Modeling the Target-Note Technique of Bebop-Style Jazz Improvisation: An Artificial Neural Network Approach
Author(s): Petri Toiviainen
Source: *Music Perception: An Interdisciplinary Journal*, Vol. 12, No. 4 (Summer, 1995), pp. 399-413
Published by: University of California Press
Stable URL: http://www.jstor.org/stable/40285674

Music Perception
Summer 1995, Vol. 12, No. 4, 399–413

Modeling the Target-Note Technique of Bebop-Style Jazz Improvisation: An Artificial Neural Network Approach

PETRI TOIVIAINEN
University of Jyväskylä, Finland

In cognitive science and research on artificial intelligence, there are two central paradigms: symbolic and analogical. Within the analogical paradigm, artificial neural networks (ANNs) have recently been successfully used to model and simulate cognitive phenomena. One of the most prominent features of ANNs is their ability to learn by example and, to a certain extent, generalize what they have learned.

Improvisation, the art of spontaneously creating music while playing or singing, fundamentally has an imitative nature. Regardless of how much one studies and analyzes, the art of improvisation is learned mostly by example. Instead of memorizing explicit rules, the student mimics the playing of other musicians. This kind of learning procedure cannot be easily modeled with rule-based symbolic systems. ANNs, on the other hand, provide an effective means of modeling and simulating this kind of imitative learning. In this article, a model of jazz improvisation that is based on supervised learning ANNs is described. Some results, achieved by simulations with the model, are presented. The simulations show that the model is able to apply the material it has learned in a new context. It can even create new melodic patterns based on the learned patterns. This kind of adaptability is a direct consequence of the fact that the knowledge resides in a distributed form in the network.

Introduction

Improvisation, the art of spontaneously creating music while playing or singing, is the basic element of nearly all musical cultures of the world. In the Western tradition, the art of improvisation reaches its highest level among jazz musicians. In jazz, one improvises on the melodic, harmonic, as well as rhythmic level. The basis of improvisation differs considerably between different styles of jazz: a Dixieland musician bases his improvisation, more than the others, on the structure of the melody of the composition; a bebop musician leans rather on the harmonic structure; whereas a free jazz group,

Requests for reprints may be sent to Petri Toiviainen, Department of Musicology, University of Jyväskylä, PL. 35, FIN-40351, Jyväskylä, Finland. (e-mail: ptoiviai@tukki.jyu.fi)

for example, may operate solely within the framework of some prearranged musical structure.

As a cognitive process, the improvisation of jazz is extremely complicated. Irrespective of the jazz style in question, the improviser has to take into account constraints on various hierarchical levels. Every musician has, too, a personal way of approaching this problem. To build an unambiguous model of jazz improvisation is, thus, probably impossible.

One way of learning to understand better the components of jazz improvisation is simulation. When constructing the model of improvisation, one has to ignore several aspects of the process; for example, many stylistic and interpretative aspects must, at least in the first stage, be ignored. On the other hand, a sufficiently simple representation of the music must be chosen.

The art of improvisation is learned mostly by example. Instead of memorizing explicit rules, the student mimics the playing of other musicians. On the basis of this material, he then forms implicit rules for the style of improvisation concerned. This kind of learning procedure cannot be easily modeled with rule-based expert systems. On the other hand, artificial neural networks, or connectionist systems, are suitable for this purpose. Neural networks are systems consisting of massively interconnected simple processing elements, or neurons. One of their most prominent features is the ability to learn by example.

Recently, neural networks have been successfully used to model and simulate cognitive phenomena of music (Bharucha & Todd, 1989; Leman, 1991; Todd, 1989). In this paper, a connectionist model of bebop-style melodic improvisation, based on harmony, is described; some results, achieved by simulating the model, are presented. The basis of the model is the target note technique, one of the cornerstones of bebop-style jazz improvisation.

About Harmony-Based Improvisation

Several factors influence the structure of an improvised jazz melody: for example, the harmonic structure, the melody of the theme, and the style of the accompaniment. Among the essential factors are, also, the instrument used—its fingering and range of pitches, the musical background of the soloist, and the musical interaction with the other players in the group while playing.

An improvised jazz melody has, in the majority of cases, a hierarchical structure. A bebop-style jazz solo is composed of choruses; that is, units having the same harmonic structure as the theme. A chorus is divided into parts—a typical structure of a bebop chorus is AABA, each part consisting of eight measures. A part of a chorus contains a few melodic phrases, which themselves are composed of melodic patterns of a few notes. Consideration of this hierarchical structure is essential when striving for a convincing model of jazz improvisation.

The target-note technique (Mehegan, 1959), a common way of explaining the microstructure of an improvised jazz melody, can be described as follows: the notes of a four-note chord and its upper structure (9, 11, 13) are regarded as principal tones; when approaching a chord, one of its principal tones is chosen as a target note; the target note is reached through a learned melodic pattern.

For the simulation model, the target-note technique can be described as a dynamic process with feedback, as presented in Figure 1. At the beginning of the process, some starting notes—here two—are fixed, as well as the present and following chords. The starting notes, together with the present chord, determine the possible melodic patterns, whereas the following chord determines the possible target notes. From the interaction of those constraints, a new melodic pattern emerges. The target notes of the pattern are then used as the starting notes of the next pattern.

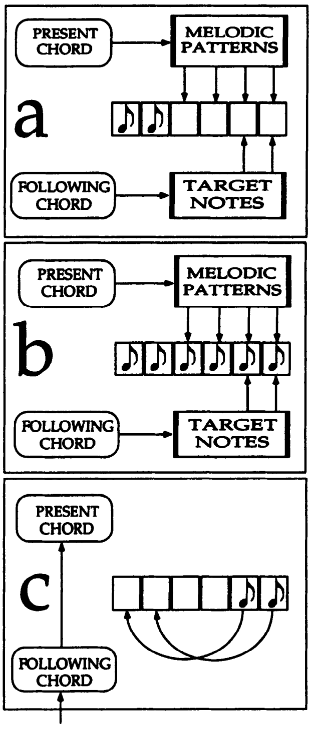


Fig. 1. The target-note technique as a dynamic process with feedback.

The Architecture of the Model

A complete representation of the melody used in the simulation model should include at least the following aspects: the pitch (pitch class and octave), the time values of the notes (as represented in traditional Western notation), phrasing, and dynamics.

An essential step in constructing a model to simulate the production of music is to decide how time is represented. When using neural networks, a natural choice is to use a representation analogous to "piano-roll" notation, used in the old player pianos. In piano-roll notation, time has been translated into an ordered spatial dimension. Accordingly, in a neural network, the melody can be represented in a two-dimensional grid of neurons, where one dimension corresponds to time and the other to pitch. In the limiting case, the entire melody could be produced by the model simultaneously. This kind of static approach does not, however, correspond to the way in which humans produce music. The production of music is, essentially, a dynamic process, during which the musician continuously interacts with the environment. The other limiting case would be the use of serial nets (Todd, 1989): nets that produce one musical event (note) at a time. In serial nets, a short-term memory is usually implemented, realized by, for example, a feedback mechanism. The target-note technique being the basis of the improvisation model in question, an intermediate form of the above-mentioned limiting cases was considered as the most suitable choice: the network processes one melodic pattern at a time, using a modified piano-roll representation.

The rhythmic structure of both the melody and harmony of bebop jazz is very regular—if the phrasing is ignored. First, an overwhelming majority of all jazz composed and performed is in $\frac{4}{4}$ time (Mehegan, 1962). Second, it can be said that all jazz from 1900 to the present day has used the following ratio of time values: a quarter-note pulse, representing the rhythmic unit of a jazz composition; a half-note pulse, representing the harmonic unit; and an eighth-note pulse, representing the melodic unit (Mehegan, 1962).

In an improvised jazz melody, one certainly uses a broad range of time values, from whole notes to thirty-second notes. The smaller time values can, however, often be regarded as an ornamentation of a frame melody proceeding in eighth-note time values. Furthermore, in jazz melodies of medium to fast tempi, the eighth note is the dominant time value. Consequently, it is justified to make the following simplifications concerning the architecture of the network: (1) only $\frac{4}{4}$ time is possible; (2) the network processes one half-measure-long melodic pattern at a time faster chord progressions are not possible; and (3) the smallest possible time value is an eighth note.

For the representation of pitch in neural networks, there are several possibilities (Bharucha, 1991). In this model, invariant pitch class representation is used. The pitches are, thus, represented relative to the root of the present chord.

In the network, melody is represented as follows: for every eighth note, 12 neurons represent the pitch classes of the chromatic scale, relative to the root of the present chord. Moreover, one neuron represents the rest and another the ligature. In order to simplify the model, the following restrictions are made: (1) the contour of the melody is ignored; this causes the output of the network to be open to various interpretations. One way of interpreting the output is always to move to the nearest tone represented by the pitch class. (2) The network produces only monophonic melodies; this is not a severe restriction (cf. wind instruments). (3) Phrasing, dynamics, and octave are ignored; the consideration of those aspects would, at this stage, excessively complicate the architecture of the network.

In the network, one eighth note corresponds, thus, to a group of 14 neurons. Because of syncopation, characteristic of jazz music, the target note is often played on the eighth note preceding the first beat. Therefore, it is reasonable to use six groups of neurons: the first corresponding to the eighth note preceding the half of the measure in question; the following four corresponding to the half of the measure in question; and the last corresponding to the eighth note following the half of the measure in question. The network consists, thus, of $6 \times 14 = 84$ neurons altogether. The neurons of the network are fully interconnected. There are two types of connections: (1) neurons belonging to the same group (column) have fixed inhibitory interconnections; this guarantees that, in the relaxed state, only one neuron is active in every group. (2) Neurons belonging to different groups have excitatory interconnections, which are modified during the learning phase, according to the Hebbian learning rule (McClelland & Rumelhart, 1988).

The network is, actually, a modified auto-associator. An auto-associator is a fully interconnected neural network, that is, it has connections from every neuron to every other neuron. It can memorize the presented patterns of activation by strengthening connections between the active neurons. When receiving noisy or incomplete patterns as input, the correct pattern is reproduced through the flow of activation between neurons. In Figure 2, the architecture of the auto-associator, together with representations of some melodic patterns, is presented.

In addition to the auto-associator module, the network contains two groups of neurons, representing the types and functions of present and following chords. These neurons feed activation into the auto-associator module. For a given chord type, neurons representing the most-often-used notes, in each column, receive the strongest external activation. This causes

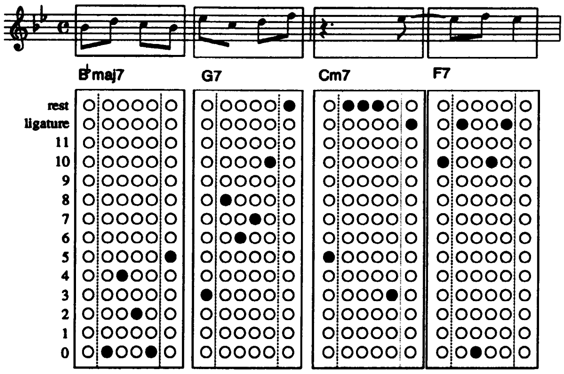


Fig. 2. The structure of the auto-associator network used in the model and representations of four melodic patterns. The network is divided into six groups, each representing an eighth note. Each group consists of 14 neurons representing the rest, ligature, and 12 pitch classes of the chromatic scale.

the most-often-used notes to be the most probable to appear in the melodic pattern produced.

The idea of functional harmony is essential in jazz improvisation. The chords in a progression are related to each other and have functions. For example, a minor seventh chord can appear at the II, III, or VI degrees, each of which is, in improvisation, treated differently. In addition to that, there are several variations for the upper structure of the dominant seventh chord. In this model, nine different chord types/functions are used: major seventh (maj7) on the I and IV degrees; minor seventh (m7) on the II, III, and VI degrees; half-diminished (m7-5); three kinds of dominant seventh chords ($7^{13\sharp 11}$ $7^{13\flat 9}$ alt.; Figure 3).

As previously mentioned, invariant pitch class representation—relative to the root of the present chord—is used in the auto-associator. The distribution of the target notes is also presented invariably—relative to the root of the following chord. The external activations containing information about the target notes must, thus, be transposed before being fed into the auto-associator.

In the model, eight distinct connection matrices are used, each corresponding to one half measure in a four-measure range. There are two main reasons for this. First, an auto-associator has a limited storage capacity: unorthogonal learning vectors tend to fuse or cause "cross-talk" (McClelland & Rumelhart, 1988). Second, this provides a simple means of obtaining higher-level structure in the melodies: a bebop jazz composition can, in the majority of cases, be divided into phrases four measures long.

During the learning phase, one melodic pattern at a time is presented to the network: the neurons representing the notes of the melodic pattern are

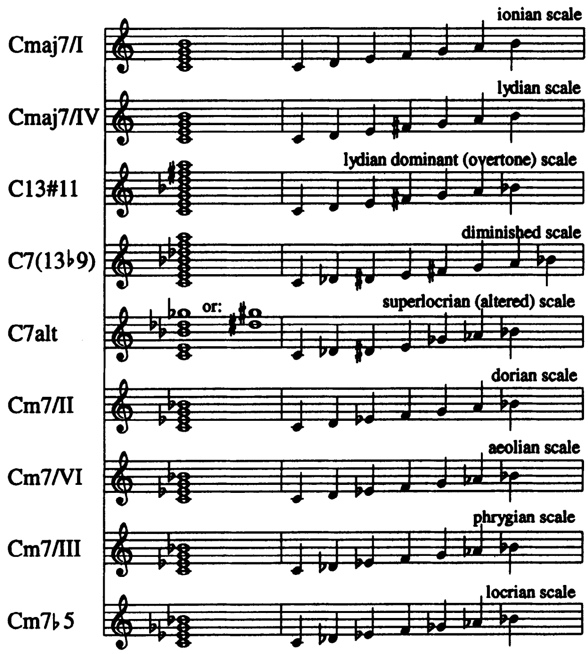


Fig. 3. The chord types and functions used in the model. For each chord type, the most likely used scale is presented.

activated, together with the neurons representing the roots and the types of the present and the following chord. The learning occurs in two distinct groups of neurons: (1) the melodic patterns are learned through strengthening the connections between the active neurons of the auto-associator, and the neuron representing the type of present chord, according to the Hebbian learning rule; (2) the target notes are learned through strengthening the connections between the neurons representing the type of the following chord, and the neurons representing the invariant target notes. Inside these two groups of neurons, differing learning rates can be used, resulting in melodies of different styles. In Figure 4, an example of the learning of one melodic pattern is presented.

During the testing phase, the activations of the neurons representing the root and the type of the present and the following chord are set to maximum. The starting note of the melody is chosen, and the corresponding neuron is activated. During the relaxation process, the activations of the neurons are updated asynchronously: the activation of one—randomly chosen—neuron is updated at a time. The updating procedure consists of: (1) calculating the incoming activation, according to the formula

$$\text{input}_j = a_j + \sum_{i \neq j} w_{ij} a_i, \tag{1}$$

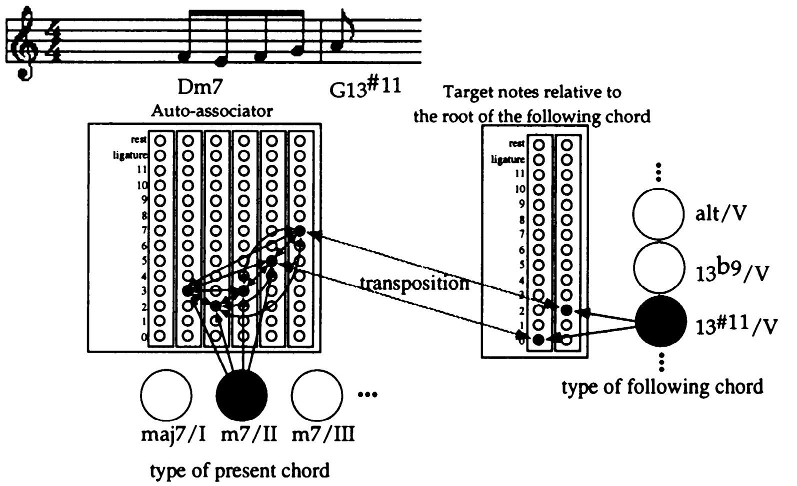


Fig. 4. An example of the learning of one melodic pattern in the network. The black arrows represent the connections, the strengths of which are increased.

where w_{ij} is the connection strength between neurons i and j, and a_i is the activation value of neuron i; and (2) calculating the new activation according to the formula

$$a(\text{input}) = \begin{cases} 0, \text{ if input} \leq 0, \\ \text{input, if } 0 < \text{input} \leq 1, \\ 1 \text{ if input} > 1 \end{cases} \tag{2}$$

As formula (1) states, the input of a neuron consists of both activation flowing from other neurons and its own activation fed back into it. The positive feedback is necessary, if the network is expected to settle down into a stable state. The activation function of this model, together with the positive feedback, resembles that of the so-called brain-state-in-a-box model (Rumelhart, Hinton, & McClelland, 1986). Using asynchronous updating has the following advantages: (1) the network can be kept out of oscillations during the relaxation process (Rumelhart et al., 1986), and (2) the system behaves, to a certain extent, stochastically: beginning from the same starting notes and chords, it can produce various melodic patterns, depending on the order of the updating. The updating procedure continues until the network has reached a stable state. In every group of neurons, there is now one active neuron. These form the melodic pattern the network has produced. After feeding back the last two notes of the melodic pattern as the starting notes of the next pattern, and updating the chords, the relaxation process is, again, started.

Observations About the Experiments

In one experiment, the network was trained on improvised solos played by trumpet player Clifford Brown. The training sets consisted of excerpts, typically 32 measures long, from four solos, taken from Baker (1982).

After the learning phase, the network was tested with *Rhythm Changes* chord progression. Based on George Gershwin's famous composition "I've Got Rhythm," it is one of the most widely used harmonic structures in bebop. There are several variations of the Rhythm Changes; the version used is as follows:

B♭ maj7 Gm7	Cm7 F7	Dm7 G7	Cm7 F7
Fm7 B♭7	E♭ maj7 A♭7	Dm7 G7	Cm7 F7
B♭ maj7 Gm7	Cm7 F7	Dm7 G7	Cm7 F7
Fm7 B♭ 7	E♭ maj7	B♭ maj7	B♭ maj7
Am7	D7	Dm7	G7
Gm7	C7	Cm7	F7
B♭ maj7 Gm7	Cm7 F7	Dm7 G7	Cm7 F7
Fm7 B♭7	E♭ maj7	B♭ maj7	B♭ maj7

When teaching the network, various learning rates were used for the melodic patterns (η_1) and the target notes (η_2). It was found that the ratio η_1/η_2 had a considerable influence on the style of the melodies produced. When η_1/η_2 was large, the network could accurately reproduce the learned melodic patterns, while it was not able to choose proper target notes, and vice versa.

The strength of the inhibitory connections, w_{inh}, of the auto-associator also had an effect on the melodies. When using strong inhibition, the network settled down quickly. In this case, the stochasticity, being part of the asynchronous updating procedure, showed itself in the melodies: the final stable state of the network depended greatly on the order of the updating. With sufficiently strong inhibitory connections ($w_{\text{inh}} \approx 0.5$), the network did not stabilize according to the learned melodic patterns, but the inhibition dominated over the excitatory connections. When using weak inhibition, the neurons of the network began to inhibit each other only with considerably high activation values. For this reason, the network stabilized slowly, and so the stochasticity of the updating had less influence on the end result. In the extreme case, when the strength of inhibition was less than 0.01, the network always produced the same melodic pattern for the same chord type.

When using proper values for the learning rates and inhibition ($\eta_1/\eta_2 \approx 2$, $w_{\text{inh}} \approx 0.3$), the network was found to be capable of producing stylistically fairly consistent melodies—on the micro level. On the other hand, it cannot deal with larger structures, like melodic phrases. In this respect, the melodies produced by the network resemble those of a beginning improviser. An interesting aspect is that the model has a sort of creativity: it can produce new melodic patterns based on the patterns it has learned.

Results of two simulations are presented in Figures 5c and 6c. The training set of the first simulation (Figures 5a and b) consisted of excerpts of

Fig. 5. First simulation. The training set consisted of excerpts from Clifford Brown's solos on (a) "All the Things You Are" and (b)"Gertrude's Bounce." (c) Output of the network; the rectangles indicate melodic patterns that occur in the training set. Letters "A" and "G" refer to the names of the compositions.

b

Fig. 5b.

Clifford Brown's solos on "All The Things You Are" and "Gertrude's Bounce," while that of the second simulation (Figures 6a and b) consisted of his solos on "Confirmation" and "Donna Lee."

Conclusions

An artificial neural network, modeling the target-note technique of bebop-style jazz improvisation, has been presented. The starting point of the model is learning by means of examples: the student is presented melodic patterns, and after having learned them, he is able to use them in a given context, (i.e., harmonic progression) and to generalize what he has learned. Instead of presenting the student explicit rules concerning the style

C

Fig. 5c.

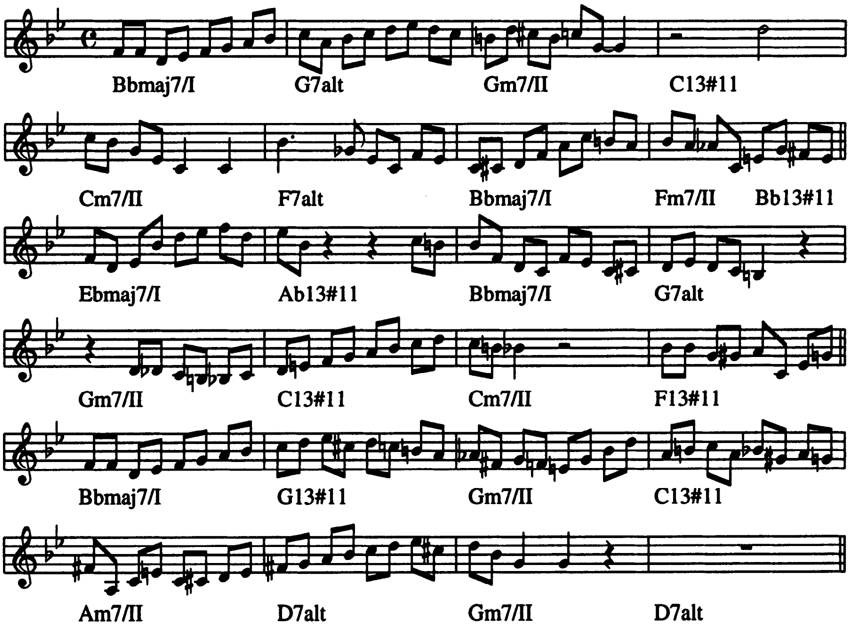
of improvisation in question, he himself builds implicit rules, based on the material learned, and applies these in new situations.

The simulations showed that the model can apply the material it has learned to a new context. If it cannot find a proper melodic pattern for a given chord progression, it can create a new pattern based on the patterns it has learned. This kind of adaptability—or creativity—is a direct consequence of the connectionist paradigm. The structure of each melodic pattern is not presented symbolically, but rather resides in a distributed form in the network—in the connection strengths between the neurons. To construct a rule-based expert system behaving in the same manner would be a

Fig. 6. Second simulation. The training set consisted of excerpts from Clifford Brown's solos on (a) "Confirmation" and (b) "Donna Lee." (c) Output of the network; the rectangles indicate melodic patterns that occur in the training set. Letters "C" and "D" refer to the names of the compositions.

laborious. The epistemological relevance of such a system, based on numerous explicit rules, would also be small.

The main shortcoming of the system described here is its limited ability to operate on higher hierarchical melodic levels, such as phrases. This might be overcome by, for example, using a higher-level network. This network would learn and produce only the target notes of a longer harmonic progression; for example, a progression four measures long. Feeding the target notes produced by this network into the lower-level network, and taking b



c

Figs. 6b and 6c.

into account the respective position in the four-measure structure, would probably yield better results.

References

Baker, D. N. (1982). *The jazz style of Clifford Brown: A musical and historical perspective.* Hialeah, FL: Studio P/R.

Bharucha, J. J., & Todd, P. M. (1989). Modeling the perception of tonal structure with neural nets. *Computer Music Journal, 13*(4), 44–53.

Bharucha, J. J. (1991). Pitch, harmony, and neural nets: A psychological perspective. In P. Todd & G. Loy (Eds.), *Connectionism and music.* Cambridge: MIT Press.

Leman, M. (1991). Tonal context by pattern integration over time. In A. Camurri (Ed.), *Proceedings of the IX Colloquio di Informatica Musicale.* Genoa: AIMI, DIST.

McClelland, J. L., & Rumelhart, D. E. (1988). *Explorations in parallel distributed processing.* Cambridge, MA: MIT Press.

Mehegan, J. (1959). *Jazz improvisation I: Tonal and rhythmic principles.* New York: Watson-Guptill Publications.

Mehegan, J. (1962). *Jazz improvisation II: Jazz rhythm and the improvised line.* New York: Watson-Guptill Publications.

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processing. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1: Foundations.* Cambridge, MA: MIT Press.

Todd, P. M. (1989). A connectionist approach to algorithmic composition. *Computer Music Journal, 13*(4), 27–43.