WOLFGANG: ‘‘Emotions’’ and Architecture which Enable Learning to Compose Music

Doug Riecken
Computer Systems Research Laboratory
Bell Laboratories - Lucent Technologies
Room 2C304
600 Mountain Avenue
Murray Hill, NJ 07974 USA
wolf@research.bell-labs.com
(908) 582-8011

Abstract: *In this paper, we will discuss, from a general perspective, the design and implementation of a system called WOLFGANG that composes tonal monodies. Our examination will focus on defining the evaluation criteria guiding WOLFGANG’s compositional processing. The thesis of the work presented is derived from the hypothesis that a system’s ‘‘innate”sense of musical sound strongly influences the development of its ‘‘perception” as well as composing habits. As the system develops its musical skills, it also develops a subjective use of a musical language biased by its sense of musical sound, and its adaptation to the cultural grammar of its environment.*

In essence, a person becomes a composer via many diverse experiences originating from emmercement in a cultural. Further, a composer continues to learn and modify composing techniques as a direct reflection of changes to the current culture. Thus, it is necessary to design systems which represent and apply knowledge learned from cultural experiences. Systems such as WOLFGANG, must be designed with extensive and flexible knowledge representations so as to support (1) strong plan reformulation functions and (2) strong abstraction capability. It is plausible to expect such systems (3) to be interactive assistants in a broad range of tasks, not just limited to composing.

Keywords: spreading-activation, K-lines, dispositions, emotions.

1. Problem Statement and General Discussion

> ‘‘For at best, the very aim of syntax oriented theories is misdirected; they aspire to describe the things that minds produce - without attempting to describe how they’re produced.’’
>
> Marvin Minsky (Minsky 1981)

By casual observation, one might consider the problem of investigation described within this paper as a study in the composition of music by computer. While the body of work reported herein describes a computer software system for composing musical compositions, the composition process is actually a secondary problem of interest; mainly, the composition process serves as a tool by which to examine the problem of how to define evaluation criteria able to guide a composition process. An investigation of this problem entails the formal representation of a set of properties which contribute to the formal structure and functioning of the system. These properties can be represented as (1) principles defining epistemological adequacy, or (2) an architecture of system components which consistently and explicitly define aspects of system behavior, or (3) both.

Throughout this paper, epistemological adequacy refers to the completeness of the corpus of knowledge encoded in system memory necessary to compose musical compositions. Due to this corpus, the system can follow, based on a first order logic embedded in the knowledge, a valid sequence of steps in composing a given musical work. System components ‘‘influence” the behavior of other related components, especially system memory, by way of specific structures and functions. Specifically, system components manage and modify system memory, effecting the access and use of knowledge encoded in memory, thus modifying existing reasoning processes.

What we hope to provide in this paper is a general architectural description of a software system called WOLFGANG which composes tonal monodies; this description is abstracted from more detailed discussion and results reported in other published works (Riecken 1992, 1989). Further, this description serves as an introductory vehicle to discuss a plausible theory that demonstrates how a computer system might develop an ability to compose music in a personalized manner, and that determines what structures and functions possibly support the development of a personalized composition process. We attempt to provide a framework for such a theory based on the continuing study and design of the WOLFGANG system.

The design goals of WOLFGANG have evolved from the premise that a considerable percentage of cited work reporting the design and implementation of composing systems are based on establishing grammatical formalisms to construct propositions (for discussion on related work refer to (Riecken 1992)). Specifically, the knowledge contents of many of these systems results from knowledge-bases deriving from some ‘‘cultural grammar”. By cultural grammar, we refer to systems that guide composition processes by way of a musical theory reflecting the musical grammar of a given culture (e.g., 19th century Western music).

Based on the assumption that a composer develops his or her skills from the grounding of some theories of music reflective of the composer’s cultural development, I now put forth the following two propositions as guides to the research reported herein. First, if WOLFGANG is to demonstrate an ability to compose music, then it must develop a musical skill that reflects its musical development within some environment. Thus, WOLFGANG would demonstrate cultural grammar based processing. Second, if WOLFGANG is to demonstrate cultural grammar based processing, then WOLFGANG must be designed with an ‘‘innate” ability to sense and attentively ‘‘perceive” musical sound; WOLFGANG must be able to utilize the musical characteristics it encounters in its environment with “some initial’’ bias.

2. History of WOLFGANG Research Objectives

The WOLFGANG system began as a research project to study the development of compositional performance. This work focused on the application of Michalski’s STAR methodology (Michalski 1983) for inductive machine learning in a knowledge-based system; it resulted in the first generation of WOLFGANG. In this first generation system, the evaluation criteria for guiding the composition process were derived from an implied grammar of Western music. This grammar was represented in terms of facts and rules in WOLFGANG’s knowledge-base.

The second generation of WOLFGANG introduces ‘‘emotional” criteria; it allows a user to request a sixty-four measure monody composition that realizes an emotional characteristic chosen by the user. The set of emotional characteristics include happiness, sadness, anger, and meditativeness. The second generation of the WOLFGANG addresses two different areas of research: (1) investigation of musical development, and (2) extension of the evaluation criteria guiding the composition process. Within this paper we will focus on the second research area.

In developing the second generation WOLFGANG system, the following plausible hypothesis for how to guide the composition process was applied. The hypothesis states that knowledge defining a system’s ‘‘innate” sense of musical sound strongly influences the development of its musical perception and composing habits. It is hypothesized further that as such a system learns, it develops a personalized profile. This profile is forged from both the system’s sense of musical sound and its adaptation to the cultural music grammar of its environment.

To examine this hypothesis, WOLFGANG was designed to guide its composition process biased by a cultural grammar of music, as well as by a disposition for crafting musical phrases such that they express a specific emotional characteristic. In this paper, the term disposition is influenced by Marvin Minsky’s use of this term in his 1980 paper on K-lines (Minsky 1980). In his paper Minsky’s states as follows: ‘‘I use ‘disposition’ to mean ‘a momentary range of possible behaviors;’ technically it is the shorter term component of the state. In a computer program, a disposition might depend upon which items are currently active in a data base, e.g., as in Doyle’s flagging of items that are ‘in’ and ‘out’ in regard to making decisions” (Doyle 1979). While the design of WOLFGANG makes use of the concept of dispositions, the current work in progress is an evolving attempt to integrate some of Minsky’s ideas; it is the author’s view that considerable work is still required.

Specifically, the evaluation criteria guiding WOLFGANG's composing process consists of: (1) a cultural grammar reflective of WOLFGANG's musical development, and (2) an ability to realize the emotive potential of musical elements represented in the respective cultural grammar.

An abstract description of WOLFGANG's composition process is as follows. Based on the grammatical context of a given compositional decision, WOLFGANG defines a set of valid solutions from its domain knowledge of music which satisfy the cultural grammar; then WOLFGANG selects from this set of valid solutions that solution which best satisfies WOLFGANG's current disposition, in order to endow the current musical phrase with a specific emotive potential.

3. Cultural Grammar and Disposition

Before examining the WOLFGANG system, a brief discussion of the terms of cultural grammar and disposition should improve our understanding on their use throughout this paper.

WOLFGANG's cultural grammar represents its complete body of domain knowledge for composing music. This knowledge is acquired and used by WOLFGANG over some period of its musical development. Primarily, the knowledge consists of two classes of elements: (1) musical ''methods" (e.g., methods of motivic development, methods of harmonization, etc.) and (2) musical ''artifacts" (e.g., musical intervals, melodic ornamentation, etc.). A third category consists of musical elements that are examples of musical artifacts that have become part of WOLFGANG's musical lexicon during the development process. Also, specific elements from WOLFGANG's corpus of musical elements (both methods and artifacts), over time tend to demonstrate a higher frequency of use in specific contextual composing scenarios; this fact establishes patterns of composing habits. We might postulate that these patterns make up WOLFGANG's compositional signature.

WOLFGANG's disposition for crafting musical phrases such that they express specific emotive characteristics is supported by two design features. First, WOLFGANG is designed with a facility to express the emoting potential of musical materials. As WOLFGANG develops its musical skills, this facility not only permits emotional expressiveness; it also influences, by subjective organization, the construction of WOLFGANG's cultural grammar. Second, WOLFGANG's disposition is defined by its ability to realize specific emotions (e.g., one of the four emotions supported by WOLFGANG). Changes in the dispositional state of WOLFGANG occur whenever the context dealt with during a composing session motivates change to a different emotion type. This design feature provides WOLFGANG with a range of possible behaviors during compositional processing.

4. System Architecture

The WOLFGANG system architecture is composed of the following five fundamental system components: (1) a corpus of E-nodes,(2) a K-line network, (3) a set of blackboard systems, (4) a disposition feedback facility, and (5) logfiles. Herein this section, we will discuss, in general terms, the basic concept and functionality for each component. From this discussion, we hope to establish a general framework by which the system architecture of WOLFGANG enables the performance and behaviors hypothesized in our study.

4.1 The Corpus of E-nodes

E-nodes (emotion-nodes) are the most fundamental system component. Informally, an E-node is a collection of information defining the emoting potential of a given primitive musical artifact. The term primitive musical artifact refers to higher-level music-structural elements such as musical modality, vertical harmonic structures (e.g., a tonic triad, a dominant seventh, etc.), horizontal harmonic structures (harmonic progressions consisting of paired vertical harmonic structures), amplitude, tempo, rhythm, and musical intervals between paired pitches. The use of the term emotive refers to a primitive musical artifact's potential to express emotion or sentiment in a demonstrative manner. The information provided by each E-node allows WOLFGANG to interpret the emoting potential of each primitive musical artifact. While the musical artifacts mentioned are actually quite complex structures (in terms of music theory), for the design of WOLFGANG they are viewed as simple (cognitive) elements, and in that sense, are taken for granted as elements of the musical imagination.

A critical issue concerning E-nodes must now be reviewed. E-nodes are not instances of musical knowledge as are primitive musical artifacts. An E-node is simply a qualification and quantification of the emotive potential of a given primitive musical artifact. In order for WOLFGANG to be knowledgeable of a given primitive musical artifact, a representation of the respective artifact must be encoded into system memory. Once this representation has been encoded, it inherits its emoting potential from the respective E-node. Thus E-nodes serve as ''innate'' system properties; they are not part of WOLFGANG's musical knowledge.

Now that we have an abstract sense of what an E-node is, let us review it in more detail. An E-node is defined as a collection of information; the actual composition of this information consists of an identifier and a set of four numeric values. Each numeric value defines the emotive potential of a particular emotion, as realized by the primitive musical artifact represented by an E-node. The four distinct emotions represented by the four numeric values assigned to each E-node are happiness, sadness, anger, and meditativeness. The design of WOLFGANG provides each primitive musical artifact with its range of emoting potential over the defined set of emotions. (The number of individual emotions supported is currently restricted to four so as to minimize the system's complexity.)

The significance of E-nodes lies in their qualification of emoting potentials for the four emotions over the complete set of primitive musical artifacts. These emotive qualifiers perform a critical role during WOLFGANG's development; they serve as operands by which emoting values are computed and assigned to new primitive musical artifacts, or to a compound musical artifact. The term development refers to the acquisition of musical knowledge by WOLFGANG to improve its performance as a composer. The term compound musical artifact refers to a musical artifact composed directly or indirectly (or both) of two or more primitive musical artifacts. Thus, E-nodes provide emotive primitives (initial properties) used to derive the emoting potential for the combinatorial development of WOLFGANG's musical knowledge.

The decision to specify the emotive potential of all musical artifacts is motivated by our task of developing evaluation criteria for guiding the composition process. As we have learned thus far, one of the metrics which guides WOLFGANG's composition process consists of composing musical phrases that satisfy some current compositional disposition. This means that WOLFGANG will attempt to compose a musical phrase provoking a specific emotion that matches the current disposition of the system; at any moment, WOLFGANG's disposition is at some level of Happiness or Sadness, Anger or Meditativeness. Therefore, the selection of musical artifacts is biased towards those musical artifacts which provide the highest emotional potential matching WOLFGANG's current disposition; consequently, the system design of WOLFGANG requires a method to specify the emoting potential of all musical artifacts.

4.2 Memory Based on the K-line Theory

The storage, access, and management of WOLFGANG's musical domain knowledge is supported by a network of interconnected musical artifacts; these artifacts reflect WOLFGANG's musical experiences and development. The principal influence in the design of WOLFGANG's memory is Minsky's K-line theory of memory (Minsky 1985, 1980).

> ''When you ''get an idea,'' or ''solve a problem,'' or have a ''memorable experience,'' you create what we shall call a K-line. This K-line gets connected to those ''mental agencies'' that were actively involved in the memorable mental event. When that K-line is later ''activated,'' it reactivates some of those mental agencies, creating a ''partial mental state'' resembling the original.''
>
> Marvin Minsky (Minsky-80)

The K-line memory theory describes the behavior of a system by the dynamic relationships of linked agents called K-lines.The complexity of a K-line can range from a representation defined
by a single instance of information, a K-node, to extremely complex representations of behavior defined by a K-line composed of a large set of linked K-lines and K-nodes. Minsky's theory explains that sets of K-lines form societies of agents providing specific mental functions, and that these societies or agents, or both, can dynamically form multiple connections (K-lines) with each other to create new K-lines, thus increasing the body of knowledge. It is the activation of sets of K-lines (societies of agents) that bring about partial mental states. A total mental state is composed of several partial mental states active at a single moment in time. System development is supported by constructing new K-line connections within or between partial mental states.

4.2.1 Significance In Application of K-line Memory Features
WOLFGANG's memory is another significant aspect of the system. In the second generation of WOLFGANG, great care was taken to design a system which demonstrated improvement in its ability to ''develop". If WOLFGANG is to develop composing skills, these skills should demonstrate some characteristic of personalized style. The design of WOLFGANG's memory attempts to capture two powerful qualities found in the K-line theory: flexibility and recall of personalized habits.

The term flexibility refers to WOLFGANG's ability to store and use diverse knowledge during its development as a composing system. The issue of flexibility provided critical motivation for the evolution from the first generation to the second generation of WOLFGANG. The first generation system consistently reached impassable stages of development. This problem resulted due to the nature common to knowledge-based systems; these systems have historically demonstrated their ability to perform quite well over restricted sets of problems, except that the acquisition and representation of a large body of knowledge elements quickly promotes conflicting assertions of facts and rules. The K-line model avoids conflicting assertions by allowing multiple patterns of diverse knowledge to represent different memorable experiences; musical experiences that provide good results during a composing session are therefore encoded as individual K-lines in WOLFGANG's memory. This representation serves to minimize the number of rules which ''govern" methods and facts, and thus to avoid the attending problems of complexity.

The concept recall of personalized habits refers to WOLFGANG's ability to develop and apply its composing skills in a subjective manner reflecting its musical learning experiences. This feature of system performance is a direct result of the explicit method by which musical knowledge is encoded in WOLFGANG's memory (as individual instances (K-lines) of successful musical methods and facts). Thus, as WOLFGANG develops, it also develops a distinct style of composing; this results from the frequent combinations of collaborating K-lines which, over time, are referenced as one single compound K-line. By attempting to model features of Minsky's K-line memory, WOLFGANG develops personal composing habits; these habits become the system's compositional signature.

4.2.2 Design and Implementation of Memory
WOLFGANG's memory is implemented as a frame-based spreading-activation network; semantic relationships within the network form individual K-lines. Each K-line is represented by a discrete set of network links. These network links interconnect supporting K-lines or K-nodes, or both, to represent some musical knowledge. All K-lines and K-nodes are implemented as frame structures known as K-line frames (KF). A KF is a structure (schema) representing specific knowledge of some object or concept; the structure associates features which are descriptive of a given object or concept. These features are represented as attributes called slots. Slot values can be either some physical value (such as a symbol or numerical value), or some process/function to be invoked to perform some task. Frames were defined by Minsky as a framework for representing knowledge (Minsky 1975).

In WOLFGANG, the slot values in each KF identify the respective KF and its supporting structures and characteristics. The overall K-line network is partitioned into two functional parts composed of: (1) method facts (e.g., methods of motivic development, methods of harmonization, etc.) and (2) facts (specific instances of intervals, harmonic structures, etc.).

Within a K-line network, K-lines are partitioned into different classes of musical artifacts, such as sets of melodic intervals, rhythmic patterns, harmonic progressions, methods for motivic development, and so on. This is done to provide effective management and efficient access of system memory; the management and access of K-line classes is supported via blackboard technologies (see glossary for discussion on blackboard technologies). Finally, each K-line class is implemented as a frame, called a musical-component-frame (MCF). Each MCF may contain an arbitrary number of slots; each slot in a MCF references a distinct KF contained within a K-line class. The MCFs are implemented as lists.

4.3 Blackboards as Knowledge Negotiators

WOLFGANG's blackboard systems serve as knowledge negotiators. They manage the interactions of K-lines from different K-line classes as they collaborate to compose a musical work. The blackboard technologies are implemented in a distributed hierarchical model. The model consists of a primary blackboard systems supported by (three) subordinate blackboard systems. The primary blackboard system, called the root blackboard (RBB), manages the overall composition process. The three subordinate blackboard systems include the melodic blackboard (MBB), the harmonic blackboard (HBB), and the rhythmic blackboard (RHBB). These three blackboard systems manage composing processes relating to melody, harmony, and rhythm respectively. Also, these three blackboard systems serve as blackboard knowledge sources (KS) to the RBB.

Each blackboard system is implemented as a shared memory allowing access to its respective K-line classes (e.g., K-line melody classes, K-line methods of motivic development classes, etc.). The activation of K-lines within these classes serve as KSs which attempt to assert information provided by each active K-line onto the blackboard. This information is evaluated by a blackboard scheduler. Each blackboard system comprises a scheduler for managing blackboard functions. In WOLFGANG, the schedulers are implemented as inference engines composed of minimal sets of rules relating to the distinct tasks of each respective blackboard. (The schedulers are the only system components within the WOLFGANG architecture whose knowledge is defined explicitly by rules.)

4.4 A Disposition Feedback Facility

The disposition feedback facility provides WOLFGANG with the ability to evaluate decisions made during a composing session. Evaluations are based on the emoting potential derived from each possible decision, and a ranking of each possible decision with regard to an ordered list of previous decisions that demonstrate high emoting potentials. This facility allows WOLFGANG to mark for future use during a given composing session specific musical artifacts that compute high emotional potentials, while satisfying the current disposition of the system.

Implementation of the disposition feedback facility consists of: (1) a feedback loop, (2) a variable called the *dispositionValue*, and (3) an ordered list of decisions called the *listOfGoodElements*.

Both variables, *dispositionValue* and *listOfGoodElements*, are asserted and maintained on the RBB. The variable *dispositionValue* maintains the current disposition of WOLFGANG as one of the four emotion types: happy, sad, angry, and meditative. The variable *listOfGoodElements* provides WOLFGANG with a list of musical elements which have been applied previously during the current composition session, and have provided high emoting potentials. This list is significant; it acts as a short term store of musical artifacts applicable in the motivic development of remaining musical phrases. Thus, during an individual composing session, WOLFGANG creates and appends to the variable *listOfGoodElements* musical features derived from decisions made during the session which provide high emoting potentials. The feedback loop is implemented as a background process that sends status messages to the RRB scheduler when an interesting event has occurred and has been posted on the RBB; the scheduler then appends the musical artifact associated with this event to the *listOfGoodElements*.

4.5 System Logfile

The logfile insures that each composition is distinct; it provides information of previously composed works, so that WOLFGANG can avoid the excessive repetition of musical artifacts by reviewing its previous composing habits during a session. The logfile consists of trace data from the last twenty sessions stored off in a hard disk file.

5. Closing Discussion

WOLFGANG provides a working model of music composition in which ‘‘dispositions” are instrumental in deciding about steps in the elaboration of tonal monodies. The research and design of WOLFGANG has resulted from a subjective view of musical composing according to which the emoting potential of high-level musical constructs is more important to the musical logic of a monody than are their syntactic features. A musical composition is thought to be an artifact which stimulates the senses and cognitive awareness of both its creator and any intended listener. We therefore view composing as a process that creates an artifact to communicate some cognitive (emotional) effect. The composing process necessitates the development of a set of musical skills, and the application of these skills based on the disposition of the composer. We might consider WOLFGANG’s compositional processing as constrained by its cultural grammar, and guided by its disposition to musically communicate some emoting quality.

Acknowledgments

The author is deeply indebted to Marvin Minsky and John Carson
for their helpful interest and comments in this research project.

References

Doyle, J. 1979.
A Truth Maintenance System. A.I. Memo No. 521, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts. Cambridge, Massachusetts.

Michalski, R. S. 1983
A Theory and Methodology of Inductive Learning. In *Machine Learning*, eds. Michalski, R.S., Carbonell, J. G. and Mitchell, T. M.,Los Altos, California: Morgan Kaufmann.

Minsky, M. 1985.
The Society of Mind. New York, New York: Simon and Schuster.

Minsky, M. 1982.
Music, Mind, and Meaning. In Music, Mind, and Brain: *The Neuropsychology of Music*, ed. M. Clynes, New York, New York: Plenum Press. (see also: Minsky, M. 1981. Music, Mind, and Meaning. A.I. Memo No. 616, Artificial Intelligence Laboratory, February, Massachusetts Institute of Technology, Cambridge, Massachusetts.)

Minsky, M. 1980.
K-Lines: A Theory of Memory. *Cognitive Science Journal*, Volume 4, Number 2, April-June, Ablex Publishing. (see also: Minsky, M. 1979. K-Lines: A Theory of Memory. A.I. Memo No. 516, Artificial Intelligence Laboratory, June, Massachusetts Institute of Technology, Cambridge, Massachusetts.)

Minsky, M. 1975.
A Framework for Representing Knowledge. In *The Psychology of Computer Vision*, ed. P. H. Winston, NewYork, New York: McGraw-Hill.

Nii, H. P. 1989.
Introduction. In *Blackboard Architectures andApplications*, eds. Jagannathan, V., Dodhiawala, R., and Baum, L., New York, New York: Academic Press.

Riecken, R. D. 1992.

WOLFGANG - A System Using Emoting Potentials To Manage Musical Design. *In Understanding Music with AI: Perspectives on Music Cognition*, eds., M. Balaban, K. Ebcioglu, and O. Laske, Cambridge, Massachusetts: AAAI Press/MIT Press.

Riecken, R. D. 1989.
Goal Formulation With Emotional Constraints: Musical Composition by Emotional Computation. In *Proceedings of AAAI - First Annual Conference on Innovative Applications of Artificial Intelligence*. Stanford University, Palo Alto, California: AAAI-IAAI.