Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009) Jyväskylä, Finland
Jukka Louhivuori, Tuomas Eerola, Suvi Saarikallio, Tommi Himberg, Päivi-Sisko Eerola (Editors)

Modeling a Melody Recognition Task using a Cohort Network

Naresh N. Vempala,[*1] Anthony S. Maida,[* #2]

[*]*The Institute of Cognitive Science and*
[#]*The Center for Advanced Computer Studies, The University of Louisiana at Lafayette, USA*
[1]nxv5312@louisiana.edu, [2]maida@cacs.louisiana.edu

ABSTRACT

Dalla Bella, Peretz, and Aronoff studied the effects of musical familiarity on melody recognition by comparing performance between musicians and nonmusicians in a melody gated-presentation (MGP) task. They identified three events in this task which were the familiarity emergence point (FEP), the isolation point (IP), and the recognition point (RP). The FEP occurred earlier in musicians than nonmusicians, but the IP occurred earlier in nonmusicians. Finally, the RP occurred slightly earlier in musicians. We simulated the qualitative results of the MGP task using a connectionist simulation of the cognitive processes underlying the emergence of these three events. We call this a *melody cohort network* (MCN). Separate neural networks modeled musicians and nonmusicians where the musician network represented a larger corpus of stored melodies. The MCN consisted of a core network which modeled the IP and meta-level networks which used the core network as input to model the FEP and RP. Our MCN captures the qualitative results of the MGP task and shows how stored memory size may affect the melody recognition process. We also used the simulation to predict the effects of two levels of severity of acquired amusia characterized by elevated thresholds for perceiving changes in pitch.

I. INTRODUCTION

Music shares several features with language including a hierarchy, temporal structure, vocabulary, and tonal properties (Limb, 2006). There are cases of disorders involving double dissociation between music and language (e.g. aphasia without amusia, and amusia without aphasia), suggesting specificity for music and language in the brain (Peretz, 2002). However, both domains share neural correlates for subcortical processing as well as a large amount of cortical processing. Neuroscientific literature supports left hemispheric dominance for language, and right hemispheric dominance for tonal music while processing melodies, in right-handed people (Maess, Koelsch, Gunter, & Friederici, 2001; Zatorre & Belin, 2001). Melody recognition involves processes similar to spoken-word recognition, such as consolidating temporal input into a higher-order percept using working memory. Hence, as a starting point, musical notes may be considered to be functionally analogous to phonemes in spoken words.

Based on this analogy to spoken-word recognition, and the existence of established experimental paradigms in spoken-word recognition, Dalla Bella, Peretz, and Aronoff (2003) used Marslen-Wilson's cohort model (Marslen-Wilson, 1987) and gating procedures (Grosjean, 1980; Cotton & Grosjean, 1984) to study melody recognition in musicians and nonmusicians. They presented gated melodies selected from a repertoire of French traditional songs (Berthier, 1979) and identified gated familiarity and recognition points consistent with the cohort model. These melodies were divided into *familiar* and *unfamiliar* melodies based on previously established norms (Peretz, Babaï, Lussier, Hébert, & Gagnon, 1995). This melody gated-presentation (MGP) task consisted of two experiments. In the first experiment participants judged whether the melody was familiar after each increment of a gated presentation. The Familiarity Emergence Point (FEP) was defined as the point at which the participant began to correctly consider that a melody was familiar. The FEP was measured by the note number in the melody where the participant first correctly judged that the melody was familiar. The FEP was measured only for *familiar* melodies. The FEP occurred earlier in musicians than nonmusicians. For the FEP to occur earlier in musicians, they must have a stronger feeling of knowing at that point in the melody than nonmusicians. Dalla Bella *et al.* explained this finding based on (a) Marslen-Wilson's cohort model, and (b) Koriat and Levy-Sadot's (2001) proposal that the *feeling of knowing* is in part based on the total amount of accessed information in long-term memory (LTM). The musician, given his or her training and greater exposure to music than the nonmusician, will have more melodies stored in LTM than the nonmusician. Thus, the musician will access a larger initial cohort of melodies, leading to a stronger feeling of knowing, and an earlier FEP.

In their second experiment, Dalla Bella *et al.* studied the time course of melody recognition. Participants sang the melody they thought was being presented beyond the presented gate, and indicated their confidence level on a scale from 1 to 7, where 7 indicated maximum confidence. For this experiment, Dalla Bella *et al.* established two points in the time-course. The Isolation Point (IP) was defined as the point at which the participant demonstrated a correct insight into the identity of the melody. This was measured by the note number at which the participant correctly sang the next three consecutive notes of the melody beyond the presented gate, and did not change his or her response for the remainder of the trial. The Recognition Point (RP) was defined as the point at which the participant was completely confident about his or her judgment. This was measured by the note number at which the participant not only sang the melody correctly, but also indicated a maximum confidence rating of 7 in his or her judgment.

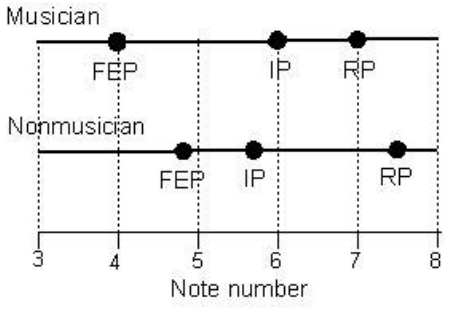

Figure 1. Time-course of melody recognition in musicians and nonmusicians (Adapted from Dalla Bella *et al.*, 2003).

On average, the IP occurred earlier in nonmusicians than in musicians but the RP occurred earlier in musicians than in nonmusicians. Dalla Bella *et al.* proposed that since musicians access a larger initial cohort than nonmusicians, they must examine and assess a larger number of candidates before isolating the correct melody. This accounts for the earlier occurrence of the IP in nonmusicians. The summarized results of Dalla Bella *et al.* are in Figure 1. We developed a simulation to model the hypothesized cognitive processes underlying the FEP, the IP, and the RP with the intention of capturing the qualitative configuration of these six points.

II. METHODOLOGY

We built a melody cohort network (MCN) for simulating the events in the time-course of the MGP task. The MCN consisted of a core network and two meta-level units. The core network of the MCN was designed by adapting a Tank & Hopfield (1987) sequence recognition neural network (SRNN) meant to detect temporal sequences. This core network was used to model the IP. A meta-level unit referred to as the familiarity unit monitored the core network to model the FEP. A second meta-level unit called the recognition unit used inputs from the core network and the familiarity unit to model the RP. The nonmusician MCN consisted of a corpus of five melodies whereas the musician MCN consisted of a corpus of 20 melodies.

A. Motivation for using the Tank & Hopfield SRNN

The Tank & Hopfield (1987) SRNN was designed to recognize noisy and time-warped sequences unfolding in time, such as phoneme sequences in spoken words. The network consisted of detectors, sequence recognition (SR) neurons, and delay filters. The detectors represented input, and were connected to a bank of delay filters which enabled the input to be stored until sequence completion and combined into one single percept. Input weights to each SR neuron were preset to match outputs of delay filters, thereby associating a specific SR neuron with the recognition of a specific input sequence.

Several influential computational models of spoken-word recognition exist, some of which are TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994), Shortlist B (Norris & McQueen, 2008), Merge (Norris, McQueen, & Cutler, 2000), NAM (Luce & Pisoni, 1998), and PARSYN (Auer & Luce, 2005). However, we used the Tank & Hopfield SRNN for modeling the MGP task because it satisfies five key requirements of the cohort model. (1) The SRNN has a parallel processing architecture enabling multiple candidates to be accessed and assessed based on input. (2) The SRNN satisfies the cohort model's LTM representational specificity requirement by assigning separate, computationally active recognition SR neurons to specific temporal sequences stored in LTM. (3) The SRNN is an activation-based network where each SR neuron is activated based on bottom-up sensory input and not top-down contextual feedback, as specified in the cohort model. (4) The cohort model requires the recognition system to be tolerant to noise in the input signal. The SRNN is designed for detecting noisy and time-warped sequences. (5) The SRNN uses a time-evolving winner-take-all (WTA) mechanism among SR neurons, satisfying the cohort model's requirement that each SR neuron should take into account the behavior of competitor neurons. The winner-take-all mechanism is implemented through inhibitory feedback connections from competitor SR neurons to each SR neuron. This allows only the SR neuron with the strongest activation to win.

B. Core Network Design and Architecture

The network was implemented in Java and its simplified architecture is shown in Figure 2. The network consisted of pitch detector neurons, delay filters, connection weights, and SR neurons. Pitch detector neurons, as shown in Figure 2 and Figure 3, detected the pitch tones in the input melody, and had binary outputs of 0 or 1 depending on the absence or the presence of the pitch tone at a particular time step in the input sequence. All input melodies fell within a two-octave range. Each octave had 12 notes: C, C#, D, D#, E, F, F#, G, G#, A, A#, B. Hence, the core network consisted of 24 pitch detectors to detect pitch tones of the stimuli, which were limited to the 24 total pitch tones. Pitch detectors for tones in the first octave were denoted by the note name followed by the suffix 1 (e.g. C1, C#1, D1). Pitch detectors for the second octave were specified by the suffix 2 (e.g. C2, C#2, D2), as shown in Figure 3. Melody sequences were represented using a notation that captured quarter note, half note, whole note, and rest durations as prefixes to the note name or rest. One limitation of the notation was its inability to capture rhythmic nuances based on onsets, such as the difference between one half note and two quarter notes. The duration of each melody sequence was computed as the sum of quarter note time steps. All melodies were transposed to the key of C.

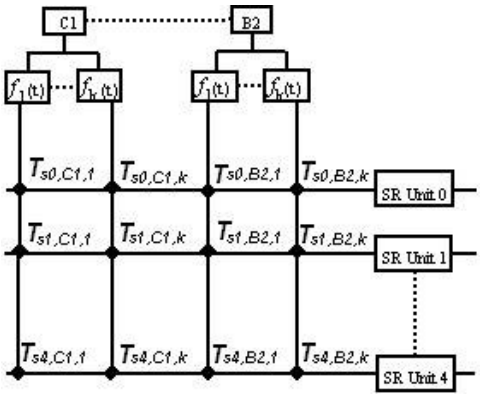

Figure 2. Core nonmusician network with pitch detectors, delay filters, connection weights, and SR neurons/units.

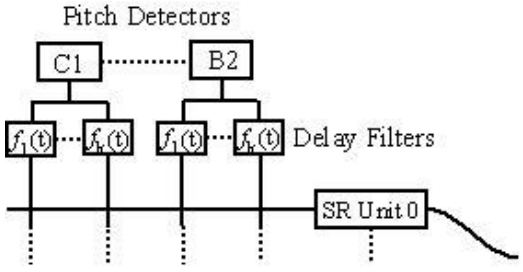

Figure 3. Pitch detectors and delay filters.

Each pitch detector was connected to a bank of smoothing and delay filters, as shown in Figure 2 and Figure 3. The delay filters provided a method of simulating short-term memory. Outputs of the delay filters were specified by continuous delay functions of the form $f_k(\tau) = \beta(\tau/k)^n e^{-n(\tau/k)}$, where k is the number of time steps remaining for the melodic sequence to complete, τ is the duration of the sequence, and β and n are constants. Tones occurring closer to the beginning of the sequence have longer durations (ks) and peak later than tones occurring closer to sequence completion. n determines the position of the peak with respect to sequence duration τ for a given k. For our simulations, n was set to 5, and β was set to e^n. The delay functions enabled the peaks of pitch tones within a melody sequence to synchronously reach their maximum at the end of the input sequence.

The core network uses a recurrent inhibitory network of leaky integrator neurons called SR neurons/units. Each SR neuron, as shown in Figure 2 and Figure 4, has an input voltage u_i and an output activation V_i. The output voltage is also the neuron's activation. It is a sigmoidal activation function of the input voltage, as shown below. It has a range between 0 and 1.

$$V_i = \frac{1}{1 + e^{-\frac{u_i}{0.5}}}$$

The RC circuit acts as a leaky integrator. The *Si* triangle applies a logistic-sigmoid transformation.

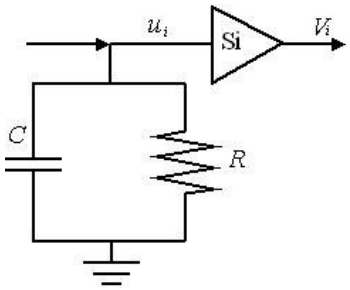

Figure 4. SR neuron/unit.

The LTM representations for each melodic sequence were coded by the connection weights positioned between the delay filters and the SR neurons, at the line intersections in Figure 2. The connection weights, $T_{i,X,k}$, for each SR neuron i, are set to 1 or 0, in order to enable the SR neuron to detect its corresponding melody sequence in combination with the delay functions. X represents the pitch tone input such as C1, D1 and so forth, and k denotes the time units remaining until sequence completion. The method used to preset connection weights may be explained with an example. Assume SR neuron i is built for recognizing the specific melody C E G C, where all tones are in the first of the two octaves, and all tones last for one time unit (quarter note). When the melody is input to the network, pitch detector C1 will have an output of 1 in the first time unit, pitch detector E1 will have an output of 1 in the second time unit, and so forth. Since C1 is the first entry in the four-note sequence, $T_{i,C1,3} = 1$. This allows a delayed input to SR neuron i centered on 3 time units. Since C1 is also in position 4 of the four-note sequence, $T_{i,C1,0} = 1$ provides additional input to SR neuron i with no delay, because k is 0. So, we set connection weights for SR neuron i as follows: $T_{i,C1,3} = T_{i,E1,2} = T_{i,G1,1} = T_{i,C1,0} = 1$. In a similar fashion, each SR neuron/unit in the musician network and the nonmusician network was preset with connection weights based on the melody it needed to recognize.

The connection weights for each SR neuron are analogous to stored long-term memories. These memories were associated with low energy states. Therefore, the dynamic evolution of the network during sensory input, represented by the circuit dynamics equation below, tended to converge to a stored memory.

$$C\frac{du_i}{dt} = -\frac{u_i}{R} - \sum_{j \neq i} \alpha V_j - \gamma + \sum_X \sum_k \int_0^\infty \beta T_{i,X,k} \quad f_k(\tau) \quad D_X(t-\tau)d\tau$$

Here, C is the membrane capacitance, and R is the membrane resistance. C was set to 1, and R was set to 0.5. The summation-integral 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 $T_{i,X,k}$ for an input tone D_X. The SR neurons in the musician and nonmusician core network are connected to each other by means of inhibitory links, represented by the dotted intersections on the left of each SR neuron in Figure 5. These inhibitory connections implement a time-evolving WTA competition among the SR neurons, provided by the summation of αV_j. α was set to 2.5. γ is a global inhibition term used to provide an activation threshold in order to prevent inappropriate melodies from activating the network. β is a term used to scale the excitatory inputs for the summation integral term in the circuit dynamics equation. γ was set to 2.5, and β was set to 0.6. This equation was implemented using the forward-Euler method using a step size of 0.1.

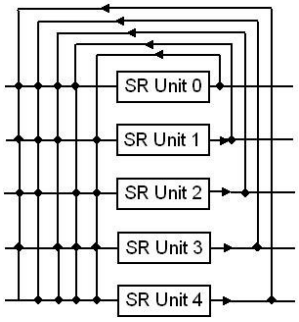

Figure 5. Inhibitory connections among SR neurons.

C. Corpus Selection for Nonmusicians and Musicians

Separate networks were used for the musician and the nonmusician to differentiate their levels of musical training. Both core networks had the same architecture but different corpus sizes in their LTM, determined by the number of melody-specific SR neurons. The musician's corpus was four times larger than the nonmusician's corpus. The nonmusician's core network consisted of five SR neurons designed to detect five popular melodies. The musician's core network consisted of 20 SR neurons. Five of these 20 SR neurons were designed to detect the same five popular melodies, common to both networks. The remaining 15 SR neurons were designed to detect 15 additional melodies in the musician's corpus. The five

common melodies were considered to be *familiar* melodies similar to the familiar French melodies in Dalla Bella *et al.*'s study. These melodies were selected on the basis of sales and popularity from best-selling artists listed in Recording Industry Association of America (2008) and Billboard.com (2008), under the assumption that both musicians and nonmusicians are familiar with these five melodies. Our intention was to use the musician core network to represent a jazz musician. We selected 15 melodies from jazz standards recommended by two sources (Schoenberg, 2002; Hal Leonard Corporation, 2002), based on the premise that a jazz musician knows several jazz standards. The corpus of 20 melodies and their corresponding SR neurons is listed in Table 1. Each melody sequence was a small subset of the song from which it was selected. It captured an essential part of the song (e.g. chorus, the initial few notes, bass line). Each sequence lasted 10-25 time units, with each time unit being set to a quarter note. All melody sequences were transposed to the key of C. During simulation, both networks were turned on for a period of 30 time units. Stimulus presentation started at time unit 3.

Table 1. Corpus and corresponding SR neurons for musician and nonmusician.

SR Neuron	Melody	Musician	Nonmusician
s0	Banana Boat Song	yes	yes
s1	Beat It	yes	yes
s2	Hound Dog	yes	yes
s3	We Don't Need No Education	yes	yes
s4	Sound of Silence	yes	yes
s5	12th Street Rag	yes	no
s6	Autumn Leaves	yes	no
s7	Blue Horizon	yes	no
s8	Boplicity	yes	no
s9	Haitian Fight Song	yes	no
s10	I Got a Right to Sing the Blues	yes	no
s11	Lullaby of Birdland	yes	no
s12	Opus 1/2	yes	no
s13	Sweet Georgia Brown	yes	no
s14	Well You Needn't	yes	no
s15	The Man I Love	yes	no
s16	West End Blues	yes	no
s17	Cottontail	yes	no
s18	Reflections	yes	no
s19	Joy Spring	yes	no

D. Meta-level Familiarity Unit

As described in Section I, Dalla Bella *et al.*'s explanation for the earlier occurrence of the FEP in musicians, was that musicians access a larger initial cohort of melodies than nonmusicians because of a having a larger corpus of stored melodies in their LTM. This results in a stronger feeling of knowing, and hence an earlier FEP. We extended the Dalla Bella *et al.* explanation by proposing that the initial cohort of accessed melodies activated a set of familiarity neurons correlated with a feeling of familiarity. Such a familiarity set, called the Familiarity Unit, would need to be activated above a certain threshold to create a sense of familiarity. If the musician accesses a larger initial cohort than a nonmusician, then the total activations of the neurons representing the initial cohort, serving as input to the familiarity unit, would be larger for the musician. This would drive the familiarity unit above threshold earlier for the musician. We extended the core network by adding a familiarity unit, as shown in Figure 6. The feeling of familiarity was evoked by the activity of this unit. The sum of output voltages V_i of the SR neurons was sent as input to the familiarity unit at each time step. The familiarity unit uses a sigmoidal activation function as shown below. The output value of the familiarity unit was in the interval (0, 1).

$$Familiarity \quad Unit \quad Output = \frac{1}{1 + e^{-(\sum V_i)}}$$

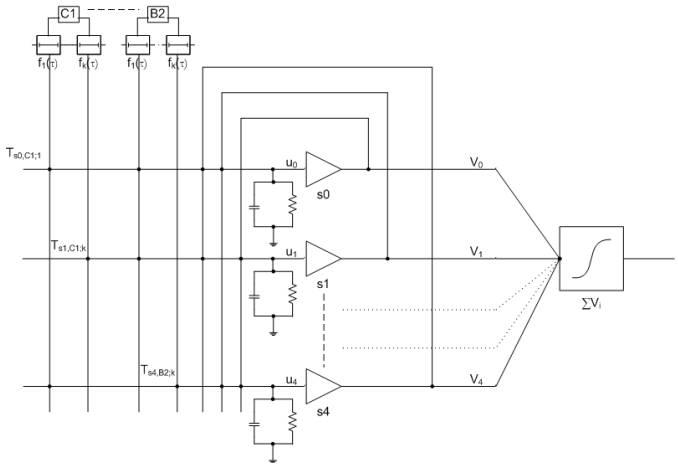

Figure 6. Melody cohort network with the added familiarity unit (Adapted from Tank & Hopfield, 1987).

E. Meta-level Recognition Unit

The RP is the point where the participant is both accurate and confident in his or her recognition of the melody. Experimental studies showing dissociations between confidence and accuracy suggest the role of other factors in determining the confidence level, besides the strength of the memory trace (Busey, Turnicliff, Loftus, & Loftus, 2000; Chua, Schacter, Rand-Giovannetti, & Sperling, 2006). In an experimental task where subjects chose an answer to a question with two alternatives, Koriat (2008) found that a higher level of familiarity about the question's domain can increase the subject's confidence level, independent of response accuracy. Based on the hypothesis that greater familiarity causes a higher level of confidence, we assumed that a meta-level recognition unit computes the

recognition point by monitoring the core WTA network used in finding the IP, as well as the familiarity unit for measuring confidence. The recognition unit computes the RP using the following equation:

$$RP = IP + ((l_i - IP)\ \alpha\ (1 - Fam))$$

Here, l_i is the length of melodic sequence presented to the network in time units. α is a constant. *Fam* is the familiarity unit output at the IP. The second term in the equation computes a cost in time units based on the strength of the familiarity unit output and the amount of evidence already presented to the network in reaching the IP. This cost is added to the IP to compute the RP. The greater the strength of the familiarity output, the lesser the cost. Likewise, the greater the amount of evidence already presented, the lesser the cost. α was set to 2.

III. MGP TASK SIMULATIONS AND RESULTS

A. Modeling the FEP

Input and output voltages of all SR neurons were initialized to 0. Test melodies were presented to the network after three time units, allowing the network time to reach a steady state. The five *familiar* melodies were presented to both MCNs. Both networks recognized all five melodies. The melody-specific SR neurons corresponding to the presented melody sequence responded most strongly with activation levels above those of their competitor neurons in both musician and nonmusician networks. V_i outputs of both networks for melody 3 are shown in Figure 7 and Figure 8. The corresponding SR neuron for melody 3 is *s2*. As shown in Figure 7, in the nonmusician network, *s2* shows a high level of activation for melody 3. In the musician network (Figure 8), in addition to a high level of activation shown by *s2*, competitor SR neurons show partial activations to melody 3. The activation levels of these competitor neurons are based specifically on how the musician's network evaluates similarity of melodies represented by these neurons to melody 3 coupled with the time-evolving WTA mechanism. Activations of the familiarity units for both networks are shown in Figure 9. The musician familiarity unit shows higher activation, thereby modeling an earlier FEP.

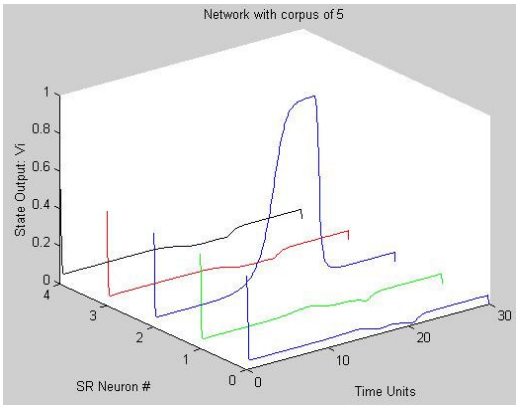

Figure 7. V_i activations of SR neurons in nonmusician for melody 3.

We used a familiarity unit output threshold of 0.65 for signaling the FEP. The results for all five tests are in Table 2. The mean FEP for the nonmusician network occurred 11 time units after the beginning of the melody. The mean FEP for the musician network occurred at 9 time units into the melody, giving the musician network an FEP 2 time units earlier than the nonmusician network. These results were comparable to Dalla Bella *et al.* who obtained results where the musician's FEP occurred 0.8 to 0.9 notes earlier than the nonmusician.

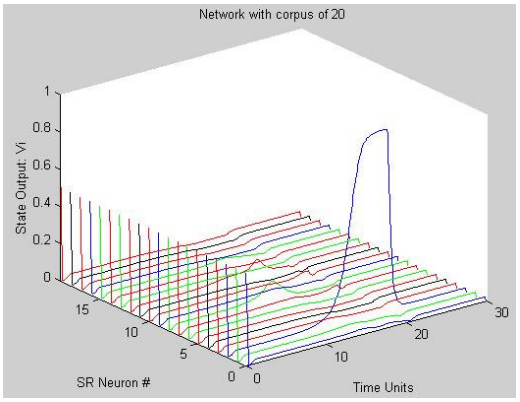

Figure 8. V_i activations of SR neurons in musician for melody 3.

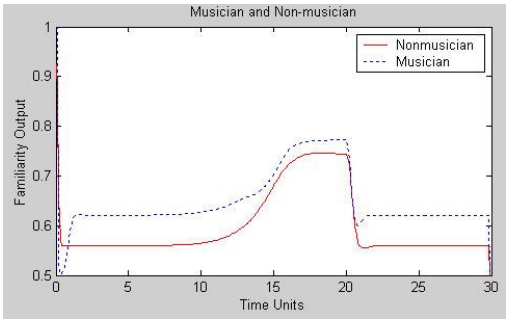

Figure 9. Familiarity unit activations in nonmusician and musician networks for melody 3.

Table 2. FEP results for musician and nonmusician networks.

	FEP measured in time units from melody commencement					
Melody Presented	**1**	**2**	**3**	**4**	**5**	**Mean FEP**
Nonmusician	10	10	11	11	13	11
Musician	9	9	9	9	9	9

B. Modeling the IP

The IP was identified in the core network by carefully comparing activation levels of SR neurons at each gate/time unit of the input. The IP was defined as the time unit at which the activation level of the correct SR unit (a) was at least 0.7 and greater than the activation levels of other competitor SR neurons in the network, and (b) continued to remain higher than its competitor neurons for the next three consecutive time units or until sequence completion. After the start of the input melody, output V_is of all SR neurons were compared to find

the SR neuron with the highest V_i, at each time unit. If the SR neuron's V_i was at least 0.7, the time unit *t1* was also noted. If the SR neuron's V_i continued to be higher than the V_is of all other SR neurons for the next three time units of melodic input or until sequence completion, then *t1* was the IP.

The results for all five melodies are in Table 3. The mean IP for the nonmusician network occurred 12.4 time units after the beginning of the melody. The mean IP for the musician network occurred at 13 time units into the melody, giving the nonmusician network an IP 0.6 time units earlier than the musician network. These results were qualitatively comparable to Dalla Bella *et al.*'s results where the nonmusician's IP occurred 0.3 to 0.4 notes earlier than the musician.

Table 3. IP results for musician and nonmusician networks.

	IP measured in time units from melody commencement					
Melody Presented	**1**	**2**	**3**	**4**	**5**	**Mean IP**
Nonmusician	11	12	12	12	15	12.4
Musician	12	12	12	13	16	13

C. Modeling the RP

The results for both networks are in Table 4. The mean RP for the musician network occurred at 14.9 time units after melody commencement. The mean RP for the nonmusician network occurred at 15 time units after melody commencement. These results indicated that the musician required lesser information than the nonmusician to reach the RP after the IP, despite reaching the IP 0.6 time units later, because of a higher level of confidence indicated by the level of familiarity.

Table 4. RP results for musician and nonmusician networks.

	RP measured in time units from melody commencement					
Melody Presented	**1**	**2**	**3**	**4**	**5**	**Mean RP**
Non-musician	13.8	13.7	14.6	14.7	18	15
Musician	13.6	13.7	14.4	14.6	18.1	14.9

D. Melody Similarity Measurements

To evaluate the core network's realization of melodic similarity, we needed a mathematical measure for computing the melodic similarity of the input melody to the melodies corresponding to SR neurons in the musician and nonmusician networks. According to a study conducted by Müllensiefen and Frieler (2004), edit distance measurements with a rich symbolic representation compared well with human similarity judgments.

Our symbolic notation captures important aspects of each melody such as pitches, pitch durations, and rests, despite its inability to capture note onsets. Hence, we computed normalized edit distances (Wagner & Fischer, 1974) of all the melodies associated with the SR neurons. These edit distances were then converted into similarity values whose output values fell within 0 and 1, where 0 indicates no similarity, and 1 indicates maximum similarity as in the case of an identical melody. A comparison between similarity values and activation levels of SR neurons in musician and nonmusician networks for input melody 5 is provided in Figure 10. As shown in the figure, the network's realization of melodic similarity matched the edit distance measurements well.

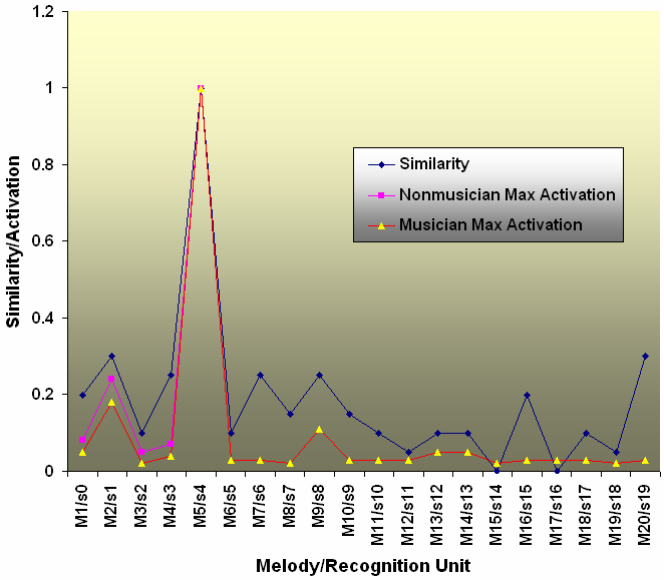

Figure 10. Similarity and activation comparison for melody 4.

E. Summary of MGP Task Simulations

The summary of MGP task simulation results for the musician and nonmusician MCNs showing the FEP, the IP, and the RP as points in the time-course of melody recognition are provided in Figure 11. A comparison between Figure 1 and Figure 11 indicates that our simulation results captured the rank ordering of the FEP, the IP, and the RP, and qualitatively matched Dalla Bella *et al.*'s results.

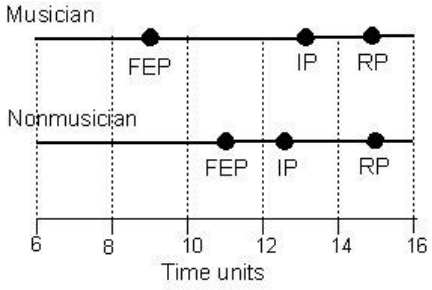

Figure 11. Time-course of simulation results in musician and nonmusician networks.

IV. EFFECTS OF INCREASING CORPUS SIZE

We conducted an initial test to see how well the MCN would scale with increased corpus size. 10 additional jazz melodies were added to the 20 melody corpus (used for the musician network), and an MCN with 30 SR neurons was built. Simulations for the FEP, the IP, and the RP as described in Section III, were run on this 30 corpus network. These results

along with the previous results of the 20 melody musician network are provided in Tables 5, 6, and 7. The results indicate that although there is no noticeable difference in the mean IP and RP values, the FEP values changed considerably with an increase in corpus size by 10 melodies. This is also reflected in the familiarity outputs of the 5-melody, 20-melody, and 30-melody corpus networks, in Figure 12.

Table 5. FEP results for 20 and 30 corpus networks.

	FEP measured in time units from melody commencement					
Melody Presented	**1**	**2**	**3**	**4**	**5**	**Mean FEP**
20 corpus MCN	9	9	9	9	9	9
30 corpus MCN	6	6	7	7	6	6.4

Table 6. IP results for 20 and 30 corpus networks.

	IP measured in time units from melody commencement					
Melody Presented	**1**	**2**	**3**	**4**	**5**	**Mean IP**
20 corpus MCN	12	12	12	13	16	13
30 corpus MCN	12	12	13	13	16	13.2

Table 7. RP results for 20 and 30 corpus networks.

	RP measured in time units from melody commencement					
Melody Presented	**1**	**2**	**3**	**4**	**5**	**Mean RP**
20 corpus MCN	13.6	13.7	14.4	14.6	18.1	14.9
30 corpus MCN	13.6	13.7	14.5	14.6	18	14.9

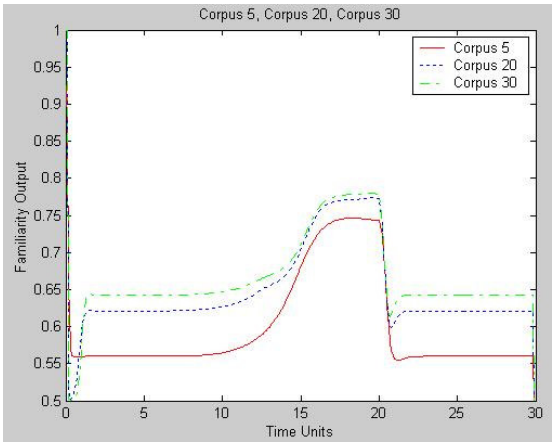

Figure 12. Familiarity unit activations in 5-melody, 20-melody, and 30-melody corpus networks for melody 3.

Increasing the corpus size increases the initial familiarity output value, even prior to stimulus presentation. Eventually, the FEP measure of 0.65 would become irrelevant for a corpus size greater than 40-50 melodies. This indicates that a future improvement for measuring the FEP would involve using a normalized total sum of V_i activations of all the SR neurons as input to the familiarity unit instead of the actual total sum of V_i activations.

V. SIMULATING ACQUIRED AMUSIA

Since we had an MCN that could model melody recognition, we wanted to use it to simulate the effects of acquired amusia on a subject with an intact set of melodies in his or her LTM which were learned and stored in LTM prior to being affected by the condition of amusia. The MCN allows us to represent previously acquired melodies stored in LTM through the melody-specific SR neurons and their connection weights. We wanted to simulate a specific kind of amusia attributed to a deficit in pitch processing (Peretz et al., 2002). One possible reason for this pitch processing deficit in amusics is impaired or elevated pitch perception thresholds than normals (Patel et al. 2008). We conducted two experiments on the 30-melody corpus network to simulate two severity levels of this kind of acquired amusia.

In the first experiment, the threshold for perceiving a change in pitch by the amusic was assumed to be greater than one semitone (Severity level 1). Therefore, for detection by the amusic, the minimum change in pitch should at least be a whole note, irrespective of direction. The five familiar melodies were again fed as input sequences to the network. Based on this severity level, pitches/features in melodies *M1*, *M2*, and *M5* are perceived to be the same as for a normal subject. However, pitches/features in *M3* and *M4* are perceived differently by the amusic. The perception of melody *M3* by the amusic prior to being affected by amusia and after being affected by amusia is provided below.
Pre-amusia *M3*:
E2, 2D#2, E2, 2D#2, E2, 2D#2, E2, 2C2, C2, 2R, C2.
Post-amusia *M3*:
E2, 2E2, E2, 2E2, E2, 2E2, E2, 2C2, C2, 2R, C2.

In the second experiment, the threshold for perceiving a change in pitch by the amusic was assumed to be greater than two semitones (Severity level 2). Therefore, for detection by the amusic, the minimum change in pitch should at least be three semitones, irrespective of direction. The five familiar melodies were again fed as input sequences to the network. Based on this severity level, all five melodies are perceived differently by the amusic. The perception of melody *M4* by the amusic prior to being affected by amusia and after being affected by amusia is provided below.
Pre-amusia *M4*:
C2, 2D2, 2D#2, 2D2, R, 2C2, D2, 2D#2, 2D2, R.
Post-amusia *M4*:
C2, 2C2, 2D#2, 2D#2, R, 2C2, C2, 2D#2, 2D#2, R.

In both experiments, the amusic-perceived sequences were input to the network. The FEP, IP, and RP results are provided in Table 8 and Table 9. For severity level 1, the FEP results from Table 8 may be compared with Table 5. The impaired perception of melody *M3* causes the FEP to occur one time unit later. Given that *M3* is 16 time units in duration, the IP results in

Table 8 when compared with Table 6, indicate that isolation of the melody barely occurred at the time of sequence completion. The IP affects the RP as well. For melody *M4,* the network takes one time unit longer for isolation, although recognition is not affected. Overall, there is a noticeable effect of severity level 1 of acquired amusia in the network's performance with respect to melody recognition.

Table 8. Results for 30 corpus amusic network for severity level 1.

Melody Presented	1	2	3	4	5
FEP	6	6	8	7	6
IP	11	12	16	14	15
RP	13.4	13.7	16	15	17.8

Table 9. Results for 30 corpus amusic network for severity level 2.

Melody Presented	1	2	3	4	5
FEP	6	6	8	6	7
IP	13	14	16	No	6
RP	14	14.5	16	No	18.5

For severity level 2, the FEP, IP, and RP results from Table 9 may be compared with Tables 5, 6, and 7. The network takes longer to reach the FEP for melodies *M3* and *M5.* However, it reaches an earlier FEP for melody *M4.* One possible reason for this could be the impaired melody *M4's* similarity to other melodies in the corpus. The impaired perception of melody *M4* may cause the network to have stronger partial matches with competitor neurons allowing them to have higher activation levels, thereby resulting in a higher summed activation, and an earlier FEP. This is illustrated in the V_i outputs of SR neurons for input melody *M4*, in Figure 13. A competitor SR neuron *s15* has a higher activation level than the corresponding melody-specific SR neuron *s3*.

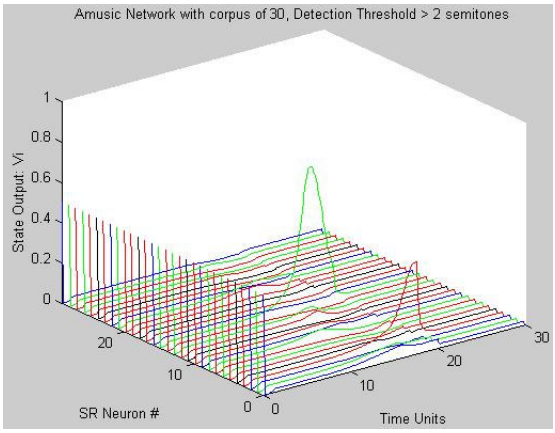

Figure 13. V_i activations of SR neurons in amusic network with severity level 2 for melody 4.

Because of this competition, the network was unable to isolate and recognize melody *M4*, as indicated in Table 9. Although the activation level of *s15* was strong, it was not strong enough for the network to incorrectly isolate melody *M16* corresponding to neuron *s15.* Again, the isolation of melody *M3* barely occurred at the time of sequence completion. Overall, there is a greater effect of severity level 2 of acquired amusia on the network's performance with respect to melody recognition.

VI. CONCLUSIONS

Our goal was to model the qualitative results of Dalla Bella *et al.*'s MGP task, using a connectionist simulation of the cognitive processes underlying the FEP, IP, and RP events in the time-course of melody recognition. We used the Tank & Hopfield model to build a melody cohort network (MCN) because of its relevance to the cohort model. We designed two MCNs representing different levels of musical training to model the IP. We extended each MCN with two meta-level units to model the FEP and the RP. Our simulation results qualitatively matched Dalla Bella *et al.*'s results by capturing the rank ordering of the FEP, the IP, and the RP. We also used the MCN to simulate melody recognition for two severity levels of acquired amusia.

In its current form the MCN has limitations, and offers room for various improvements. All melody sequences in LTM have the same strength. Higher probabilities need to be associated with melodies that are heard more often. This may be implemented in the form of higher activation rates for melodies with higher prior probabilities. The MCN does not show the facilitatory effects of top-down melodic context based on syntactic information. In addition, each melody is processed in a single layer of the network, as one single percept instead of a combination of melodic chunks. To account for these effects, the network design needs to be refined to allow hierarchical stages of processing. Another limitation is concerned with acquisition of melodies. Although the network models the recognition of melodies by SR units for an existing LTM state, it does not learn novel melodies. The connection weights in their current low energy state may be considered to be stable states of a learning algorithm. New melodies may be acquired by randomizing the initial weights, and applying the learning algorithm to train the weights to their final low energy states.

REFERENCES

Auer, E. T., & Luce, P. A. (2005). Probabilistic phonotactics and spoken word recognition. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of Speech Perception* (pp. 610-630). New York, NY: Blackwell.

Berthier, J. E. (1979). *1000 chants* [1000 songs]. Paris : Presses de l'lle-de-France.

Billboard.com. (2008). Retrieved January, 2008, from http://www.billboard.com.

Busey, T. A., Turnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin and Review, 7*, 26–48.

Chua, E. F., Schacter, D. L., Rand-Giovannetti, E., & Sperling, R. A. (2006). Understanding metamemory: Neural correlates of the cognitive process and subjective level of confidence in recognition memory," *NeuroImage, 29*, 1150–1160.

Cotton, S., & Grosjean, F. (1984). The gating paradigm: A comparison of successive and individual presentation formats. *Perception and Psychophysics, 35*, 41–48.

Dalla Bella, S., Peretz, I., & Aronoff, N. (2003). Time course of melody recognition: A gating paradigm study. *Perception and Psychophysics, 65*, 1019-1028.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics, 28*, 267–283.

Hal Leonard Corporation (2002). *Vol. 7 - Essential Jazz standards: Jazz play-along series*. Milwaukee, WI: Author.

Koriat, 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*, 34–53.

Koriat, A. (2008). When confidence in a choice is independent of which choice is made. *Psychonomic Bulletin and Review, 15*, 997-1001.

Limb, C. J. (2006). Structural and functional neural correlates of music perception. *The Anatomical Record Part A, 288A*, 435-446.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1–36.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience, 4*, 540-545.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*, 71-102.

McClelland, J. L., & Elman, J. L. (1986). The trace model of speech perception. *Cognitive Psychology, 18*, 1–86.

Müllensiefen, D., & Frieler, K. (2004). Measuring melodic similarity: Human vs. Algorithmic judgments. In *Proceedings of the Conference on Interdisciplinary Musicology*. Austria: Graz.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189–234.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*, 357–395.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*, 299–370.

Peretz, I. (2002). Brain specialization for music. *The Neuroscientist, 8*, 374-382.

Peretz, I., Babaï, M., Lussier, I., Hébert, S., & Gagnon, L. (1995). Corpus d'extraits musicaux: Indices relatifs à la familiarité, à l'âge d'acquisition et aux évocations verbales. *Canadian Journal of Experimental Psychology, 49*, 211–239.

Recording Industry Association of America. (2008). Retrieved January, 2008, from http://www.riaa.com.

Schoenberg, L. (2002). *The NPR curious listeners guide to Jazz*. New York, NY: Perigree.

Tank, D. W., & Hopfield, J. J. (1987). Neural computation by concentrating information in time. *Proceedings of the National Academy of Sciences, 84*, 1896–1900.

Wagner, R., & Fischer, M. (1974). The string-to-string correction problem. *Journal of the ACM, 21*, 168–173.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11*, 946-953.