# Data Abstraction In Cognitive Models for Compositional Design in Music

Linda Dzuris [*] James Peterson [†]

[*]Department of Performing Arts, Brooks Center BOX 0525, Clemson University, Clemson SC 29634-0525, email: `ldzuris@clemson.edu`

[†]Department of Mathematical Sciences, Martin Hall BOX 0975, Clemson University, Clemson SC 29634-0975, email: `petersj@clemson.edu`

**Abstract**

In this paper, we discuss the design of musical data, which is neutral emotionally, for use in the training and building of models of cognition. The data will provide training examples for auditory cortex and emotion modules. There are two goals to this research: first, model the process of musical design for eventual use in the development of an autonomous musical composition program; and second, use the musical model as a quantitative means of training the auditory cortex portion and associated emotional circuits of a general model of cognition.

# 1 Introduction:

In this paper, we discuss the design of musical data for use in the training and building of models of cognition which consist of many parts such as cortex (auditory, visual and associative) and emotions. There are two goals to our research: the first, model the process of musical compositional design for eventual use in the development of an autonomous musical composition program; and the second, use the musical model as a quantitative means of training the auditory cortex portion and associated emotional circuits of a general model of cognition.

In Section 1.1, we show how pivotal experiments from psychophysiology give us valuable clues about how to organize data arising from visual images into emotional categories. From this, we infer which musical data sets are useful to generate and we show that this paper, which concentrates on the development of emotionally neutral data, is of central importance as it is the cornerstone around which the emotionally tagged data are built.

The complete description of how we generate the neutral musical data and rules which we used for our compositional elements are found in Section 2. The connection between grammars and music is discussed and motivated in Section 2.1. In 2.2, we discuss a rapid prototyping technique used in the $18^{th}$ century know as a Würfelspiel matrix, Cope (2001), and indicate, in general, how such matrices can be built. Finally, we show how to design such matrices of data in Section 2.3 and finish with samples of the generated compositional designs, Section 2.4.

In Section 3, we discuss how this work is placed within the more general area of cognitive modeling and how a model of musical composition provides us with a useful means of validating complicated and complex cognition models. We conclude this section with a short discussion of some of the clues and ideas from biology and cognitive science that partially shaped our thinking on how to build useful abstract models.

In Section 4 and 4.1, we lay out the basics of the cognitive modeling architectural design we use for modeling the music composition process. In essence, we need to capture why certain notes in a valid neutral composition are preferred over other choices. The discussion is based on the use of suitable encodings of the musical notes into mathematical forms useful for model building. Since we do not actually build these models in this paper, the discussion is necessarily brief. In our brains, pathways that are useful are often given an enhanced probability of use via a process called excitation; similarly, pathways of limited usefulness are actively discriminated against using a process called inhibition. In Section 4.1, we discuss an overview of the modeling process that uses inhibition and excitation to ensure that only valid notes are selected in our neutral music samples. Hence, we train our model so that it captures how notes are generated in the data faithfully. Then, section 4.2, delineates the procedure by which we construct the transitional mappings between opening, middle and closing phrases. Finally, in 4.3 and 4.4, we show how actual musical fragments can be generated by constructing the analog of sentences in this grammar. We close with a short discussion with the role the musical data plays within the context of a full cognitive model.

### 1.1 Emotion Models:

In a sequence of seminal papers, (Lang, Bradley, Fitzmmons, Cuthbert, Scott, Moulder & Nangia, 1998); (Codispotti, Bradley & Lang, 2001); and (Cuthbert, Bradley & Lang, 1996), it has been shown that people respond to emotionally tagged or affective images in a semi-quantitative manner. Human volunteers were shown various images and their physiological responses were recorded in two ways. One was a skin galvanic response and the other a fMRI parameter. Typical results are plotted in Figure 1. In this database of images, extreme images always generated a large plus or minus response while neutral images such as those of an infant generated null or near origin results.

**Figure 1 goes here**

If we followed the original intent and spirit of the Affective Image research, we would like to develop or generate Würfelspiel type arrays for each of nine primary emotional states as indicated in Figure 1. These would be emotional states that correspond to the following nine locations on the two dimensional grid:

| 2D Coordinates | Physiological Responses | Image Type |
|---|---|---|
| (High, High) | high galvanic and high fMRI response | Thrills |
| (High, Null) | high galvanic and flat fMRI response | |
| (High, Low) | high galvanic and low fMRI response | Murder |
| (Null, High) | flat galvanic and high fMRI response | |
| (Null, Null) | flat galvanic and flat fMRI response | |
| (Null, Low) | flat galvanic and low fMRI response | |
| (Low, High) | low galvanic and high fMRI response | Flowers |
| (Low, Null) | low galvanic and flat fMRI response | |
| (Low, Low) | low galvanic and low fMRI responses | Cemeteries |

Clearly, the emotional tags associated with the images in the affective image database are not cleanly separated into primary emotions such as anger, sadness and happiness. However, we can infer that the center (Null, Null) state is associated with images that have no emotional tag. Also, the images do cleanly map to distinct 2D locations on the grid when the emotional contents of the images differ. Hence, we will assume that if a database of images separated into states of anger, sadness, happiness and neutrality were presented to human subjects, we would see a similar separation of response. Our hypothetical response would be captured in the emotion triangle of Figure 2.

**Figure 2 goes here**

In both the musical and painting compositional domain, we will therefore design Würfelspiel matrices for the four positions marked in Figure 2. In addition, we can identify the intangibles *antagonistic* with the emotional attribute *anger*, *demoralized* with *sad* and *contented* with *happy*. We then use a similar triangle to design job scheduling data in that setting. The motivations and arguments that explain these mappings are explained in our job scheduling papers (Kurz & Peterson, 2003b and 2003c). In this paper, we will concentrate on how we generated the musical data that is interpreted as *neutral* in the emotion triangle.

# 2 Generating Neutral Musical Data:

We design musical data for these purposes using a grammatical approach to musical composition. This is, of course, simplistic, but we want a mechanism for quickly and repeatedly generating large volumes of data that can be used to train the auditory cortex of the cognitive model.

## 2.1 A Musical Grammar:

In the literature, there are many attempts to model musical compositional designs. Several theorists discuss music in terms of large chunks or sections and overall function. One example is found in Caplin (1998), "*Classical Form: A Theory of Formal Functions For The Instrumental Music of Haydn, Mozart and Beethoven*", there is a focus on larger musical structures such as those found in the symphony movements of Haydn, Mozart, and Beethoven. Most of the discussion is not relevant to the design of musical fragments of the type we desire, though there is mention of the idea of a musical sentence. He states that the sentence is usually eight measures in length, but that the fundamental melodic material is presented in the opening two measures. Caplin admits that within the "*basic idea*" contained in these first measures, there may be several distinct *motives*. Since motives are in and of themselves identifiable morsels of music, we can logically compose shorter sentences than those Caplin is analyzing.

Approaching the matter of function from the opposite direction, small units building up to larger ones, we have Davie (1953), "*Musical Structure and Design*", who compares musical sounds to words in terms of clauses, sentences, and paragraphs leading to larger structures of form. He refers to musical cadences in terms of various punctuation marks. Depending on the type used, impressions of rest, incompleteness and surprise might be conveyed. He suggests the use of musical cadences which are perfect and plagal for completeness (by which Caplin means V-I with soprano on scale degree 1 or IV-I); imperfect for incompleteness (V-I with soprano on scale degree 3 or 5) and interrupted for surprise (deceptive cadences are V or V7-VI). Davie goes on to examine ways to extend phrases. He gives examples of grammatical clauses followed by phrases lengthened by adding adjectives, etc. The musical examples are extended by sequential repetition, additional measures of the sequential repetition, stretching the cadence and repetition of the stretched cadence

These musical techniques do mimic the end result of adding adjectives in grammar. The sentence becomes longer. We see how the comparison breaks down when we see that an adjective further distinguishes one noun from any other noun. Repetition of a musical fragment is merely a restatement. Sequential repetition can be compared to restatement of a sentence by a different person (different starting pitch).

In defining a musical sentence, Davie believes it is the combination of two or more phrases that are necessary for balance. In other words, there is a question phrase and a response phrase. All possible three measure outcomes from our matrix will be complete musical sentences. The opening fragment functions as a question phrase and the closing fragment functions as the response. We have added a connecting phrase between the two. This insertion extends our musical line, providing a smooth transition between question/answer. In terms of speech, the exchange flows more naturally by sounding less abrupt.

Some, such as Berry (1966), "*Form in Music: An Examination Of Traditional Techniques Of Musical Structure and Their Application in Historical and Contemporary Styles*", believe music to be essentially abstract. However, he does believe it has the ability to "*impart a sense, a mood, impression of states or qualities.*" Here, too, cadences are considered *musical equivalents of punctuation.* Where Davie focuses on scale degree and harmony (real or implied) to determine degrees of finality, Berry goes a step further. He points out the effect cadences have on the rhythmic motion of a line. A cadence does one of two things, either interrupts rhythmic motion or conveys closure by stopping it all together. Berry's explanation of smaller structural units also supports our three measure phrases as musical sentences. He states that a musical phrase (Berry interchanges the terms phrase and statement) may be compared to a clause that contains at least a subject and a predicate (question/answer). He goes on to say that it has a distinct beginning, a clear course of direction and an ending (cadence).

Though not a common form of analysis, relating nouns and verbs to tonic and dominants has been discussed by theorists. An example from Cope (1991), "*Computers and Musical Style*", is that V7-I equals a simple verb-object motion, with the tonic coming as a consequence to the dominant motion. This is similar to the idea of the object of a sentence being a consequence of the verb's action. Further, a clear explanation of the role of tonic in tonality is provided by Vincent D'Indy,

> "the ensemble of musical phenomena which human understanding is able to appreciate by direct comparison with a constant element-the tonic"

Cope goes on to say that the ensemble of musical pitches and chords typically relates in terms of major and minor keys. Also, these phenomena function in relation to the key name pitch and chord which are called "tonic" by theorists. In C major, the ensemble of notes would be CDEFGABC. The tonic note is C and the tonic triad (three-note chord) is CEG. There is a detailed section on parsing, which is a technique for diagramming relationships between sentence parts in language. A sentence is broken into two basic parts, a noun phrase and a verb phrase. From there, each can be broken into smaller pieces (article plus noun/adverb plus verb). Cope shows a parse of the C-major scale that shows the basis for the melodic movement used in our own examples. First, he presents the sentence as CDEFGABC. Next, he identifies noun equivalents,C (tonic) and F (subdominant). The G (dominant) functions as a verb, leaving the articles and their modifiers: A (Prolongation), E (Adverb), D (Prolongation) and B.

Logical motions in music are systematically described by Cope in the following way. SPEAC equal to Statement, Preparatic, Extension, Antecedents and Consequent. The desired line of succession then looks like this: S followed by PEA; P by SAC; E by SPAC; A by EC; and C by SPEA Cope believes that a "*generated hierarchy*" of notes is necessary to produce music that makes sense. He reasons that a random substitution of a word type that would produce a nonsense sentence (noun - verb - noun could equal horse grabs sky), would have the same effect in music. We agree with Cope that tonality itself is a hierarchy where certain notes or scale degrees have specific functions. Additionally, a set of generalities for tonal melodies is given. Cope based this list on examples from composers across generations. These are basic composition rules taught in your average music theory classes: use mostly stepwise motion; follow melodic skips with stepwise motion in the opposite direction; use one or more notes per beat agreeing with harmony; and usually begin and end on tonic or dominant chord members.

Additional detailed references for works indicating that a language/music comparison is potentially useful are included in Cope's books. Here we simply remark that important prior work includes that of William Wood (1970) in the development of Augmented Transition Networks (ATNs), which was based on the work of Bobrow and Fraser (1969). Also, Pereira and Warren (1980) used functional natural language systems and Johnson (1983) computational linguistics. Further, ATNs were used in question-answering systems to model learning and visual processing by Bolc (1983) and McTear (1987). On the basis of this prior work and his own, Cope writes in summary,

> "With proper selection of elemental representations (i.e.,nouns are tonics) and careful coding of appropriate local semantics, the same program that produces sentences can produce music" .

Further,

> "The advantage of ATN grammars is that several types of nodes within the automaton allow very flexible parsing or generating behavior to be implemented" .

Kohs (1973). "*Musical Form: Studies in Analysis and Synthesis*" stresses that the human mind will organize music into sections based on repetition and contrast and that distinctions exist between functional and decorative (nonharmonic) tones. Then, in making his own connection between speech and music, Koh brings in another factor, **emotion**. We will study the connections of emotion to music more completely in (Dzuris & Peterson, 2003a). Kohs connects to emotional states as follows:

> "In prehistory, vocal melody was probably developed as a form of emotionally inflected speech. Melody has been associated with words since the earliest times, and wordless vocalization has always been a rather rare phenomenon. Thus it is not surprising that some of the characteristics of melody are derived from speech, and that some of the melodic forms are related to the forms of prose and poetry" .

Poetry can be analyzed in terms of meter, its pattern of accented and unaccented syllables. Musical meter is analyzed the same way; in terms of strong and weak beats. Kohs compares nonmetrical music to the less structured rhythm that is characteristic to prose. Again, we have a writer stress that the key to a sentence or musical phrase is that both words and musical elements are put together based on the function each will carry out. Kohs is explicit about music having

> "its own special kind of grammar and syntax. Successive tones may be grouped to form a musical phrase having a sense of completion and unity similar to that found in a verbal sentence. Some tones, like the verbal subject and predicate, are essential to the musical structure; others are decorative. Suspensions, appogiaturas, neighboring tones and similar decorations cannot stand alone without resolution any more than an adjective may stand without a noun or pronoun. Musical phrases may be simple or complex; a short musical idea may be expanded by a variety of means, such as parenthetical insertions, or extensions at the beginning or the end"

He follows with an example comparison of a sentence and a musical phrase, in which both are transformed by the addition of words and tones respectively. The additions to the basic musical line are tried and true techniques taught to all students of composition.

We have gleamed a primitive view of how we wish to abstract structure from music by building on these works. Although the mapping between a grammatical view of music and the actual way a person composes is imperfect, it pervades many of the discussions on composition, orchestration and so forth in the musical literature. We have thus decided that our first attempt at autonomous music creation will be based on a simplistic grammatical approach: we will try to create a collection of short musical fragments which embody or encapsulate a notion of good compositional style.

## 2.2 The Würfelspiel Approach:

We will start by using an $18^{th}$ century historical idea called The Musicalisches Würfelspiel. In the 1700's, fragments of music could be rapidly prototyped by using a matrix $\mathcal{A}$ of possibilities. We show an abstract version of a typical Musicalisches Würfelspiel matrix in Equation 1. It consists of P rows and three columns. In the first column are placed the opening phrases or nouns; in the third column, are placed the closing phrases or objects; and in the second column, are placed the transitional phrases or verbs. Each phrase consisted of L notes and the composer's duty was to make sure that any opening, transitional and closing (or noun, verb and object) was both viable and pleasing for the musical style that the composer was attempting to achieve.

$$\mathcal{A} = \begin{bmatrix} \text{Opening 0} & \text{Transition 0} & \text{Closing 0} \\ \text{Opening 1} & \text{Transition 1} & \text{Closing 1} \\ \vdots & \vdots & \vdots \\ \text{Opening P-1} & \text{Transition P-1} & \text{Closing P-1} \end{bmatrix} \tag{1}$$

Thus, a musical stream could be formed by concatenating these fragments together: picking the $i^{th}$ Opening, the $j^{th}$ Transition and the $k^{th}$ Closing phrases would form a musical sentence. Since we would get a different musical sentence for each choice of the indices $i$, $j$ and $k$ (where each index can take on the values 0 to $P-1$), we can label the sentences that are constructed by using the subscript $i, j, k$ as follows:

$$S_{i,j,k} = \text{Opening i} + \text{Transition j} + \text{Closing k}$$

Note that there are $P^3$ possible musical sentences that can be formed in this manner. If each opening, transition and closing fragment is four beats long, we can build $P^3$ different twelve beat sentences.

It takes musical talent to create such a The Musicalisches Würfelspiel array, but once created, it can be used in the process of learning fundamental principles of the music compositional process. We will eventually use musical fragments which are tagged with a specific emotional color to build a model which can assemble musical fragments which have a specific emotional content. However, we will start with a proof of concept using a Musicalisches Würfelspiel matrix with emotionally neutral examples.

## 2.3 Neutral Music Data Design:

We will start with simple compositional patterns and ideas; hence, we will be using only quarter and half notes permitted in any of the phrases. Further, we do not want the musical fragments to be too long, so for now each fragment consists of four beats in 4/4 time. We will begin opening phrases and end closing or cadence phrases on tonic C, approaching or leaving by step or tonic chord leap. Finally, the middle phrases are centered around a third or a fifth. Using these guidelines, we created a Musicalisches Würfelspiel array using opening phrases as nouns, middle phrases as verbs and closing or cadence phrases as objects. We chose the noun phrases as shown in Figure 3.

**Figure 3 goes here**

Now the last note in each of four opening phrases must be able to be played right before any of the first notes in a middle phrase. Correct combinations are not random choices and so the musical composer's skill is captured to some extent in the choices that are made for the middle phrases. Our design alphabet can be encoded as $\mathcal{H}_1 = \{c,\ d,\ e,\ f,\ g,\ a,\ b,\ C\}$, and $\mathcal{H}_2 = \{c^2,\ d^2,\ e^2,\ f^2,\ g^2,\ a^2,\ b^2,\ C^2\}$, where $C$ being the C above middle C. All the letters in $\mathcal{H}_1$ label quarter notes from the middle C octave and the notes in the $\mathcal{H}_2$ alphabet denote half notes. Our alphabet is thus $\{\mathcal{H}_1, \mathcal{H}_2\}$ which has cardinality 16. Within this alphabet, the second opening, *cedf*, can be written as the matrix

$$n_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

There are similar matrices for each of the other opening phrases. Each opening phrase is thus a $4 \times 16$ matrix that has very special property: a row can only have one 1. A given middle phrase will have a similar structure and only some of the possible middle phrase matrices we could use will be acceptable. Our middle phrase design choices for emotionally flat music are shown in Figure 4.

**Figure 4 goes here**

In essence, when we design these two sets of matrices, we provide samples for a hypothetical mapping from the set of $4 \times 16$ opening matrices to the set of $4 \times 16$ middle matrices. These samples provide enough information for us to approximate the opening to middle transition function using a blend of neurally inspired architectures and function approximation techniques. Finally, the cadence or closing phrases, as shown in Figure 5, were designed so that they would sound pleasing when attached to any of the possible opening - middle phrase combinations. Again, our design gives us a set of appropriate $4 \times 16$ matrices which encode valid closing phrases. These closing phrases need to be coupled to the end notes of any middle phrase. The closing data we design then gives us enough samples to approximate the middle to ending transformation.

**Figure 5 goes here**

In addition, our opening data gives us four examples of starting notes for neutral musical twelve note sequences. Now there are nine possible start notes for each opening phrase and the fact that we do not choose some of them is important. Also, each four note sequence in any of the three phrases, opening, middle and closing, is order dependent. Given a note in any phrase, the selection of the next note that follows is *not* random. The actual note sequence that appears in each phrase also gives sample data that constrains the phrase to phrase transformations. We can use this information to effectively approximate our mappings using excitation/ inhibition neurally inspired architectures. Roughly speaking, if a given subset of notes are good choices to follow another note, then the notes not selected to follow should be actively inhibited while the acceptable notes should be actively encouraged or enhanced in their activity. The complete Musicalisches Würfelspiel matrix, as seen in Figure 6, thus consists of four rows and three columns that will provide a total of 64 distinct musical fragments that are intended to model neutral musical sentence design.

**Figure 6 goes here**

Currently, these fragments are rather short since they are just four beats (i.e. $L = 4$) for each grammatical element. This is enough to help with the prototype development of this stage. However, when we generate the additional Musicalisches Würfelspiel matrices that will correspond to the other emotional shadings, we may find it necessary to use longer note sequences (i.e. increase $L$) in order to capture the desired emotional colorings. If that turns out to be true, we can easily return to this neutral case and redesign the neutral Musicalisches Würfelspiel matrix to use longer note sequences.

## 2.4 The Generated Musical Phrases:

From our $4 \times 3$ Musicalisches Würfelspiel matrix, we can generate 64 musical selections of twelve beats each. In Figure 7, we show the selections generated using opening one, all the possible middle phrases and the first cadence phrase. The intent here is that all 64 selections we so generate will be devoid of emotional content. You should try playing these pieces on the piano ( and compare them to our later selections that are purported to have angry, sad and happy overtones that are displayed as detailed in (Dzuris et al., 2003a).

**Figure 7 goes here**

The fragments shown in Figure 7 show only a few of the possibilities. For example, the selections for opening two, verb one and all the cadence phrases are displayed in Figure 8. Again, we invite you to play the pieces. The important point is that all 64 pieces we can so generate using the Würfelspiel approach are **equally** valid emotionally neutral choices. Hence, the Musicalisches Würfelspiel matrix we have created *captures* the essence of the solution to emotionally neutral music compositional design problems.

**Figure 8 goes here**

# 3 Cognitive Modeling:

We are developing models of abstract compositional design in the three separate domains of music composition, painting composition and job scheduling. All three share similar structure despite great differences in culture and background. We use variants of the historical compositional prototyping mechanism from the $18^{th}$ century known as Toss of the Dice (Würfelspiel) matrices to construct training data for examples of good compositional design in which any particular sample of training data is equally valid as a choice. This can be done in a number of emotionally distinct flavors following the affective image psychophysiological literature although the notion of emotional attribute must be extended to more general notions of intangibles in job scheduling using genetic algorithms.

Our models of cognitive processing are based on distributed models of computation in which ensembles of interacting computational modules are linked to create larger functional units. The linked units are then used to model cognitive functions such as emotions and, more importantly, disturbances in emotional processing.

Cognitive models are built by finding an appropriate level of abstraction for known cellular biological and neurobiological data that allows us to find generic principles of biological information processing. One of the hardest problems we face in our attempt to develop software models of a high level cognitive process is validation. We mean validate in this sense: the model should behave in the way we expect the cognitive function to behave. This is not generally the way a cognitive model is evaluated. If one builds a model of memory (hippocampus) with a lot of low level biological detail, one might have $10,000$ neurons, each in several interconnected modules. We know that if you take a slice through the hippocampus, we can get that slice to live in the artificial environment of a petri dish for some time. Moreover, we can take measurements from an ensemble of reading points in that tissue – perhaps 100 or more – and get voltage and/ or current maps of that local patch of tissue. One type of validation is to get the model just built to match these maps. However, there is a lot of debate as to what this means. Those sorts of local readings are clearly not what that complicated system of interrelated modules actually does. There will always be questions as to whether or not matching these readings really is validation of the model. There is a large and unexplained gap between validation of the sort proposed by matching voltage and current maps to validating that the computations performed by the modules are to the external human witness similar to what we would see in a real human performing tasks. Further, validation of the type performed by the human witness is usually done from the perspective of psychology or psychobiology, but most of these validations are more qualitative than quantitative.

To address this problem, we decided that in addition to developing cognitive models that are measured in traditional ways, we would also develop models that can be assessed by experts in other fields for validity. There are three such models we are attempting to build: one is a model of music composition in which short stanzas of music that are emotionally colored or tagged are generated autonomously; second, a model of painting composition in which primitive scenes comprised of background, foreground and primary visual elements are generated autonomously that also have emotional attributes; and third, a model of job scheduling design using genetic algorithms in which the optimal solutions are colored by intangibles which are similar to emotional attributes and could therefore be classified as meta level outputs of an emotional subsystem. This paper will focus on the important task of designing emotionally neutral musical data. In other papers that are in preparation, we address neutral data generation for painting, (Peterson & Dzuris, 2003f), and job scheduling, (Kurz et al., 2003b). The task of generating emotionally labeled musical data is presented in (Dzuris et al., 2003a) while the problems of generating emotionally labeled paintings and the intangibles of job scheduling are presented in (Peterson & Dzuris, 2003g) and (Kurz et al., 2003c), respectively.

### 3.1 Clues To Our Models From The Literature:

Clues as to how to set up these abstract models can be found in a variety of sources from the open literature. The evolution of nervous systems and in particular, the large order structures of the brain, are very informative. Discussions in (Redies & Puelles, 2001), "*Modularity in vertebrate brain development and evolution*", help us understand the modularity of the underlying neural structures for information processing. This has influenced our software design by helping us determine the minimal module architecture which will allow interesting cognitive response.

Motivations for the modeling of the core emotional and cognitive dysfunction engines are varied. A proper model of the computational outputs we interpret as emotional states or qualia requires a concomitant model of how cognitive processes develop flaws. Hence, an understanding of models of cognitive dysfunction such as depression is closely connected to models of emotional processing. Key resources include those that are neurophysiological in nature such as detailed in (Deadwyler & Hampson, 1997), "*The Significance of neural Ensemble Codes During Behavior and Cognition*". Some of the actual circuitry that may per-mit emotional responses to be generated are discussed in Drevets (1998), "*Neuroanatomical Circuits in Depression: Implications for Treatment Mechanisms*". These potential neural architectures lead us to ask whether or not there are specialized places within the brain which correspond to specific emotions and/ or cognitive states. These techniques are used to target questions relevant to cognitive dysfunction in (Honey, Fletcher & Bullmore, 2002), "*Functional brain mapping of psychopathology*".

As we mentioned earlier, there is also a rich literature in the field of psychophysiology which provides a way to assign quantitative measures to certain kinds of emotional or affective outputs. The psychophysiological literature includes acoustical and visual studies. The acoustic studies of (Bradley & Lang, 2000), "*Affective reactions to acoustic stimuli*", concern the physiological responses of listeners to certain specific auditory tones or probes. This is not a response to music per se, but it gives valuable data as to normal human responses. There are also studies that measure the response of human subjects when briefly exposed to samples of pictures from a carefully selected database of images whose emotional content is varied. For example, the results of these studies are included in (Codispotti et al., 2001), "*Affective reactions to briefly presented pictures*". We can also examine patients with cognitive dysfunction such as schizophrenia via imaging tools as in (Lang, et al., 1998), "*Emotional arousal and activation of the visual cortex: An fMRI analysis*".

Finally, there is little in the physiological literature which is directed toward how a human responds to music. An overview which summarizes some of the relevant physiology is given in Peretz (2001), "*Listen to the brain: a biological perspective on musical emotions*". However, there is little said specifically about musically correlated neural cognates.

### 3.2 The Basic Model:

Our basic neural model is based on abstractions from neurobiology. A model of isocortex is motivated by recent models of cortical processing outlined in (Raizada & Grossberg, 2003). This article uses clues from visual processing to gain insight into how virgin cortical tissue (isocortex) is wired to allow for its shaping via environmental input. Clues and theoretical models for auditory cortex can then be found in the survey paper of Merzenich (2001).

The individual neural objects in our cortex will be abstractions of neural ensembles, their behaviors and outputs gleaned from both low level biological processing and high level psychopharmacology data.

## 4 Data Set Design:

The neutral music data set contains examples of equally valid solutions to a compositional design process. In other papers, we discuss musical data that is emotionally tagged (Dzuris et al., 2003a), neutral and emotionally labeled paintings (Peterson et al., 2003f and 2003g) and job scheduling data (Kurz et al., 2003b and 2003c). Each of these papers discuss the generation of 64 examples of solutions to compositional design tasks that are equally acceptable according to some measure. So, can we *learn* from this kind of data how the experts that designed these data samples did their job? Essentially, buried in these data sets are important *clues* about what makes a great design. How do we begin to understand the underlying compositional design principles?

Each data set that is encoded into a Würfelspiel matrix, whether using music, art or abstract optimization languages, therefore contains crucial information about equally valid examples of data in different emotional/ intangible modalities. From this data, we can build mappings that tell us which sequences of choices are valid and which are not from the perspective of the expert who has designed the examples. In a very real sense, a cognitive model built from this data is automatically partially validated from a psychological point of view.

Each data set has an associated alphabet with which we can express noun, verb and object units. For our purposes, let's say that each of the nouns, verbs or objects is a finite list of actions from an alphabet of $R$ symbols, where the meaning of the symbols is of course highly dependent on the context of our example. In a simple music example, the list of actions might be a sequence of four beats and the alphabet could be the choices from the C major scale. Thus, we will assume each of the $P$ nouns consists of a list of length $L$ from an alphabet of $R$ symbols. Since the general alphabet has $R$ symbols, each element of a noun can be thought of as a vector of size $R$ whose components are all 0 except for a single 1. Let the letters in the alphabet be denoted by $a_0$ through $a_{R-1}$. Then a noun has components 0 to $L-1$ where component $n[j]$ is the letter $a_j^n$. The letter $a_j^n$ is then encoded into a vector of length $R$ all of whose entries are 0 except for the entry in slot $j$.

## 4.1 Inhibition and Excitation:

The raw inputs we see as the noun vector $n$ are thus normally processed into a specialized vector for use in algorithms that model the compositional process. For example, in the generation of music, we write an opening phrase in standard notation (the kind we would see on sheet music). This is the raw input $n$. We can choose to encode this data in a form more amenable to computation and data processing in many ways. In our work, we have encoded the raw data into an abstract grammar, thereby generating the feature vector $N$. In general, the input nouns are preprocessed to create output noun states denoted by $N$. In a similar fashion, we would preprocess verb and object inputs to create verb and object output states denoted by $V$ and $O$, respectively. The preprocessing is carried out by mappings $f_n$, $f_v$ and $f_o$, respectively. For example, the mapping from raw data to the feature vector form for nouns is represented by Equation 2:

$$\begin{bmatrix} a_0^n \\ \vdots \\ a_{L-1}^n \end{bmatrix} \rightarrow_{f_n} \begin{bmatrix} N_0 \\ N_1 \\ \vdots \\ N_{L-1} \end{bmatrix}$$

Our association of a noun $n$ with the feature vector $N$ is thus but one example of the mapping $f_n$. Now, a primitive object in our purported compositional grammar has length $L$. We can denote such an input noun object as $n_i$ and a corresponding output noun object as $N_i$. Our mapping problem is thus to determine the rule behind the mapping from the noun feature vectors $N$ to the verb feature vectors $V$, $g_{NV}$, and from the verb feature vectors $V$ to the object feature vectors $O$, $g_{VO}$. We can express this mathematically by Equation 2:

$$\begin{bmatrix} N_0 \\ N_1 \\ \vdots \\ N_{L-1} \end{bmatrix} \rightarrow_{g_{NV}} \begin{bmatrix} V_0 \\ V_1 \\ \vdots \\ V_{L-1} \end{bmatrix} \text{ and } \begin{bmatrix} V_0 \\ V_1 \\ \vdots \\ V_{L-1} \end{bmatrix} \rightarrow_{g_{VO}} \begin{bmatrix} O_0 \\ O_1 \\ \vdots \\ O_{L-1} \end{bmatrix} \tag{2}$$

We can combine these processing steps into the diagram shown in Figure 9.

**Figure 9 goes here**

## 4.2 Noun to Verb Processing:

The data we are given is order dependent. For example, if we are given a noun, $n_i$ of form $\{a_{i0}^n, \ldots, a_{i,L-1}^n\}$, then we attend to the letters of this noun sequentially as $a_{i0}^n \rightarrow a_{i1}^n \rightarrow \ldots \rightarrow a_{i,L-1}^n$. We are given that each noun $n_i$ is associated with a set of possible verbs, $\{v_j\}$ equal to $\{a_{j0}^v, \ldots, a_{j,L-1}^v\}$ for $0 \leq j < P-1$ and one task is thus to understand the noun to verb mapping. However, another task is to understand how to generate the original noun sequence. Why are some noun sequences useful or pleasing in this context and others are not? To generate a noun sequence $n$ equal to $\{a^n_0, \ldots, a^n_{L-1}\}$ means we choose a random start letter $a^n_0$ and then from that preferred sequences are generated while non-interesting words are biased against. Hence, we think of a mapping, the Noun Generating Map or NGM as accepting an input, $a^n_0$ and generating a preferred second note $a^n_1$. Then $a^n_1$ is used as an input to generate a preferred third note, $a^n_2$ and so on until the full string of letters is finished. To model this mapping, we start by using the information about useful noun strings we have. Given letter $a^n_{i0}$, we know that $a^n_{i1}$ is preferred.

We embed the original data into an analog vector by converting each letter $a^n_{i0}$ of the noun, which is a 0 or a 1 in our initial encoding, into a real number $\xi^n_{i0}$. To set the value of the real number $\xi^n_{i0}$, we choose a tolerance, $\epsilon$, in the interval $(0, 0.25)$ and choose a real number $y$ randomly from the interval $[-0.5\epsilon,\ 0.5\epsilon]$ and then set the value of $\xi_{i0}$ to be $\epsilon \pm y$. Hence, the value of $\xi^n_{i0}$ lies in the interval $[0.5\epsilon,\ 1.5\epsilon]$. For example, if $\epsilon$ was chosen to be 0.20, then for all indices in the binary encoding of the letter $a^n_{i0}$ that are 0, we would randomly choose $y$ from $[-0.1,\ 0.1]$ generating $\xi^n_{i0}$ values that lie in $[0.10,\ 0.30]$. We will call the number $\epsilon$ our analog threshold. The entry in the binary encoded letter that corresponds to a 1 will be randomly chosen from $[1 - 1.5\epsilon,\ 1 - 0.5\epsilon]$. Hence, for $\epsilon = 0.2$, entry with a 1 will be assigned a real number in the interval $[0.7, 0.9]$. Hence, the raw binary noun, verb and object data is mapped into a new analog representation in which each entry is a real number chosen as above.

We know that only certain letters should follow $a^n_{i0}$. Hence, only certain analog states $\xi^n_{i1}$ are permissible given a start state of $\xi^n_{i0}$. We infer from this that there is an unknown mapping $h^{01}$ which maps the analog encoding of letter 0 to the analog encoding of letter 1, $h^{01}(\xi^n_{i0}) = \xi^n_{i1}$. This mapping has special characteristics: our data tells us that only certain letters that can follow letter 0. Each acceptable second letter is a vector in $R$ dimensional space whose components are are analog zero except the one the corresponds to the second letter. That component is an analog one. We have at most $P$ examples of acceptable second letters. Hence, we have $R - P$ second letters that are not acceptable. In other word, the preferred output for a given noun is a $R$ row matrix formed from the acceptable verbs that has at most $P$ columns.

We can do this for all of the nouns in our data set. Hence, we will have at most $P$ first letter choices and each of these will have at most $R - P$ unacceptable second letters. Let $T$ and $T'$ denote the set of all acceptable and unacceptable outputs respectively. We model the mapping $h^{01}$ as a chained feed forward network, Peterson (1998), with feedforward and feedback connections between artificial neurons. This mapping takes the first letter of a noun and outputs a set of acceptable second letters. Training is done by matching input to output using excitation and inhibition. We know which elements in the analog output vector should be close to one and which should be close to zero for the analog input. We initialize all of the tuneable parameters to be small positive if the connection from component $k$ in the input to component $j$ on the output is between two analog ones. All other connections are initialized to small negative numbers. For each first letter we have data for, we do the following: pick the initial first letter in our data set and compute the relevant output for the first associated second letter. Increase the connection weights on any path between a high input and a high output and decrease the connection weight on any other paths. Cycle to the next second letter and redo until all possibilities are exhausted. We thus continue this process until every input generates an output with a high component value in the location that corresponds to the index for the second letter. At this point, we say we have trained our nonlinear mapping $h^{01}$ so that first letters in our data noun sequences are biased to connect to their corresponding second letters. A second letter is then chosen randomly from the set of acceptable second letters via an additional input line which is a sense is a coarse model of *creativity*.

If we let the set of all the generated weights be the matrix $W^{01}$, we note this is an

$R \times PR$ size matrix. We can develop a similar mapping for the second to third letter, $h^{12}$ with weights $W^{12}$, the third to fourth letter, $h^{23}$ with weights $W^{23}$, and finally, the mapping from letter $L-1$ to letter $L$, $h^{L-2,L-1}$ with weights $W^{L-2,L-1}$.

The procedure for creating a valid noun sequence can now be given. Choose a valid starting letter for a noun, $a^n_{i0}$ and we map it to its analog form, $\xi^n_{i0}$. Then, applying the first to second letter map, we find an acceptable second letter set by the computation $h^{01}(\xi^n_{i0}) = \{\xi^n_{i1}\}$. These second letters can be used as inputs into the next map, generating an acceptable third letters. Hence, the composite map, $h^{12}h^{01}$ takes a valid first letter and creates a set of three letter analog sequence defined by Equation 4.2:

$$\begin{bmatrix} \xi^n_{i0} \\ \{\xi^n_{i1}\} \\ \{\xi^n_{i2}\} \end{bmatrix} \quad in \quad \begin{bmatrix} \xi^n_{i0} \\ h^{01}(\xi^n_{i0}) \\ h^{12}(\, h^{01}(\xi^n_{i0})) \end{bmatrix} \tag{3}$$

The analog sequences are then mapping into three letter sequences by assigning an analog value to either a one or zero using a threshold tolerance $\tau$. This means we map a component whose value is above $\tau$ to 1 and one whose value is below $\tau$ to 0. This can of course generate invalid sequences as we are only supposed to have a single 1 assigned from any analog sequence. We do have to make sure that our developed map does not allow this. For example, for $\tau = .6$, the vector

$$\begin{bmatrix} 0.83 \\ 0.55 \\ 0.35 \end{bmatrix} \quad \rightarrow \quad \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

and the example below generates an invalid binary sequence:

$$\begin{bmatrix} 0.83 \\ 0.55 \\ 0.35 \end{bmatrix} \quad \rightarrow \quad \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

which we would not know how to interpret as part of a noun. Nevertheless, despite these obvious caveats, the procedure above learns how to generate all acceptable three letter nouns given an initial start letter. There will be at most $P^2$ possible second and third letters in this set. Since this set of possibilities will grow rapidly, after generating the letter two set of possibilities, we randomly choose one of the columns of the letter two matrix as the second letter choice and apply the $h^{12}$ mapping to that letter. We then randomly choose one of the columns of the letter three matrix as the third letter choice.

We then extend this procedure to the generation of all $P$ letters by the concatenation $h^{P-2,P-1} \cdots h^{12}h^{01}$ which we will denote by the symbol $H^n$, where the superscript indicates this is the mapping we will use for noun sequences. The mapping $H^n$ is the Noun Generating Map or NGM that we seek. It generates a set of $P^{P-1}$ letter two to letter $P$ sequences. By making a random column choice at each letter, we generate one random $P$ letter noun sequence for each initial letter we use. We can do something similar for the verb and object data generating the Verb and Object Generating Maps $H^v$ and $H^o$, respectively.

These three mappings are the noun, verb and object generator mappings we were seeking. Then we need to connect nouns to verbs and verbs to objects. The mapping from $N$ to $V$ is where the real processing lies. The Würfelspiel matrix training data approach tells us that it is permissible for certain nouns to be linked to certain verbs. While we could memorize a look-up table based on this data, that is not what we wish to do. We want to determine underlying rules behind these associations as emergent behavior in a complex system of interacting agents. Thus, each output noun $N_i$ in the the collection of P nouns $\{N_i : 0 \leq i < P\}$ should activate any of the P verbs $\{V_j : 0 \leq j < P\}$ via the map $g_{NV}$. Further, each verb $V_j$ in $\{V_j : 0 \leq j < P\}$ should activate the output nouns $\{O_i : 0 \leq i < P\}$ by the action of the map $g_{VO}$.

To build the mappings $g_{NV}$ and $g_{VO}$, we will use a more sophisticated model than a simple chained feedforward network with feedback and feedforward connections. Our computational model will consist of four fully connected abstract neural objects capable of receiving input as is shown in Figure 10.

**Figure 10 goes here**

In such a cluster of four nodes, organized in a square fashion there are six possible node to node connections. Each of these neural objects will consist of nine individual abstract neurons as shown in Figure 11. These nine node internal clusters will also be fully interconnected. The output of a given internal neuron is a feature vector whose exact values can be shaped by pharmacological inputs so as to phase lock with either none of the other neurons or some subset of the nine internal nodes. Since there are nine internal neurons, there are many possibilities for phase locked subclusters. Each of the four ensemble nodes has three interconnections to other ensemble nodes. Rather than simply weighting such interconnections with a scalar value as is usual in naive connectionist models, we instead will determine whether a connection is active and its weight based on which phase locked subclusters are available. One such possible mapping can be set up as follows: there are 130 phase locked clusters of size 6 to 9: if such a phase locked cluster is available, this activates connection one; there are 252 phase locked clusters of size 4 and 5; any such cluster activates connection two; there are 130 remaining phase locked clusters: any of these being available activates connection 3. Hence, shared activity chooses activation of a connection pathway and the weight of the ensuing synaptic connection can be set via standard Hebbian protocols.

**Figure 11 goes here**

## 4.3 Sentence Construction:

There are then two ways to create a valid musical sentence. The first does not use the mappings $g_{NV}$ and $g_{VO}$. A random choice of starting phrase or noun is chosen to begin the sentence selection process. This input generates a valid noun or opening phrase. Then, a randomly chosen starting note for the middle phrase is then used to generate a valid verb. Finally, a random starting note for the closing phrase generates the final four beat sequence of our valid neutral phrase. The output of the cognitive module is a short sentence of the type we have discussed. The second method is more interesting. The randomly generated noun $N$ generates a valid verb $g_{NV}(N)$ and the valid object is generated by the concatenation $g_{VO}(g_{NV}(N))$. Thus, the composite map $g_{VO}\, g_{NV}$ provides a sentence generator.

Once we can generate sentences, we note that we can move to the generation of streams consisting of sentences concatenated to other sentences once we create an Object to Noun mapping. This is done in a way that is similar to what we have done before using a Würfelspiel array approach. Other possibilities then come to mind; for example, in the context of music composition, since key changes are logical transitions, we can create arbitrarily long musical streams punctuated by appropriate key changes by using the Würfelspiel array approach to model which key changes between given key signatures are pleasing between two musical streams.

## 4.4 The Cognitive Models:

The model we have generated so far, creates valid neutral musical phrases. In other papers, we address the manner in which we generate valid emotionally tagged musical phrases following the design of the triangle shown in 2. While the full discussion of how a cognitive model of emotions is built from this type of data is relegated to other papers (Peterson & Charlesworth, 2003d and 2003e), it is worthwhile to mention the process briefly for completeness.

In other papers, we have discussed this process for painting and genetic algorithm designs for job scheduling. Hence, if we use the superscripts $\alpha$ and $\beta$ to denote emotional modality and data type, respectively, we can label the mappings in the form $\{g_{NV}^{\alpha\beta},\, g_{VO}^{\alpha\beta}\}$. We let $\alpha = 0, 1, 2 3$ denote neutral, happy (contented), sad (demoralized) and angry (antagonistic) and $\beta = 0, 1, 2$ indicate music, painting and genetic algorithm optimization. We thus have a collection of mappings

| | | | | |
|---|---|---|---|---|
| $\{g_{NV}^{00},\, g_{VO}^{00}\}$ | $\{g_{NV}^{01},\, g_{VO}^{01}\}$ | $\{g_{NV}^{02},\, g_{VO}^{02}\}$ | $\{g_{NV}^{03},\, g_{VO}^{03}\}$ | music |
| $\{g_{NV}^{10},\, g_{VO}^{10}\}$ | $\{g_{NV}^{11},\, g_{VO}^{11}\}$ | $\{g_{NV}^{12},\, g_{VO}^{12}\}$ | $\{g_{NV}^{13},\, g_{VO}^{13}\}$ | painting |
| $\{g_{NV}^{20},\, g_{VO}^{20}\}$ | $\{g_{NV}^{21},\, g_{VO}^{21}\}$ | $\{g_{NV}^{22},\, g_{VO}^{22}\}$ | $\{g_{NV}^{23},\, g_{VO}^{23}\}$ | GA |

for the generation of musical, painting and optimization compositional designs.

For each emotional modality, our musical, painting and job scheduling Würfelspiel data provides 64 equally valid data points. Consider the 64 "sad" data points for music. We know as humans that this data is "sad". Each such "sad" data point provides auditory cortex training data and 64 examples of the "sad" emotional attribute. The painting data gives us 64 examples of visual cortex training data as well as 64 additional "sad" emotional attributes. Hence, we have 128 examples of "sadness" split between the sensory pathways of hearing and vision. In this way, we build 128 examples of each emotional attribute from music and painting split equally between hearing and vision. The intangibly tagged data "demoralized" from job scheduling provides 64 examples of a higher level cognitive output. We know from studies of neural processing, that the front of the auditory and visual cortex is closely aligned to sensory data and as you progress into more interior layers of cortex, neural ensembles begin to respond to progressively higher and more abstract patterns. For example, in the auditory cortex, initially the nerve cells respond to simple phonemes of perhaps 20 mS duration and higher levels are responsive to words, then sentences and so forth. We can make similar analogies to processing in the visual cortex. Outputs from primary sensory cortex are fed into higher level associative cortex where more abstract processing is performed. The intangible “demoralized” tagged data set, then provides 64 examples of a particular type of valid compositional design for the training of the outputs of the higher level associative cortex. Hence, our data provides a validating pathway for two types of primary sensory cortex as well as a primitive model of higher level associative cortex. We can design algorithms to train our cortical tissue models using laminar cortical processing via on-surround excitation/ off-surround inhibition and Hebbian based connection strengthening.

As discussed for neutral data in the context of music, given a random starting note say from column one of a Würfelspiel music matrix of given emotional modality, our model will generate an entire valid musical composition. Note that this output should actually be interpreted as two separate pieces of information: one, as a musical composition and two, as an emotional state. Our data thus provides training for the correct output of a model of emotions as well as we have a total of 192 happy, sad and angry input/ output training samples. In addition, we have 64 input/ output samples each for contented, demoralized and antagonistic emotional states. We can thus generate valid compositional designs that have a specific emotional tag for both the auditory, visual and associative cortex pathways.

We will start with an emotional model which outputs two parameters (loosely based on the psychophysiological data experiments). The first is actually a skin conductance parameter and the second is a complicated computed value that arises from certain fMRI measurements. The interesting thing about these values is that in experiments with human subjects, when people saw pictures with emotional contents such as “sad”, “angry” and so forth, the two parameters mentioned above determined a x-y coordinate system in which different emotional attributes were placed experimentally in very different parts of this plane.

For example, “sad” images might go to quadrant 2 and “angry” images might be mapped to quadrant 3. This is an over simplification of course, but the idea that images of different emotional attributes would be separate in the plane is powerful. Our 128 examples of each emotional attribute thus give us 128 data points which should all be mapped to the same decision region of this two dimensional plane. Thus, we have data that unambiguously gives us a desired two dimensional output for our emotional model.

The emotional model we will use will be based on four fully connected software agents. Each software agent will consist of nine artificial neurons that accept neuro transmitter inputs that can modulate their outputs in the three time frames we have mentioned (200 mS, 1 - 10 S and days) via some interesting second messenger interactions. The outputs of these neurons can phase lock. Depending on how many phase lock, the output of the software agent can be directed to be high toward another software agent. For example, if anywhere from six to nine neurons are phase locked, this would indicate that the output signal from the software agent would be sent to its neighbor to the east. It is easy to devise many strategies for agent connection that are based on cooperating ensembles of neural activity. This mimics many aspects of what we know about motor and behavior control.

The overall output of the four interacting software agents that model our emotional engine is then trained via Hebbian principles to match the data we have provided. At this point, we have a model that given an abstract auditory and visual input stream from the data will generate both a musical and painting composition and a two dimensional emotional attribute vector. We can then turn the system around if you like, by noting that each data set corresponds to a certain decision region in ”emotional space”. Further, recognize that we have a coupled model of sensory processing and emotional computation. We have an auditory and visual agent which given input from a known emotion decision region, generates a musical and artistic stream of a given emotional attribute. We model this as a three software agent construction: auditory, visual and emotional. Each of these agents accepts inputs from the others. We have enough data to develop a first pass at a model which given a two dimensional emotional input and a visual and auditory random start, will generate an emotional tagged auditory and visual steam. We can then *validate* this model easily by just letting anyone listen or look at our output and tell us if it is good. Hence, we are validating the whole model instead of just the equivalent of a local patch of hippocampal tissue in a slice. Indeed, it should also be possible to include validation in the learning algorithm.

# 5 Conclusions:

The design for the generation of neutral musical data presented here provides us with as much data as we wish for the training of the models of cognition that we have described in outline. We note that the way this data is designed captures primitive elements of good musical compositional design in a convenient matrix form that is ideally suited for use in the training of a variety of artificial neural models that can be used to build larger models of cognition. Further, this data, in conjunction with emotionally tagged musical data (Dzuris et al., 2003a), allows us to train both the auditory cortex portion and some of the emotional pathways of the cognitive model. We can leverage these models into two directions: the first, a model of the musical compositional process that is based on fundamental models of cognition and not based on statistical analysis of a composer's work and pastiches such as presented in Cope (2001); the second, true models of cognition proper in which the musical data provides just one part of the overall model. The data generated from our work with art primitives used to assemble painting compositions, detailed in (Peterson et al., 2003f and 2003g), trains the visual cortex and the data generated from our work with the intangibles associated with complex optimization algorithms trains the higher level associative cortex (Kurz et al., 2003b and 2003c). Together, these are all necessary components to complete our model.

# 6 Acknowledgments:

This research was partially supported by the National Science Foundation grant DBI 0119171, "*Asynchronous Methods on Heterogeneous Computer Networks for Abstracting High Level Biological Meaning*", from the Division of Biological Infrastructure; Biological Databases and Bioinformatics. In addition, we are grateful for the many conversations on the interface between music, art, mathematics, computer science and cognition that we have had with our research colleagues M. Kurz and S. Charlesworth which have helped shape this work.

# 7 References:

1. Berry, W. (1966). *Form In Music: An Examination Of Traditional Techniques Of Musical Structure and Their Application in Historical and Contemporary Styles.* Prentice-Hall, Inc.

2. Bradley, M. & P. Lang. (2000). Affective reactions to acoustic stimuli. *Psychophysiology*, 37, 204 - 215.

3. Caplin, W. (1998). *Classical Form: A Theory of Formal Functions For The Instrumental Music of Haydn, Mozart and Beethoven.* Oxford University Press.

4. Codispotti, M., M. Bradley & P. Lang. (2001). Affective reactions to briefly presented pictures. *Psychophysiology*, 38, 474 - 478.

5. Cope, D. (1991). *Computers and Musical Style.* A-R Editions, Inc., 1991.

6. Cope, D. (2001). *Virtual Music: Computer Synthesis of Musical Style.* MIT press.

7. Cuthbert, B., M. Bradley & P. Lang. (1996). Probing Picture Perception: Activation and Emotion. *Psychophysiology*, 33, 103 - 111.

8. Davie, C. (1953). *Musical Structure and Design*, Dover Publications, Inc.

9. Deadwyler, S. & R. Hampson. (1997). The Significance of neural Ensemble Codes During Behavior and Cognition. In W. Cowan, E. Shooter, C. Stevens & R. Thompson, (Eds.), *Annual Review of Neuroscience*, Vol. 20, 217 - 244.

10. Drevets, W. (2000). Neuroimaging Studies of Mood Disorders. *Biological Psychiatry*, 48, 813 - 829.

11. Dzuris, L. & Peterson, J. (2003a) Data Abstraction In Emotionally Tagged Models for Compositional Design in Music. In preparation.

12. Honey, G., P. Fletcher & E. Bullmore. (2002). Functional brain mapping of psychopathology *Journal of Neurological Neurosurgical Psychiatry*, 72, 432 - 439.

13. Kohs, Elias. (1973). *Musical Form: Studies in Analysis and Synthesis*, Houghton Mifflin Company.

14. M. Kurz & Peterson, J. (2003b). Cognitive Models for Compositional Design in Genetic Algorithms: Data Abstraction. In preparation.

15. M. Kurz & Peterson, J. (2003c). Compositional Design in Genetic Algorithms with Intangibles: Data Abstraction. In preparation.

16. Lang, P., M. Bradley, J. Fitzmmons, B. Cuthbert, J. Scott, B. Moulder & V. Nangia. (1998). Emotional arousal and activation of the visual cortex: An fMRI analysis. *Psychophysiology*, 35, 199 - 210.

17. Merzenich, M. (2001). Cortical Plasticity Contributing to Child Development. In J. McClelland & R. Siegler (Eds.), *Mechanisms of Cognitive Development: Behavioral and Neural Perspectives*, (pp. 67 - 96). Lawrence Erlbaum Associates, Publishers.

18. Peretz, I. (2001). Listen to the brain: a biological perspective on musical emotions. In Juslin, P. & J. Sloboda, (Eds.), *Music and Emotion: theory and research* (pp. 105 - 130). Oxford University Press. .

19. Peterson, J. (1998). General Feed Forward Architectures: Volume 1: Theory. Available at
http://www.ces.clemson.edu/∼/petersj/NeuralCodes/ffntheory.

20. Peterson, J. & Charlesworth, S. (2003d). Multi-Agent Cognitive Models of Compositional Design. In preparation.

21. Peterson, J. & Charlesworth, S. (2003e). Multi-Agent Cognitive Models of Emotionally Tagged Compositional Design. In preparation.

22. Peterson, J. & Dzuris, L. (2003f). Cognitive Models for Compositional Design in Painting: Data Abstraction. In preparation.

23. Peterson, J. & Dzuris, L. (2003g). Emotionally Tagged Models for Compositional Design in Painting: Data Abstraction. In preparation.

24. Raizada, R & S. Grossberg. (2003). Toward a Theory of the Laminar Architecture of Cerebral Cortex: Computational Clues from the Visual System. *Cerebral Cortex*, Jan. 2003, 100 - 113.

25. Redies, C. & L. Puelles. (2001). Modularity in vertebrate brain development and evolution. *Bioessays*, 23, 1100 - 1111.

# List of Figures

## 8 Figures:

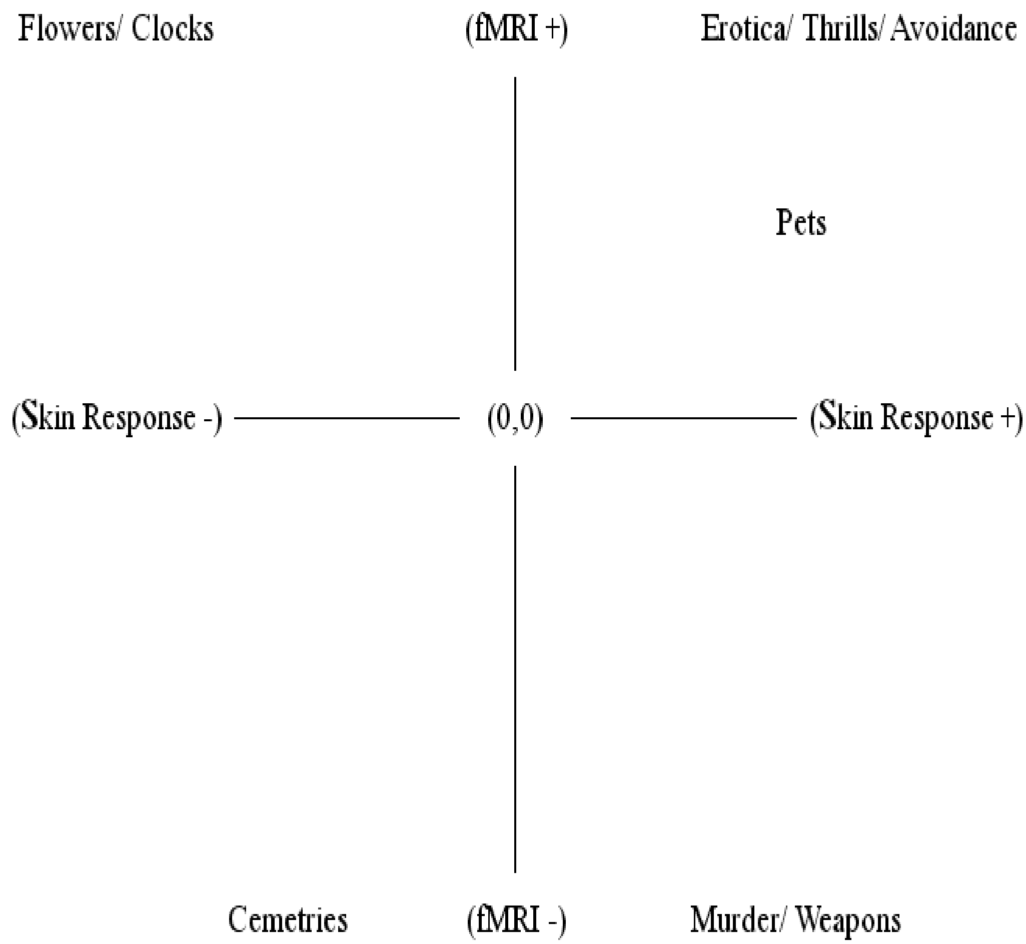



Figure 1: Human Response To Emotionally Charged Picture Data

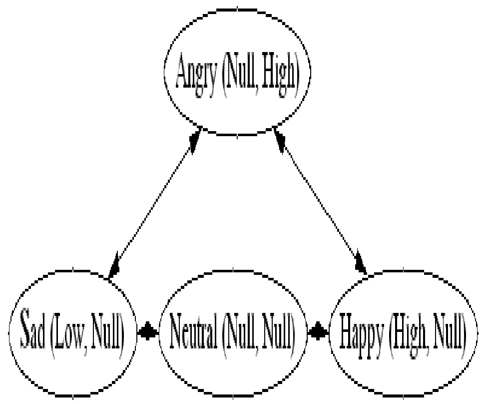



Figure 2: Emotionally Charged Compositional Data Design

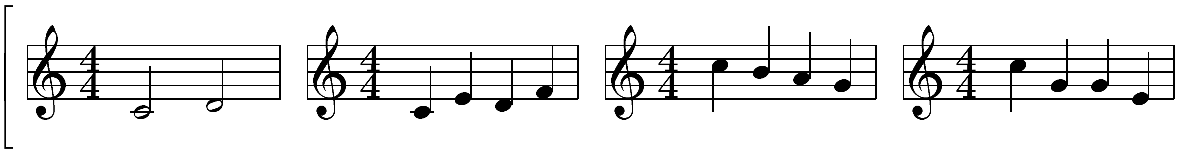


Figure 3: Musical Opening Phrases

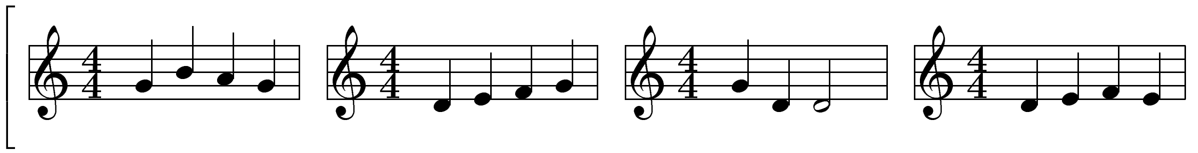


Figure 4: Musical Middle Phrases

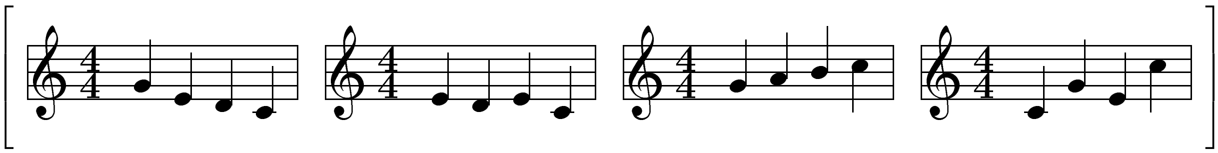


Figure 5: Musical Closing Phrases

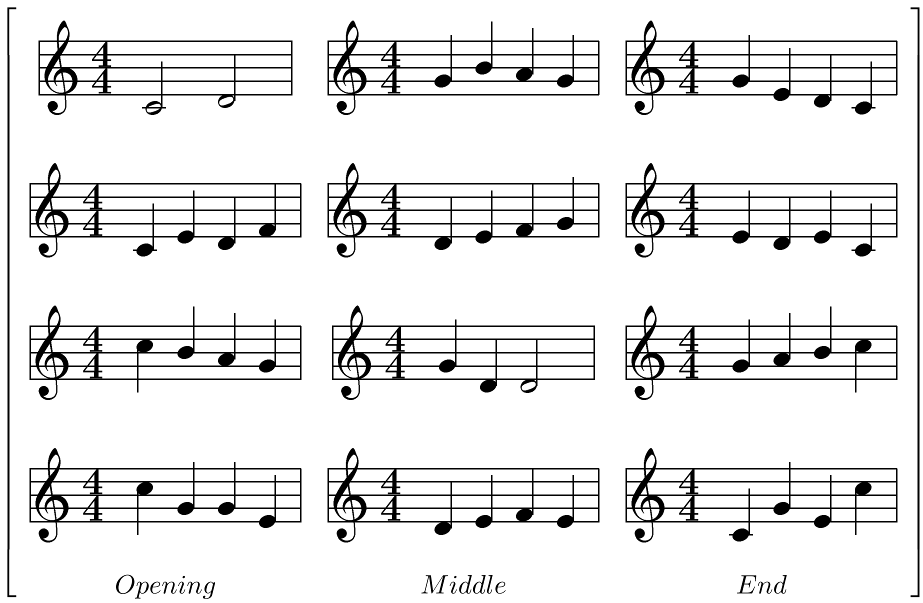
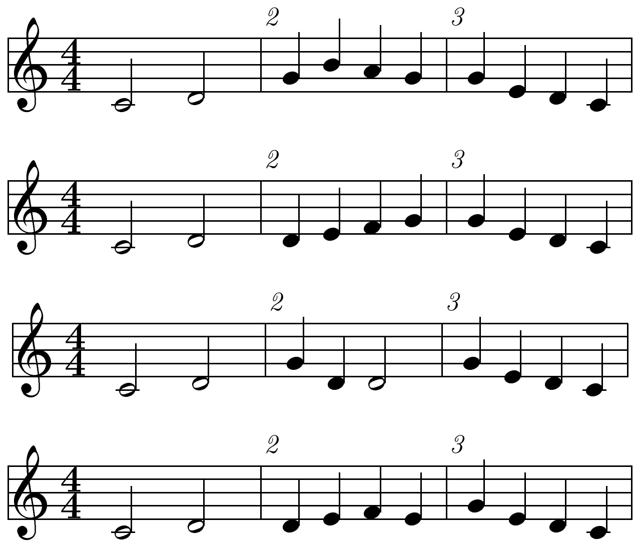



Figure 6: The Neutral Music Matrix



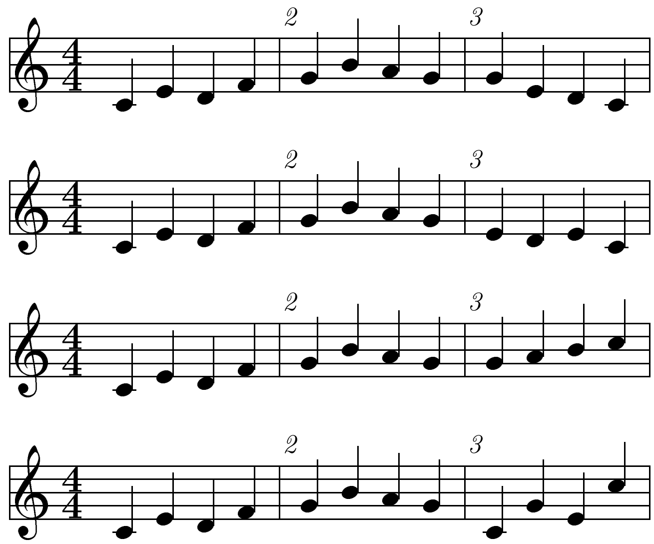

Phrase 111

Phrase 121

Phrase 131

Phrase 141

Figure 7: Neutral musical fragments that have been generated using the Neutral Musical Matrix using the first opening phrase, all the middle phrases and the first ending phrase. The first column of the figure provides a label of the form $xyz$ where $x$ indicates the opening used; $y$, the middle phrase used; and $z$, the ending phrase. Thus, 131 is the fragment built from the first opening,the third middle and the first ending.




Phrase 211

Phrase 212

Phrase 213

Phrase 214

Figure 8: Neutral musical fragments that have been generated using the neutral Musical Matrix using the second opening phrase, the first middle phrase and all the ending phrases. The first column of the figure provides a label of the form $xyz$ where $x$ indicates the opening used; $y$, the middle phrase used; and $z$, the ending phrase. Thus, 213 is the fragment built from the second opening,the first middle and the third ending.

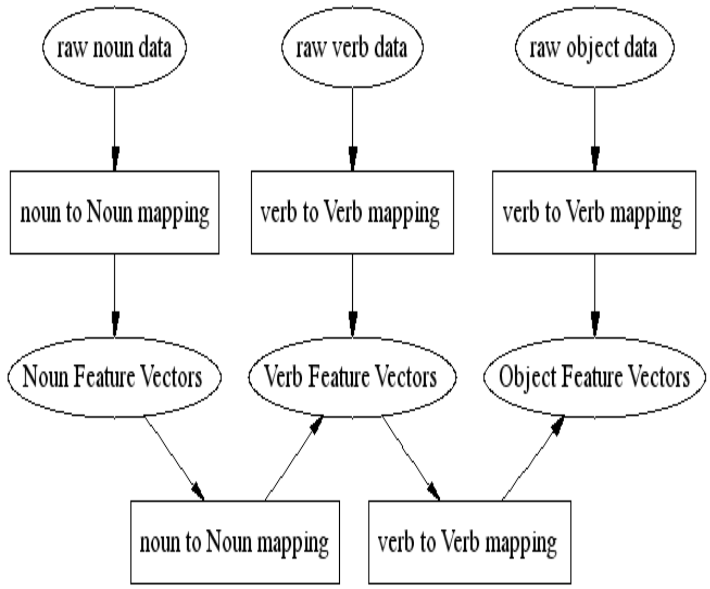



Figure 9: Raw Sentence to Feature Vector Processsing

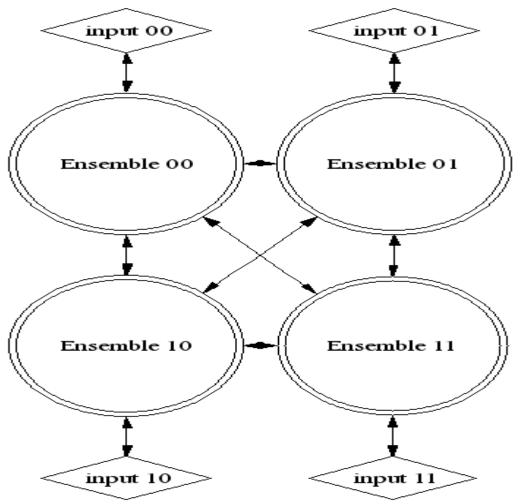



Figure 10: Four Neural Ensemble Structure

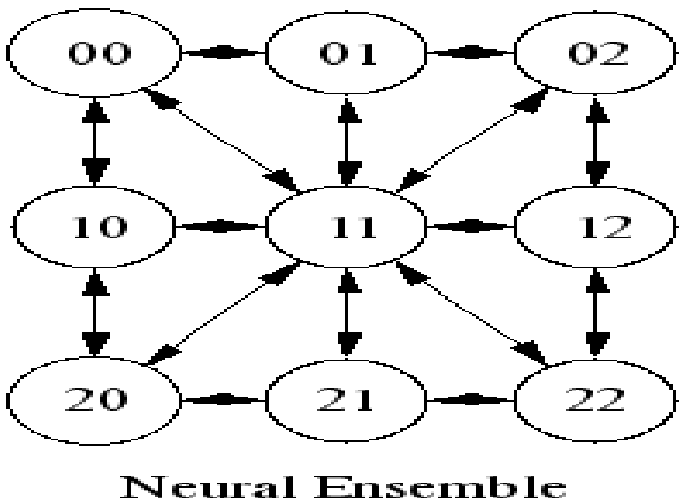



Figure 11: A Typical Neural Ensemble