A Computational Model of Music Cognition Based on Interacting Primitive Agents

Yuzuru Hiraga
University of Library and Information Science
1-2 Kasuga, Tsukuba, 305 JAPAN
hiraga@u.tsukuba.ac.jp

Abstract
This paper presents a conceptual framework of a computational model of music cognition, in which cognition is characterized through the interaction of simple primitive agents. The topic is centered on 'coarse-receptor' agents, which recognize the skeletal structure of a melody congruent with that of the Implication-Realization Model by Narmour.

1 Introduction
A remarkable aspect of music cognition (and our cognition in general) is that while it can detect the finest of subtleties, it also allows for broad generalizations of ensembling associations among most distantly related phenomena. Such complexity obviously suggests a multi-layered, multi-modal mode of processing. One way of viewing this, and the one adopted in this paper, is that the basic processing units are simplistic but also numerous; and behavioral complexity arises from their mode of interaction, but not from individual intricacy. Such processing units will be dubbed as agents, and conforms with the general methodological framework put forth by Minsky[1985]. To substantiate this approach, it is necessary to identify what the functions of individual agents are (or should be), and what modes of interaction they enable or are possible.

This paper outlines the initial steps of work in this direction. The scope here is limited to the incremental recognition (through segmentation and grouping) of a single-line melody consisting of dissonated tones. The emphasis here is that such a simple setting already gives rise to fundamental issues in computational modeling of cognition. The paper, due to space limitation, will focus on the basic ideas of the framework, and will not present detailed description or extensive discussion.

2 Background
In a computationally oriented approach, the goal is to provide a formal and well-grounded model/theory of music cognition; with all operations rigidly determined, based on a few general principles. This is best illustrated by comparing two extreme cases of music cognition research, namely that of Lerdahl and Jackendoff [1983] and the works by Longuet-Higgins (see [Longuet-Higgins, 1987]; but essentially [Longuet-Higgins and Lee, 1982]).

Longuet-Higgins' work is a 'natural' computational approach — while being limited in scope and capability, the frameworks are simple and clean, enabling testing and evaluation on a rigid basis. On the other hand, Lerdahl and Jackendoff provide a far broader and thorough treatment. However, the framework is not formalized in detail, which causes severe problems when building a computer implementation.

The aim here is to maintain the formality and well-groundedness in the former but retaining the breadth and totality of the latter. In particular, the present framework seeks its conceptual basis in, and can be construed to be subsumed by the Implication-Realization Model extended by Narmour [1990, 1992]. A point of note, though, is that this is not an attempt to implement these 'grand' theories. The strategy is to work bottom-up, following the well-groundedness requirement.

In our particular framework, we further impose the following requirements and restrictions.

- Incremental processing: Processing should proceed with the flow of music, deriving (partial) results at every instant.
- Robustness: The model should be able to cope with disruptive change, performance errors, or even unfamiliar style. That is, it should not be governed by a presupposition of particular style or cultural context.
- Minimal framework: The current framework deals with only pitch and onset time of a one-line melody as input. Other infor-
tion are intended to be incorporated in the incremental development of the framework.

3 Primitives

3.1 Basic Concepts

An agent is a cognitive entity that responds and reacts to a particular, primitive relationship between notes. Here, we consider only single-line melodies, with each note having only pitch and onset time as properties. The time interval between successive onsets are called a span. Thus the relations are either intervalic (difference between pitch) or temporal (difference between spans).

We also restrict our scope to agents with wide receptive ranges—in other words, coarse reception. Agents are dynamically generated and invoked by the relation they respond to; though can be recognized only if the notes are present. Adjacent notes are by definition proximal, and we shall return to the notion of proximity below.

If the same relation holds for successive notes, then the scope of the agent is transitively extended to the whole sequence. That is, for a relation \( R \) and successive notes \( n_1 \) and \( n_2 \), if \( R(n_1, n_2) \) and \( R(n_2, n_3) \), then \( R(n_1, n_2, n_3) \) is said to hold, and the corresponding agent recognizes the sequence as a single entity (the term "transitive sequence" is used to mean both the sequence itself, and the agent that recognizes it). Transitive extensions are the basis for forming groups.

3.2 Intervalic Relations

Intervals are further decomposed into direction (up, down, same) and magnitude (counted e.g. by half-tones). The following are the primitive intervalic relations incorporated here:

- same (two pitches are the same)
- direction: up or down
- magnitude: scalar (sequential), chordal (remote), octave

For example, a transitive extension over an uprelation corresponds to an ascending sequence.

Note that magnitude relations are coarse-classified. A sequential-relation corresponds to a scale-step motion, including half- and whole-tones but further extended to minor (and possibly major) thirds. Chordal (remote)-relations correspond to wider intervals, like thirds, fourths, as found e.g. in a triad. An agent recognizing a sequential (remote) interval will not further distinguish the exact interval size. Thus, a tune played in

\(1\) Duration, i.e. the actual length of a note is also not considered.

\(2\) Though "transitive" itself is coarse-defined.

3.3 Temporal Relations

For the temporal relation over spans, only one type of primitive relation, span equality, is incorporated.

Span equality is also coarse-defined, and a range of tolerance is allowed. An obvious bound that two spans cannot be considered equal is when the shorter note can have two (or more) spans included in a longer note's span—i.e. the ratio is smaller than 1:2. Reversing this, two spans are considered to fulfill the span-equality relation if the shortening ratio is greater than 1:2. Practically, a more narrower tolerance range would be more feasible, although this is sufficient for our present purpose. Such logarithmic basis for comparing note length is actually prevalent, as in the common music notation.

3.4 Proximity

A note is considered to be active (present) during its span. But its perceptual trace, or influence, will remain salient even after the note is over. Intuitively, this salience will be retained longer for a longer note, i.e. influence is a function of span.

The influence of a note is operationally defined as extending one span-length before and after its span boundaries. That is, for a note with span \( s \) on onset \( T \), its influence ranges from \( T - s \) to \( T + s \) (or \( T + 2s \)). Upon the introduction of influence, the notion of proximity is defined as follows: a note \( n \) is proximal to note \( n_o \) if its onset falls within the influence of \( n_o \).

Two remarks are in order. First, proximity is not a symmetric relation. A (shorter) note may be proximal to another (longer) note, although the converse may not hold. Second, by extending the influence before the onset, we have deviated from the notion of perceptual trace. The true moti-
4 Interaction and Grouping

The primitive agents presented above can interact with each other, which become the basis of deriving various music structure suites. The modes of interaction can be one of the following:

- **Reinforcement**: Agents of same mode (temporal/intervallic, direction/magnitude) are in mutual agreement on a structural decision.
- **Cooperativity**: Agents of different mode are in agreement.
- **Repeatability**: Agents of same mode are in contradictory judgement.
- **Suppression**: Agents of different mode are in disagreement.
- **Competitiveness**: Agents (same or different mode) compete for inclusion of the same note.

Some examples of how these are applied are given below.

4.1 Meter

To cope with metrical structure, we introduce a meter agent which is similar to a primitive temporal agent, except that it reacts to a fixed span and not the note span themselves. Since the influence criterion allows looking two beats (sepaus) ahead, a meter agent can circumvent subdivisions and dotted or doubled notes.

Two meter agents are reinforcing (congruent) if the set of notes extrapolated from their respond-

68.5 ICMC Proceedings 1993
cases lead to essential ambiguity, some which may be deliberately implanted by the composer. On the other hand, if all agents agree on a segmentation, the decision is decisive but the music flow will be severely disrupted.

The claim here is that in ordinary music, disambiguation of such cases is possible through the intervention of other agents, and by a relatively simple scheme. The above case, for example, can be disambiguated e.g. if a meter agent suggests which of note 2 or note 3 is on a stronger beat. For this to be possible, the meter agent must be somehow established beforehand. What the claim states is that that will exactly be the case.

Note that large intervalic magnitude does not imply a point of segmentation, and that magnitude relation types should be treated on equal grounds. This is the case in Figure 2, where the smaller intervals (as in notes 2-3) actually are strong candidates for segmentation. In this case, judgement is derived from the ascending sequence in dotted lines (a hierarchical construct) and replicated (conformant) patterns in notes 1-2, 3-4, and 5-6. This may further be overridden if the strong beat is on note 2 and on. Thus all judgements are dynamic and context dependent, although they will be possible at that point. A particular point to note is with same-pitch relations. Namely, same-pitch suggests strong linkage, but the fact that they are divided (and not played as one note) is also (strongly) indicative of a point of segmentation.

In the temporal domain, equal-span sequences also suggest groups, but the situation is a bit more complex. Here we draw on the proximity criteria, requiring that notes included in a note's influence suggest grouping. In particular, if a short note sequence follows a long note, then those falling in the long note's proximity are grouped with (actually, governed by) it. Conversely, a short sequence followed by a long note will be blocked and cannot extend their influence any further, thus segmentation will occur (either before or including the long note).

4.3 Hierarchical Structure

Recognized groups can be mutually related to form larger structures. At a lower level, two modes of structuring can be identified. One is that a particular member is promoted to represent the whole group, inheriting its entire span and thus extending its influence. The other is concatenation, in which continuity at the surface structure serves to link the groups.

Hierarchical structures (as in [Lerdahl and Jackendoff, 1983]) are essentially a promotion-based way of constructing larger-order structures. Although their significance in general is unquestionable, the present framework sees this as part of the whole extent of conceivable structures. Their extent of applicability can also be questioned, although this will not be discussed here.

5 Remarks and Conclusion

This paper presented a brief outline of an agent-based model of music cognition. This can be characterized, in effect, as an attempt to approach the Implication-Realization Model of Narmour from a computationally based standpoint.

There are many issues that have not been mentioned and/or incorporated into the current framework. One is the detection of conformance, or the similarity of sequences. Basically, the judgement itself can be reduced to the reinforcing interaction between a previous and current sequence over some relation. However, a precise, formal account of what counts as conformance, and how such processing is invoked in the first place, present intricate issues to be solved. Other issues include implication, or predictive/predemptive mode of processing, and toality. Although for the latter, the present stance is that a good deal of results can be obtained without appealing to toality.

An initial implementation on computer has been made and experimented on simple folk tunes. The results show a general approval of the overall scheme, although also revealed many points at which further extension and refinement are necessary. These results and a more full version of the scheme will be reported elsewhere.

References


