GODEL TUNE: FORMAL MODELS IN MUSIC RECOGNITION SYSTEMS

By Dominique M. Richard, Ph.D.
18 rue du Parc de Clagny
78000 VERSAILLES
FRANCE

ABSTRACT: Formal models of music have been used as a design principle of music recognition systems. These models range from statistical to generative/transformational models. However, the use of formal models in recognition problems implicates new data structures and algorithmic processes to the mechanism of perception and can result in inconsistent or extremely limited systems. By first modeling the perception mechanism as a statistical decision process merging sensory data and memory information, we can more readily identify the data elements and computational structures required for the recognition task and integrate formal representation(s) of music in a dynamically changing system.

"C'est l'exception qui confirme la règle" (Anonymous)

INTRODUCTION:
Formal models, defined by a choice of representation and a finite set of axioms and rules (Goldstien, 1979), have been used in automatic language recognition systems to complement the purely acoustical recognition of an utterance. Music recognition systems have typically followed a similar structure. In most cases, however, formal models have been used as a posteriori corrections to force a formal structure on a string of symbols derived from an acoustical stream (Levinson, 1980). Unfortunately, the resulting systems are extremely limited as they are too rigid to address a significant domain or are too related to yield consistent results. These consequences mirror Gödel's conclusion which states that formal systems cannot be at once complete and consistent. The use of formal systems in recognition tasks must then be limited to locally consistent systems, dynamically learned from experience.

FORMAL MODELS OF MUSIC THEORY:
Music and language have been formally and informally described, within probabilistic, generative and semantic paradigms in a structural description (e.g. Schank, 1954; Leerdahl and Jackendoff, 1983) or temporal implication description (e.g. Meyer, 1973; Narmo, 1977). In probabilistic models (Chomsky and Miller, 1958) the "language" structure is represented by a series of matrices describing the transition probabilities between symbols or groups of symbols for example (p. notes (Hiller, 1959). In generative and semantic models the "language" is represented by a finite set of production rules. The generative approach (Chomsky, 1975; Schank, 1972) derives its rules from a primal formalism while the semantic approach derives them from experience through the recursive application of arbitrary transformational rules. Expressions of these formalisms are found in Chomsky's "universal grammar" based on the work of Chomsky (1955), in Schank's "universal conceptual language" based on semantic analysis in Levi-Strauss's "universal myth" (1969) based on the interpretation of some 400 South American tales; in Barlow and Jaccobson's work which is based on the analysis of Bach's chorales (1978).

LIMITATIONS OF FORMAL MODELS:
However, the probabilistic, generative and semantic approaches to language/music formalism cannot appropriately represent natural "language." The probabilistic approach was demonstrated as inadequate by Chomsky (1956) and recognized as a simple heuristic for implementing constrained recognition tasks (Bahl & Jelineck, 1983). Despite a heightened concern for the specific structure of language/music the generative and semantic rules of representation were also found inadequate. First of all, since the different theories vary in their premises and in their "universal conclusion," the very concept of "universality" is put into question. In fact, since all these theories are summarized in a finite set of IF-THEN rules, they are defined as formal systems which are inherently incomplete. Furthermore, we note that with the accumulation of several theories into a single system (Emund and al., 1980), a hierarchy must be assumed to avoid the chaining of heuristics into a conflicting conclusion. Increasing the size of a production rule system greatly increases the complexity of the presence of consistency (Godel, 1931). As Lent noted, heuristics derived from formal systems have "something akin to Heisenberg's principle" (1982, 222) they cannot exhaustively represent knowledge as Russell (1902) had hoped.

ICMC GLASGOW 90 PROCEEDINGS

338
REQUIREMENTS FOR MUSIC RECOGNITION:

Because "each new composition tends to employ a unique set of syntactic premises (Meyer, 1967, 237), a music recognition system should be able to adapt to a specific piece, style or interpretation and still rest on solid listening practices. However, these requirements are formally contradictory since they require knowledge of standard practice in the form of production rules (e.g. Winograd, 1968) but also demand the ability to change these very rules. The preference rules introduced by Lordahl and Jakobson (1983) in their generative theory of tonal music are symptomatic of this fact. In his most recent contribution, Narmour (1989) suggests an alternate rule structure for the representation of heuristics by proposing a replacement of the IF...THEN formalism by an IF...THEN...EXCEPT formalism. This new structure allows for the separation of basic production rules, the Gestalt (Jakobson, 1985), from the exception rules which can re-direct or even inhibit the formal rules of perception. This shift in formalism indicates a fundamental change in the analysis by displacing the focus from complete/universal theories to consistent/local theories, concentrating on implicational and learning algorithms rather than on all encompassing models.

CORRELATIVE STRUCTURE OF A MUSIC RECOGNITION SYSTEM:

Jelinek's model of speech recognition (1976), Golstein's model of pitch perception (1978) and Narmour's theory of implication/realization (1977, 1989) can be combined in a comprehensive system of music recognition designed with this "consistent/local" formalism (Richard, 1989). The architecture of such a system is directly inspired from Jelinek's model: its implementation is derived from Golstein's and Narmour's theories. It includes a coder (= acoustic processor) which translates the acoustic stream into a "vector" of pitches conditioned on the measured spectrum, a decoder (= language processor) which calculates the a posteriori distribution of pitches given the a priori current formalism of the piece and algorithms allowing learning based on experience (Figure 1). The implementation of this system uses "subjective" probabilistic constructs to operate on discrete distributions of objects (pitch) and combinations of objects (harmonies, keys, etc.) and parallel distributed processing (PDP) structures to enable detection and dynamic learning (McClelland and Rumelhart, 1986). The basis of its engine is Narmour's implication/realization theory which operates on scales of pitches, intervals, harmonies, keys and other musical constructs to monitor the fluctuations of the IF...THEN rules of Gestalt and to detect the closure of processing and substructures which constitute the exceptions to the current Gestalt rule of perception. The learning algorithm thus establishes the EXCEPT criteria for Gestalt inhibition.

REFERENCES:


ICMC GLASGOW 1990 PROCEEDINGS