# *Neural net Architectures in Modeling Compositional Syntax: Prediction and Perception of Continuity in Minimalist Works by Phillip Glass & Louis Andriessen*

Yayoi Uno, College of Music, and Michael C. Mozer, Department of
Computer Science, University of Colorado, Boulder CO 80309.
email address: uno@spot.colorado.edu; mozer@neuron.cs.colorado.edu.

Acknowledgment is made to Georg Wenzel, Department of Computer
Science, CU-Boulder, for programming and computational assistance.

Abstract: This paper explores the application of neural net architectures in modeling the compositional syntax, prediction, continuity, and perception in two contrasting minimalist works by Phillip Glass and Louis Andriessen. The neural net architecture is trained by *back propagation*, involving various sizes of *hidden units* and *windows*. In this preliminary investigation, we demonstrate how the "peaks" in *prediction error* model syntactical changes in the music and explore how the patterns of change in the *prediction errors* correlate with one's perception of continuity and discontinuity.

***I. Introduction.*** Recent studies in music theory offer diverse approaches toward the assessment of formal continuity and segmentation of contemporary art musics (Hanninen, 1996; Taavola & Lefkowitz, 1993). Algorithms have been developed, based on Gestalt principles, to measure the strength of various musical parameters in creating *disjunction* at local and global levels of musical structure (Tenney & Polansky, 1981; Uno & Hübscher, 1994).

Neural networks have been generally used to predict outcomes based on previous samples. In this paper, *prediction errors* generated by neural-net architectures are used to model the syntactical changes that occur in the temporal unfolding of a musical work: to what extent can we correlate the change in *prediction errorr* with the change in compositional syntax and our perception of continuity in the music? As a preliminary investigation, we apply this procedure to two contrasting minimalist works by Phillip Glass and Louis Andriessen.

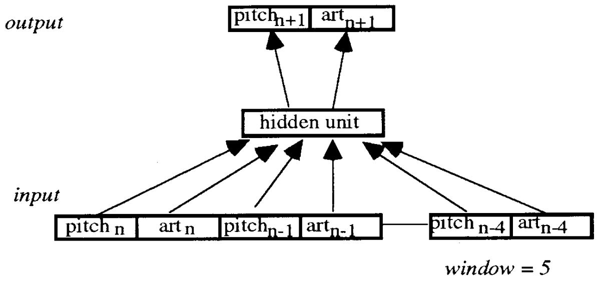
## ***II. Neural-net Architecture.***

A) *Architecture. Neural networks* are pattern recognition devices modeled loosely after the architecture of the brain. A neural network consists of a large number of simple neuron-like processing units, massively interconnected. Neural networks come in several different varieties, called *architectures.* A *feedforward* architecture acts as an associative memory; when given an input, it produces the associated output. From a statistical perspective, feedforward architectures can be viewed as nonlinear regression or classification models (Bishop, 1995; Ripley, 1996). Neural networks have been used in the past for temporal sequence prediction (Elman, 1990; Mozer, 1994).

For the moment, think of the neural network as a black box whose input is a sequence of notes from the piece, comprising of given parameters, pitch and articulation (art) as illustrated under **Fig. 1.** The output is a prediction of what note comes next in the piece. For example, we might present an input "window" of five consecutive notes, and ask the neural network to predict the sixth note; the window could then be shifted to include note six, allowing note seven to be predicted from notes two through six, and so forth.

Each processing unit in a neural network has an *activity level,* a scalar value from 0 to 1. Inputs and outputs are represented as patterns of activity over processing units. In a *feedforward* architecture, units are arranged in layers, and activity is propagated from the input layer to the output layer through a *hidden unit.* The architecture we used is a standard three-layer network with sigmoidal units and direct input-output connections trained by *back propagation* (Rumelhart, Hinton, & Williams, 1986).

**Fig. 1**



B) Representation. In representing musical data for these excerpts, each note is characterized by a pitch and a type of articulation. Since both Glass' and Andriessen's works maintain an eighth-note duration as a constant, duration was omitted as a variable. To present a sequence of notes to the network, each note needs to be encoded as a pattern of activity over processing units. We used one unit for each alternative value of the articulation. In the case of

Glass' "Strung Out," articulation is determined by the phrase grouping; the note that begins each slur requires a heavier attack, therefore, it was given the accentual value of 1, and all other notes the value of 0. In the case of Andriessen's work, three types of articulation, i.e., accent with marcato, accent alone, and no accent, are assigned the diminishing values of 2, 1, 0, respectively.

We used one unit for each possible pitch, and represented a particular pitch by activating its corresponding unit plus one neighbor on either side. This builds some of the linear structure of the pitch continuum into the network, essentially telling the network that an $E_4$ is similar to $Eb_4$ and $F_4$.

Rather than having to be programmed by hand, neural networks are *trained* from a set of examples. Each example consists of an input-output pair.
Following training, when the neural net is given the input, it produces the associated output. Beyond reproducing the training outputs, neural networks can generalize to novel inputs that they have never previously been exposed to. This happens because the network learns about statistical regularities in the training examples, and it can use these regularities to make appropriate responses to novel inputs.

Ordinarily, a network would be trained on, say, three-fourth of a sequence (the *training set*) and asked to predict the remaining one-fourth (called the *test set*). However, to use a neural network in this manner requires an assumption of *stationarity*, i.e., that the structure of the *test set* is the same as the structure of the *training set*. This was not true of the pieces we studied. Therefore, we followed a somewhat non-standard procedure: We trained the network on an entire excerpt given, and then used the *training set* as a *test set*. This is valid because we do not care about the network's generalization performance as much as we are focusing on aspects of compositional syntax of each piece that the network can and can't learn.

After training the network, we can examine its predictions for each note of the piece, and compute a *prediction error* (PDE)--a measure of how far off the prediction is from the actual note. Although one might expect the PDE to be zero, since the network was shown the entire piece during training, the network does not have the capacity to memorize the entire piece. What happens instead is that it picks up on the strongest regularities in the piece, and then produces a small error when the actual note could be generated from the previous notes based on the underlying rules of the piece, and a large error when the actual note violates the underlying rules and is, therefore, unexpected or "surprising." For example, at the start of a new musical phrase, one would expect the PDE to be high, because there is a break in the continuity of the piece; however, in an ascending scale, one would expect the error to be low since a simple rule describes the progression of notes.

From the point of view of perception, the procedure we followed is roughly analogous to having a person listen to a piece of music repeatedly, with sufficient replays that the he/she becomes familiar with the music, but not so familiar that the piece has been memorized. Even with great familiarity, some points in the piece will be unexpected, while others will seem entirely routine and uninteresting.

Two decisions we had to make in building the network were: (1) the number of *hidden units*, i.e., processing units interspersed between the input and output layer, and (2) the size of the input *window*, i.e., the number of notes used to predict the next note. Rather than arbitrarily choosing a value for these two parameters, we trained twelve networks, specified by the Cartesian product of five, ten, twenty, or forty *hidden units*, and a context of five, ten, or twenty notes. We then averaged the predictions of the twelve networks and compared to the actual note to compute a prediction error. The idea of using an ensemble of networks corresponds roughly to having a collection of experts, some of whom are are making predictions based on only simple rules and local bits of structure while others are making predictions based on more complex rules and broader musical context.

### *III. Application to two minimalist works by Phillip Glass and Louis Andriessen.*

The two works under comparison, "Strung Out" (1976) by Phillip Glass and "Hoketus" (1992) by Louis Andriessen, differ radically in their syntactical organization that to categorize both under the common rubric of minimalism seems highly reductive. These two works were chosen, for our preliminary investigation, due to the monothematic, steady eighth-note pulse--characteristics that allowed us to focus exclusively on pitch and articulation as the variables in assessing the syntactical changes that take place in the music.

For each work, the neural network generated the PDEs based on pitch, articulation, and the combination of pitch and articulation; in determining the latter, we controlled the weighting of pitch and articulation at the ratio of 3:1. **Fig. 3 & 4** display the combined PDEs of pitch and articulation: the upper graph display the oscillation in contours (low to high) of pitch based on the input. The horizontal axis shows the index of notes from the beginning to the end of the sampled excerpts. The lower graph displays the combined PDEs of pitch and articulation generated by the neural networks; the twenty highest "peaks" in the prediction errors are highlighted by a cross ("x").

a) "Strung Out" (1976) by Phillip Glass. This work for solo amplified violin illustrates Glass' minimalist syntax based on an additive and subtractive techniques involving a diatonic pattern in C (**Ex. 3**). Notice how the initial melodic pattern, $E_4$-$G_4$, $E_5$-$D_5$-$C_5$, is transformed through elongation, truncation, and change in the slurring of notes. Rests appear briefly in the middle passage, between note entries #704 to #950.

**Fig. 3:** Glass' "Strung Out"

*pitch / articulation*

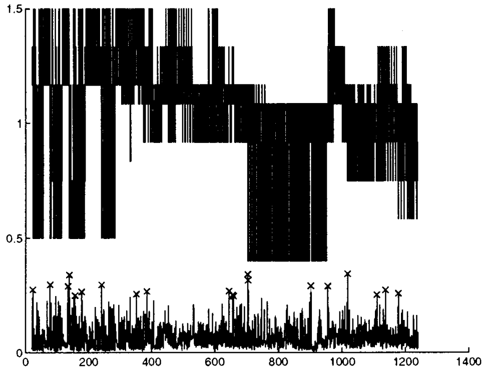
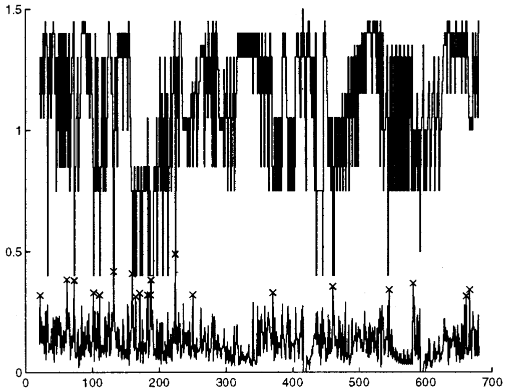
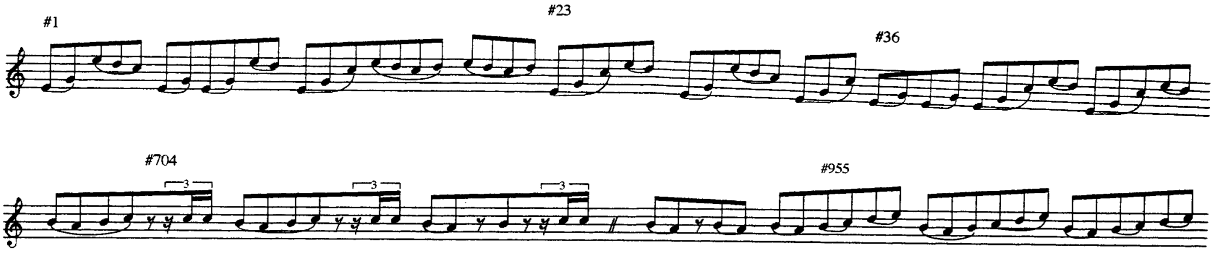


**Fig. 4:** Andriessen's "Hoketus" (reh. E)

*pitch/ articulation*

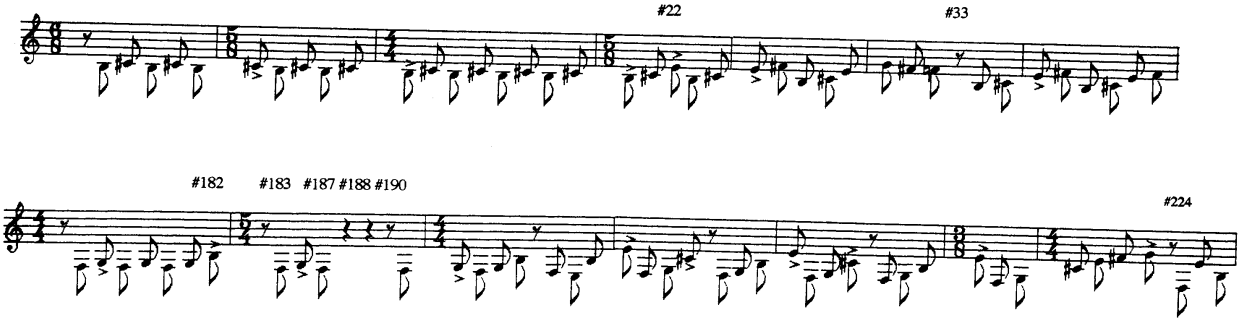


**Ex. 3:** musical excerpts from Glass' "Strung Out"



**Ex. 4:** musical excerpts from Andriessen's "Hoketus"

*upper stem = group one*
*lower stem = group two*



The overall contour of the PDEs for "Strung Out" (**Fig. 3**) displays a relatively dense concentration of "peaks" in PDEs in the opening passage (#1-#400). In general, PDEs peak at points where there is a sudden disruption or deviation from a pattern based on: 1) an additive/subtractive process or 2) static repetition. For instance, the initial peaks occur when the repetition of $E_5$-$D_5$-$C_5$-$D_5$ becomes disrupted (see #23) by the reappearance of the initial motive or when the additive process becomes disrupted (see #36) by a contraction in the pitch range to $E_4$-$G_4$. The following region, note entries #401 and #600, maintains relatively low PDEs, as the melodic pattern "locks" in to a narrower pitch range governed by predictable processes of change. The next passsage, #704-#1017, is framed by the two highest peaks in PDEs. Note entry #704 corresponds to the beginning of the passage where rests and a triplet subdivision of the pulse appear for the first time (see **Ex. 3**). The "peak" at #704 correlates with the high degree of surprise experienced perceptually by the sudden intervention of rests (which were not heard before in the context of this work). As this new pattern is established and the pitch range is reduced to the alternating notes $B_4$-$A_4$, PDEs dip down, only to peak again at #955 where an ascending pattern emerges without rests (see **Ex. 3**). The highest peak at #1014 corresponds to the point where a descending pattern emerges that picks up $G_4$, a note which had been left out of the melodic range for the previous eight-hundred note entries.

b) "Hoketus" (1992) by Louis Andriessen. This work is scored for a set of two pan flutes, pianos, electric piano, bass guitars, congas, and alto saxophones. **Ex. 4** shows the composite melodic pattern formed by the two groups in two different passages. In contrast to Glass, Andriessen's syntax is defined by greater melodic range, angularity in pitch contour, and by his strategic use of accents and rests; the syntactical changes are less predictable, as they do not follow a systematic process of elongation and truncation. Accents and rests demarcate the melodic patterns with greater articulative force in "Hoketus" than the slurring that demarcates the melodic grouping in Glass' work.

The graphic contours of pitch and combined effect of PDEs (**Fig. 4**), indeed, show a much greater oscillation and range than the PDEs displayed for Glass' work. Here the peaks are located in the opening third of the piece, where the maximum PDE occurs at #224. In the opening passage (**Ex. 4**), the initial peaks are formed where the pattern breaks out of the alternating dyad, $B_4$-$C\#_4$, and lands on an accented $E_4$ (#22), and where a rest appears for the first time (#33). The highest concentration of peaks are found between #182-190, where Andriessen disrupts that dyadic oscillation between $F_3$-$G_3$, with frequent rests--a gesture that had not been exploited previously. The PDEs dip down to a low region in the passage between #262-347, where the pattern stabilizes to a dyadic interchange between the two groups. For the remaining portion of this work, peaks in PDEs occur sporadically, corresponding generally to points in the work where an accent or rest ocassionally disrupts the continuity.

## *IV. Summary and Future Considerations.*

Based on our preliminary investigation, we summarize that: 1) passages defined by predictable processes and/or repetition correlate with relatively low degrees of PDEs, while disruption in the exisitng pattern or an introduction of new pattern generates a sudden peak in PDE; 2) there is a direct correlation between one's perception of continuity and PDEs: high measure of PDE corresponds with a disruption in the continuity, and vice versa; 3) articulation in Andriessen's work exerts a strong force in controlling the PDEs, while rests play a strong force in controlling the PDEs for both.

Future considerations for refinement include, but are not limited to: 1) comparing PDEs based on different representation of input data, i.e., symbolic vs. numerical; 2) adjusting the weighting of variables in normalizing the combined output of PDEs; 3) examine the *variance measure*--numerical measure of how much the PDE varies from one note to another--as an evaluative criterion. In addition, we will extend the application to works that are based on stochastic, random, and/or chance-based procedures, to further explore the correlation between PDEs, syntactical changes, and issues pertaining to perception.

## *References:*

Bishop, Christopher M. (1995). *Neural Networks for Pattern Recognition.* Oxford : Clarendon Press.

Elman, J. L. (1990). "Finding structure in time." *Cognitive Science* 14: 179-212.

Hanninen, Dora A. (1996). *A General Theory for Context-Sensitive Musical Analysis: Application to 4 works by Contemp. American Composers.* Ph. D. dissertation, NY: University of Rochester.

Mozer, Michael C. (1993). "Neural Net Architectures for Temporal Sequence Processing." *Time Series Prediction: Forecasting the Future and Understanding the Past.* Eds., A. S. Weigend and N. A. Gershenfeld. Addison-Wesley: 243-264.

Ripley, Brian. (1996). *Pattern Recognition and Neural Networks.* New York : Cambridge University Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). "Learning internal representations by error propagation." In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume I: Foundations*:318-362. Cambridge, MA: MIT Press,

Taavola, Kristin & Lefkowitz, David. (1993). *Generat -ing Segmentation: A Piece-Specific Approach.* New England Society of Music Theory, Tufts University.

Tenney, James and Polansky, Larry. (1980). "Temporal Gestalt Perception in Music." *Journal of Music Theory* 24/2: 205-242.

Uno, Yayoi and Hübscher, Roland. (Denmark,1994). "Temporal-Gestalt Segmentation—Extensions for Compound Monophonic and Simple Polyphonic Musical Contexts: Applications to Works by Boulez, Cage, Xenakis, and Ligeti. " *Proceedings of the International Computer Music Conference*:4-7.