delays.

7 Conclusion

The prototype neuronet AN object has demonstrated potential for altering the nature of the output from a granular synthesiser, but also the need for further work to optimise its sonic potential. The resultant sound is derived from the network only in terms of grain scheduling and is limited in scope, although interesting rhythmically. This is due in large part to the limited implementation of spike codifying to generate grain parameters and this will be addressed in further modifications to the system. Although the synthesiser engine (multigrain~) is capable of generating five different waveforms, four different grain envelopes of grain durations from 10-100ms and harnessing panning and amplitude data for each grain, this has not been exploited at the time of writing of this paper. Neural network-derived spike coding of some or all of these parameters could give very interesting sonic results. For instance, a harmonically rich waveform could be low-pass filtered with cut-off frequency modulated by the firing rate of a functional neuron group, or the same group firing rate might determine proportions of the fundamental and a number of harmonics, a higher firing rate equating to a higher high-frequency harmonic content, etc. Future versions of the described objects will incorporate plasticity to allow the sonic potential of topographical considerations and stimulus-driven network adaptation to be investigated. There is also the possibility of the synthesiser 'learning' from its input stimulus or its own output. The intent of this research is to produce a musically intuitive instrument, which provides considerable real-time player control of timbre. Much work needs to be done on the most tonally appropriate means of mapping network spiking behaviour to grain parameter assignment. Further refinements to the system will maximise grain polyphony via code optimisation and dynamic oscillator resource allocation and a simple, intuitive user-interface will need to be developed. A device such as an iPad, capable of tracking multiple (half-a-dozen, or more) finder-tip positions, might serve as an excellent finger-actuated grain parameter controller.

References

1. Miranda, E.R. Wanderley, M.M.: New Digital Musical Instruments: Control and Interaction Beyond the Keyboard, The Computer Music and Digital Audio Series. Middleton, Wis.: A-R Editions (2006).
2. Morrison R.: Liverpool ends year on a cultural high with The Fragmented Orchestra. The Times, London, 13th December (2008).
3. Hickling, A.: Noises Off, The Guardian, 19 December (2008).
4. Ferguson E.: The Future of Music? It might be… The Observer, London, February 25th (2009).
5. Miranda, E.R. Matthias, J.R.: Music Neurotechnology for Sound Synthesis. Leonardo. 42:439-442 (2009).
6. Matthias, J.R., Ryan, N.: Cortical Songs. CD. Nonclassical Records, London (2008).
7. Grant, J., Matthias, J. R., Ryan, N., Jones, D., Hodgson, T., Outram, N.: The Fragmented Orchestra. www.thefragmentedorchestra.com (2009).
8. Izhikevich, E. M.: A Simple Model of Spiking Neurons. IEEE Transactions on Neural Networks. 14: 1469 (2003).
9. Roads, C.: Microsound. Cambridge: MIT Press: 14-16 (2004).
10. Hodgkin, A. L., Huxley, A. F.: A Quantitative Description of Ion Currents and its Applications to Conduction and Excitation in Nerve Membranes. J. Physiol. (Lond.), 117:500-544 (1952).
11. Lubenov, E.V., Siapas, A. G.: Neuron. 58:118-131 (2008).
12. Izhikevich E.M.: Which Model to Use for Cortical Spiking Neurons? IEEE Transactions on Neural Networks. 15:1063-1070 (2004).
13. Izhikevich E.M., Gally, J.A., Edelman, G.M.: Spike-timing Dynamics of Neuronal Groups. Cerebral Cortex. 14:933–944 (2004).
14. Bi, G., Poo, M.: Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type. Journal of Neuroscience. 18 (24): 10464-10472 (1998).
15. Murray, J., Miranda, E.R., Matthias, J.: Real-time Granular Synthesis with Spiking Neurons. Proceedings of Consciousness Reframed - 8th International Conference, University of Plymouth, Plymouth, UK (Invited talk) (2006).
16. Roads, C.: Microsound. Cambridge: MIT Press: 383-388 (2004).
17. Truax, B.: Real-Time Granular Synthesis with the DMX-100. Computer Music Journal, Vol 12, No. 2: MIT Press (1988).
18. Bokesoy, S., Pape, G.: Stochos: Software for Real-Time Synthesis of Stochastic Music. Computer Music Journal, Vol. 27, No. 3 pp. 33-43 (2003).