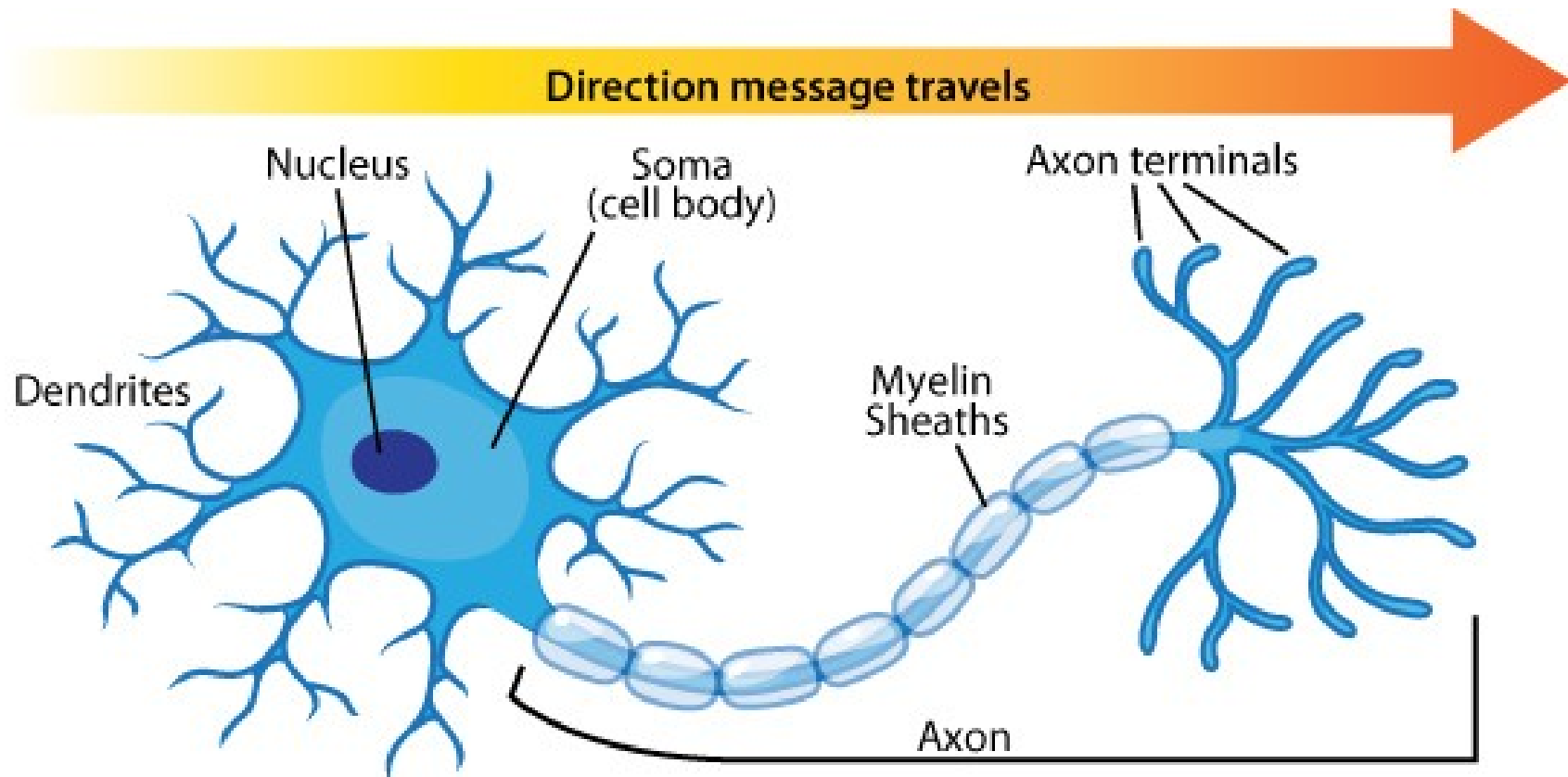


Onset Detection with Spiking Neural Network

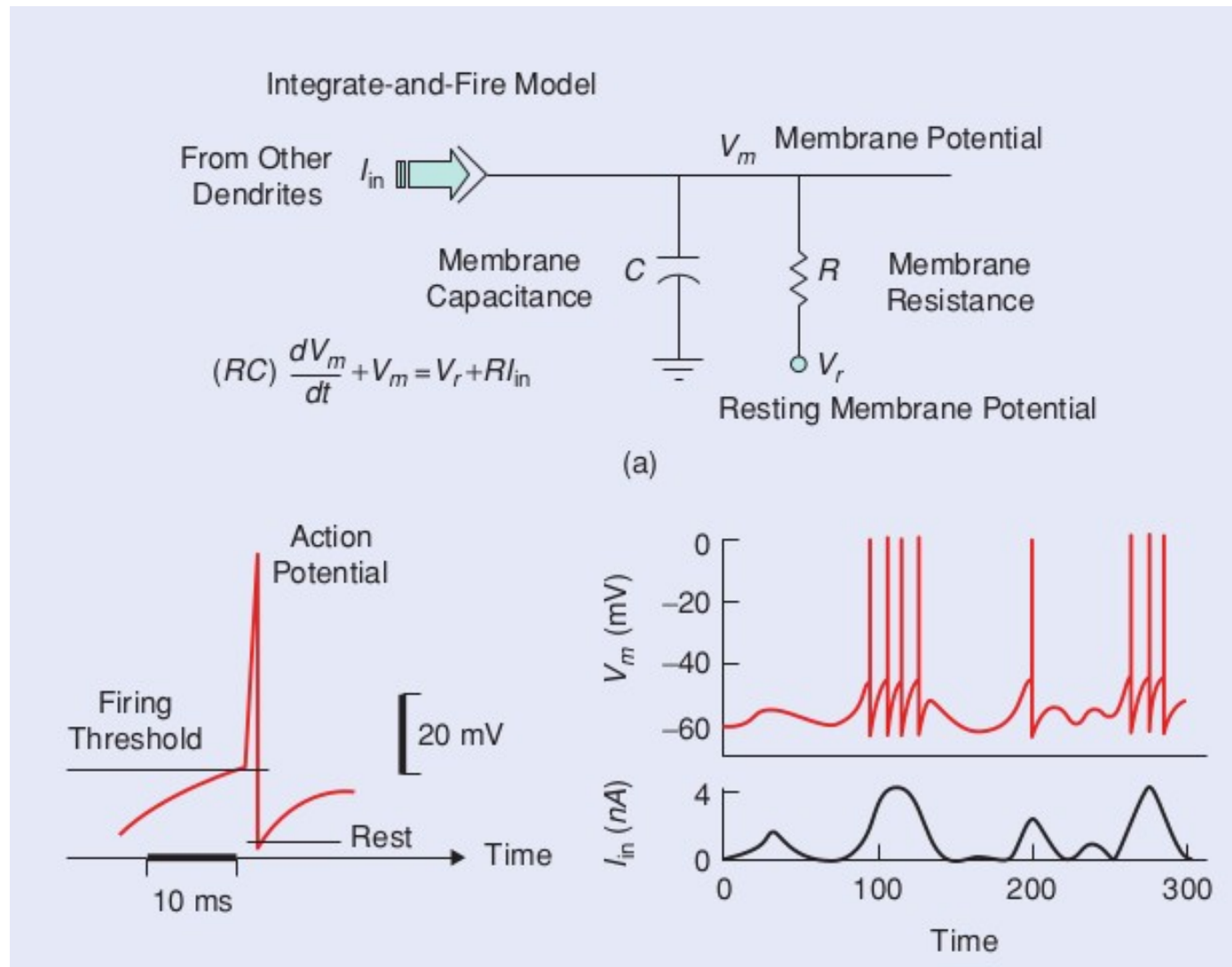
Huang-Yu, Yao
Institute of Systems Neuroscience

2019/06/25

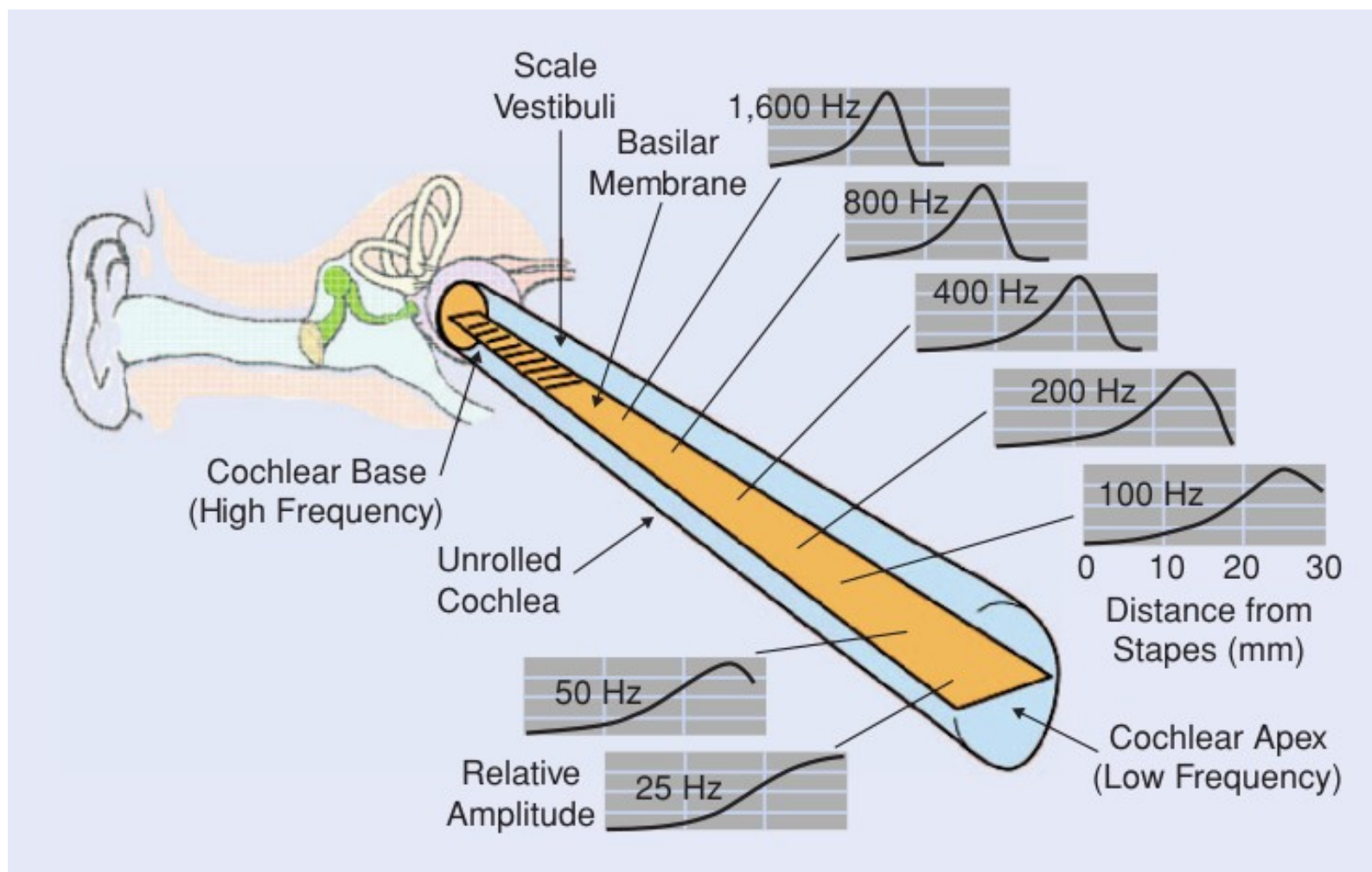
Biological neuron



Spiking neural model



Auditory periphery



Previous work

IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 15, NO. 5, SEPTEMBER 2004

1125

Robust Sound Onset Detection Using Leaky Integrate-and-Fire Neurons With Depressing Synapses

Leslie S. Smith, *Senior Member, IEEE*, and Dagmar S. Fraser

Abstract—A biologically inspired technique for detecting onsets in sound is presented. Outputs from a cochlea-like filter are spike coded, in a way similar to the auditory nerve (AN). These AN-like spikes are presented to a leaky integrate-and-fire neuron through a depressing synapse. Onsets are detected with essentially zero latency relative to these AN spikes. Onset detection results for a tone burst, musical sounds and the DARPA/NIST TIMIT speech corpus are presented.

Index Terms—Depressing synapse, integrate-and-fire neuron, onset detection.

I. INTRODUCTION

THIS PAPER describes a biologically inspired technique for *onset detection*. Onsets occur at the start of certain perceptible changes in a sound. In [1], the term *onset detection* refers to the detection of discrete events in acoustic sig-

in the cochlear nucleus spiking strongly at stimulus start [2], [3]. Therefore, modeling aspects of the early auditory system (the cochlea, auditory nerve (AN), and cochlear nucleus) might offer engineering insight into early auditory processing. From an ecological perspective there are good reasons to believe that onsets provide a useful cue. The onset comes at the start of the sound (or at the beginning of some change in the sound), and is, therefore, useful for priming a response. Initial onsets are relatively undamaged by reverberation, since the first onset in the received signal will normally be from the direct path, and those onsets caused by reflections will generally be smaller. Indeed, these are normally ignored by animals when they estimate the location of a sound source. (This is known as the precedence effect, or law of the first wavefront [4].) Other cues such as offsets are severely smeared out in time in reverberant environments.

Tools

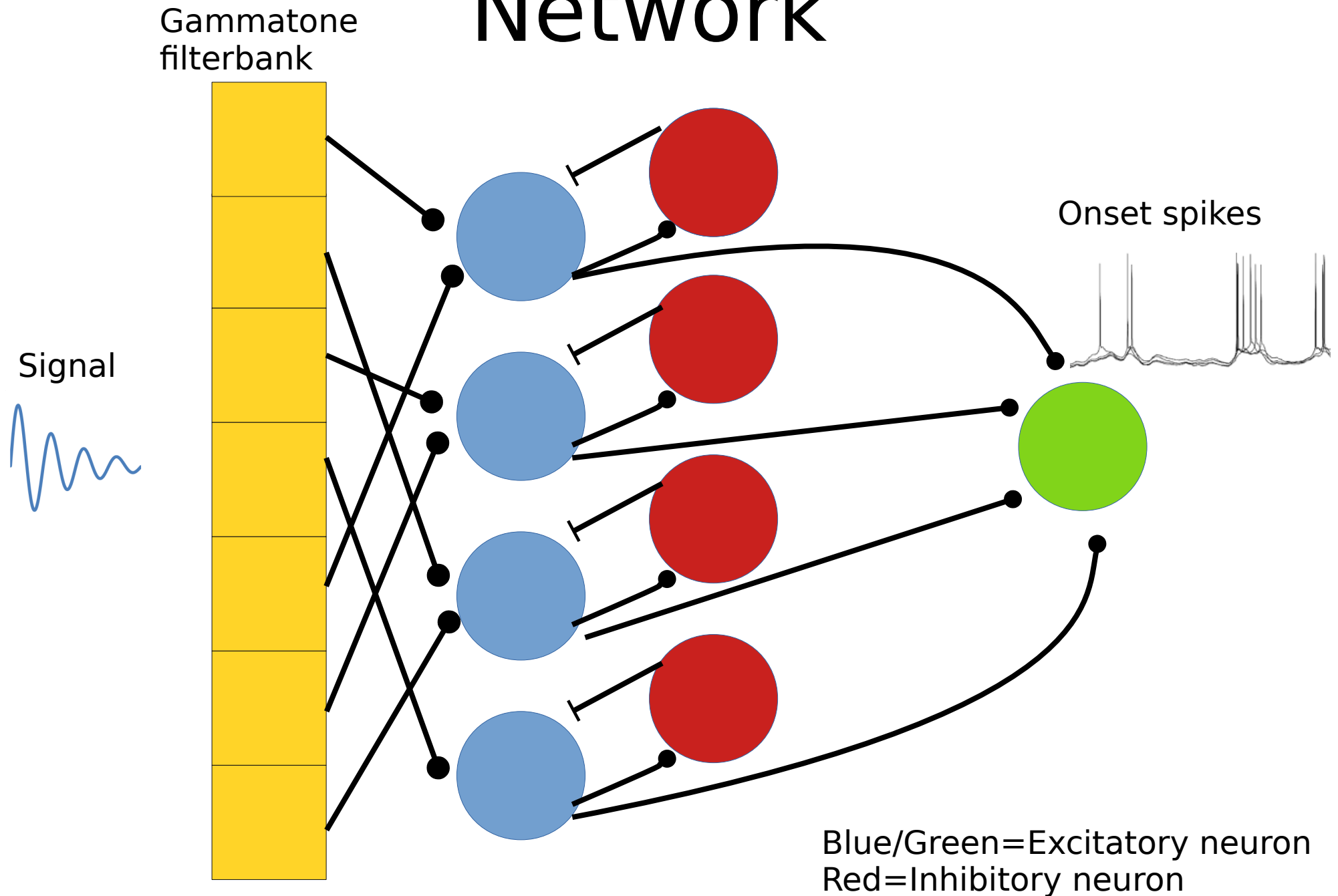
- Gammatone filter – cochlea and auditory nerve fiber



- SNN simulator – higher level neural circuit



Network



Parameters

- Gammatone filterbank

frequency range: 20Hz-16000Hz, divided into 32 bands

- Auditory nerve

The positive values from the filterbank are viewed as input current I to the following model:

$$\frac{dV}{dt} = (I - V) + x_i \times 0.2 \sqrt{2}$$

x_i is a random number generator, v is membrane potential. And spike threshold is 1, reset potential is 0, refractory period equals 5 ms.

Parameters (con't)

- Other downstream neurons

NEST model: iaf_psc_alpha default except stated

Excitatory: $I_e=180$, $\tau_m=20$, $t_{ref}=15$

Inhibitory: $I_e=300$

- Connection weights

Blue \rightarrow Red: 180, delay=1.0

Red \rightarrow Blue: -200, delay=1.0

Blue \rightarrow Green: 9, delay=0.1

Dataset

- **ODB** (onset detection database)
 - 19 real recordings in wav format and their onset positions in text format.

Evaluation

- We consider an onset to be correctly matched if a detected onset is reported within 50 ms of the ground truth onset time.

c : number of correct detections, FP: false positive, FN: false negative

$$P = \frac{c}{c + FP} \quad R = \frac{c}{c + FN} \quad F = \frac{2PR}{P + R}$$

- Merged rate, doubled rate and mean deviation are also provided.

Results

File: 15-tamerlano_act_i_track_15b.txt

OK= 79

FP= 114

FN= 60

Doubled= 9

Merged= 9

MergedRate= 15

DoubledRate= 7.89474

MeanDeviation= 0.0068562

Prec= 0.409326

Rec= 0.568345

Fmeasure= 0.475904

File: 25-rujero.txt

OK= 56

FP= 63

FN= 38

Doubled= 3

Merged= 3

MergedRate= 7.89474

DoubledRate= 4.7619

MeanDeviation= 0.0140039

Prec= 0.470588

Rec= 0.595745

Fmeasure= 0.525822

Results (con't)

File: 2-artificial.txt

OK= 115

FP= 169

FN= 18

Doubled= 22

Merged= 22

MergedRate= 122.222

DoubledRate= 13.0178

MeanDeviation= 0.005436

Prec= 0.40493

Rec= 0.864662

Fmeasure= 0.551559

File: 2-uncle_mean.txt

OK= 137

FP= 99

FN= 41

Doubled= 8

Merged= 9

MergedRate= 21.9512

DoubledRate= 8.08081

MeanDeviation= 0.00673569

Prec= 0.580508

Rec= 0.769663

Fmeasure= 0.661836

Results (con't)

File: 3-long_gone.txt

OK= 109

FP= 134

FN= 34

Doubled= 14

Merged= 14

MergedRate= 41.1765

DoubledRate= 10.4478

MeanDeviation= 0.00716404

Prec= 0.44856

Rec= 0.762238

Fmeasure= 0.564767

File: 3-you_think_too_muchb.txt

OK= 149

FP= 130

FN= 36

Doubled= 24

Merged= 24

MergedRate= 66.6667

DoubledRate= 18.4615

MeanDeviation= 0.00451691

Prec= 0.53405

Rec= 0.805405

Fmeasure= 0.642241

Results (con't)

File: 6-three.txt

OK= 143

FP= 143

FN= 12

Doubled= 20

Merged= 20

MergedRate= 166.667

DoubledRate= 13.986

MeanDeviation= 0.00189189

Prec= 0.5

Rec= 0.922581

Fmeasure= 0.648526

File: 8-ambrielb.txt

OK= 95

FP= 108

FN= 52

Doubled= 11

Merged= 11

MergedRate= 21.1538

DoubledRate= 10.1852

MeanDeviation= 0.00613316

Prec= 0.46798

Rec= 0.646259

Fmeasure= 0.542857

Results (con't)

File: its_alright_for_you_o.txt

OK= 73

FP= 55

FN= 11

Doubled= 6

Merged= 6

MergedRate= 54.5455

DoubledRate= 10.9091

MeanDeviation= 0.00874986

Prec= 0.570312

Rec= 0.869048

Fmeasure= 0.688679

File: realorgan3.txt

OK= 10

FP= 218

FN= 5

Doubled= 1

Merged= 1

MergedRate= 20

DoubledRate= 0.458716

MeanDeviation= -0.005159

Prec= 0.0438596

Rec= 0.666667

Fmeasure= 0.0823045

Results (con't)

File: RM-C002.txt

OK= 54

FP= 85

FN= 53

Doubled= 9

Merged= 9

MergedRate= 16.9811

DoubledRate= 10.5882

MeanDeviation= -0.005635

Prec= 0.388489

Rec= 0.504673

Fmeasure= 0.439024

File: RM-C003.txt

OK= 19

FP= 104

FN= 37

Doubled= 5

Merged= 5

MergedRate= 13.5135

DoubledRate= 4.80769

MeanDeviation= 0.00445684

Prec= 0.154472

Rec= 0.339286

Fmeasure= 0.212291

Results (con't)

File: RM-C026.txt

OK= 31

FP= 121

FN= 5

Doubled= 1

Merged= 1

MergedRate= 20

DoubledRate= 0.826446

MeanDeviation= 0.0154542

Prec= 0.203947

Rec= 0.861111

Fmeasure= 0.329787

File: RM-C027.txt

OK= 126

FP= 108

FN= 97

Doubled= 5

Merged= 5

MergedRate= 5.15464

DoubledRate= 4.62963

MeanDeviation= 0.00728849

Prec= 0.538462

Rec= 0.565022

Fmeasure= 0.551422

Results (con't)

File: RM-C036.txt

OK= 20

FP= 61

FN= 25

Doubled= 1

Merged= 1

MergedRate= 4

DoubledRate= 1.63934

MeanDeviation= 0.015671

Prec= 0.246914

Rec= 0.444444

Fmeasure= 0.31746

File: RM-C038.txt

OK= 85

FP= 68

FN= 109

Doubled= 6

Merged= 6

MergedRate= 5.50459

DoubledRate= 8.82353

MeanDeviation= 0.00770341

Prec= 0.555556

Rec= 0.438144

Fmeasure= 0.489914

Results (con't)

File: RM-G008.txt

OK= 58

FP= 83

FN= 12

Doubled= 8

Merged= 8

MergedRate= 66.6667

DoubledRate= 9.63855

MeanDeviation= 0.00401655

Prec= 0.411348

Rec= 0.828571

Fmeasure= 0.549763

File: RM-J001.txt

OK= 78

FP= 134

FN= 35

Doubled= 8

Merged= 8

MergedRate= 22.8571

DoubledRate= 5.97015

MeanDeviation= 0.0056541

Prec= 0.367925

Rec= 0.690265

Fmeasure= 0.48

Results (con't)

File: tiersen11.txt

OK= 15

FP= 42

FN= 23

Doubled= 3

Merged= 3

MergedRate= 13.0435

DoubledRate= 7.14286

MeanDeviation= 0.004846

Prec= 0.263158

Rec= 0.394737

Fmeasure= 0.315789

Overall:

OK=1452

FP=2039

FN=703

Prec=0.41564

Rec=0.67378

Fmeasure=0.51413

Discussion

- False positive rate seems very high. The network should be more silent.
- Network fine-tuning should start from more naive music clips)