A real time model of the formulation and realization of musical expectations

Jonathan Berger & Dan Gang

Stanford University

Address:

Jonathan Berger and Dan Gang
The Center for Computer Research in Music and Acoustics
Department of Music
Stanford University
Stanford, CA 94305

e-mail: brg@ccrma.stanford.edu, dang@ccrma.stanford.edu,

telephone: 650 723 4971

Abstract:

In this paper a model of musical listening is described. The model provides a visualization of the formulation and realization of musical expectations as a listener hears (or imagines) functional tonal music. The model provides a framework for categorizing and evaluating expectations and their associated realizations.

The model is a modular sequential recurrent neural network that incorporates interdependent sub-nets for harmonic progressions and meter. Using this model we examine the interplay of metric inference and functional tonal harmony. The model visualizes the mutual influence of these musical characteristics and provides visualizations of normative and disruptive listening situations.

In addition to visualizing processes of expectations and realization a number of cognitive implications are discussed. Musical listening involves sub-symbolic learning through experience. Acquiring metric and harmonic schema are an emergent property of a listener's exposure to metered tonal harmonic progressions. Harmony and meter are mutually influential in creating a combined context for prediction. From these predictions interpretation of the metric schema and harmonic expectations are formulated. Expectations can be described in terms of strength and specificity, the affect of which attributes result in specific, ambiguous and vague expectations. Under normative situations expectations are realized while in irregular situations there is a difference between the expectation and the actual heard events. The degree to which the expectation is realized (termed here, the DRE) corresponds to the affect of surprise.

The computational model simulates some of the cognitive processes involved in musical listening. It provides a visualization of the complex dynamic processes involved in musical listening, and suggests methods to qualify and quantify aspects of these processes.

"As the director of an orchestra, I could make experiments,
observing what elicits or weakens an impression and accordingly
correct, add, delete, take risks."

- Joseph Haydn

1. Introduction

Franz Joseph Haydn was perhaps the first composer to articulate his craft in terms of risk taking vis a vis his audience. These risks were strategically placed disruptions of listeners' musical expectations. It is therefore fitting that this work addresses theoretical aspects of musical expectations from the standpoint of Haydn's audience[1].

Can a musical analysis express the extraordinary playfulness one senses when listening to the following musical example (see fig. 1)?

< insert fig. 1 - Haydn, Piano sonata H. XVI:50, third movement, mm. 1-11>

Labels such as $^{V}/_{III}$ or $^{V/}_{V/vi}$ to describe the chord in the penultimate measure infer functional relationships to the event. But these relationships are neither present nor relevant at the time the chord is heard.

Terms such as elision or *takterstikung* to describe the ambiguity of phrase in measure four address the compositional technique but say nothing about its perceptual affect.

Musical treatises based upon rhetoric describe methods with which composers willfully confuse the listener. Most analytical and descriptive methods provide some means to evaluate where and when a musical event conforms to the global norms of style and genre, and the local norms of function and context. Fewer provide a sense of the degree to which norms are

disrupted. Fewer still consider the perceptual affect of perturbations of the norm.

Listener's expectations are implicitly addressed in some analytical approaches. In Hugo Riemann's pedagogical work on tonicization, for example, the author distinguishes situations in which the target resolution of the tonicization is absent. For such situations Riemann (1916) introduces square-bracket notation to represent the 'expected' chord along with the actual sounded chord.

Gjerdingen (1988) identifies melodic archetypes and relates stylistic evolution to deviation from schematic norms. Narmour (1990) describes melodic tendencies that create implications for specific realizations. Some of these tendencies have been validated experimentally (Cuddy & Lunney, 1995).

Narmour (1990) proposes that expectations result from both bottom-up and top-down processes. Bottom-up processes are independent of prior knowledge and include principles relating to the size and direction of a melodic . Top-down processes incorporate experience with genre as well as of the history of the particular piece as it is heard (extra and intra-opus knowledge). Bharucha (1987) describes a connectionist framework for modeling the interaction between top-down and bottom-up processes.

Lerdahl and Jackendoff (1983) apply perceptually based rules of preference and reduction to identify hierarchical structure in a musical score. Meyer (1973) proposed that listeners, aware of the implications of an event, assess the probability of what will follow. For Meyer, musical expectations (a term he later supplanted by the term *implications*) contribute greatly to emotive response to music.

However, with the exception of Hasty (1997) who describes the projective implications

of metric pulse, these theories do not describe the experience of listening as it unfolds in real time.

Jackendoff (1992) addresses real-time musical parsing and proposes three models of parsing that provide means of examining the interpretive choices a listener makes for each musical event she hears. The numerous analytical indeterminacies that arise make this a problematic endeavor. The author proposes that the most likely model is one in which all analyses are performed simultaneously. This requires that all possible interpretations that persist above a 'threshold of plausibility' remain active. The author proposes that a selection function is applied to continuously compare the plausibility of active interpretations giving preference to the more salient. If, as the author suggests, multiple interpretations remain active it seems likely that competing interpretations not only exist in parallel but also influence one another.

In this paper we introduce a theory of musical listening that categorizes and evaluates the percept of the listener in terms of the expectations she formulates and the degree to which these expectations are realized. We provide an account of a listener's real-time experience while listening to tonal music.

2. The Musical Problem

2. 1 Regularity and irregularity

In considering musical expectations we are concerned with how a listener perceives and processes the musical stream of events as she hears it. Particularly, we seek to distinguish between regular and irregular musical events in the context of the listener's prior experiences.

In order to describe percepts of regularity and irregularity we return to fig. 1. We focus on three features of this brief excerpt, the change in pattern in measures 4-6, the harmony in measure 10, and the silence in measure 11.

2.2 measures 4-6

By the time a listener hears the upbeat to measure four there is a strong and quite specific expectation that the leading tone will resolve, the tonic will arrive, and that the phrase will end on the following beat.

And so it does. However the embellishing dominant harmony of the following beat (measure 4, beat 2) disrupts the recurring pattern of change in the preceding measures creating a sense of surprise in the listener.

The second beat of measure four implies continuity. This is a surprise which can often directly identify the ambiguity. In this case it subverts the realization of the preceding beat but does not create a single specific alternative. It is only at the next metrically stressed position that the surprise of the preceding two beats is interpreted as instantiating an elided phrase (see fig. 2).

The surprise here occurs in the context of expected regularity. The *temporal shift*, that is, the occurrence of familiar chords in sequentially familiar positions but shifted to occur at an unexpected point in time is a common technique of surprise in music. Regularity provides the context for Haydn's 'risk taking'. By subverting the context of 'familiar' the composer can elicit surprise.

<insert fig. 2 - Elision of phrase in m4-6 of ex 1>

2.3 measures 10-11

The first inversion B major chord in measure 10 constitutes one of Haydn's more risky moments. Measure 10 parallels measure 3, and, with the exception of the embellishing chromatic lower neighbor in measure 9 that replaces its diatonic counterpart in measure 2, there is no warning that the dominant will be replaced. In the context of the music that immediately precedes it the event is syntactically rare (in the context of C major that is) and creates a harsh affect. The dissonant cross relations that arise between measures 9 and 10, and the subito forte marking make this surprise all the more pronounced.

The affect of this shock is augmented in the silence that follows. The silence is surprising in that it joins measure 10 to subvert an expected cadence. By isolating measure 10 with silence, and subsequently subverting metric expectations by prolonging the silence (by placing a fermata over the rest) Haydn simultaneously contradicts multiple expectations.

The contextualized silence of measure 11 serves as an example of the dynamic nature of expectations and realizations. The silence draws attention to the expectation that precedes it, and thus causes us to ponder retrospectively. The silence however turns vague as it extends beyond its expected duration demanding that the listener attend to the future. At this event a single silence forces the listener not only to expect the future but to attend to the past.

Before it turns ambiguous, the rest's surprise is attributable to the failure of the strongly expected resolution of the leading tone to materialize.

2.4 The need for theoretical constructs to describe and evaluate expectations

The verbosity of the above description of such risks underscores the need for a theory of expectations. The characteristics of strength and specificity of an expectation, and a measure of the degree of correspondence between expectation and realization serve as the primary tools of the theory. These characteristics constitute the sensations of different amounts of ambiguity, and vagueness, and the corresponding degree of surprise. This goal is problematic due to the obscurity of processes that originate in the listener's mind during the audition to music. A description of these processes must capture the complexities of dynamically changing contexts.

A listener constantly imagines continuations and constructs predictions for how the music she hears will proceed. In this sense, as is the case with performance, active listening is an instance of non-transcribed composition.

The composer relies upon the listener's facility to build expectations and manipulates audience attention and emotional response by satisfying or denying the realization of these expectations.

Tracking these processes may add to our understanding of the intentions and choices made by the composer. We hope to gain insight into these fundamental issues by introducing a representation that visualizes the context in which expectations are created, the attention that they draw, and the way in which they are realized or subverted.

Given the complex dynamic nature of listening, the multi dimensionality and dependencies of music, the high level of abstraction and vagueries of the phenomena it is not

surprising that there is a lack of adequate tools with which to construct an analytical theory of musical listening. To this goal a computational model using recurrent neural networks (RNN) is developed.

Computational models provide the ability to address multi dimensionality, abstraction, and complex interdependencies. Neural networks provide a means of capturing processes which are difficult to formulate by rules. The success of the model's behavior is used to validate intuitions about perception, and to refine the theory.

A neural network architecture that contains recurrent connections facilitates the incorporation of the past history of the sequence. In this way a dynamic context is built which contains the necessary information for learning the temporal correlation of data involved in real time musical processes.

Because of this the RNN is suitable for modeling temporal processes. Such a model can provide a visual representation of dynamic processes of formulation and realization of expectations. The visualization provides a means of classifying conditions under which expectations are formed, and to interpret listening states. The model also provides a means of qualifying and perhaps quantifying the correspondence between the listener’s expectation and the actual heard event.

3. Goals

We propose a general theoretical model of music cognition that describes states and attributes of a listener’s formulation and realization of expectations.

To establish the theory we introduce a method of identifying and describing listening states

when regular patterns are disrupted. This method is implemented by a predictive model that provides the means for qualitative and quantitative assessment of these states. Errors in the model's prediction designate different perceptual states of expectation formulation and realization. These include *ambiguity*, *vagueness*, and *surprise*. The model simulates a listener's real-time interpretation of normative and disruptive musical events. The input to the model is a representation of tonal harmonic events with associated metric positions.

The model serves to describe the contexts in which expectations are formulated and the affect of an expectation's associated realization. In normative situations there is a high degree of correspondence between expectation and realization. The model's predictive error will therefore be small. When the predictive error is large we designate the musical affect to be a surprise. We call the quantification of error the degree of realized expectation (DRE).

In the model harmonic events represented by a vector of pitch classes serve as musical input. The model's predictions (i.e., the listener's expectations) are represented in terms of activation strengths of each vector element. Since the activations represent individual strengths of each element the expectation can be evaluated in terms of which pitch classes are expected, and how strongly they are expected. The functional tonal idiom ascribes specific meaning to particular triadic combinations of pitch classes. The degree to which these combinations can be interpreted as harmonic events defines the amount of *specificity* of expectation. If the interpretation is singular the expected event is considered to be specific. We thus arrive at metrics for describing *strength* and the *specificity* of expectation. The DRE reflects the degree of correspondence between the vector of the expectation and that of the actual sounded event.

Thus the theoretical goal of designating and specifying various conditions and states of expectations and their associated realizations are inextricably bound to the model that serves to visualize these concepts and provide means to qualify and (to a degree) quantify these situations.

The specific goal of this study is to build and refine the model and evaluate its results using abstractions of actual musical excerpts.

4. **Computational Approach**

The computational model deals with a highly reductive representation of short segments of diatonic functional tonal music. We consider diatonic tonal harmonic progressions and their associated metric placement (harmonic rhythm) as a reduced yet meaningful representation of music audition. Harmonic progressions are comprised of hierarchical and functionally related events that generate expectations of what events can follow others. However, a listener not only predicts *what* will occur next, but also, *when* it will occur. Listeners use simple periodic patterns to organize the temporal dimension. 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 harmonic expectations are mutually influential. Thus, although the model uses a greatly simplified representation of music the interactions of meter and harmonic rhythm constitute complex behaviors as a consequence of their mutual interdependence.

The Jordan sequential network (Jordan, 86) provides an appropriate modeling method to cope with real-time musical processes. The Jordan sequential net is a simple form of a

RNN. It is a feed forward architecture with feedback connections between the output and a pool of units in the input layer. The feedback connections have fixed values (i.e., they do not learn). This simple modification enables the net to retain the back propagation learning algorithm (Rumelhart, Hinton & Williams, 1986) in dealing with temporal processes.

A pool of units in the input-layer represents the context of the sequence. The context at time t+1 is the accrued history of the sequence up to time t, plus the predicted event or the target event at time t. The history decays exponentially according to an empirically established decay parameter.

The Jordan sequential network provides the ability to learn sequences of musical events (in our case metered harmonic progressions) and to establish contexts within which predictions for the harmonies and their associated metric positions are formed. We interpret these predictions as harmonic expectations and cognizance of metric schema. Todd (1991) demonstrated the Jordan sequential net's ability to generate original melodies whose properties are extrapolated from a set of learned melodies. One of the authors (Gang, Lehmann & Wagner, 1998) used a modular approach in building Jordan sequential nets to realize real-time harmonization of melodies.

To cope with the mutual interdependencies of meter and harmony we develop a modular approach to building the Jordan sequential net (Berger & Gang, 1996; Gang & Beger, 1996; Gang & Berger 1997). Sub-nets of meter and harmony are integrated together and connected by a common hidden layer. Each sub-net has its own decay parameter and a unique strategy for updating its context. The metric context is fed by the output by way of the feedback connections and the harmonic context is fed by the actual heard event. The

integration models the mutual influences of meter and harmony. This modular approach enables the model to deal with more complex tasks than the simple Jordan sequential net.

5. Design and architecture

5.1 Corpus

The corpus consists of fifty functional tonal progressions, all in major keys. These patterns were evenly divided into quadruple and triple metered progressions each four complete measures with no upbeat. The triple metered progressions were padded with zeros at the end so that all examples contain sixteen events. Harmonic rhythm in the corpus ranges from one chord per beat to one chord per measure, with most of the examples having one or two chord changes per measure. The examples are typical perfect authentic cadential formulae commonly found in late Eighteenth century music and in contemporary harmony text books.

The learning set consists of forty examples randomly selected from the corpus. The learning set should reflect the statistical distribution of structural properties and regularities that constitute idiomatic 'norms'.

The generalization set comprises the remaining ten examples from the corpus. These examples were used to tune the net as explained in section 6.3.

In addition to the examples taken from the corpus, three additional examples were added to the generalization set. These examples were excerpted from works by Haydn and are described in section 7.

5.2 Architecture and Representation

The computational 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 normalized pitch class (pc). A general view of this architecture is presented in fig. 3. 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. The model integrates two sub-networks that represent distinct yet mutually-influential entities.

These entities, harmony and meter, are intertwined in a complex manner and combine in the hidden units to formulate context. The mutual influence of these contextual entities are established and learned during the course of formulating corresponding harmonic and metric predictions.

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. This 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. In the case of the pc units the context is updated with the target instead of 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 event and not the expectations. We are currently investigating the implications of incorporating

harmonic expectations into the context. Although this would seem to be a cognitively relevant approach we found that it does not affect the performance of the model in predicting element t for context t. This would, however, be a useful approach when trying to predict the element at time t+n for n>0. In this case the expectations would serve to bridge the temporal gap between element t and element t+n and would constitute an imagined musical context.

As a result of these considerations update rules for harmony and meter are derived as follows:

For harmony:

$$context_h(t+1) = decay_h * context_h(t) + target_h(t)$$

For meter:

$$context_m(t+1) = decay_m * context_m(t) + Output_m(t)$$

Both $decay_h$ and $decay_m$ are bounded real numbers between 0 and 1. In this way, the context of the harmony ($context_m$) and the context of meter ($context_m$) at time t+1 are iteratively built and are, respectively, an exponential decay of the history of the harmony and meter. The decay parameters are found to be important in terms of the quality of the model's performance and were empirically drawn from the tuning process described in section 6.3.

The metric pool of units are fully connected to the hidden layer together with the pool of pcs, implementing the integration of the mutual influences of meter and harmony. The hidden units are fully connected to the output layer (see fig. 3).

<insert fig. 3: the network architecture>

6. Running the network

6. 1 Learning

In the model expectations are not directly learned but rather emergent properties of the process of learning specific harmonic progressions. Bharucha & Todd (1991) suggested using harmonic progressions to capture schematic expectations using a three layer back propagation neural network. In the learning phase the network is trained with the forty harmonic progressions of the learning set. Each harmony of each progression is sequentially fed into the harmonic context thus simulating real time listening. The disparities between each actual output and the corresponding target of the harmony and metric index is computed. These errors were used to derive the iterative process of setting the weights of the net.

6.2 Generalization

In the generalization phase the net is introduced to the ten new metered harmonic progressions of the generalization set. 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 prediction 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.

The network was trained and tuned with normative examples and thus successfully

generalizes in normative situations. When a new or noisy situation is introduced to the network, the model retains much of its capability of meaningful generalization. To examine the model's response to non-normative situations two of the musical examples that augmented the generalization set included the musical 'problems' described in sections 2.2 and 2.3.

6.3 Tuning the network

Decay parameters and the number of hidden units used in the net were arrived at by using the ten normative generalization examples mentioned in section 5.1. In principle, the network should be able to predict the examples with relative accuracy since these examples conform to the idiomatic norms of the learning set, that is, they contain the most frequently recurring chords and their associated metric positions.

We search for the architecture that will produce the least error by adjusting the number of hidden units and decay parameters. The error between the actual output and the target is computed as their mean square difference. In some states the model reflects the presence of multiple chords at a particular state. This ambiguity can occur both in normative and disruptive situations. If the context is too short to provide a single definition for its continuation ambiguity will arise. Thus, there will be an error in situations in which the idiom allows for greater variety.

Using this strategy, we first performed a wide search for harmony and meter decay parameters and the number of hidden units. We evaluate the net's performance by computing the net's error produced from the ten normative examples from the generalization set. We validate the performance by judging the results according to musical criteria. The aim of the wide search was to prune the broad parameter space. This left a more restrictive set of values to be

evaluated in terms of optimal performance.

The procedure for producing a result from the net is a cycle composed of four stages: training on the learning set; reproducing one of the learning states; introducing the new patterns from the generalization set, and evaluating the results. The quality of this result is a function of the initial weights set with small random values at the beginning of the learning process. Each run cycle will produce a varied result.

To statistically infer the quality of performance of a specific architecture the network is run ten times. (In statistical terms, the quality of performance is the random variable). The error of each run computed by the mean square difference between the outputs and the targets provides one sample of the random variable. The average and standard deviation of the samples of each random variable are computed. As expected, a significant difference between the average of the performance of different architectures was found to correspond to judgement according to musical criteria. The best architectures which produce the least error directs our search for the optimal model which was, n the end, determined by musical judgement.

7. Experiments

Three excerpts from two works by Haydn were used in the experiment phase. These include: the opening four measures of the second movement of the F major string quartet op. 3 no. 5, Andante cantabile (see fig.4), the opening five measures of the C major piano sonata, H XVI/60, third movement (Allegro molto), and measures 8 11 of H. XVI/60: III.

<insert fig. 4 -Haydn String quartet>

These examples were selected to represent a diversity of types of expectations and realizations. They were chosen to examine how the computational model represents these cognitive processes when normative and disruptive musical situations are encountered in real-time.

The model visualizes cognitive listening processes by providing a graphical representation of harmonic expectations and metric interpretation (fig. 5). The output of the model is represented as four components of sixteen successive columns. These sixteen columns represent the network's predictions at sixteen discrete time steps. Each time step is associated with a single musical event defined by the beat level resolution. Hence, to interpret fig. 5 consider each event as a quarter note unit.

In the graphical visualization the two top components represents pitch class members of the target and the predicted harmony, respectively. Below this are two components representing the model's inference of metric organization in terms of triple and quadruple meters. Within each component are squares of varying sizes. Each square represents an activation of the pitch or metric vectors of the output. The larger the square, the stronger the activation.

<put fig.5 figure net output of Haydn-normative>

7.1 Haydn: String quartet, F major, op. 3, no. 5, second mvt., mm. 1-4.

The first event, representing the first beat shows strong activations for pcs [0 4 7]. Since the entire corpus commences on a tonic downbeat there is no ambiguity in the net's initial harmonic prediction. There is, however, ambiguity in the model's metric interpretation as evident in the activations of down beats in both metric pools. This expected behavior is the result of training with an evenly divided number of quadruple and triple examples in the corpus.

This metric ambiguity will persist until there is a change in harmony. Since all of the examples in the corpus have a single tonic harmony in the opening measure, and none of the examples has a harmony that crosses over a measure boundary, the model is keenly sensitive to the correlation between harmonic rate of change and metric inference. The fourth beat is thus a critical event information in terms of metric and harmonic prediction. If the model would unambiguously infer a triple meter schema for the first three events, then it should predict a harmonic change at beat four. However, since the model is undecided as to a metric interpretation, it activates pcs [0,4,7,9] with a weak activation of pc[2]. These activations conform simultaneously to the anticipation of harmonic change in the case of a down beat in triple meter and to the anticipation of continuation of the tonic in the case of a fourth beat in quadruple meter.

Since the target (i.e., actual sounded event) at beat four is a tonic triad this squelches any interpretation of triple meter from beat five. Since the model now has sufficient information to contextualize quadruple meter, it predicts a change to a subdominant at beat

five. A subtle harmonic ambiguity is evident here in weak activations of pcs [4] and [5] leaving the possible interpretations of vi or IV. This ambiguity disappears in the following weak beat where the model correctly predicts IV. However in the following strong beat the weak activation of pc[2] suggests the possibility of a change to the supertonic.

For the rest of the example the predictions are strong, specific and consistently correct[2]. Beats nine and eleven display weak activations adding possible interpretations for a supertonic harmony (beat nine) and an added seventh (beat eleven), to the strong anticipation of the dominant.

With a normative example the model is expected to successfully and specifically predict the harmonic events in their correct sequence and temporal position. The model is also expected to resolve the initial metric ambiguity within a few preliminary time steps. In the event that ambiguities arise the model reflects the statistical distribution of the learned examples of the corpus. In the case of a slightly erroneous prediction the model updates its context and expected to correct its prediction for the next event.

The ambiguities and errors discussed in regards to fig. 5 do not reflect overt musical surprise. They do however reflect the interpretive decisions and choices a listener makes when hearing music. Even the most 'normative' music demands: relying on context to build expectations; relying on the correlation between expectation and realization to provide cues needed to establish a metric framework; and, conversely, relying on the metric interpretation to influence expectations.

7.2 Haydn: C major piano sonata, H XVI/60, third movement (Allegro molto) mm. 1-4.

<insert fig. 6 net output of Haydn - surprise 1 (3/4)>

Next we return to the challenging musical example that opens this paper. The perceptual affect of the metric shift described in section 2.2 is clearly visualized in fig. 6. Here the model reacts to the unanticipated dominant on beat eleven. The unrealized strong and specific prediction for a tonic in beat eleven represents a surprise. In section 7.5 we describe a means of quantifying the amount of surprise based upon the dissimilarity between the prediction and the actual heard event.

The unusual placement of V_7 at beat eleven updates the context with an irregularity that will influence the prediction of beat twelve. This is evident in the lack of strong and specific prediction for a tonic in the final beat.

In fig. 6 the disruption of harmonic rhythm produces a surprise represented by the inability to predict the event of beat eleven and the weak and unspecific activations at beat twelve. Nevertheless, the metric inference remains strong and specific as the interpretation of triple meter is unwavering. In section 7.3 we describe another model in which surprise is reflected in the disruption of metric regularity as well as in the prediction of harmony.

7.3 Haydn: C major piano sonata, H XVI/60, third movement (Allegro molto) m. 8-11

< insert fig. 7 net output of Haydn - surprise 2 (3/4) >

After disrupting the opening phrase with an internal digression Haydn gives the listener reason to believe that, from measure 9, a repetition of the opening measures at the octave will bring the music back to the regularity established by the beginning. As at the opening, the established regularity creates a strong expectation that the cadential dominant will arrive at the downbeat of measure 10. However Haydn surprises his audience at this point with an unprepared and uncontextualized B major chord.

Fig. 7 visualizes the neural network's output when fed this example. The return to regularity described above implies that the listener has an already established sense of temporal correlations at the levels of beat, meter and phrase.

To account for this, an extended model would need to be introduced to longer musical examples to learn long term temporal correlations. The fact that in measure 9 the metric schema is already established in the listener's mind can be implemented by biasing the metric interpretation of the model. We demonstrate the use of bias in section 7.4.

The B major chord is rare in that it is not in the lexicon of the purely diatonic corpus. Because of this it provides an irregular context to the model. The model persistently predicts G major for all three beats of the measure even though the context is updated by each repetition of the B major triad. This constitutes a surprise that results from the disparity between the prediction and the actual heard event.

The irregular harmonic events do not immediately affect the metric interpretation. However, after three successive updates of the context, the irregularity results in metric ambiguity at beat ten.

Haydn follows this shocking surprise with another - a prolonged silence. Correspondingly the model updates the context with a vector of zeros representing silence resulting in irregularity of metric interpretation. Despite the target's vector of zeros, and the erratic metric interpretation the model persists in strong and specific expectations for a dominant harmony.

The behavior of this model reflects characteristics that result from numerous parameters including random weights, decay parameters, number of hidden units, the content of the learning examples and the order in which they are introduced. Training two identical networks each with differing initial random weights will result in unique dynamic behaviors. For each trial run the net will learn different characteristic properties of the data. Thus, one trial may appear to focus upon one conception of regularity while another may reflect others. In both cases, however, these different results may each reflect coherent behaviors.

< insert fig. 8 Haydn, surprise version 2>

To illustrate these differences, fig. 8 presents the results of the same musical example in a different trial. In contrast to the previous results, this model reflects the surprises of the B major chord and the ensuing silence of beats seven through twelve differently. In this case, the surprise is reflected both in the harmonic expectations and metric interpretations.

A cognitive model of listening must, by nature of its task, assume broad generalizations about listeners. The 'idealized listener' denies individualities of each listening experience whether by a single listener rehearing a piece or a larger audience experiencing the same music. As demonstrated above, multiple trials or different sets of values for the learning parameters can be used to simulate multiple listenings, be they repeated hearings by a single listener, or listenings by multiple listeners.

7.4 Haydn: C major piano sonata, H XVI/60, third movement (Allegro molto) mm. 1-2. - Metric bias for 4/4

< insert fig.. 9(a b & c): net's output of Haydn, the bias varsion>

In the final example we return to the opening two measures of fig. 1 and present three models reflecting the affect of premonitory metric bias on expectations.

In fig. 9a the network produces activations in both meter pools in the first event. This reflects the fact that the corpus is evenly divided between triple and quadruple examples that are introduced to the net in random order during the training phase. The metric ambiguity of the opening events persists until the harmonic change at beat four. Since all the examples in the corpus commence with a complete measure of tonic harmony the prediction for pcs [0 4 7] remains strong and specific until beat four. In beat four the metric ambiguity posits a downbeat in 3/4 against a weak beat in 4/4. The triple meter interpretation influences the

harmonic prediction for change to a subdominant harmony, while the interpretation of 4/4 is reflected in the continued prediction for the tonic. The actual heard event is a supertonic harmony. The updated context extinguishes the quadruple interpretation.

In fig. 9b the model reflects a strong bias towards triple meter. There is no metric ambiguity. This results in a different prediction for beat four. Here, the expectation for a move away from the tonic is evident in the absence of activation for pc[7]. Instead the activations of pcs[0 4 5 9] suggest expectations of vi or IV.

In fig. 9c the network strongly and specifically interprets the opening four beats of the music in $^{4}/_{4}$ meter. The model's strong and specific metric interpretation influences a similarly strong prediction for a tonic in beat four. There is only a small activation of pc[9] corresponding to the harmonic change that would reflect triple meter. The change to the super-tonic in beat four causes the context to shift the metric interpretation to triple meter. From this point on the prediction is strong, specific, and consistently correct.

The premonitory metric bias visualized by the model is a result of the order of training during the learning phase. When the corpus is trained with a randomly ordered set of training examples the network does not reflect significant bias for one interpretation over the other. However, by training the network first with the set of examples in one meter followed by its complimentary set in the second meter, we can bias the network's initial metric prediction towards the later. The affect of this bias on the model's consequential prediction is discussed in detail in (Berger & Gang, 1998).

The cognitive implication of bias is important. If a listener has strong metric expectations at the start of a piece it can affect the way expectations are built. Premonitory metric bias can

result from extra-opus experiences including title (e.g. minuet), context (the second movement of a string quartet), genre (relatively few popular tunes are in 3/4) and from other factors.

Assuming, as we do, that performance cues are not supporting initial interpretations, (as for example in the situation described here, in which the harmonic rhythm is static within the first measure) there is nothing beside premonitory bias to assist in interpretation. For this reason beat four in (the unbiased) fig. 5 is ambiguous, while beat four in (biased) fig. 9 is a surprise.

7.5 Measuring the degree of realized expectations (DRE)

In predictive models the error between the actual and the desired outputs might be used in a meaningful way to demonstrate some of the model's properties and temporal correlations of the data. In natural language processing the sequential positions of errors produced by recurrent neural network models have, for example, been used to distinguish boundaries between linguistic units (i.e.., distinguishing word endings from a stream of letters). By graphing the error over time, cues for segmentation can be inferred. Onsets of new events will produce high error. In the case of word boundary segmentation, for example, these errors will subdue over time as the predictability of the sequence increases as more letters of the word are revealed. The decrease in the error marks the end of the distinguished word (Elman, 1990).

Following this idea we calculate the errors produced by the model for each output and the correspondent desired output. To this aim we defined distance functions to compute different

aspects of correspondence.

Fig. 10a contains a graph of the DRE for mm. 1-4 of fig 1. The surprise of measure 4 beat 2 was replaced by a tonic in order to provide a normative example. Fig. 10b is a similar graph of mm. 8-11 of fig. 1 (corresponding to fig. 7). The DRE is computed by the square of the differences between the target and the output.

The descriptions of section 7.1-7.4 provide qualitative insights regarding expectations and their corresponding realizations. Measuring the DRE suggests a quantitative description of these insights. Refinement of meaningful methods to compute the DRE demand validation by other experimental data and consideration of other distance functions.

<insert: fig. 10a and 10b. - DRE)

8. Conclusions and further directions

Strunk (1933) expresses the rhetorical framework for manipulating expectations:
"Only when the roles of the game are well established is it feasible for the composer to play upon the expectation of his listener. And even then, to play on expectation he must first arouse it. To secure emphasis he must first exercise self control. He cannot afford to be continually surprising to his listener. He must be simple before he is complex, regular before he is irregular, straightforward before he is startling.".

Haydn's risk taking involved experiments with manipulating various types and degrees of musical surprise. That these surprises have been felt and appreciated by Haydn's audience

goes without saying. However, traditional analytical techniques lack the tools and terminology to qualify and quantify these basic musical affects.

In this paper we addressed these issues by proposing a computational model that simulates some of the cognitive processes involved in musical listening. Specifically, we examined the interplay of metric inference and functional tonal harmony. The model visualizes the mutual influence of these musical characteristics and provides visualizations of normative and disruptive listening situations. The RNN model provides an approach to visualization of expectations and their associated realizations. Various phenomena and situations relating to musical expectations can be identified with these visualizations. The model offers methods to describe qualitative attributes of expectations and realizations. The interpretation of error in the predictive model suggests methods of quantifying attributes of expectation and the degree of its realization.

Our work suggests a number of cognitive implications. Musical listening involves sub-symbolic learning through experience. Acquiring metric and harmonic schemas are an emergent property of a listener's exposure to metered tonal harmonic progressions. Harmony is perceived as part of the musical stream while meter is inferred from it. Harmony and meter interact with one another to produce tonal harmonic expectations and metric interpretations. Harmony and meter are mutually influential in creating a combined context for prediction. From these predictions interpretation of the metric schema and harmonic expectations are formulated. Expectations can be described in terms of strength and specificity. The affect of these attributes result in specific, ambiguous and vague expectations. Under normative situations expectations are realized while in irregular situations there is a difference between

the expectation and the actual heard events. The degree to which the expectation is realized (DRE) corresponds to the affect of surprise.

The computational model can be refined and expanded in a number of ways. These include the incorporation of a more extensive corpus, the use of longer musical examples, a richer and more encompassing representation of music, and further investigation of distance functions as a representative measurement of DRE. Empirical findings from psychological experiments with human subjects can provide data to validate the performance of the model. Probe tone studies of expectations for functional tonal harmonic progressions (Schmuckler, 1989), (Krumhansl & Shepard, 1979) and ERP studies (Patel, Gibson, Ratner, Besson & Holcomb, 1996), and (Cohen & Erez, 1991) provide some preliminary approaches for validation.

The ability to simulate temporal processes makes the Jordan sequential neural network an appropriate foundation for modeling musical listening. The added modularity of sub-nets to the architecture provide a framework for modeling the mutual influence of heard music and inferred meter (Gang & Berger, 1999). In general, the integration of sub-nets provides a natural means to extend the model by adding more musical parameters. A more sophisticated model should allow for expectations for occurrences farther in the future than the next event. Such a model should also be able to retrospectively re-evaluate contexts after surprise.

Using a RNN approach different models can be derived by changing learning parameters such as the number of hidden units, decay parameters, or different trials of the same architecture. Every model discovers different types of patterns and regularity and generates a coherent solution for a task. This is potentially problematic in terms of what model to chose

and what constitutes the more appropriate model for a given task. However, an ensemble of networks each trying to solve a problem (Sharkey, 1999) can merge the individual attributes of single networks to provide a flexible modeling methodology. An ensemble of neural nets can provide a fertile laboratory for studying differences between a single individual hearing of a musical work, multiple hearings of the same work, and group listening experiences.

How a listener reacts and adapts to music in real-time involves enormous complexity and subtlety. This study hopes to provide a modest first step towards understanding the processes involved, and suggest a computational approach with which to delve deeper into this domain.

9. References

Berger J. & 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*, Montreal.

Berger, J. & Gang D. (1998) A computational model of meter cognition during the audition of functional tonal music: Modeling a-priori bias in meter cognition. In *Proceedings of the International Computer Music Conference*, Ann Arbor.

Bharucha J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception.* 5:1, 1-30.

Bharucha, J. J. & Todd P. M. (1991). Modeling the perception of tonal structure with neural nets. In P. M. Todd, & D. G. Loy (Eds.) *Music and Connectionism*, Cambridg:MIT Press.

Cuddy, L. L. & Lunney, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Perception and Psychophysics*, 57, 451-462.

Cohen, D. & Erez, A. (1991). Event-related-potential measurements of cognitive cpomponents in response to pitch patterns. *Music Perception*, 8, 4. pp. 405-430.

Elman, J. L., (1990). Finding structure in time. *Cognitive Science*, 14, 179-211.

Gang, D. & Berger, J. (1996) Modeling the degree of realized expectation in functional tonal music: A study of perceptual and cognitive modeling using neural networks. In *Proceedings of the International Computer Music*

Conference, Hong Kong.

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

Gang, D. & Berger, J. (1999). A unified neurosymbolic model of the mutual influence of memory, context and prediction of time ordered sequential events during the audition of tonal music. In *Spring Symposium on Hybrid Systems and AI* , Stanford. AAAI.

Gang, D., Lehmann, D. & Wagner, N. (1998). Tuning a neural network for harmonizing melodies in real-time. In *Proceedings of the International Computer Music Conference*, Ann Arbor.

Gjerdingen, R. O. (1988). *A classic turn of phrase : music and the psychology of convention*. Philadelphia : University of Pennsylvania Press.

Hasty, C. (1997). *Meter as rhythm*. New York: Oxford University Press.

Jackendoff, R. (1992). Musical parsing and musical affect. In M. R. Jones & S. Holleran (Eds.) *Cognitive Bases of Musical Communication* (pp. 51-68). Columbus: American Psychological Association of The Ohio State University.

Jordan, M. I., (1986). Attractor dynamics and parallelism in a connectionist sequential machine. *Proceedings of The Eighth Annual Conference of the Cognitive Science Society*, New Jersey: Hillsdale.

Krumhansl, C. L., & Shepard, R. N. (1979). Quantification of the hierarchy of tonal unctions within a diatonic context. *Journal of Experimental Psychology:Human Perception and Performance*, 5, 579-594.

Lerdahl, F. & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge MA:MIT Press.

Meyer, L.B. (1973). *Explaining music: Essays and explorations*. Berkeley: University of California Press.

Narmour, E. (1990). The analysis and cognition of basic melodic structures: The implication-realization model. Chicago: University of Chicago Press.

Patel, A. D., Gibson, E., Ratner, J., Besson, M. & Holcomb, P. J. (1996) Processing grammatical relations in music and language: An event-related potential (ERP) study. In Proceedings of the Fourth International Conference on Music Perception and Cognition (pp.337-342). Montreal: McGill University, Faculty of Music.

Riemann, H. (1916). *Harmony simplified: The theory of the tonal functions of chords* (pp. 130). London:Augener,

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Vol. 1* (pp. 318-362). Cambridge: M.I.T. Press.

Schmuckler, M. A. (1989). Expectation in music: Investigation of melodic and harmonic processes. *Music Perception*, 7, 109-149.

Sharkey, A. J. C. (1999) (Ed.). *Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems*. London: Springer-Verlag.

Strunk, O. (1933). Haydn. In D. Ewen, (ed.), *From Bach to Stravinsky*, (pp 78- 82). New York: W.W. Norton.

Todd, P.M. (1991) A connectionist approach to algorithmic composition. In P. M. Todd & D. G. Loy (Eds.) *Music and Connectionism*, Cambridg:MIT Press.

Notes

1. We use the phrase ’Haydn’s audience’ in its broad sense. We assume the listener to have prior listening experience with Western diatonic functional tonal music. We make no assumptions about a listener’s musical literacy or training nor do we assume familiarity with any specific work of music including those used in the model. Issues relating to extra and intra-opus experience are cursorily touched on in this paper. More literal consideration of ’Haydn’s audience’ as opposed to our generalization are relevant and engaging but beyond the scope of this paper.

2. The opening of the slow movement of op. 3 no. 5 presents a four measure phrase that is normative in its progression and regular in its rate of harmonic change. Although the opening period is extended to six measures by the additional cadence of measures 5-6, the listener arrives at measure four with little sense of disruption or surprise. Overall the model produces strong and specific predictions that are correctly realized. The ambiguities that occur result either from a lack of context (metric ambivalence in first four beats, and harmonic ambiguity in beat four), or from occurring at points at which the statistical distribution of ’schematic norms’ is less clearly articulated (e.g. the downbeat of measure 3).

Fig. 1 Haydn: Piano sonata, C major, H. XVI/50, 3rd mvt., mm. 1-11

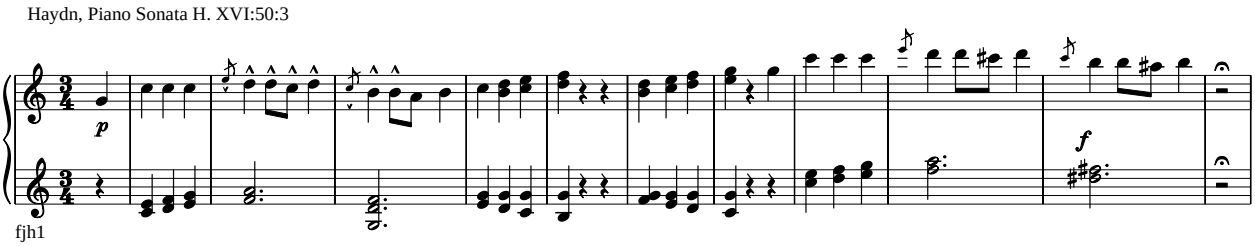

Fig. 2. Haydn: Piano sonata, C major, H. XVI/50, 3rd mvt., mm. 1-11, elided phrase

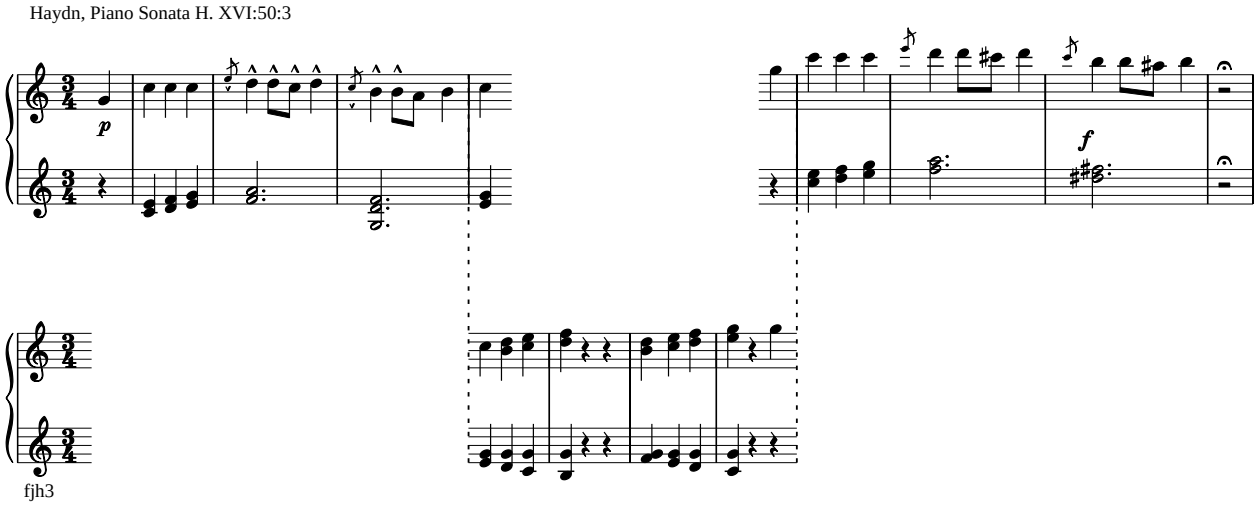

Fig. 3 Neural network architecture

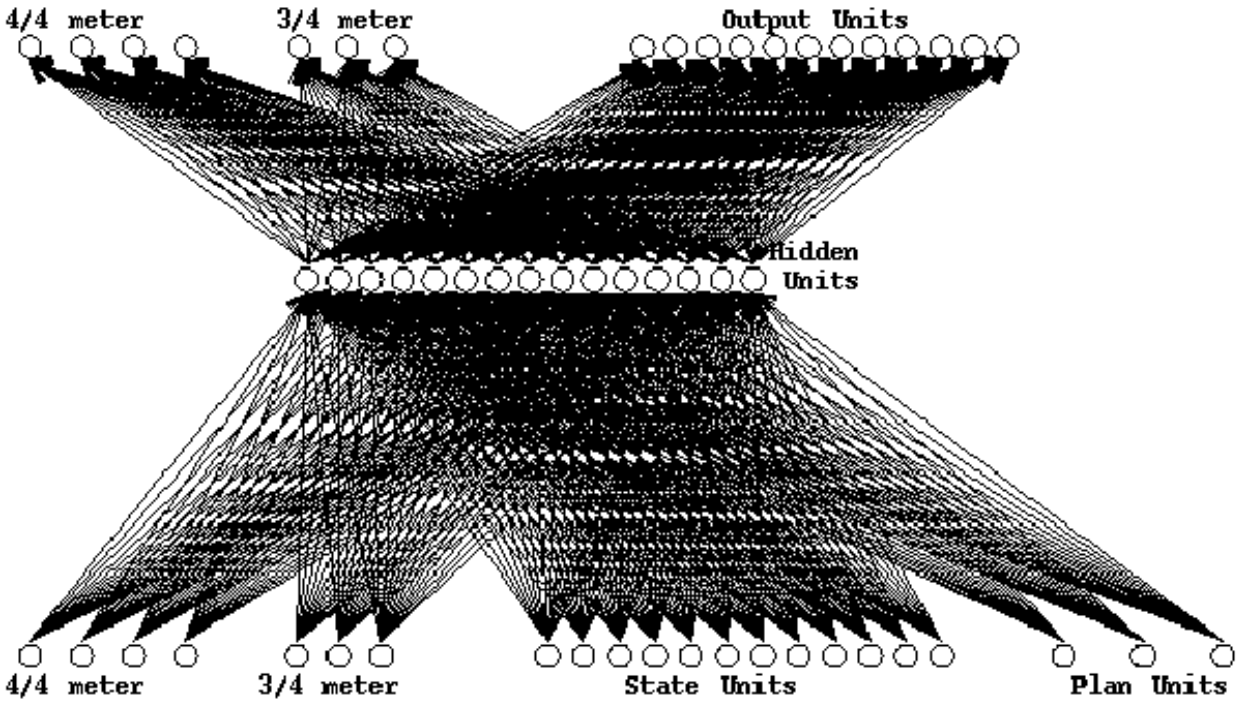

Fig. 4. Haydn: String quartet, op. 3, no. 5, second movement (Andante cantabile) mm. 1-4.

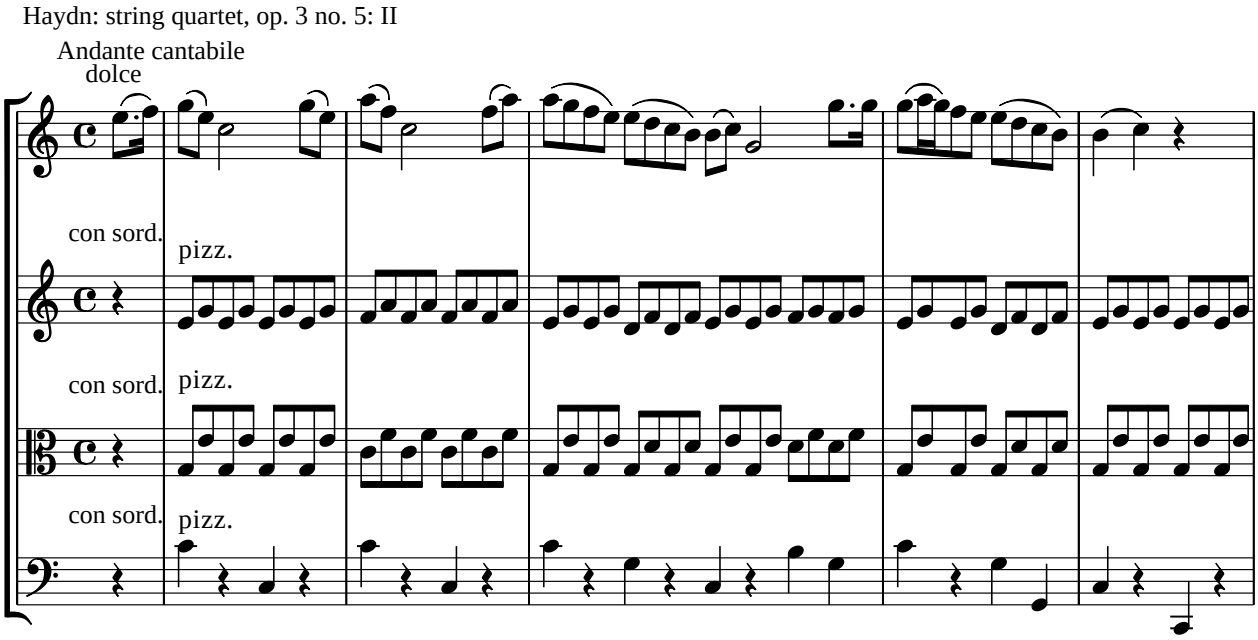

Fig. 5. Haydn: String quartet, op. 3, no. 5, second movement (Andante cantabile) mm. 1-4. Visualization of harmonic expectations and metric interpretation.

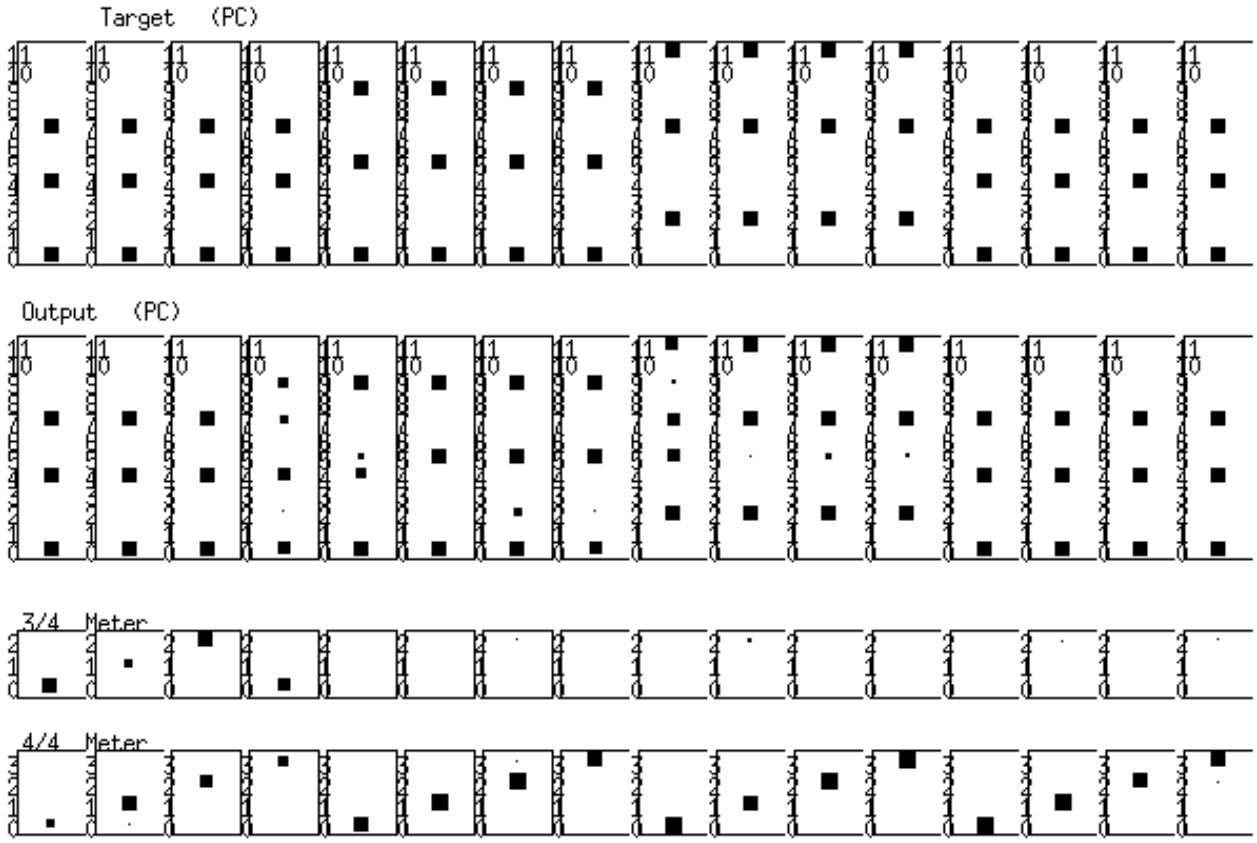

Fig 6. Haydn: C major piano sonata, H XVI/60, third movement (Allegro molto) mm. 1-4.

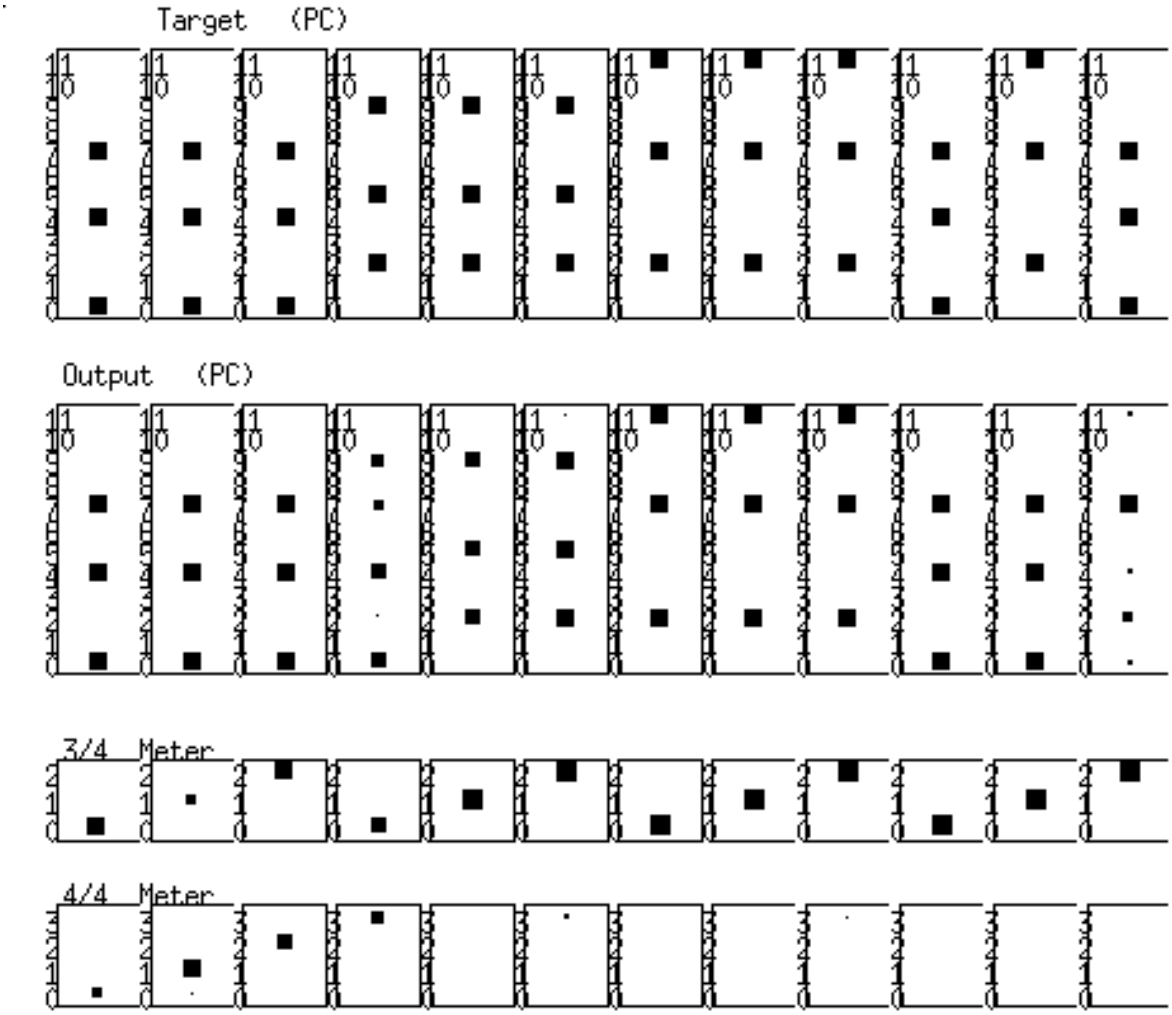

Fig 7. Haydn: C major piano sonata, H XVI/60, third movement (Allegro molto) mm. 1-4.

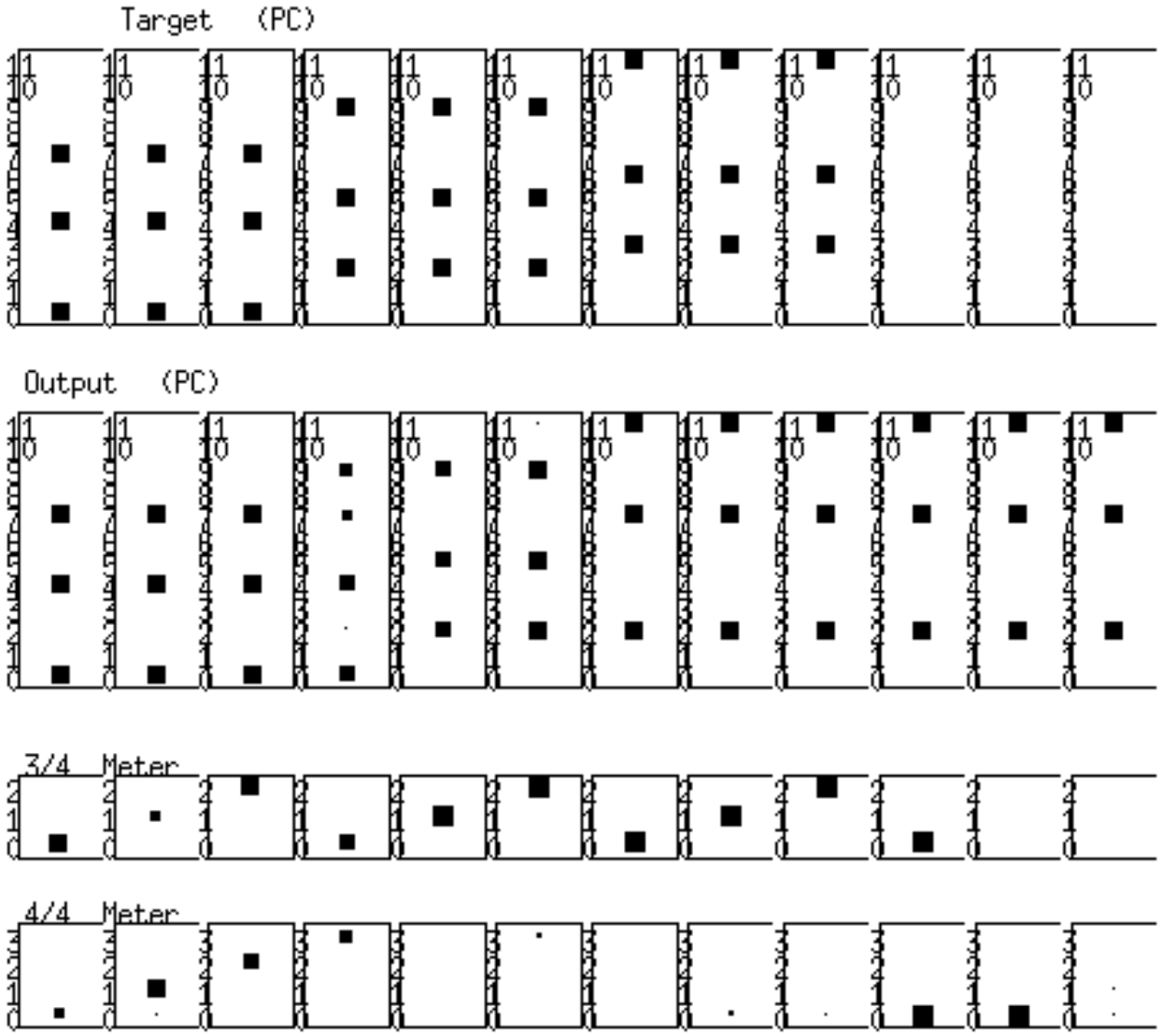

Fig 8. Haydn: C major piano sonata, H XVI/60, third movement (Allegro molto) mm. 8-11.

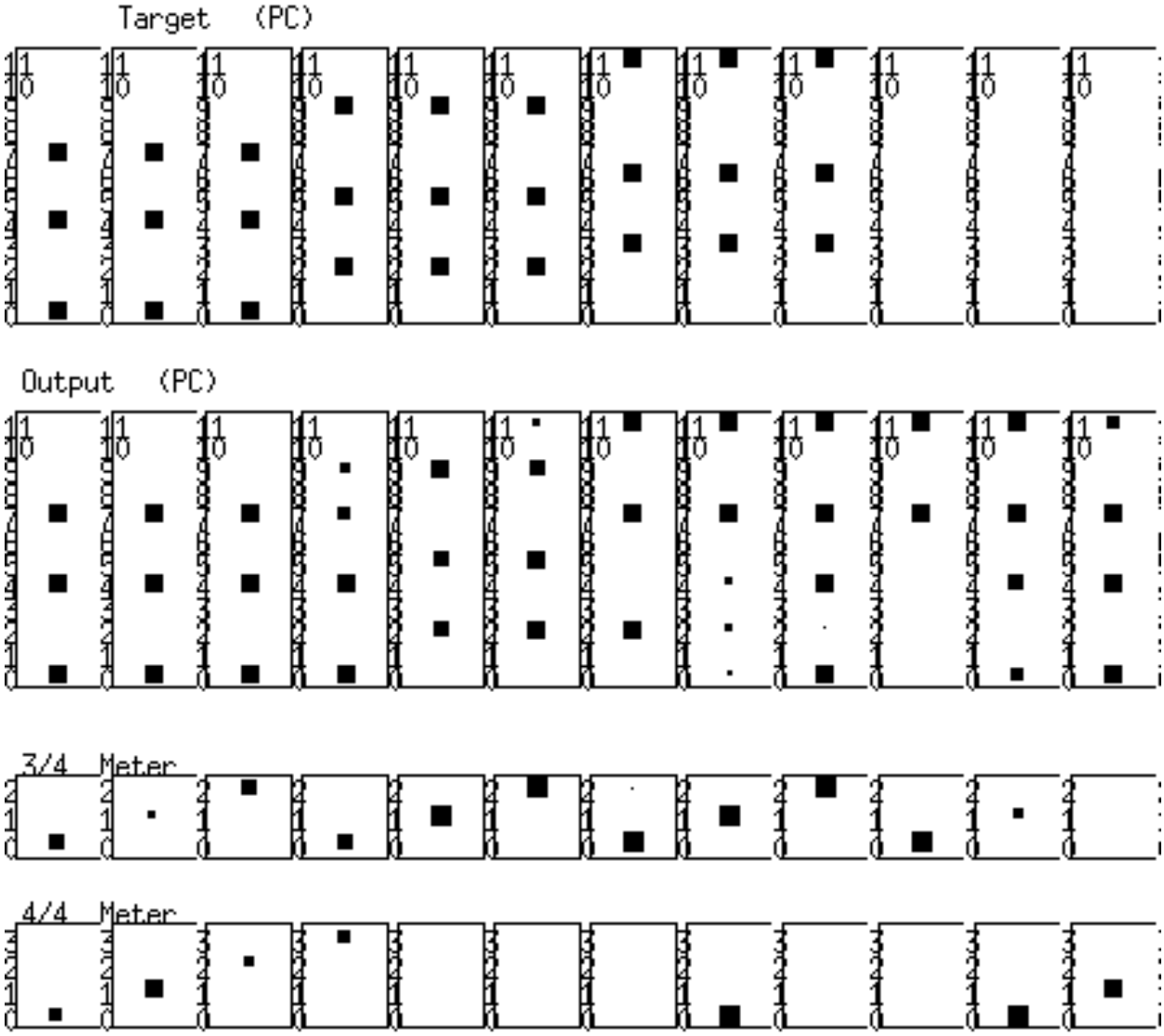

fig 9a. Haydn: C major piano sonata, H XVI/60, third movement (Allegro molto) mm. 1-2. - No metric bias.

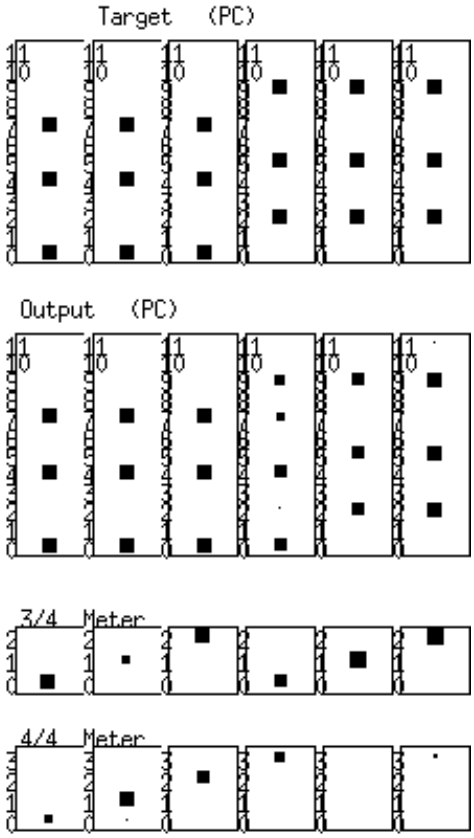

fig 9b. Haydn: C major piano sonata, H XVI/60, third movement (Allegro molto) mm. 1-2. - Metric bias for 3/4

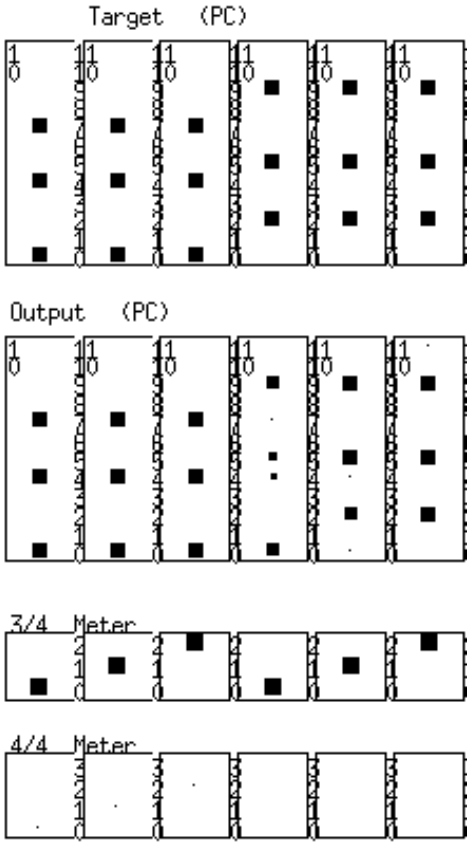

fig 9c. Haydn: C major piano sonata, H XVI/60, third movement (Allegro molto) mm. 1-2. - Metric bias for 4/4

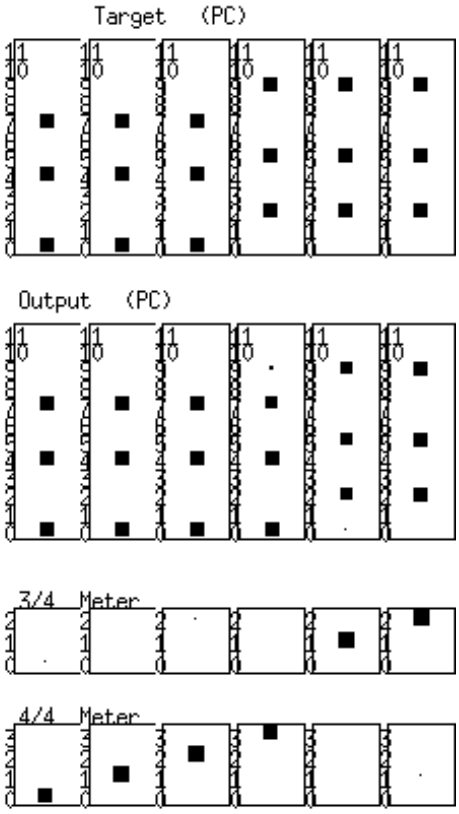

Fig 10. DRE of Haydn, H.XVI/50, 3rd mvt. mm, 1-4 and mm. 8-11