In E. Miranda (Ed.), *Readings in Music and Artificial Intelligence*. Amsterdam: Harwood Academic Publishers (2000).

Symbolic AI Versus Connectionism in Music Research

Petri Toiviainen
Department of Musicology
University of Jyväskylä

Abstract

In cognitive science and research on artificial intelligence, there are two central paradigms: the symbolic and the analogical. Within the analogical paradigm, interest in artificial neural networks, or connectionism, has experienced a resurgence during the last decade; this change has also been reflected in the field of musical modeling. This article provides a general survey of the relationship between symbolic AI and connectionism, both on a general level and from the point of view of music research. This is followed by a short introduction to artificial neural networks, which includes a description the main principles of their structure and function as well as a presentation examples of their use in the field of music.

Introduction

During the last few decades, attempts to model human musical activity have mostly been based on traditional AI techniques, which rely on logical manipulation of symbols. While a great deal of impressive results have been obtained within this paradigm (see, e.g., Balaban, Ebcioglu & Laske, 1992), these studies have failed to shed light on certain important areas of music cognition, such as those related to perception, motor action, and performance interpretation. Depending critically upon verbalization and introspection, they have proven ineffective for the investigation of the inarticulate aspects of musical activity. Since the 80s, connectionism, or modeling with artificial neural networks, have gained popularity among music researchers as a tool for exploring such tacit musical knowledge.

Artificial neural networks (ANNs), also referred to as connectionist or parallel distributed processing (PDP) systems, can be characterized as strongly idealized models of networks formed by biological neurons: although the basic principles of their mechanisms and structures have been adopted from biological neural networks, they are generally not intended to model the physiological processes in the neural tissue. Technically, they are nonlinear dynamical systems consisting of a multitude of simple, interconnected processing units. They have three important properties. First, they are parallel, i.e., the processing units interact simultaneously and independent of each other. Second, they are distributed, i.e., their knowledge resides in the strengths of interneuronal connections; and the data manipulated by them are represented as patterns of neuronal activation. Third, they are adaptive, i.e., when exposed to data from the environment, they are capable of learning by adjusting the strengths of their interneuronal connections. Development of mathematical models of ANNs began four decades ago with the work of, e.g., McCulloch and Pitts (1943), Hebb (1949), and Rosenblatt (1959). More recent work by Hopfield (1984), Rumelhart and McClelland (1986), Grossberg

(1986), Kohonen (1989), and others has led to a resurgence of interest in the field.

This paper is organized as follows: in the first part, the paradigms of symbolic AI and connectionism are compared in detail. This includes comparing the premises and discussing the strengths and weaknesses of the two approaches. The consequences of the latter to music research is discussed, i.e., in which sectors of music research connectionism may be a better choice than symbolic AI. The next section will provide a short introduction to ANNs. This is followed by a review of connectionist research on music, with pointers to some important references. Finally, the work of the present author on timbre recognition with Kohonen self-organizing maps is briefly described.

Relationships Between Symbolic AI and Connectionism

Premises

The premises of symbolic AI and ANNs are fundamentally different. The paradigm of symbolic AI is based on the standard von Neumann-style serial processing under the control of a single powerful central processing unit, whereas ANNs employ brain-style computation, i.e., simultaneous interaction of a large number of simple processing units. Moreover, symbolic AI systems function by applying inference chains to logical variables — for instance, "X mother of Y" & "Y parent of Z" ⇒ "X grandmother of Z" — while ANNs apply evolutive rules to numerical variables. To put it in an other way, symbolic AI relies on logic, while ANNs rely on (nonlinear) dynamics. AI systems work under a centralized control, i.e. a collection of so-called meta-rules used to determine which rules of inference (or production rules) to apply in a given computational stage. An ANN lacks such a centralized control: all its units are simultaneously interacting.

With respect to the representation of data, these two paradigms differ significantly from each other. Symbolic AI systems tend to operate on a high conceptual level, using dedicated symbols to represent each concept. They thus adopt a local representation. It is possible to use such a representation in ANN models also. According to a general view, however, a distributed (or subsymbolic) representation seems to better exploit the strengths of ANNs than a local one. A distributed representation can be defined from several points of view. For instance, it may imply representing the data as a vector, each component of which stands for a specific microfeature of the represented domain. Or, it may denote representing many items at once over the same set of processing units or connection weights. It must be noted that there is no clear distinction between local and distributed representation, but rather there exists a continuum between these two extremes. If an item is represented as a collection of microfeatures, each of these features can again be represented either in a local or a distributed manner. For example, the most local way of representing chords in an ANN is to designate one neuron for representing each chord. A more distributed representation is obtained by representing each chord as the tones it is composed of. The degree of distribution can be further increased by representing each of those tones as a harmonic or subharmonic series of frequencies.

Knowledge, like data, is represented differently in the two approaches. Schematically, symbolic AI utilizes explicit knowledge represented in a local way, while ANNs employ distributedly represented implicit knowledge. Symbolic AI systems rely on sets of predefined inference rules provided by an external programmer. In ANNs, on the other hand, knowledge resides in a distributed form in the strengths of interneuronal connections. While these connection strengths can also be hard-wired by an external programmer, a more usual way of establishing them is using suitable learning algorithms together with a set of training data. ANNs, thus, build their knowledge by means of automated learning.

Strengths and weaknesses

Due to their fundamentally different premises, symbolic AI systems and ANNs display rather unlike qualities. It can be stated that ANNs have proven effective in tasks were symbolic AI systems have problems, and vice versa. The main strengths of ANNs can be summarized as follows (McClelland, Rumelhart & Hinton, 1986; Rumelhart & McClelland, 1986; Gutknecht, 1992):

- *Learning capability*. ANNs are usually not programmed but trained by presenting them with examples. On the basis of those examples, they adapt to their environment by means of specific learning algorithms. The use of automated learning reduces the need for "knowledge engineering". The latter means the process of defining the representation formalisms, data structures and correct rules to be used, and is a crucial task in developing a symbolic AI application. Knowledge engineering is often difficult, unnatural, and error prone, because it requires making explicit of the knowledge which is often implicit. It is also time-consuming and thus expensive. It must be noted that within symbolic AI, research in automated learning is an area of great interest. The learning techniques developed so far in this field, however, require a very accurate choice of examples and often do not tolerate contradictory examples. When used, for instance, for composing music, ANNs can free the composer from the often tedious enterprise of defining the rules which describe a given style of music.
- *Generalization capability*. ANNs are capable of extracting significant features from the training set and using them to process a novel input pattern. For instance, an ANN trained to recognize the timbre of a musical instrument is probably capable of recognizing it even if it is played with a slightly different articulation.
- *Tolerance towards noise and contradictions*. ANNs are robust with respect to the presence of noisy or contradictory data. In the field of music this is a valuable feature, since musical phenomena very often do not follow any consistent set of rules.
- *Tolerance towards overloading of information*. ANNs do not have a fixed storage capacity: when a network is overloaded with input data, similar components of information tend to blend together, resulting in generalization of features.

Some of the main problems of the connectionist approach are listed below (Serra & Zanarini 1990; Gutknecht, 1992);

- *Limitation to "toy-size" problems*. At least in the domain of music, ANNs have so far been applied only to small, strongly stylized problems. The application to real size problems is only in its initial phases, one notable example being the work of Leman & Carreras (1996).
- *Difficulties with long inference chains*. Symbolic AI systems are built specifically for dealing with this type of reasoning, and it is not realistic to expect ANNs to be as efficient in this sector.
- *Limited explanation capabilities*. Attempts have been made to explain the behavior of connectionist networks by, for instance, analyzing the structure of the connection strength matrix learned. These explanations are, however, at the level of primitive features of the network, such as activation functions and energy landscapes. Explanations on a higher level of knowledge are difficult to achieve. An ANN can be trained to harmonize melodies in a given style, but it is usually difficult to analyze why the network arrived at a particular solution.
- *Difficulties with structured representation*. Structured knowledge, such as concept hierarchies or inference nets, is difficult to represent in ANNs, contrary to traditional AI models. This shortcoming manifests itself, for instance, in ANNs used for algorithmic composition: while they are capable of

successfully capturing the surface structure of a melodic passage and produce new melodies on the basis of the thus acquired knowledge, they mostly fail to pick up the higher-level features of music, such as those related to phrasing or tonal functions.

The Appeal of Connectionism in Music Research

ANNs seem to be better suited than symbolic AI systems for dealing with low level forms of knowledge, such as analysis and recognition of signals. Many tasks of musical activity involve low level processing of often noisy or distorted sound data. These include, for instance, the perception of pitch and timbre, and the localization and segregation of sound sources. Attempts to model these kinds of processes with traditional AI systems may prove to be intricate: in order to be capable of properly dealing with noisy or distorted input data, such models should probably be equipped with a multitude of rules. Even relatively simple ANNs, on the other hand, have been shown to tolerate noise and distortions present in such low level tasks of musical cognition.

From the point of view of modeling musical cognition, ANNs are appealing because they provide a convincing model of learning. They thus make it possible to simulate both the action of a musical system and its development. With self-organizing networks, for instance, it is possible to shed light on how schemata of perceptual learning develop through adaptation to the musical environment (Leman, 1995). Used in combination with advanced auditory modeling techniques, they enable one to study music starting from acoustical signals instead of symbols. This makes it possible to adopt the attitude of ecological modeling, where the programmers restricts their role to specifying the interactions between the musical environment and the model, incorporating the environment into the computer, and formalizing the dynamics (Leman 1996). This is in contrast with the classical modeling situation, where the programmers are situated between the world and the model, i.e., they interpret the musical environment, form a representation of it, and use this knowledge to explicitly formalize a conceptual structure.

Connectionist methods can also be utilized in studying higher level musical activities. To construct a rule-based expert system of, e.g., composition, analysis, or performance of music, an expert is needed who can verbalize the rules which define the solutions of the problem in question. These rules must be correct and consistent. A large part of our musical activities is, however, not verbalizable. The ability to play music, for instance, is learned to a great extent by example or through mimicking other players, rather than through memorizing explicit rules concerning the musical style in question. Furthermore, musicians and composers themselves often find it difficult or even impossible to analyze their own works. ANNs, being capable of extracting implicit knowledge from examples, may offer a more plausible alternative for modeling these kinds of musical processes.

A Short Introduction to ANNs

An ANN consists of a set of simple processing units (nodes, neurons) which are linked to each other by a set of weighted connections (see Fig. 1). Each unit receives numerical inputs from other units and produces a single numeric output, which is usually some simple nonlinear function of the inputs. The output is then passed on the output connections to still other neurons for further processing. What makes an ANN capable of performing interesting computations is that each connection is associated with a connection strength which indicates how much influence the sending unit has on the receiving unit. Depending on whether the connection strength is positive or negative, the influence can be characterized as excitatory or inhibitory, respectively.

Units within a network can be divided into three basic categories: input, hidden, and output units. The role of input units is to receive information from the

environment; the activation pattern of the output units corresponds to the final result of the network's processing. Units belonging to neither of these categories are called hidden units; they have an important role in the representation of knowledge in the network.

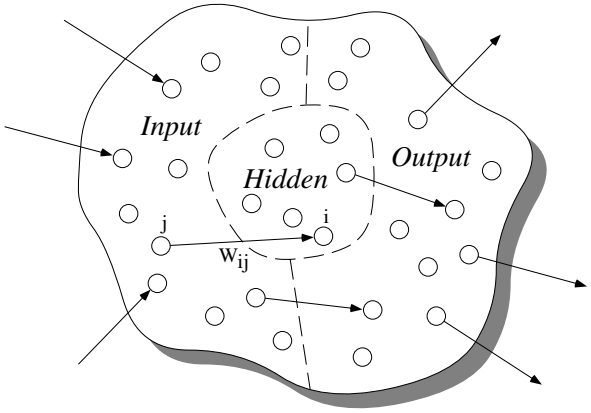

Figure 1. Generalized structure of an ANN, showing the input, hidden, and output neurons and some interneuronal connections.

The vector of activation values of all the units of the network at time *t*, $\mathbf{A}(t) = (a_i(t))$, corresponds to the information currently being processed by the network, whereas the matrix of connection strengths, $\mathbf{W}(t) = (w_{ij}(t))$, represents the current state of knowledge of the network. The network can adapt to the environment by adjusting its connection matrix; various learning algorithms have been developed for this purpose.

A variety of types of ANNs have been developed for a wide range of purposes. They can be categorized, for instance, on the basis of their architecture, the dynamics, the type of data they process, or the learning algorithm. Two important partitions will be briefly discussed here, namely dynamic vs. static networks and supervised vs. unsupervised learning.

The behavior of static networks is characterized by equations that are memoryless. Their output is a function of the current input only, not of past inputs or outputs. This category embraces, for instance, the perceptron (Rosenblatt, 1958) and the multilayer perceptron (Rumelhart, Hinton & Williams, 1986) network. The latter, also referred to as the back-propagation network, is perhaps the most used and studied type of ANN. In such a network, the neurons are organized in layers — an input layer, one or more hidden layers, and an output layer — and have feedforward connections from each layer to the next one (see Fig. 2.a). A non-linear activation function (see Fig. 2.b) is necessary at least in the hidden layer(s) in order to obtain good performance; a common choice is the sigmoid function, defined by

$$f(\alpha) = \frac{1}{1 + e^{-(\alpha - \theta_i)}}, \qquad (1)$$

where α is the input and θ_i the internal threshold of node i. Such a network is usually trained with the back-propagation of error algorithm (Rumelhart, Hinton & Williams, 1986), in order to produce a desired mapping from the input space to the output space. This algorithm is actually an optimization procedure: the goal is to minimize the error between desired and obtained outputs by means of the gradient descent method (see Fig. 3).

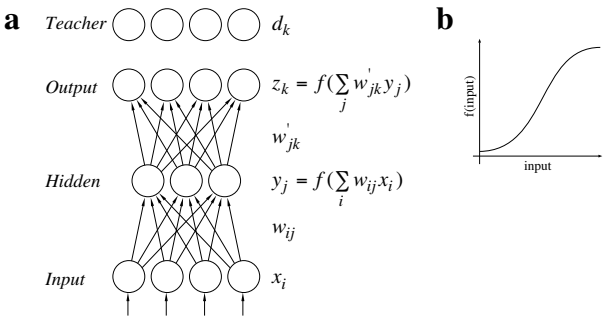

Figure 2. a) Structure of a multilayer perceptron network. In the formulas, x_i, y_j, and z_k are the activation values of the neurons in input, hidden, and output layers, respectively. The teacher layer contains the current desired output, d_k, during the training phase. b) A typical nonlinear activation function.

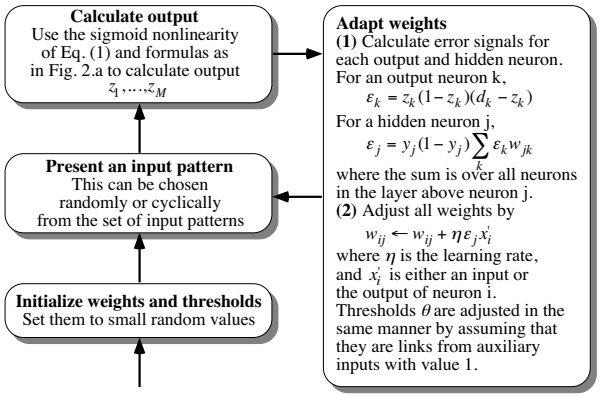

Figure 3. The back-propagation of error algorithm, used for training a multilayer perceptron network. For the explanation of mathematical symbols, see Fig. 2.

Dynamic networks are systems with memory. Their dynamical behavior is described by differential or difference equations. Dynamic networks can be further divided into two subcategories: networks with output feedback and those with state feedback.

Networks with output feedback can be used for processing sequential data. Under this category belongs the network architecture proposed by Jordan (1986): a multilayer perceptron network which has feedback connections from the output layer to the input layer. Variants of this architecture have been used for musical applications (see below).

Networks with state feedback are typically single-layer networks with feedback connections between nodes. In the most extreme case the neurons are completely interconnected, i.e., every neuron is connected to every other. Networks with state feedback can be used, for instance, for pattern completion through auto-association. When being presented with a noisy or distorted version of a memorized pattern, these networks are capable of recollecting the original pattern by means of a relaxation process. Examples of networks with state feedback are the interactive activation model (Rumelhart, Hinton & McClelland, 1986), Hopfield network (Hopfield, 1984), and Boltzmann machine (Hinton & Sejnowski, 1986).

ANNs can be further categorized, on the basis of the type of learning they employ, into supervised and unsupervised learning networks. The networks discussed so far belong to the former category: they are intended to produce a desired output from a given input. The latter category, unsupervised learning networks, is probably the most interesting class of ANNs, because their way of extracting knowledge about the environment resembles that of biological systems. Such networks are presented with only input samples; these are grouped into classes which are self-similar through a process referred to as self-organization. Examples of self-organizing networks are the Adaptive Resonance Theory (ART) networks — ART 2 (Carpenter & Grossberg, 1987) and ART 3 (Carpenter & Grossberg, 1990) — and the Kohonen self-organizing map (Kohonen, 1989). Whereas the ART networks perform an automatic categorization of the input data set, the Kohonen network maps the input vectors onto a two-dimensional surface while retaining their topological relationships.

Connectionism in Music Research: Some Pointers to the Literature

For the reader wishing to learn more about connectionist research on music, collections by Todd & Loy (1991) and Griffith & Todd (1994) provide good overviews of the field.

General discussions

General directions of connectionist modeling of music have been outlined at least by Bharucha (1988), Leman (1988, 1989), Lischka (1991), and Loy (1991). Bharucha (1988) discusses several models of music cognition, including a constraint satisfaction network for Western harmony, an auto-associative network simulating cross-cultural differences in tonal implications, and back-propagation networks that learn sequential musical schemata and specific musical sequences. Leman deals with the question how sequential musical information can be stored and processed in a connectionist network (1988), and outlines a general background for the appli-

cation of connectionist systems to music (1989), as well as presenting examples of musical applications of spreading activation networks, constraint-satisfaction networks, supervised learning networks, and self-organizing networks. A survey of current research paradigms in cognitive musicology is provided by Lischka (1991), followed by a critique of their basic assumptions and a suggestion for an alternative, more biologically-based, approach. Loy (1991) presents an informal overview of some of the traditional interests and problems of music research in the computer music community and describes the influence of connectionist theories on them.

Specialized topics

Taylor & Greenhough (1994) have modeled the perception of pitch. They utilize a self-organizing network architecture called ARTMAP, which is based on adaptive resonance theory (ART) networks. According to them, their model is capable of developing a great insensitivity to phase, timbre, and loudness when classifying pitch.

Connectionist studies on the perception of harmony, key, or tonality have been carried out by a number of researchers. Leman (1990, 1991) outlines a model for the study of the ontogenesis of tonal functions. He uses a distributed representation of chords, based on Terhardt's psychoacoustical theory of tone perception (Terhardt, Stoll & Seewann, 1982), and the Kohonen self-organizing map (Kohonen, 1989). The network is found to organize basically in terms of the circle of fifths. While chords in the aforementioned studies are considered as static, time-independent objects, in later studies Leman (1992a, 1992b) adopts a dynamic approach: from a stream of acoustic input data, he creates a tone context by temporal integration. When trained on the thus obtained input vectors, the Kohonen map is found to develop a response structure which correlates strongly with Krumhansl's (1990) psychological data. This approach has recently been extended to a realistic environment (Leman & Carreras, 1996) with the use of a recording of Book 1 of Bach's Das Wohltemperierte Klavier as input data.

For timbre recognition and classification, connectionist systems have been designed by De Poli, Prandoni & Tonella (1993), Cosi, De Poli & Lauzzana (1994), and Feiten & Günzel (1994). De Poli et al. (1993) used a three-dimensional version of the Kohonen map. As input they used the sound stimuli of Grey (1975), in order to reconstruct Grey's timbre space. Cosi et al. (1994) used an auditory model and the Kohonen network to map the sounds of 12 acoustic instruments in both clean and noisy conditions. The obtained map showed a topological organization which was found to agree with subjective classification of those sounds. Moreover, the Kohonen network was able to recognize noisy versions of the sounds. Employing two hierarchical Kohonen networks, Feiten & Günzel (1994) treated dynamic sounds as sequences of steady-state components. The first Kohonen network mapped the steady-state spectra; the trajectories obtained were then used as input to the second Kohonen network. The work of the present author on timbre is described below.

Desain & Honing (1989) have developed a connectionist quantizer which is capable of inferring the meter from input data containing timing variations. Their model works to adjust perceived inter-onset intervals so that every pair of these intervals is adjusted toward an integer ratio, if it is already close to one. Large & Kolen (1994) introduce a novel connectionist unit capable of phase- and frequency-locking to periodic components of incoming rhythmic patterns. Networks of these units can self-organize temporally structured responses to rhythmic patterns.

Todd's (1989) approach to algorithmic composition is based on Jordan's (1986) sequential network architecture: a three-layer back-propagation network processing one note at a time, with feedback connections from the output layer to the input layer. Temporal context is provided by integrating the activation values of

the input units. The compositions produced by this network suffer from lack of global structure; to overcome that, Todd (1991) suggests an architecture with two hierarchically connected sequential networks. Mozer (1991, 1994) utilizes a recurrent network trained by a variation of the back-propagation of error algorithm, referred to as the unfolding of time procedure (Rumelhart, Hinton & Williams, 1986), for note-by-note composition. In his model, he uses the distributed representation of pitch suggested by Shepard (1982): pitch is represented in a five-dimensional space. In this space, each pitch is specified by the pitch height as well as points on the chroma circle and the circle of fifths.

Self-organizing models of the perception and generation of musical sequences have been proposed by Page (1994) and Kaipainen (1994). Page (1994) criticizes previous connectionist approaches to algorithmic composition for being inappropriate as models of perception. His own approach is based on hierarchical ART 2 networks (Carpenter & Grossberg, 1987) furnished with masking fields (Cohen & Grossberg, 1987). Having been trained on simple nursery-rhyme melodies, the network was probed with short musical sequences; the elicited musical expectations were found to correspond strongly with those suggested by the training set. Kaipainen (1994) utilizes his self-organizing model, MuSeq, for demonstrating his dynamic theory of musical knowledge ecology. He postulates two directions of interaction, i.e., knowledge-acquisition and knowledge-use, and two kinds of knowledge, i.e., ''knowing-what'' for the recognition of the current musical situation and ''knowing-how'' for the determination of the consequences that actualize music.

Timbre Classification by the Kohonen Self-Organizing Map

A more detailed description of this work can be found in Toiviainen, Kaipainen & Louhivuori (1995) and Toiviainen (1996).

The Kohonen map

In the central nervous system there is a tendency to reduce the dimensionality of the incoming data. It is possible to identify various kinds of ordered feature maps (e.g., somatosensory maps connected with the sense of touch and the movement of the muscles; tonotopic mapping in the primary auditory cortex; and log-polar mapping of the retina onto primary visual cortex). The feature maps are compressed representations of the observed signals, containing information about the most relevant features and their interrelationships.

It is commonly believed that the cortical feature maps originate from self-organization. There exists a well demonstrated computational theory of self-organization (Kohonen, 1989, 1995), which is based on the assumption that lateral inhibition and redistribution of synaptic resources are responsible for self-organization in biological systems. Kohonen has formalized his theory of self-organization into a simple, yet effective, numerical algorithm. Given a set of input vectors in a multidimensional vector space, the Kohonen self-organizing map (KSOM) identifies the most salient features, i.e., the dimensions with highest variance, of the input set, and maps those features onto a two-dimensional space, while retaining the topological relationships of the input vectors.

The KSOM consists of (1) n input neurons, each having a specified activation level v_i. The input to the network is, thus, an n-dimensional vector $\mathbf{v} = (v_1,\ldots,v_n)$; (2) m output neurons receiving activation from the input neurons. The output neurons usually form a planar array. Each output neuron is identified by its location in the array; and (3) connections from each input neuron to each output neuron (see Fig. 4). A weight w_{ij} is associated to the connection from input neuron i to output neuron j. The connections to output neuron j can thus be represented by an n-dimensional vector $\mathbf{w}_j = (w_{1j},\ldots,w_{nj})$. There are several variants of

the Kohonen learning algorithm; a frequently used version is presented in Figure 5.

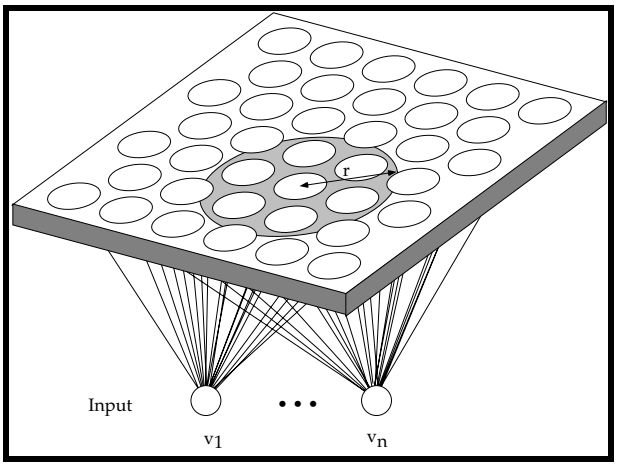

Figure 4. Structure of a Kohonen self-organizing map. The grey circle depicts a topological neighborhood of an output neuron with radius *r*.

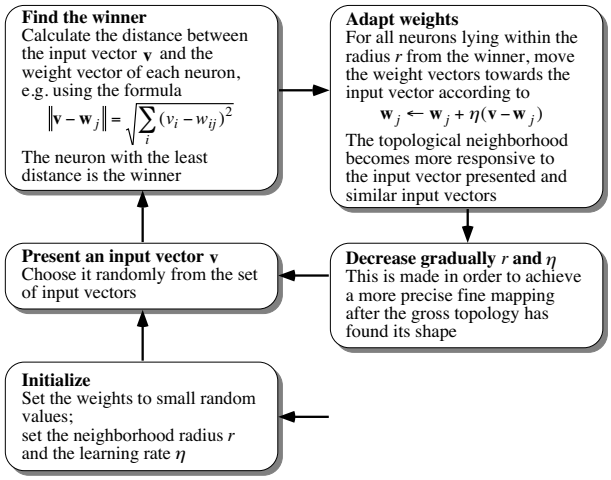

Figure 5. The Kohonen learning algorithm.

Timbre

Timbre is defined by the American Standards Association as ''that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar''. This definition is not very useful, since it actually defines what timbre is not rather than what it is.

There have been a number of studies aiming at extracting the most salient acoustic attributes affecting the perception of timbre. A widely used method is similarity rating (SR): subjects are asked to rate, on a given scale, the similarity of all possible pairs in the set of stimuli. Multidimensional scaling (MDS) is then used to map the tones into a low-dimensional space — frequently referred to as the timbre space. By examining MDS maps it has been found that the spectral energy distribution in the steady-state portion of a tone is one of the main contributors to the perception of timbre; also dynamic attributes, mostly in the onset portion of a tone, have found to have an important role.

A timbre space can also be constructed from a set of acoustical signals by means of connectionist models, e.g., the KSOM. In such models, it is necessary to extract the most significant parameters of the incoming sound signal by means of a preprocessing stage; this can be based on such methods as the Short-Time Fourier Transform, Cepstrum, Linear Predictive Coding or Advanced Auditory Modeling.

Many experiments have demonstrated that, in the cochlear nuclei and the inferior colliculus, there are cells which give particularly strong responses to certain dynamic features of time-varying signals, such as frequency and amplitude modulations as well as onsets and offsets on a given frequency range. It is unclear, however, which of these features, and to what extent, are important in timbre perception and identification.

Goals and methods

The goals of the study were

- to examine how different preprocessing strategies affect the final configuration of a self-organized timbre map. This was carried out by constructing sets of auditory images, in which the degree of emphasis on the onset of tones was varied, and combining spectral and gradient images, i.e., images which were supposed to qualitatively represent responses of FM- and AM-sensitive neurons as well as neurons encoding the spectral gradient.

- to examine how different distance metrics used in the training phase affect on the final configuration of a self-organized timbre map.
- to explore to what extent the KSOM is capable of maintaining the metrical relations between the used sound stimuli when projecting them onto two dimensions.

The overview of the experiments is presented in Figure 6. Using different preprocessing strategies, a set of various self-organized timbre maps were constructed. These were then be compared with SR data obtained using the same set of timbres by calculating correlation coefficient values. The results of these comparisons were to shed light on which features are essential in the perception of timbre. The overall aim was to find the auditory image and distance metric which would yield the highest correlation with the SR data.

Materials

The tone material used in the experiments consisted of 27 tones, produced by additive synthesis. The synthesis algorithm was controlled by six variable parameters, chosen so that they produced a wide range of different tones, including tones resembling piano, strings, brass, and woodwind instruments.

Similarity rating experiment

Nine subjects participated in the similarity rating experiment. They heard each possible pair of tones, and were asked to rate the similarity of each pair on a scale between zero (for completely similar) and twelve (for very dissimilar). Since the order in which tones of a pair are presented in similarity rating experiments has been found to have little effect, it was ignored. The experiment yielded, thus, a 27 x 27 triangular SR matrix for each subject. In this study, the SR matrix obtained by averaging the SR matrices of each subject was used.

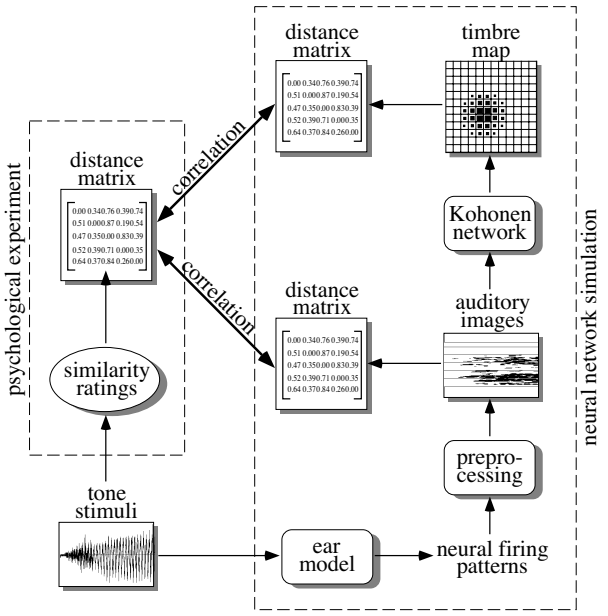

Figure 6. Overview of the timbre classification experiment.

Neural network simulations

To obtain auditory images of the tones, the stimuli were preprocessed, using the peripheral part of an auditory model by Van Immerseel and Martens (1992), modified by Leman (1994) for musical purposes. The model takes into account the filtering of the outer and middle ear, the dynamics of the basilar membrane, the mechanical response of hair cells, and the electrical response of auditory nerve fibres. The spectral images were constructed by sampling the output of the auditory model 20 times during a period of 500 ms. The onset portions were emphasized by locating the sampling points more densely near the onset of the tones; the amount of emphasis was controlled by a parameter. The method for constructing the gradient images is beyond the scope of this paper (see Toiviainen 1996).

Using various sets of auditory images as input, a series of KSOM simulations was carried out. In each simulation, the input data consisted of 27 vectors, whose dimension was 400. For comparing the input and weight vectors during the training procedure, the Minkowski metric was used. In each simulation, a 12 x 12 Kohonen network was trained for 50000 cycles, using a constant learning rate of 0.05. In the beginning,

the radius of the topological neighborhood was 6, and it was linearly decreased after every cycle so as to reach zero at the end of the training. From the obtained timbre maps, distance matrices were calculated.

Results

In the first stage of the study, the pre-Kohonen correlations, i.e., correlations between the SR matrix and distance matrices calculated from the auditory images, were examined. Using Euclidean metric, no emphasis on the onsets, and no gradient images, a correlation of 0.677 was obtained. The maximum correlation, 0.882, was obtained using Minkowski metric (see Toiviainen 1996) with $\lambda = 1$, a proper amount of emphasis on the onsets, and gradient images. The main contributor to the increase of correlation was found to be the emphasizing of onset; adding gradient images did not have any significant effect.

The post-Kohonen correlations, i.e., correlations between the SR matrix and distance matrices calculated from the timbre map, were found to depend on the respective pre-Kohonen correlations: high values of the latter tended to imply high values of the former, and vice versa. Furthermore, the post-Kohonen correlations were regularly lower than the respective pre-Kohonen correlations. This is probably due to the fact that it is not possible to project the timbre space onto two dimensions without significantly distorting the metrical relationships between the timbres. The obtained correlations varied between 0.597 and 0.748. Figure 7 displays an x-y-scatter plot of the similarity ratings vs. respective response distances on the KSOM in the case of maximum correlation.

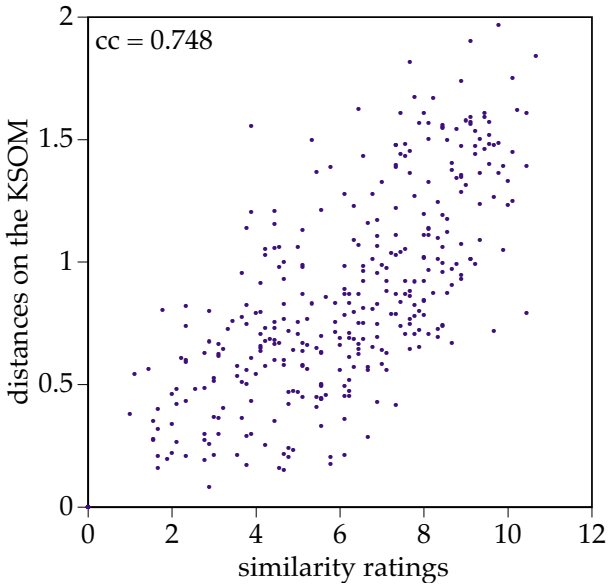

Figure 7. Response distances on the Kohonen map, obtained in one experiment, plotted against respective similarity ratings. Each dot represents a pair of sound stimuli: the abscissa is the average SR value for that pair, while the ordinate is the response distance on the KSOM. The correlation is 0.748.

Concluding Comments

We have made a general survey of what artificial neural networks are and how they relate to symbolic AI systems. Many areas of music cognition, such as music perception, motor action, and performance interpretation, are difficult to study because of the often unverbalizable knowledge involved. Leaning on continuous nonlinear dynamics instead of logical manipulation of symbols, ANNs provide a tool for the study of such inarticulate musical activities.

Symbolic AI and connectionism may be seen as complementary approaches to modelling human cognition in general or musical action in particular. Instead of debating which one of these approaches offers the 'correct' model of musical mind, one may adopt a 'postmodern' attitude and try to exploit the strengths of both methods in hybrid symbolic-connectionist systems (Gutknecht, 1992). An impressive example of this kind of an approach in the field of arts is the HARP system

by Camurri and others (Camurri, Catorcini, Innocenti & Massari, 1995).

References

Balaban, M., Ebcioglu K. & Laske, O. (1992) *Understanding Music with AI*. Cambridge, MA: MIT Press

Bharucha, J. J. (1988) Neural net modeling of music. In *Proceedings of the first workshop on AI and music,* pp. 173–182. Minneapolis/St. Paul: AAAI-88

Camurri, A., Catorcini, A., Innocenti, C. & Massari, A. (1995) Music and multimedia knowledge representation and reasoning: the HARP system. *Computer Music Journal,* 19(2), 34–58

Carpenter, G. A. & Grossberg, S. (1987) ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics,* 26, 4919–4930

Carpenter, G. A. & Grossberg, S. (1990) ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. Neural Networks, 3, 129–152

Cohen, M. A. & Grossberg, S. (1987) Masking fields: a massively parallel neural architecture for learning, recognizing, and predicting multiple groupings of patterned data. *Applied Optics,* 26, 1866–1891

Cosi, P., De Poli, G. & Lauzzana, G. (1994) Auditory modelling and self-organizing neural networks for timbre classification. *Journal of New Music Research,* 23, 71–98

De Poli, G., Prandoni, P. & Tonella, P. (1993) Timbre clustering by self-organizing neural networks. In *Proc. of X Colloquium on Musical Informatics,* edited by G. Haus & I. Pighi, pp. 102–108. Milan: AIMI

Desain, P. & Honing, H. (1989) The quantization of musical time: a connectionist approach. *Computer Music Journal,* 13(3), 56–66. Reprinted in P. M. Todd & D. G. Loy (Eds.) 1991, *Music and Connectionism*. Cambridge, MA: MIT Press, 150–160

Feiten, B. & Günzel, S. (1994) Automatic Indexing of a sound database using self-organizing neural nets. *Computer Music Journal,* 18(3), 53–65

Grey, J. M. (1975) *An Exploration of Musical Timbre*. Ph. D. dissertation, Stanford, CA: Stanford University

Griffith, N. & Todd, P. (Eds.) (1994) *Connection Science,* 6(2-3). Special issue: music and creativity

Grossberg, S. (1986) *The Adaptive Brain I: Cognition, Learning, Reinforcement, and Rhythm,* and *The Adaptive Brain II: Vision, Speech, Language, and Motor Control*. Amsterdam: Elsevier/North-Holland

Gutknecht, M. (1992) The 'postmodern mind': hybrid models of cognition. Connection Science, 4(3–4), 339–364

Hebb, D. O. (1949) *The Organization of Behavior*. New York: Wiley & Sons

Hinton, G. E. & Sejnowski, T. J. (1986) Learning and relearning in Boltzmann machines. In *Parallel Distributed Processing: Explorations in The Microstructure Of Cognition. Vol. 1: Foundations,* edited by D. E. Rumelhart & J. L. McClelland, pp. 282–317. Cambridge, MA: MIT Press

Hopfield, J. J. (1984) Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA,* 81, 3088–3092

Jordan, M. I. (1986) *Serial Order: a Parallel Distributed Processing Approach*. Technical report ICS–8604. La Jolla: University of California, Institute for Cognitive Science

Kaipainen, M. (1994) *Dynamics of Musical Knowledge Ecology. Knowing-What and Knowing-How in*

the World of Sounds. Ph. D. dissertation. University of Helsinki

Kohonen, T. (1989) *Self-Organization and Associative Memory*. (2nd Ed.) Berlin: Springer-Verlag

Kohonen, T. (1995) *Self-organizing Maps*. Berlin: Springer-Verlag

Krumhansl, C. (1990) *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press

Large, E. W. & Kolen, J. F. (1994) Resonance and the perception of musical meter. *Connection Science,* 6(2-3), 177–208

Leman, M. (1992a) Tone context by pattern integration over time. In *Readings in Computer Generated Music,* edited by D. L. Baggi, pp. 117–138. Los Alamitos, CA: IEEE Computer Society Press

Leman, M. (1992b) The theory of tone semantics: concept, foundation, and application. *Minds and Machines,* 2, 345–363

Leman, M. (1988) Sequential (musical) information processing with PDP-networks. In *Proceedings of the First Workshop on AI and Music*, pp. 163–172. Minneapolis/St. Paul: AAAI-88

Leman, M. (1989) *Artificial Neural Networks in Music Research*. Reports from the seminar of musicology — Institute for psychoacoustics and electronic music. University of Ghent

Leman, M. (1990) Emergent properties of tonality functions by self-organization. *Interface,* 19(2–3), 85–106

Leman, M. (1991) The ontogenesis of tonal semantics: results of a computer study. In *Music and Connectionism,* edited by P. M. Todd & D. G. Loy, pp. 100–127. Cambridge, MA: MIT Press

Leman, M. (1994) Schema-based tone center recognition of musical signals. *Journal of New Music Research,* 23, 169–204

Leman, M. (1995) *Music and Schema Theory*. Berlin: Springer-Verlag

Leman, M. (1996) The convergence paradigm in music research. In *Proc. of the First Meeting of the NFWO Research Society on Foundations of Music Research,* edited by M. Leman. Ghent: University of Ghent

Leman, M. & Carreras, F. (1996) The self-organization of stable perceptual maps in a realistic musical environment. In *Proc. of the Journées d'Informatique Musicale 17-18 mai 1996,* edited by G. Assayah, pp. 156–169. Caen: University of Caen/IRCAM

Lischka, C. (1991) Understanding music cognition: a connectionist view. In *Representations of Musical Signals,* edited by G. De Poli, A. Piccialli & C. Roads, pp. 417–446. Cambridge, MA: MIT Press

Loy, D. G. (1991) Connectionism and musiconomy. In *Music and Connectionism,* edited by P. M. Todd & D. G. Loy, pp. 20–36. Cambridge, MA: MIT Press

McClelland, J. L., Rumelhart, D. E. & Hinton, G. E. (1986) The appeal of parallel distributed processing. In *Parallel distributed processing: explorations in the microstructure of cognition. Vol. 1: Foundations,* edited by D. E. Rumelhart & J. L. McClelland, pp. 3–44. Cambridge, MA: MIT Press

McCulloch, W. S. & Pitts, W. (1943) A logical calculus of the ideas imminent in nervous activity. *Bulletin of Mathematical Biophysics,* 5, 115–133

Mozer, M. C. (1991) Connectionist music composition based on melodic, stylistic and psychophysical constraints. In *Music and Connectionism,* edited by P. M. Todd & D. G. Loy, pp. 195–212. Cambridge, MA: MIT Press

Mozer, M. C. (1994) Neural network music composition by prediction: Exploring the benefits of

psychoacoustic constraints and multi-scale processing. *Connection Science,* 6(2-3), 247–280

Page, M. P. A. (1994) Modelling the perception of musical sequences with self-organizing neural networks. *Connection Science,* 6(2-3), 223–246

Rosenblatt, F. (1958) The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review,* 65, 386–408

Rosenblatt, F. (1959) *Principles of Neurodynamics*. New York: Spartan Books

Rumelhart, D. E., Hinton, G. E. & McClelland, J. L. (1986) A general framework for parallel distributed processing. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations,* edited by D. E. Rumelhart & J. L. McClelland, pp. 45–76. Cambridge, MA: MIT Press

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. 1986. Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations,* edited by D. E. Rumelhart & J. L. McClelland, pp. 318–362. Cambridge, MA: MIT Press

Rumelhart, D. E. & McClelland, J. L. (Eds.) (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA: MIT Press

Serra, R. & Zanarini, G. (1990) *Complex Systems and Cognitive Processes*. Berlin: Springer-Verlag

Shepard, R. N. (1982) Geometrical approximations to the structure of musical pitch. *Psychological Review,* 89, 305–333

Taylor, I. & Greenhough, M. (1994) Modelling pitch perception with adaptive resonance theory artificial neural networks. *Connection Science,* 6(2-3), 135–154

Terhardt, E., Stoll, G. & Seewann, M. (1982) Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America,* 71(3), 679–688

Todd, P. M. (1989) A connectionistic approach to algorithmic composition. *Computer Music Journal,* 13(4), 27–43. Reprinted in P. M. Todd & D. G. Loy (Eds.) 1991, *Music and Connectionism*. Cambridge, MA: MIT Press, 173–189

Todd, P. M. (1991) Addendum to: A connectionist approach to algorithmic composition. In *Music and Connectionism,* edited by P. M. Todd & D. G. Loy, pp. 190–194. Cambridge, MA: MIT Press

Todd, P. M. & Loy, D. G. (Eds.) (1991) *Music and Connectionism*. Cambridge, MA: MIT Press

Toiviainen, P. (1996) Optimizing auditory images and distance metrics for self-organizing timbre maps. *Journal of New Music Research,* 25(1), 1-30

Toiviainen, P., Kaipainen, M. & Louhivuori, J. (1995) Musical timbre: similarity ratings correlate with computational feature space distances. *Journal of New Music Research,* 24(3), 282–298

Van Immerseel, L. M. & Martens, J.-P. (1992) Pitch and voiced/unvoiced determination with an auditory model. *Journal of the Acoustical Society of America,* 91(6), 3511–3526