

Cognitive Models of Music and Painting

James Peterson Department of Mathematical Sciences
 Clemson University, Clemson, SC 29634

Linda Dzuris Department of Performing Arts
 Clemson University, Clemson, SC 29634

Abstract

A simplified version of information processing in the brain minimally requires cortex and limbic processing modules. In this paper, we discuss how to construct training data for auditory and visual cortex models using a 17th century construction technique called a Würfelspiel matrix. Traditionally, this was used to develop many equally valid musical compositions for use as templates on which the actual composition of the musician would be based. For example, a musician would compose 10 openings, 10 transitions and 10 closings which would be fitted into a 10×3 matrix. From this a total of 10^3 musical prototypes could rapidly be assembled by combining an entry from column 0 to one from column 1 and then ending with an entry from column 2. The artistry of the musician was essentially captured in the 30 fragments which could be combined in such a combinatorial fashion. This idea can be used to develop musical data for use in training a model of auditory cortex. In addition, the notion of Würfelspiel matrices can be extended to the design of painting data for use in the training of the visual cortex. In this paper, we discuss how these specialized training sets provide us with the data to construct interesting cortical models which can eventually be used to create models of musical and painting composition. In addition, since the data can be generated with different emotional modalities, there is also a potential for building limbic processing modules.

1 Introduction

In a sequence of seminal papers ([9], [2]) and [3]), 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 an fMRI parameter. Hence, each image could be plotted on a two dimensional grid using the skin galvanic response as the horizontal axis and the fMRI response as the vertical component. Images with violent/ sexual content (extreme images) were always found to generate a response placing them far from

the origin of this coordinate system while neutral images such as those of an infant generated null or near origin results. Clearly, the emotional tags associated with these affective images were not cleanly separated into primary emotions such as anger, sadness and happiness. However, we can infer that the center (Null, Null) state was associated with images that have no emotional tag. Also, the images did cleanly map to distinct 2D locations on the grid when the emotional contents of the im-

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

Figure 1: Music Data Matrix

ages differed. 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. Using these ideas, we will design auditory (musical) and visual (painting) data for happy, sad, angry and

emotionally neutral states. In the 1700's, fragments of music could be rapidly prototyped by using a matrix \mathcal{A} of possibilities called a Würfelspiel matrix. We show an abstract version of a typical Musicalisches Würfelspiel matrix in Figure 1. It consists of P rows and three columns. In the first column are placed the opening phrases (nouns); in the third column, the closing phrases (objects); and in the second column; the transitional phrases (verbs). Each phrase consisted of L beats 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 chosen. We will modify the Würfelspiel matrix idea to generate both emotionally labeled musical and painting data to allow us to train a model of auditory cortex and visual cortex which can then be used to train a model of associative cortex to be used for sensor fusion.

2 Auditory Data

In the literature, there are many attempts to model musical compositional designs. A theorist such as (Caplin, [1]) discusses music in terms of large chunks or sections and overall function, while (Davie, [5])

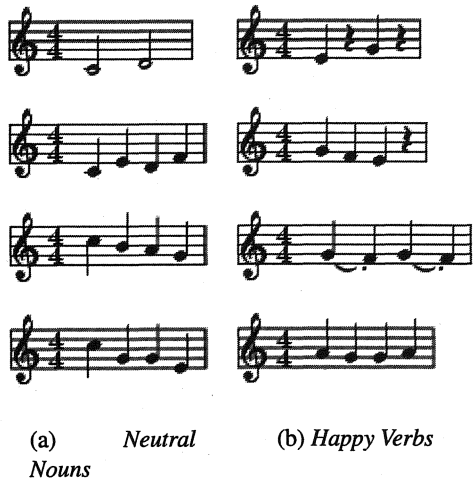


Figure 2: Music Data

discusses the matter of function from the opposite direction; small units building up to larger ones. It is very convenient to think of the structure of the Würfelspiel matrix shown in Equation 1 in terms of a traditional sentence structure by equating opening phrases to nouns, middle phrases to verbs and closing or cadence phrases to objects. Our full reasons for this choice of a grammatical infrastructure as well as our guiding principles for the generation of data matrices for neutral, happy, sad and angry music are detailed in ([6] and [7]). In the neutral music case, we use simple compositional patterns and ideas that are expressible using only quarter and half notes. 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. 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. Thus, our opening data gives 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.

To illustrate how these data samples might look, we show neutral music nouns and verbs are shown in Figure 2(a). The complete neutral Würfelspiel matrix would add the objects as well and would consist of four rows and three columns providing a total of 64 distinct musical fragments that are intended to model neutral musical sentence design. To generate musical data matrices with emotional labelings, the underlying goal in building each matrix was to remain as basic as possible. We therefore decided to work within a monophonic

texture; i.e. melody line only. Note values were restricted to quarter notes and half notes in quadruple meter. Quarter rests were also allowed, but used sparingly. All four matrices (neutral, happy, sad, angry) are structurally similar. Each consists of three columns with four fragment choices that are one measure in length. Any fragment from column one from any of the matrices is designed to function as an opening phrase. Note, in neutral music, we define a tonic center in the opening phrase, but in emotional music such as perceived as angry, our goal is to make the tonic center vague. Hence, we may or may not choose to start on a tonic center. For example, in the first opening in the angry matrix of Figure 3, we start on an A-flat. All fragments in column two of any of the matrices are then designed to function as a transition phrase.



Figure 3: Angry Music Matrix

As the label implies, these transition phrases serve as connectors between a choice from column one and a choice from column three. It is in these middle phrases that movement away from the tonic is made or continued. This movement is necessary for forward progress of a melody. Therefore, each transition phrase is now highlighting a secondary pitch, one other than the tonic note established by the opening phrase. To close our melodic lines, an ending phrase is chosen. Any fragment from column three of any of the matrices will function in the same manner. We designed each to move back to the tonic note in such a way as to produce a quality of closure to our melodic lines. To produce *emotion-deprived* or *neutral* fragments, individual characteristics

that have been documented by researchers as being contributing factors of basic emotion in music have been neutralized. Further, we use even rhythms and exact note durations in the neutral context. Fragments intended to be emotionally tagged as *happy* had individual characteristics which entail choosing a major mode, a very quick tempo of 250 and the use of staccato. The verbs of a typical happy matrix are displayed in Figure 2(b). Fragments tagged as *sad* use a minor mode with a slow tempo of 70 along with slurs, legato and using the bass clef to put us in a lower register. To emotionally tag fragments as *angry*, we use a minor mode, a moderate tempo of 180, faster than used for the sad melodies, but slightly slower than the tempo used for the happy melodies with increased variation of articulation (slurs, accents). There are also more repeated notes and the use of an ambiguous fragment where the mode is not clearly established in opening phrase. The musical data uses a rich set of notes and *articulation* attached to the notes to construct grammatical objects. We can think of the added articulation as punctuation marks as *slurs* (one note and multiple note), *staccato* and *marcato* accents are attached to various notes in our examples to add emotional quality. Our design alphabet can be encoded as $\mathcal{H} = \{c, d, e, f, g, a, b, r\}$ where each note in this alphabet is now thought of as a *musical object* with a set of defining characteristics. Here r is rest. For our purposes, the attributes of a note are choices from a small set of possibilities from the list $\mathcal{A} = \{p, b, s, a\}$. The index p indicates what pitch we are using for the note; the letter b tells us how many beats the note is held; the length of the slur is given by the value of s ; and a denotes the type of articulation used on the note. A given note n is thus a collection which can be denoted by $n_{p,b,s,a}$ where the attributes take on any of there allowable values. Our alphabet is thus \mathcal{H} which has cardinality 8. Each letter has a finite set of associated attributes and each opening, middle or closing phrase is thus a sequence of 4 musical entities. A typical angry music matrix is shown in Figure 3.

3 Visual Data /normalsize

A basic organizational plan of the human brain is presented in Figure 4. Auditory input goes to area 41 in the parietal cortex and visual input is sent to area 17 in the occipital cortex. This primary information

is processed further by areas 7 and 42 and areas 18 and 19 in the parietal and occipital cortex, respectively. The results of this processing are sent to the temporal cortex through areas 20, 21, 22 and finally 37.

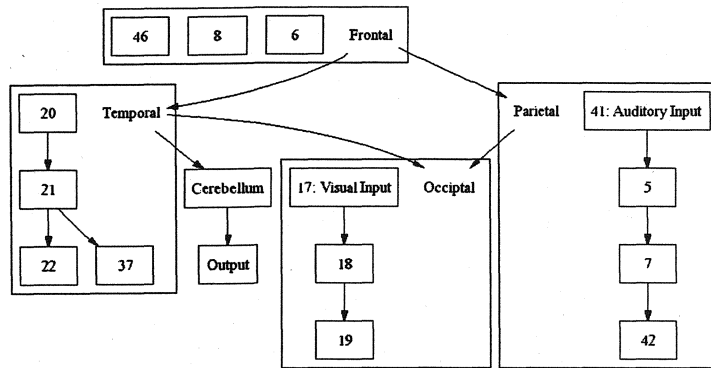


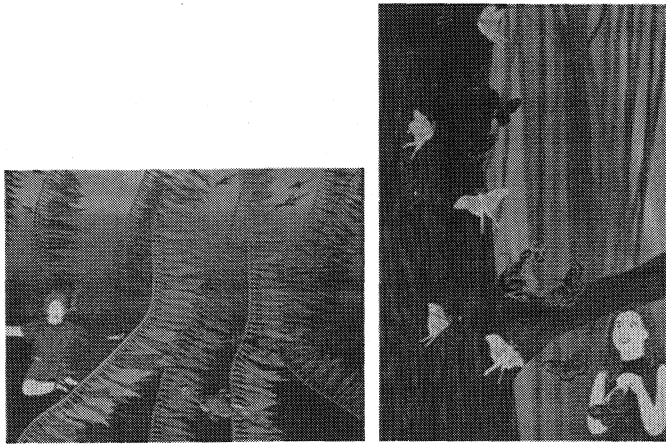
Figure 4: Brain Cortical Subdivisions

areas in each cortex that are devoted to secondary processing of primary sensory information and which are not connected directly to output pathways. These areas are called associative cortex and are primarily defined by function, not a special cell structure. In the parietal cortex, the association areas are 5 and 7; in the temporal cortex, areas 20, 21, 22 and 37; and in the frontal, areas 6 and 8. Hence, human brains have evolved to increase the amount of associative cortex available for what can be considered symbolic processing needs. Our ability to process symbolic information is thus probably due to changes in the human brain that have occurred over evolutionary time. It is noted in [8], that the increase in associative parietal cortex in area 7 probably occurred approximately 3 million years ago. Therefore, the capability of symbolic reasoning probably steadily evolved even though the concrete evidence of cave art and so forth does not occur until really quite recently. However, our point is that the creation of ‘art’ is intimately tied up with the symbolic processing capabilities that must underlie any model of cognition. The creation of appropriate sets of visual data is therefore essential to the training of a cognitive model. Our painting model uses a compositional scheme in which a valid painting is constructed by three layers: *background* (BG), *midground* (MG) and *foreground* (FG). A painting is assembled by first displaying the BG, then overlaying the MG which occludes some portions of the BG image and finally adding the FG image. The final FG layer hides any portions of the previous layers that lie underneath it. This simplistic scheme captures in broad detail the physical process of painting. When we start a painting, we know that if we paint the foreground images first, it will be technically difficult and aesthetically displeasing to paint midground and background images after the foreground. A classical example is painting a detailed tree in the foreground and then realizing that we still have to paint the sky. The brush strokes in the paint medium will inevitably show wrong directions if we do this, because we can not perform graceful side to side, long brush strokes since the facial foreground image is already there. Hence, a painter organizes the compositional design into abstract physical layers – roughly speaking, organized with the background to foreground layers corresponding to how far these elements are away from the viewer’s eye.

Consider the two paintings, Figure 5(a) and Figure 5(b), which have fairly standard compositional designs. Each was painted starting with the background and then successive layers of detail were added one at a time. As usual, the design elements farthest from the viewer’s eye are painted first. The other layers are then assembled in a farthest to nearest order. The painting seen in Figure 5(a) started with the background. This used a gradient of blue, ranging from very dark, almost black, at the bottom, to very light, almost white, at the top. There are, of course, many different shades and hues of blue as brushes are used to create interesting blending effects with the various blues that are used. However, we could abstract the background

Area 37 is where sensory information from multiple modalities is fused into higher level constructs. The top boundary of area 17 in the occipital cortex is marked by a fold in the surface of the brain called the lunate sulcus. This sulcus occurs much higher in a primate such as a chimpanzee. Effectively, human like brains have been reorganized so that the percentage of cortex allotted to vision has been reduced. Comparative studies show that the human area 17 is 121% smaller than it should be if its size was proportionate to other primates. The lost portion of area 17 has been reallocated to area 7 of the parietal cortex. There are special

to a simple blue background and capture the basic compositional design element. The many kelp plants are all painted in different planes. The kelp farthest from the viewer are very dark to indicate distance, while the plants closest to the viewer use brighter greens with variegated hues. We note that we could abstract the full detail of the kelp into several intermediate midground layers: perhaps, the farthest midground layer might be one kelp plant that is colored in dark green with the second, closest midground layer,



(a) *Painting One*

(b) *Painting Two*

Figure 5: Two Paintings

There are then two more midground images: the whimsical dragon figure on the tree branch and the human figure positioned in front of the tree. Finally, there are a large number of Baltimore butterflies and Luna moths which are essentially foreground images.

a bright green kelp plant. The human figure is placed between kelp layers, so we can capture this compositional design element by placing a third midground layer between the two midground kelp plant layers. Finally, there are many seadragons in foreground layers at various distances from the viewer. We could simplify this to a single foreground layer with one seadragon painted in a bright red. Hence, the abstract compositional design of the painting in Figure 5(a) is as shown in Figure 6(a). In a similar fashion, we can analyze Figure 5(b). The background in this painting is a large collection of softly defined trees. These are deliberately not sharply defined so that they seem to be far from the viewer. We can abstract this compositional design as shown in Figure 6(b). The midground image is the very large tree that runs from the bottom to the top of the painting.

Layer	Description
Background One	Blue gradient; dark to light
Midground One	Very dark green kelp plant
Midground Two	Human figure
Midground Three	Bright green kelp plant
Foreground	Bright red sea dragon

(a) *Abstract Seadragons Design*

Layer	Description
Background One	Fuzzy brown trees
Midground One	Large tree (brighter browns)
Midground Two	Dragon (red)
Midground Three	Human figure
Foreground	Butterfly (black); moth (green)

(b) *Abstract Tree Painting Design*

Figure 6: Abstract Painting Designs

The paintings shown in Figure 5(a) and Figure 5(b) are much more complicated than the simple abstract designs. However, we can capture the essence of the compositional design in these tables. We note that, in principle, a simpler description in terms of one background, one midground and one foreground is also possible. For example, we could redo the abstract designs of Figure 5(a) and Figure 5(b) as shown in Figure 7(a) and Figure 7(b). These new designs do not capture as much of the full complexity of the paintings as before, but we believe they do still provide the essential details. Our simple painting model is thus based on Würfelspiel matrices similar to those used in music compositions with a painting (BG, MG, FG) constructed to give the overall impression of a given emotional state.

Layer	Description
Background	Blue gradient; dark green kelp
Midground	Human figure; bright green kelp
Foreground	Bright red sea dragon

(a) *The Three Element Abstract Seadragons Design*

Layer	Description
Background	Fuzzy brown trees
Midground	Large tree; dragon
Foreground	Human figure; butterfly; moth

(b) *The Three Element Abstract Tree Painting Design*

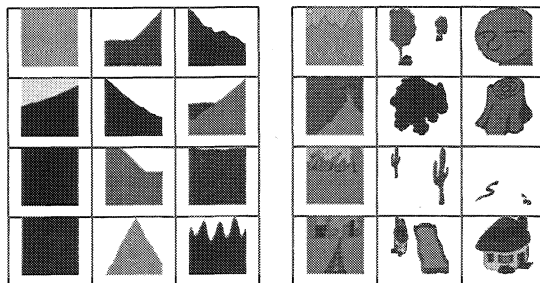
Figure 7: Three Layer Abstractions

The Würfelspiel matrix we obtain from four kinds of neutral background, midground and foreground images is shown in Figure 8(a) and a happy matrix constructed in the same way is shown in Figure 8(b). Our abstract painting compositions can be encoded as the triple $\{b, m, f\}$ where b denotes the background, m , the midground and f the foreground layer, respectively. Each of these layers is then modeled with a collection of graphical objects with the following attributes: inside color, c_i ; boundary color, c_b ; and a boundary curve, $\partial\Omega$, described as an ordered array $\{(x_i, y_i)\}$ of position coordinates. We can then use this alphabet to encode Würfelspiel painting matrices into data for use in training the visual cortex of the cognitive model.

4 Conclusions

It follows from the discussion in Section , that a reasonable sensor fusion model will require models of cortical processing. For example, in [10], it is noted that the first layer of auditory cortex is bathed in an environment where sound is chunked or batched into pieces of 200 mS length which is the approximate size of the phonemes of a person’s native language. Hence, the first layer of cortex develops

circuitry specialized to this time constant. The second layer of cortex then naturally develops a chunk size focus that is substantially larger, perhaps on the order of 1000 mS to 10000 mS . As processing is further removed from the auditory cortex via mylenated pathways, additional meta level concepts (tied to even longer time constants) are developed. We will therefore model auditory and visual cortex with three layers based on cortical models proposed in [12]. Our third layer of cortex is then an abstraction of the additional anatomical layers of cortex as well as appropriate mylenated pathways which conduct upper layer processing results to other cognitive modules. We will use the musical data to imprint the first two layers of our model of auditory cortex and the painting data to imprint the first two layers of our visual cortex models. The first step in building our models is to use the musical and painting



(a) *A Neutral Matrix*

(b) *A Happy Matrix*

Figure 8: The Neutral and Happy Matrix

to constrain or “train” a model of the associative cortex. In general, our model takes this specialized sensory input and generates a high level output. The musical data provides the kind of associated output that might come from area 37 of the temporal cortex. The low level inputs that start the creation of a music phrase correspond to the auditory sensory inputs into area 41 of the parietal cortex which are then processed through areas 5, 7 and 42 before being sent to the further associative level processing in the temporal cortex. The painting data then provides a similar kind of associated input into area 37 from the occipital cortex. Inputs

that create the paintings correspond to the visual sensory inputs into area 17 of the occipital cortex which are then further processed by area 18 and 19 before being sent to the temporal cortex for additional higher level processing. The musical and painting data are currently being used to generate models of compositional design for music and painting and more general models of cognition. High level details of the cortical modeling process are presented in [11].

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. Further, we thank Quinn Peterson for his careful development of the painting matrices.

References

- [1]. W. Caplin, **Classical Form: A Theory of Formal Functions For The Instrumental Music of Haydn, Mozart and Beethoven**, Oxford University Press, 1998.
- [2]. M. Codispotti, M. Bradley and P. Lang, *Affective reactions to briefly presented pictures*, *Psychophysiology*, 38, pages 474 – 478.
- [3]. B. Cuthbert, M. Bradley and P. Lang, *Probing Picture Perception: Activation and Emotion*, *Psychophysiology*, 33, 1996, pages 103 – 111.
- [4]. M. Conkey, *A History of the Interpretation of European ‘Paleolithic art’: magic, mythogram, and metaphors for modernity*, in A. Lock and C. Peters, editors, **Handbook of Human Symbolic Evolution**, Blackwell Publishers, Massachusetts, 1999.
- [5]. C. Davie, **Musical Structure and Design**, Dover Publications, Inc., 1953.
- [6]. L. Dzuris and J. Peterson, *Data Abstraction In Cognitive Models for Compositional Design in Music*, Technical Report, 2003.
- [7]. L. Dzuris and J. Peterson, *Emotionally Tagged Models for Compositional Design in Music: Data Abstraction*, submitted to *Journal of New Music Research*, 2003.
- [8]. R. Holloway, *Evolution of the Human Brain*, in A. Lock and C. Peters, editors, **Handbook of Human Symbolic Evolution**, Blackwell Publishers, Massachusetts, 1999.
- [9]. P. Lang, M. Bradley, J. Fitzmmons, B. Cuthbert, J. Scott, B. Moulder and V. Nangia, *Emotional arousal and activation of the visual cortex: An fMRI analysis*, *Psychophysiology*, 35, 1998, pages = 199 - 210.
- [10]. M. Merzenich, *Cortical Plasticity Contributing to Child Development*, in J. McClelland and R. Siegler, editors, **Mechanisms of Cognitive Development: Behavioral and Neural Perspectives**, pages 67 – 96, Lawrence Erlbaum Associates, Publishers, 2001.
- [11]. J. Peterson, *Polymodal Information Processing Via Temporal Cortex Area 37 Modeling*, in Kevin Priddy, editor, **Intelligent Computing: Theory and Applications II**, Proceedings of the SPIE, Volume 5421, 2004, pages 149 - 160.
- [12]. R. Raizada and S. Grossberg, *Towards a Theory of the Laminar Architecture of Cerebral Cortex: Computational Clues from the Visual System*, *Cerebral Cortex*, pages 100 – 113, 2003.

