

Modeling melody familiarity judgments in musicians and nonmusicians

Naresh N. Vempala¹ & Anthony S. Maida^{1,2}

Institute of Cognitive Science¹ and Center for Advanced Computer Studies²
University of Louisiana at Lafayette, Louisiana, USA 70504

Correspondence: nrv1@hotmail.com

Introduction

A musician possesses an expanded long-term memory [LTM] component for music relative to a nonmusician because of having experienced several years of training and performance. This study used a sequence recognition neural network (SRNN) to model the effect of musical training on the time course of melody recognition with respect to familiarity judgments.

Dalla Bella, Peretz, and Aronoff [1] adapted Marslen-Wilson's cohort model [3] for studying effects of musical training and familiarity on the time course of melody recognition. Participants, grouped as musicians and nonmusicians based on their musical training, judged whether a gated melody was familiar or unfamiliar, to obtain a Familiarity Emergence Point (FEP). FEP was defined as the point at which the listener began to correctly consider that a melody was familiar. Dalla Bella et al. found that on average, the FEP occurred earlier in musicians. They explained this finding using the proposal that the "feeling of knowing" is based on the total amount of accessed information in LTM. The musician, on account of having more melodies stored in LTM than the nonmusician, would have accessed a larger initial cohort of melodies. We extended this explanation by proposing that the initial cohort of accessed melodies activated a set of familiarity neurons above a certain threshold to create a sense of familiarity in the listener. In such a case the total sum of activations of the neurons representing the initial cohort, serving as input to the familiarity neurons, would be larger for the musician, thereby driving the familiarity neurons above threshold earlier for the musician. Our objective was to model this computationally.

Methodology

Tank and Hopfield [4] designed a neural network to recognize temporal sequences. We used this network to model melody familiarity judgments because it offers an activation-based, parallel processing architecture consistent with the cohort model, allowing specific sequence recognition (SR) neurons to be assigned to detect specific melodies stored in LTM.

We used separate SRNNs for the musician and nonmusician to differentiate the levels of musical training between the two. We used a recurrent inhibitory network of leaky integrator neurons (SR neurons) with a sigmoidal activation function. Inhibitory connections implemented a graded winner-take-all (WTA) competition among the SR neurons. Elements of the input sequence were run through temporal delay filters whose peaks were tuned to synchronously reach their maximum at the end of the input sequence. The weighted outputs of these filters drove the SR neurons. Weights coding the LTM for the sequence were positioned between the input filters and the SR neurons. Stored memories were associated with low energy states of the network, and the network was preset with these stored long-term memories.

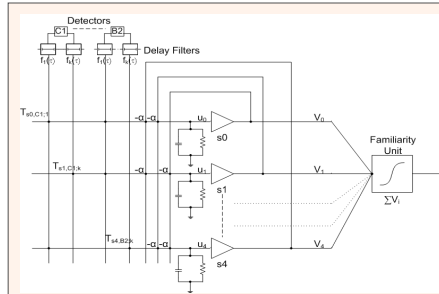


Fig. 1. Nonmusician SRNN consisting of:
Pitch Detectors: 24 pitch detectors to process melodic input into tones. Output of 0 or 1 indicating absence or presence of pitch tone input at that time step.
Delay Filters: For short-term storage of previous tones.
SR Neurons and Connection weights: Each SR neuron is associated with a specific stored melody. In the circuit it has an input voltage u_i and an output voltage V_i . The connection weights for each SR neuron i are represented as T_{iX_k} . They are set to 1 or 0, to enable the SR neuron to detect its corresponding melodic sequence in combination with the continuous delay functions. Here, X represents the pitch tone input such as C1, D1 etc., and k denotes the time units remaining until sequence completion.
Familiarity Unit: The Tank & Hopfield SRNN was extended by adding a familiarity unit which represented higher-order familiarity neurons. The feeling of familiarity was evoked by activity of this unit. Output of this unit was computed as a sigmoid function of the input.

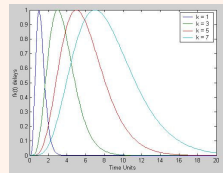


Fig. 2. Delay filter outputs are specified by continuous delay functions of the form
 $f_k(t) = \beta(t/k)^n e^{-\alpha t/k}$

The circuit dynamics of the SRNN is represented as the rate of change of the input potential for an SR neuron i as shown below:

$$C \frac{du_i}{dt} = -\frac{u_i}{R} - \sum_{j \neq i} \alpha V_j - \gamma + \sum_X \sum_k T_{iX_k} f_k(\tau) D_X(\tau - t) \tau$$

C is the membrane capacitance, and R is the membrane resistance. The summation-integral term contributes to an increase in the input voltage u_i when input matches the neuron's expectations. The neuron's expectations are based on the strength of the connection weights T_{iX_k} for an input tone D_X .

Table 1. The nonmusician SRNN had 5 neurons for 5 familiar melodies. The musician SRNN had 15 additional neurons for 15 jazz standards, representing a larger LTM. Each melody was in the key of C, and lasted 10-25 time units (quarter notes). SRNNs were turned on for 30 time units.

SR Neuron	Melody	Musician	Nonmusician
#0	France Boat Song	✓	✓
#1	How to	✓	✓
#2	Round Dog	✓	✓
#3	We Don't Need No Education	✓	✓
#4	Stand of Silence	✓	✓
#5	7th Street Rag	✓	x
#6	Autumn Leaves	✓	x
#7	Blue Horizon	✓	x
#8	Bojilary	✓	x
#9	Indiana Right Song	✓	x
#10	I Die a Little to Sing the Blues	✓	x
#11	Lullaby of Birdland	✓	x
#12	Open Up	✓	x
#13	Swing Changes, Boom	✓	x
#14	Swag You Swag	✓	x
#15	The Man I Love	✓	x
#16	When Did I See	✓	x
#17	Cherokee	✓	x
#18	Beethoven	✓	x
#19	Top Spring	✓	x

Results

Experiment 1 tested the performance of both SRNNs on each of the 5 familiar melodies. Both performed well, and the corresponding SR neuron was active above the level of all the other neurons. We compared activations of the familiarity units of both SRNNs. The musician unit shows faster activation resulting in an earlier FEP.

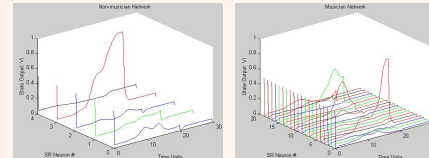


Fig. 3. V_i activations in nonmusician (left) and musician (right) SRNNs for melody 4.

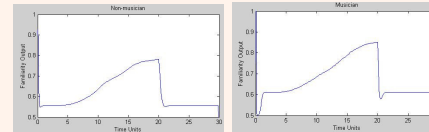


Fig. 4. Familiarity unit activations in nonmusician (left) and musician (right) SRNNs for melody 4.

Table 2. FEP results for musician and nonmusician SRNNs (FEP = Familiarity unit output of 0.7).

Melody Presented	FEP measured in time units from melody commencement					Mean FEP	Mean FEP without melody 5
	1	2	3	4	5		
Nonmusician	9.6	10.5	9.5	10.1	17.5	11.4	9.9
Musician	7.9	7.8	8.9	7.8	8.7	8.2	8.1

Experiment 2 tested both SRNNs with three unfamiliar melodies, to test the prediction that musicians would find unfamiliar melodies more familiar than nonmusicians during the early stages of input because of higher likelihood of false positive partial matches.

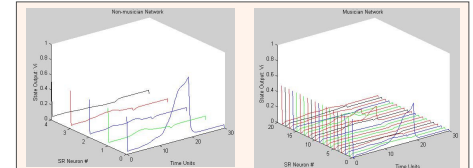


Fig. 5. V_i activations in nonmusician (left) and musician (right) SRNNs for unfamiliar melody 2. The musician SRNN had more partial matches, as predicted.

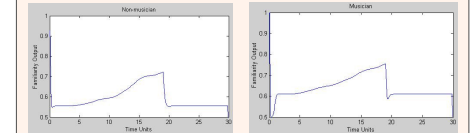


Fig. 6. Familiarity unit activations in nonmusician (left) and musician (right) SRNNs for unfamiliar melody 2 showing a faster rate of activation for the musician SRNN.

Conclusions

Based on [4] we designed two SRNNs representing different levels of musical training, and extended each SRNN by adding a familiarity unit to test the hypothesis that activation of a larger initial cohort results in a greater feeling of familiarity. We successfully modeled familiarity judgments in favor of this hypothesis. Although the current model was built with a realistic, yet limited corpus of melodies, it provides a framework for extending this study to the modeling of other aspects in the time course of melody recognition such as isolation point and recognition point. The current model offered insights by way of its limitations. Inclusion of frequency effects (as occurs in the cohort model) would allow us to associate higher prior probabilities to melodies that are heard more often. The model should be improved by including melodic context-based effects through semantic and syntactic information. Perception and recognition should be separated in the form of a more elaborate hierarchy.

References

- [1] Dalla Bella, S., Peretz, I., & Aronoff, N. (2003). Time course of melody recognition: A gating paradigm study. *Perception and Psychophysics*, 65(7), 1019-1028.
- [2] Koriati, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology - Learning, Memory, and Cognition*, 27(1), 34-53.
- [3] Marslen-Wilson, W. D. (1987). Functional parallelism in spoken-word recognition. *Cognition*, 25, 71-102.
- [4] Tank, D. W., & Hopfield, J. J. (1987). Neural computation by concentrating information in time. *Proceedings of the National Academy of Sciences*, 84, 1896-1900.