From: AAAI Technical Report SS-99-05. Compilation copyright © 1999, AAAI (www.aaai.org). All rights reserved.

A Unified Neurosymbolic Model of the Mutual Influence of Memory, Context and Prediction of Time Ordered Sequential Events During the Audition of Tonal Music

Dan Gang * and **Jonathan Berger**
CCRMA
Stanford University
Stanford, CA 94305, U.S.A
dang,brg@ccrma.stanford.edu

Abstract

We describe a hybrid system to model context formulation and resulting expectations created by a listener while attending to tonal music. The model is hybrid in that we use modular sub-networks to simulate the distinct yet mutually influential schemas involved in constructing expectations for sequential events and the temporal cyclical grid that creates metrical support for these expectations. Using a unified neurosymbolic approach we visualize the fluctuations of musical expectations that arise as a consequence of the dynamically changing musical context.

Introduction

Artificial intelligence techniques find a rich exploratory domain in music. Symbolic systems (e.g., the EMI system for style replication described in (Cope 1991; 1996); expert systems (e.g. (Ebcioglu 1992)'s first order predicate calculus rule based system for Bach chorale style composition), and sub-symbolic approaches (e.g., the use of a Jordan sequential network for generating music described in (Todd 1991)) have been used to create generative models of a variety of musical activities. However the multidimensionality of music deems complexities beyond the reach of a single paradigm. With this in mind, hybrid systems present attractive directions for music based AI research. Nevertheless, relatively little research has been done with this promising approach. In this paper we present our current research in modeling the audition of functional tonal music using a unified neurosymbolic integrative approach.

Neurosymbolic integration can be classified into two strategies: unified strategies and hybrid strategies (Hilario 1995) . Unified strategies enrich neural networks with symbolic capabilities. Hybrid strategies combine neural networks and symbolic approaches at different levels. We use a unified neurosymbolic approach to build a neural network model to visualize the fluctuations of musical expectations that arise as a consequence of the dynamically changing musical context. The model is hybrid in that it uses modular subnetworks to simulate the distinct yet mutually influential schemas involved in constructing expectations for sequential events and the temporal cyclical grid that creates metrical support for these expectations.

In this paper we describe a series of experiments using a version of a Jordan sequential network (Jordan 1986) to model the creation of contexts and expectations along with the perception and cognition of musical meter (that is, the imposition of a periodic grid of accentuated and non accentuated pulses to support and assist music cognition) during the audition of harmonic progressions of common practice, hierarchically related, functional tonal music.

Modeling Expectations of Musical Listening

The sequential and temporal nature of music

Music is, by nature, sequential and multidimensional. Musical listening involves processing a stream of sequential information. Within the stream are hierarchical relationships that occur simultaneously at different scopes from local to global. The sequences are information rich often involving multidimensional complexity. The dimensions can be independent, can influence other dimensions or can be mutually influential. Listening to music involves building contexts that are built dynamically and involve memory. Memory has a dynamic nature of its own - it can be short or long term, and can decay at variable rates and over varying time spans. Memory conditions how we understand the present, and, in turn, how we predict the future. Contexts are used to predict the next elements in the sequence. These predictions are influenced by, as well as affect dynamic changes of context. Musical expectations involve formulating and interpreting the predictions, and substantiating these interpretations by comparing with the sounded event.

Although musical information is sequential, listening associates a temporal dimension to the sequence. A listener not only predicts what will occur next, but also, when it will occur. Listeners use simple periodic pat-

Dan Gang is supported by Fulbright Post-Doctoral Fellowship from the U.S.-Israel Educational Foundation

terns to organize the temporal dimension. These patterns impose an imaginary periodic grid of accentuated and non accentuated pulses. These metric groups direct the prediction of 'when' the next event will occur. The sequential event influences the preference of a given metric organization over others. This means that metric organization and prediction are mutually influential.

Sequential neural network modeling of contextualization and expectations

Listening, performing, and some other musical activities can be represented using a sequential stream of information. The choice of Jordan's sequential net (Jordan 1986) is appealing in such cases. Jordan's sequential net is a version of the back-propagation algorithm (Rumelhart, Hinton, & Williams 1986). Using the learning algorithm, the sequential net is able to learn and predict sequential elements (such as the sequence of a melody's notes or harmonic progression).

The sequential net contains three layers. In our specific case the layers are fully connected. The first layer contains a pool of state units and plan units. The second layer is the hidden layer, and the third layer is the output layer. The output layer is fed back into the state units of the first layer for the computation of the next sequential element.

The value of a state unit at time t is the sum of its value at time $t - 1$ multiplied by some decay parameter (the value of the decay parameter is between 0 and 1) and the value of the corresponding output unit at time $t - 1$. The state units represent the context of the current sequential element, and the output layer represents the prediction of the net for the next sequential element. The feature of feedback distinguishes the Jordan sequential net as a version of back propagation. The implications of this distinction results in the ability to incorporate some sort of the history of the sequence in predicting the next element.

By so doing, a context is recreated from the start of the sequence as each new element is introduced. Furthermore, the plan units in the first layer are used to associate labels for sub sets of the sequences by encoding different values in the plan vectors. Interpolation and extrapolation of the values represented in the plan units can be used to generate new sequences. Each of these new sequences shares interesting properties with previously learned sequences. These variants of the original sequence can be interpreted as creative analogies to the learned sequences. Todd (Todd 1991) describes previous work that applies this strategy to compose melodies that share common features with one another. This principle of melodic variation is a pervasive device throughout the history of western music.

We demonstrate how processing a sequential stream of information in a version of a Jordan sequential network represents listening to music. The Jordan sequential network provides the ability to:

- learn sequences of melodies, notes or harmonic progressions in the learning phase (in our case harmonic progressions). We use the term sequence here to mean an ordered stream and not the musical concept of sequence. Harmonic progressions are sequences of simultaneously sounding pitches sharing common syntactic conventions and hierarchically related to produce sensations of increasing tension and repose.
- establish dynamic contexts in the state units.
- predict the next element for a specific element in the generalization phase (which we interpret as expectations)

In addition to the above properties, the extension of the Jordan sequential network presented here integrates modular subnets. This integrated approach simulates the distinct yet mutually influential schemas involved in building contexts and expectations. In this case, harmonic sequences are learned by the Jordan sequential net. A metric subnet is integrated with the harmonic subnet to provide a periodic iterative index that creates a framework for temporal organization. In music cognition the rate of change of harmonic elements and the hierarchical relations between adjacent chords create the aforementioned percepts of tension and repose. This rate of change is measured against a periodic metric grid. The sensation of this metric grid is commonly expressed by physiological responses such as foot tapping or clapping. The decision as to the pulse rate and pattern of accentuation that determines the meter results from real time pattern analysis by the listener. Once the metric pattern is determined, the imagined periodic repetition of pulsed patterns provides a framework within which a listener can comprehend the control of perceived tension in harmonic progressions. Cognition of meter involves an interpretation of the speed of pulsation (a beat) and a pattern of accentuation. Accentual patterns can be pairs (duple or quadruple meter) or triples (triple meter) distinguished by an accented pulse followed by one or more unaccented (or lesser accented) pulses.

Task and Design

The task of this work is to model cognitive processes involved in listening to music by using the unified approach. The model integrates two sub networks that represent distinct yet mutually influential and complexly intertwined entities, that of harmony and of meter. These two schematic entities combine to formulate context. The mutual influence of these contextual entities are established and learned during the course of formulating corresponding harmonic and metric predictions. These predictions are output in three distinct vectors, one with twelve activations corresponding to each pitch-class, and two for metric pulse units in vectors of three (representing triple meter) and four (representing quadruple meter) units.

We use a learning set of functional tonal harmonic patterns, all in major keys (that is, a single collection of available pitches). The patterns were evenly divided into duple and triple meter progressions. Harmonic rhythm in the learning set ranged from one chord per measure to one chord per beat, although the weighting was on one and two chord changes per measure for both duple and triple patterns.

In the model expectations are not directly learned but rather are an emergent property of the process of learning specific harmonic progressions. In the learning phase the network is trained with 30 metered harmonic progression. The errors are computed by comparing the actual output and the target of a specific metric index and harmony. These errors were used to derive the iterative process of setting the weights of the net.

In the generalization phase we introduce the net with five new metered harmonic progression. In this phase the metric target does not exist, while the harmonic target is established by the actual harmony heard. Nevertheless, we are interested in the actual predictions of the net, more specifically, in the distribution of the activation of the units in the output, which are by-products of the learning process.

Architecture and Representation

In previous publications (Berger & Gang 1996; 1997) we describe a neural network model of the interaction of duple and triple metric schemas with isochronous harmonic progressions. In this model we trained a sequential neural network with a repertoire of metered tonal progressions in duple and triple meter. We then introduced unambiguous, ambiguous and anomalous progressions to our artificial listener and studied the interaction and mutual influence of metric and harmonic expectations.

A general view of the neural network architecture is shown in Figure 1. Our model uses a sequential neural network with two pools of metric units (3 units for triple and 4 units for quadruple meter) and a pool of 12 units representing the pallet of available musical notes that can be combined to create a chord. These notes are represented as normalized pitch class (PC) that is the notes are relative to a common first scale degree and are represented in such a way that the order of their appearance in the chord is irrelevant. The state layer is composed of the two pools of metric units and the pool of PC's. The state units are used to establish a context that influences the prediction of the next element of the sequential information. The output layer contains the same pools of units as the state layer. The metric units represent the prediction of the net for the current metric position. The 12 PC units in the output layer represent the prediction for the subsequent chord tones. In the case of the metric units the output is fed back into the corresponding pool in the metric state and added to the context. In the case of the PC units the context is updated with the target instead of the actual output. The metric pool of units are fully connected to the hidden layer together with the pool of PCs actually implementing the integration of the mutual influences of meter and harmony. The hidden units are fully connected to the output layer (see Figure 1).

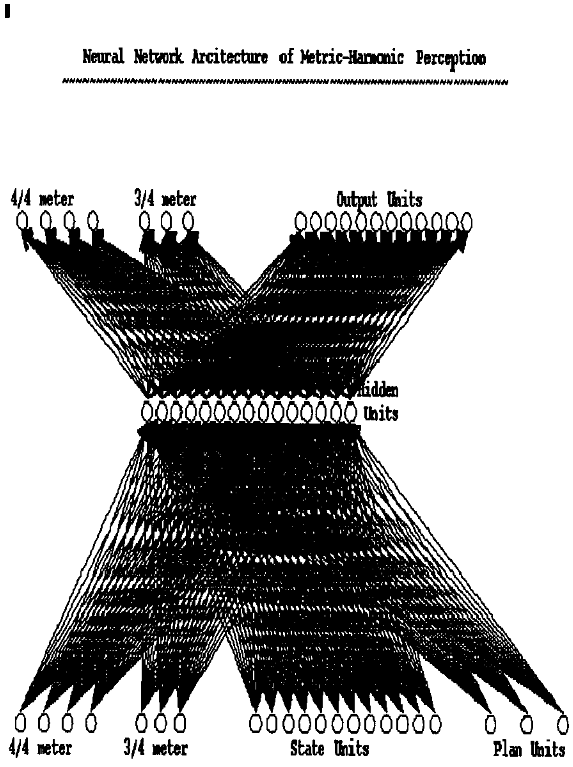

Figure 1: The Neural Network Architecture

The update rule for the metric state units dictates that the metric state is fed by the actual output. This rule simulates the fact that the listener is unassisted in her metric interpretation. In the learning phase we fed back the actual output but used the target meter to train the net. In the generalization phase the meter is unknown, hence there is no target. The update rule for the PC state pool dictates that the harmonic state is fed by the target (and not by the actual output). This rule simulates the fact that the listener is concurrently processing the present chord and expecting the chord to follow. Thus we feed the actual sounded chord and not the expectations of the chord into the state.

On the Hybrid Nature of Context and Prediction

Our modeling approach is hybrid in its integration of a continuous musical signal and of discrete metric patterns. This integration represents the interaction of two distinct but mutually independent schemas that work together to create expectations. Cognition involves segmenting the continuous musical signal into discrete events which can be evaluated both in terms of their (discrete) sequential position and of their temporal placement within the continuum. The division between the discrete and the continuous aspects of music is part of the very nature of the musical experience. We struggle to deal with the continuum of the ever fleeting musical present by evoking discrete schematic memories, building discrete frames of context, and setting discrete goals of expectation. In the visual domain schemas help us distinguish between figure and ground. In the auditory world schemas act as life preservers tossed into a continually flowing river in an attempt to grab, hold and associate moments in time.

Context

In our current architecture we have distinguished between four distinct types of long term memory(LTM): The connections (HarmonyHidden and MeterHidden) which connect the harmony and the meter in the input layer (ContextHarmony and ContextMeter) to the Hidden layer, respectively; and the connections (HiddenHarmony and HiddenMeter) which connect the Hidden layer to the harmony pool of units and the meter pool of units in the output layer (OutputHarmony and OutputMeter). In our model 'context' is created by the exponential decay of the history (i.e. the entire sequence up to the given event). The context is seen to be a part of a recursive and mutually influential process in which the four long-term memory types of connections affect and are affected by the context. Among the important factors in contextualization the order and synchronicity of these influences (along with their possible cognitive implications) are considered. The following formulation demonstrates these mutual recursive features of the net.

To the above notation we add: ExpectHarmony and ExpectMeter to denote the actual outputs of harmony and meter, respectively. DecayHarmony and DecayMeter denote the decay parameters for the harmony and meter, respectively. The value of a decay parameter is between 0 and 1. TargetHarmony and TargetMeter denote the harmonic and metric targets, respectively. The notation layer-name(t) means the value of the activation of the units in the layer at time t.

Following the terminology of our simulator (Miyata 1991) we define the two operators: forward and activation. The operator forward(layer-name1, connection-nam1name2) sends activations forward from layer-name1 to layer-name2. The activations are weighted by the specified connections and are added to the input of the layer-name2. The operator activation(layer-name) computes the activation values of the units in layer-name from their net inputs, using the appropriate activation function for those units (e.g., a non-linear logistic function).

We use lisp-like formulation to define the metric expectation at time t as a function of the harmonic context and the harmonic expectations at time t as a function of the metric expectation.

Formula of metric expectations as a function of harmonic context:

1. ExpectMeter(t) = activation(forward(activation(forward(ContextHarmony(t), HarmonyHidden)), HiddenMeter))
2. ContextHarmony(t) = DecayHarmony * ContextHarmony(t-1) + TargetHarmony(t-1)

   Remark: ContextHarmony(t-1) is the sum of exponentially decayed targets (up to t-2) + the target at time t-2 according to the recursive definition above. For reasons explained above we do not incorporate OutputHarmony(t-1) however we describe it below for completeness:
3. OutputHarmony(t-1) = activation(forward(activation(forward(ContextHarmony(t-1), HarmonyHidden) + forward(ContextMeter(t-1), MeterHidden)), HiddenHarmony))

Formula of harmonic expectations as a function of metric context:

1. ExpectHarmony(t) = activation(forward(activation(forward(ContextMeter(t), MeterHidden)), HiddenHarmony))
2. ContextMeter(t) = DecayMeter * ContextMeter(t-1) + OutputMeter(t-1)
3. OutputMeter(t-1) = activation(forward(activation(forward(ContextHarmony(t-1), HarmonyHidden) + forward(ContextMeter(t-1), MeterHidden)), HiddenMeter))

Error Interpretation

In the learning phase we use the error as a means of driving the learning. In the generalization phase we interpret the net's prediction and the error of the output as related to the actual heard event. In so doing we create a model that visualizes fluctuations in expectations during the course of listening to music.

When a given interpretation is singular and, proves to be correct, expectations are realized. We call this a normative state. Inability to distinguish a singular interpretation is called vagueness or ambiguity; non-realization of the expected goal creates a surprise. We describe and visualize these phenomena in our model.

In processing the output activations in terms of relative strengths and distributions within the output vector we distinguish between the strength and the specificity of a prediction. These two indicators, in their various combinations, present particular classes of predictive states that can be measured in terms of the degree of realized expectation (DRE), which is comprised of:

- the degree of ambiguity (DA), (a measure of the distribution of activations in terms of the degree to which the prediction is specified), and
- the degree of surprise (DSp), (a comparative indicator of the disparity between the expectation and the actual event that follows).

Thus, error measurement in terms of strength and specificity of activations provides parameters for symbolic functions to compute the DRE.

The harmonic expectations which are described by the DSp and the DA, together with the actual heard event are used to measure a continuous scale of the DRE. In a similar way the DA of the metric expectations can be evaluated however, because of the lack of actual heard metric event we can not measure the metric DSp. We use the strength and specificity of the metric pool of units to interpret the metric schema in discrete terms of the index of the beat in a duple or triple meter.

Analyzing the Results

Our model produces graphic visualizations of dynamically changing musical contexts, and a means for qualifying and quantifying the fluctuations of expectations that result from and affect these changes. Our representation allows us to visualize parallel interpretations of tonal expectation as well as multiple interpretations of metric organization.

We first describe cognition of a tonal sequence that is unambiguous both in its harmonic progression as well as in the metric placement of harmonic events. In these cases our model visualizes strong, specific and correct metric and harmonic interpretations.

We proceed to introduce various types of anomalous and ambiguous situations and demonstrate that our system convincingly expresses these surprises and ambiguities. We qualify these cases of surprise and ambiguity in terms of the strength and specificity described above. Ambiguities result from strong or weak and unspecific expectations, while surprise is qualified by strong and specific but incorrect predictions.

Figure 2 shows an example of the output of the PC and meter pools. Relative strengths of activations are represented by the size of the square associated with each of the pitch classes in the harmonic pool and beat positions for each metric pool. The Roman numeral notation above represents the target. Roman numeral notation represents the scale degree upon which the collection of simultaneously sounding pitches (i.e., chord) is rooted. By comparing the output activations to the chord tone members implied in the associated target chord we visualize the listener's prediction and the degree of similarity between the prediction and the target. Furthermore, the metric pools visualize the listener's inference of metric schema and periodic index.

This example represents the output of a standard four measure progression in triple meter. The progression should show a high DRE. The role of the metric subnet is critical in the network's agility in detecting the correct harmonic rhythm by beat five. The recurrence of *vi* (i.e., PCs - 0, 4, 9) on beat 5 squelches the continuation of metric expectation of duple meter after the weak activation of a downbeat. This plausible listening strategy is entirely consistent with the harmonic rhythm since the network is not trained with harmonic rhythms that cross measure boundaries. Of particular interest in this example is the distribution of activations at beat 3 in the harmonic pool. The implied harmonic

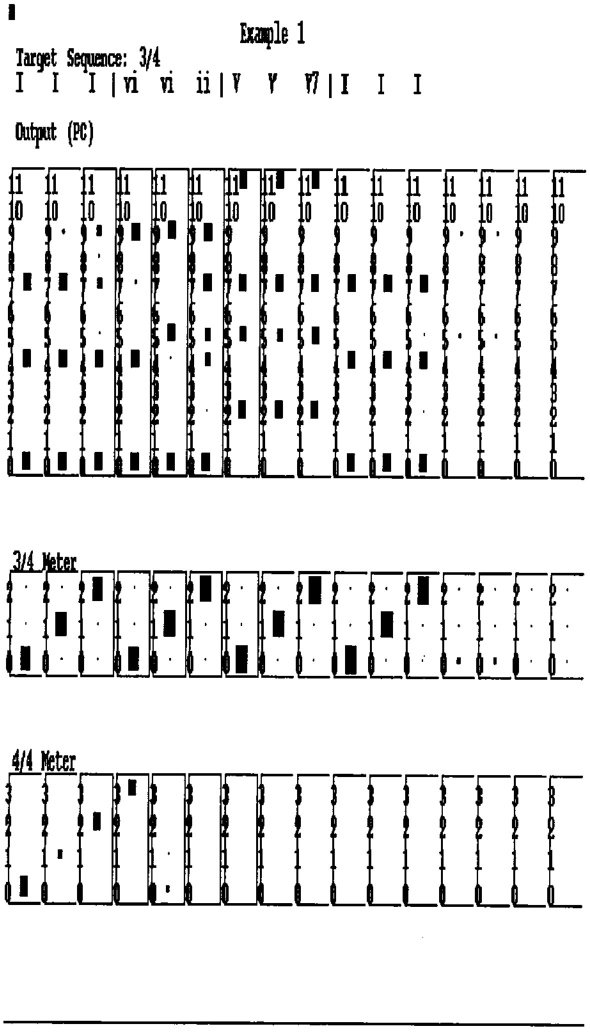

Figure 2: Simulation of expectancies.
From bottom-up: the first row represents a periodic index of duple meter, the second row represents triple meter, the 12 squares in the upper row represent the harmonic expectations by 12 PCs. The size of the squares is proportional to the strength of the units' activity. The location and the size of the squares visualize the net's prediction for the meter and harmony. Time proceeds from left to right. The Roman numeral notation above represents the target. The progression is:
[3/4 I I I — vi vi ii — V V V7 — I I I]

change is a result of the quadruple meter interpretation. The change to a submediant (*vi*) in beat 4 weakens the plausibility of quadruple meter. The repetition of this harmony in beat 5 completely obliterates activations in the quadruple pool.

Discussion and Summary

Using a unified neurosymbolic approach we visualize the fluctuations of musical expectations that arise as a consequence of the dynamically changing musical context. In our model expectations are not directly learned but rather are an emergent property of the process of learning harmonic progressions. The model uses modular subnetworks to simulate the distinct yet mutually influential schemas involved in constructing expectations for sequential events and the temporal cyclical grid that creates metrical support for these expectations.

In the learning phase we train the network with specific metered harmonic progression (all in major keys and in duple or triple meter). In the generalization phase we incorporate symbolic processing by applying a function on the network's prediction. This facilitates quantification of the degree of realized harmonic expectations (DRE) and of the corresponding inference of metric schema.

We introduced normative harmonic progressions, progressions with harmonic anomalies, and normative sequences with temporally offset elements. The model visualizes variations in activations, which we measure by strength and specificity of prediction. These measures provide qualitative means of describing the dynamic fluctuations of the DRE. A high DRE results when expectations are satisfied in a normative situation. DRE is reduced when a surprise or ambiguity appears in anomalous situations.

Distributing the tasks of predicting sequential elements and temporal organization into distinct modular subnetworks provides a means of studying the mutually influential nature of musical expectations and metric awareness.

Acknowledgments

We would like to thank to the Center for Computer Assisted Research in the Humanitics(CCARH) at Stanford University.

References

Berger, J., and Gang, D. 1996. Modeling musical expectations: A neural network model of dynamic changes of expectation in the audition of functional tonal music. In *Proceedings of the Fourth International Conference on Music Perception and Cognition.*

Berger, J., and Gang, D. 1997. A neural network model of metric perception and cognition in the audition of functional tonal music. In *Proceedings of the International Computer Music Association.*

Cope, D. 1991. *Computers and Musical Styles.* A-R editions, Inc.

Cope, D. 1996. *Experiments in Musical Intelligence.* A-R editions, Inc.

Ebcioglu, K. 1992. An expert system for harmonizing chorales in the style of j. s. bach. In Balaban, M.; Ebcioglu, K.; and Laske, O., eds., *Understanding Music with AI.* AAAI Press/ MIT Press. 294–333.

Hilario, M. 1995. An overview of strategies for neurosymbolic integration. In *Workshop on Connectionist-Symbolic Integration:From Unified to Hybrid Approaches at the Fourteenth International Joint Conference on Artificial Intelligence.*

Jordan, M. 1986. Attractor dynamics and parallelism in a connectionist sequential machine. In *Proceedings of The Eighth Annual Conference of the Cognitive Science Society.*

Miyata, Y. 1991. *A User's Guide to PlaNet Version 5.6.* Computer Science Dept., University of Colorado, Boulder.

Rumelhart, D. E.; Hinton, G. E.; and Williams, R. J. 1986. Learning internal representations by error propagation. In Rumelhart, D. E., and McClelland, J. L., eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Vol. 1.* M.I.T.

Todd, P. M. 1991. A connectionist approach to algorithmic composition. In Todd, P. M., and Loy, D. G., eds., *Music and Connectionism.* M.I.T.