Acquisition of Human Feelings in Music Arrangement

Masayuki Numao **Masashi Kobayashi** **Katsuyuki Sakaniwa***
Department of Computer Science, Tokyo Institute of Technology
2-12-1, O-okayama, Meguro-ku 152, JAPAN
Email: numao@cs.titech.ac.jp

Abstract

We often make decisions based on our feelings, which are implicit and very difficult to express as knowledge. This paper details an attempt to acquire feelings automatically. We assume that some relations or constraints exist between impressions felt and *situations*, which consist of an object and its environment. For example, in music arrangement, the object is a music score and its environment contains listeners, etc. Our project validates this assumption through three levels of experiments. At the first level, a program simply mimics human arrangements in order to transfer their impressions to another arrangement. This implies that the program is capable of distinguishing patterns that result in some impressions. At the second level, in order to produce a music recognition model, the program locates relations and constraints between a music score and its impressions, by which we show that machine learning techniques may provide a powerful tool for composing music and analyzing human feelings. Finally, we examine the generality of the model by modifying some arrangements to provide the subjects with a specified impression.

1 Introduction

KANSEI (Japanese; lit. *human feelings*) [Tsuji, 1995] has been analyzed using quantitative psychological analysis methods, such as the semantical differential (SD) method and multivariate analysis, which analyze human feelings. Such analyses can isolate feelings associated with known objects, but cannot predict feelings for a new object, nor create a new object for the purpose of generating in the subject a specific feeling. We would like to introduce and discuss a system capable of predicting feelings and creating new objects based on *seed* structures that have been extracted and are perceived as favorable by the test subject, such as patterns and colors for pictures, or spectrums and their transition for sounds. Until now, such emergent structures have been obtained only through a random and intractable combination of elements. In this paper, we explain how this difficulty is overcome using machine learning techniques.

As a representative medium, we focus on a MIDI-based music arrangement system which is applied to automatic selection or arrangement in an online KARAOKE system. This system is used for automatic downloading of MIDI data as requested by a user, or can be used more generally for information retrieval or filtering based on emotional data.

*Presently at Matsushita Electric Industrial Co., Ltd., Kadoma, Osaka, 571, JAPAN

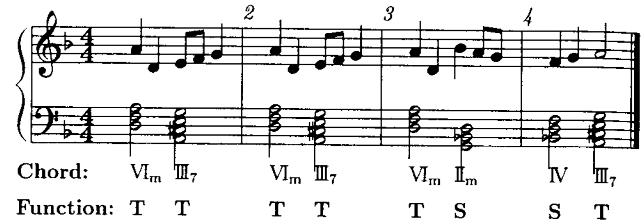

Figure 1: Melody, Chords and their Functions

2 Melody and Chords

We attempt to extract a musical structure based on melody and chords as shown in Figure 1. In a musical piece, a function — *tonic* (T), *dominant* (D), *subdominant* (S) or *subdominant minor* (SDm) — is assigned to each chord. This paper discusses the extraction of two aspects of the structure (i.e., each chord and a sequence of functions) from which the system derives constraints for assigning chords to a melody (supplemented by functions).

3 Mimicking Arrangements

To investigate the feasibility of generating arrangements automatically, the authors constructed a system that mimics human arrangements as shown in Figure 2.

The chord analyzer assigns a function to each chord by parsing chord progression, and translates scores arranged by corresponding human composers into training examples. Each example consists of a chord and its

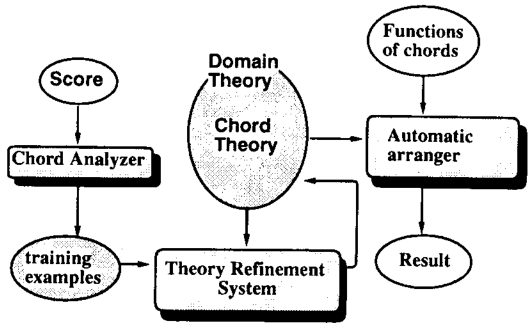

Figure 2: Automatic Arrangement System

corresponding function, from which the system learns (derives) constraints for automatic arrangement. The system is initially provided with a general theory of harmony, which is refined by a learner called a *theory refinement system* [Tangkitvanich and Shimura, 1992; Numao and Shimura, 1989; Mooney and Ourston, 1994]. Each note in the score has a function, to which the automatic arranger assigns a chord.

Figure 3 shows an example of such an arrangement. The system can assign chords based on the general theory of harmony, even though they are mediocre. Theory refinement introduces some *decorations* to the chords and refines the techniques used in the training examples, improving the arrangement.

The present authors prepared a set of training and test examples by the same composer. Figure 4 compares the arrangements produced by the initial and refined theories with those of the human composer, which shows that refinement based on the training examples increases the number of matched chords in the test examples.

Verbal reports from the subjects characterized the arrangements produced by the initial theory as *simple*, *not interesting* and *flavorless*, while those produced using the refined theory are described as *refreshing* and *novel.* Learning using multiple scores enhances *variety*, but the results lack uniformity.

There is more demand today for professional MIDI arrangers than there was several years ago, since digital synthesizers, electronic pianos, piano players, online KARAOKE systems, etc. have become more popular. This type of arrangement system will satisfy these needs by learning and rivaling the skills of professional arrangers.

4 Music Recognition Model

The system outlined above only mimics arranged scores and does not consider human feeling, so we have attempted to develop a system for acquiring music recognition model. Such a model is necessary to achieve automatic arrangement based on situation-dependent human feelings. The authors prepared some musical pieces of 8 bars, and then played them for some subjects in order to get their impressions. This was accomplished using the semantic differential (SD) method. The results were

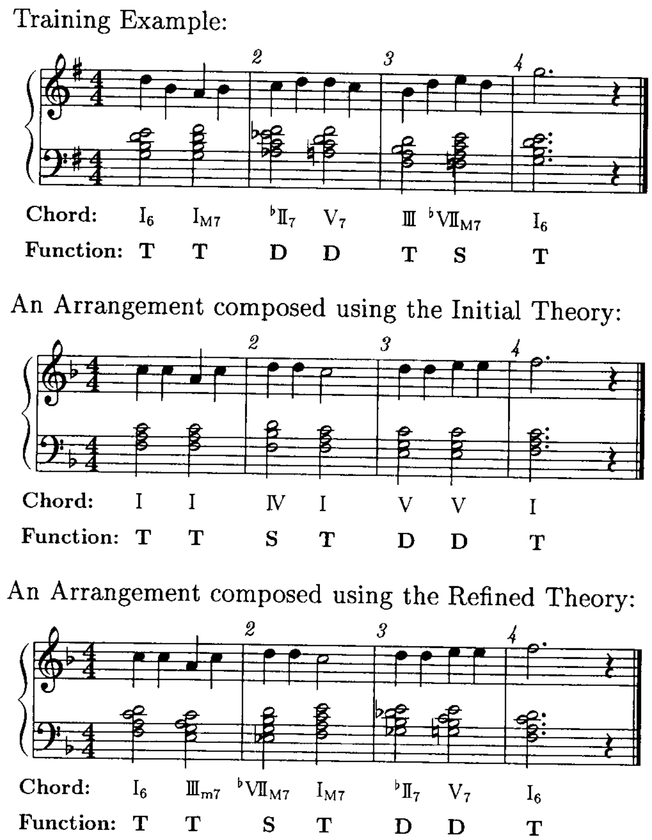

Figure 3: An Example of Arrangement

utilized by an inductive logic programming (ILP) system [DeRaedt, 1996] capable of deriving a model for music recognition.

4.1 Training Examples

In order to categorize responses by subjects based on the musical pieces they listened to, the authors selected the following 5 pairs of complementary adjectives: *bright - dark*, *clear - unclear*, *fast - slow*, *favorable - unfavorable*, *stable - unstable.* A musical piece is evaluated as one of 7 grades for each pair, such that 1 indicates most *bright*, and 7 indicates most *dark* in the spectrum presented by a bright-dark pair.

Let evaluation of a piece P for a pair of adjectives A be E_A^P for a subject ($1 \leq E_A^P \leq 7$). The system generalizes melody by analyzing notes in each bar based on background knowledge described in Section 4.2. For each piece P consisting of 8 bars: $Bar_1, Bar_2, \cdots, Bar_8$, the system learns to evaluate A from the following 8 training examples:

$$(Bar_1, E_A^P), (Bar_2, E_A^P), \ldots, (Bar_8, E_A^P)$$

where Bar_i is a sequence of (*pitch_name*, *length*). An example is considered positive if $5 \leq E_A^P \leq 7$, and negative if $1 \leq E_A^P \leq 4$. The system also weighs each learned clause according to the values E_A^P of the examples used.

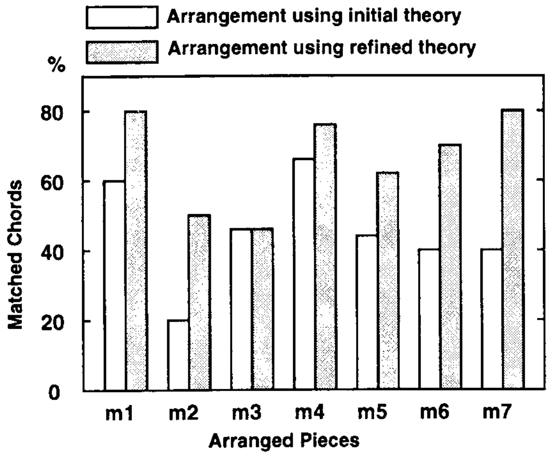

Figure 4: Matched Chords

The system generalizes chord progression by analyzing successive chords. For each piece P with a sequence of n chords: $Chord_1, Chord_2, \cdots, Chord_n$, the system learns to evaluate based on the following $n-1$ training examples:

$$
\begin{gathered}
(Chord_1, Chord_2, E_A^P) \\
(Chord_2, Chord_3, E_A^P) \\
\cdots \\
(Chord_{n-1}, Chord_n, E_A^P)
\end{gathered}
$$

where $Chord_i$ is a combination of the following notes and their function:

Root: i, ♭ii, ii, ♭iii, iii, iv, ♭v, v, ♭vi, vi ♭vii, vii

Third: major, minor, suspended fourth

Fifth: fifth, augmented, diminished

Seventh: none, major sixth, minor seventh, major seventh

Function: T, D, S, SDm

which constructs $12 \times 3 \times 3 \times 4 \times 4 = 1728$ combinations. The model describes the relationship among two successive chords and their functions. The high number combinations and relations results in neither neural networks nor decision trees being appropriate as learning tools. Instead, what is required is inductive logic programming that learns not only attributes and propositional descriptions, but also predicates for finding useful relations in obscure combinations. In the experiment, the authors use a learner similar to FOIL [Quinlan, 1990] except that its background knowledge may also be described using Horn clauses.

4.2 Background Knowledge for ILP

The model is constructed based on background knowledge — definitions of predicates that describe melody and chords. Melody is considered to be a sequence of notes that have pitch and duration described by the following predicates:

- Average duration
- Minimum (lowest) pitch
- Maximum (highest) pitch
- Difference between minimum and maximum pitch
- Pitch transition (rising or falling)

Chords are analyzed based on background knowledge as follows:

`root_i(Chord)` : The root of `Chord` is i.

`root_v(Chord)` : The root of `Chord` is v.

`major(Chord)` : `Chord` has the major third.

`fifth(Chord)` : `Chord` has the perfect fifth.

`seventh(Chord)` : `Chord` has the minor seventh.

`tonic(Chord)` : The function of `Chord` is T.

`dom(Chord)` : The function of `Chord` is D.

`subdom(Chord)` : The function of `Chord` is S.

`succ(Chord1,Chord2)` : `Chord1` and `Chord2` are successive chords.

Using the background knowledge for each adjective pair, some predicates *adjective_pair(Example, Weight)* are derived. E.g. `bright-dark(`*Example of Chords, Weight*`)` to detect bright or dark bar is learned as follows:

```
bright-dark((C1,C2,_),7)
  :- succ(C1,C2),major(C1),subdom(C1),dom(C2).
bright-dark((C1,_,_),6)
  :- major(C1),dom(C1),root_v(C1).
```

4.3 Predicting an evaluation

After the system creates a recognition model for a subject by processing the results of his/her evaluation of various pieces, it predicts the subject's evaluation of subsequent pieces, which are transformed into a set of examples for analyzing melody and chords. If an *adjective_pair(Example,* W_j*)* is satisfied by k_j examples, the evaluation for the adjective pair is calculated by

$$\frac{\sum_j W_j k_j}{\sum_j k_j}.$$

4.4 Experiments in Recognition

The present authors prepared 100 well-known music pieces, from which they extracted 8 successive bars without modulation. The subject evaluates 85 of the 100 pieces, and the results for adjective pairs are studied by the system to predict evaluations for the other 15 pieces. They then take an average of the results by 8 different subjects.

Let F_A^p be a prediction of the evaluation of a piece p for adjective pair A, and n be the number of pieces. The difference between the prediction and the evaluation by a subject is shown to be:

$$diff_A = \frac{1}{n} \sum_{p=1}^{n} |E_A^p - F_A^p|.$$

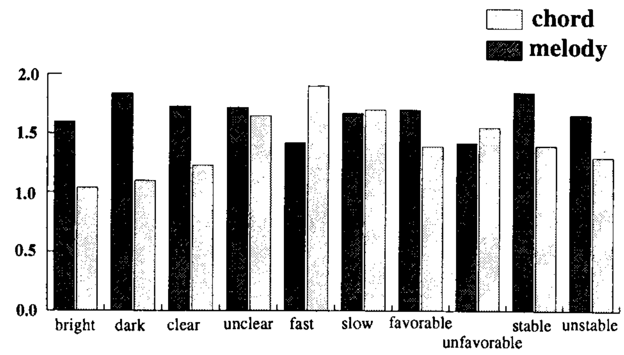

Figure 5: Average Difference

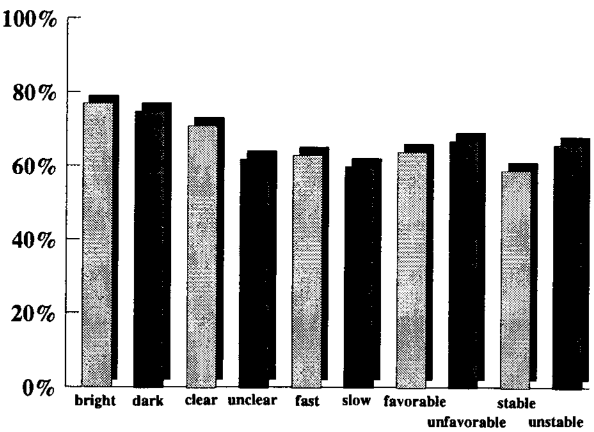

Figure 6: Prediction Accuracy

Figure 5 shows the average difference for the subjects: $\frac{1}{8}\sum_{s=1}^{8} \mathit{diff}_A$, when the prediction is based on chords or melody. Figure 6 shows the percentage of correct predictions. The results show that the differences are between 1.0 and 2.0 in the 7-grade evaluation, and that on average 70% of predictions are correct, which indicates that the system is capable of predicting evaluation of new pieces by the subjects very well. For criteria *fast*, *slow* and *unfavorable*, the difference increases when the prediction is based on chords. This means that these impressions are based mainly on melody and particularly from the length of each note. The system predicts responses within the criteria *bright* and *dark* from chords very well. These impressions are dependent mainly on whether the third is major or minor. Although the system cannot discern these impressions in melody, the authors are now attempting to detect them by analyzing sequences of notes.

Figure 7 shows the variance of impressions among the subjects, which is relatively large in the adjective pairs fast-slow and stable-unstable. These subject-dependent pairs are learned by the system very well according to Figures 5 and 6.

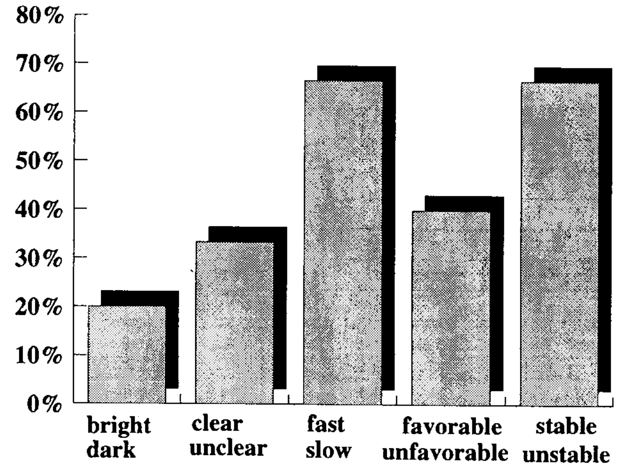

Figure 7: Pieces with Variance > 1

This level of arrangement will be employed in KARAOKE systems for generating a user model of emotional data to extract an appropriate piece from a MIDI database, or to select suitable background pictures or video for each title.

5 Arrangement based on the Model

Utilizing the music recognition model, the system controls the arrangement process based on its overall mood. To improve the arrangement, we assume that a human arranger (or the system shown in section 3) composes the original arrangement, which is then modified slightly to result in a specified impression.

5.1 Modifying a Score to Change its Impression

Before arrangement, the system develops a music recognition model for the user. Information about the mood of the user is provided in the form of a 7-grade scale for each of the adjective pairs. The chord progression of a score is modified according to the following priorities:

1. Remove any chord or chord progression whose impression is described by the opposite adjective in the pairs.
2. To minimize differences from the original score, avoid:
   (a) modifying any chord and chord progression that satisfy the given adjective,
   (b) changing any functions, and
   (c) modifying any chord and chord progression that is not in opposition to the given adjective.
3. Modify chords to minimize differences in the evaluation.

5.2 Experiments in Arrangement

The authors prepared 94 well-known music pieces without modulation, from which they extracted 4 or 8 suc-

Table 1: Percentage of pieces with matching impressions

subject	bright	stable	favorable	average
A	45	55	50	50
B	50	75	60	63
C	71	71	50	65
D	78	54	33	55
E	91	64	78	77
F	67	55	58	58
G	45	54	70	56
H	56	83	80	74
average	67	62	61	63
standard deviation	16.2	10.7	14.9	8.7

Table 2: Percentage of pieces with improved impressions

subject	brightness	stable	favorable	average
A	60	33	37	41
B	45	54	63	54
C	50	78	44	57
D	70	64	43	61
E	82	56	56	68
F	57	50	75	62
G	40	75	75	64
H	50	100	71	74
average	58	62	59	60
standard deviation	13.0	18.9	14.3	9.3

cessive bars. The subject evaluates 80 of the 94 pieces in 3 pairs of adjectives: *bright - dark*, *favorable - unfavorable*, *stable-unstable*, and the results of the evaluations are processed by the system to modify the chord progressions of 6 of the 14 pieces in 6 criteria — bright, dark, favorable, unfavorable, stable and unstable. The subject evaluates the modified $6 \times 6 = 36$ pieces and the original 6 pieces without being notified of how each piece has been modified.

The present authors repeated the above experiment using 8 subjects. Table 1 shows the percentage of arrangements whose modifications corresponded to the intended changes, as well as their standard deviation. According to the table, intended arrangements are produced on average 60% of the time. Table 2 shows the percentage of arrangements for which the subjects' impression is improved and the standard deviation.

Figure 8 shows the average percentage of pieces for which the impression is improved by the modification. According to the figure, 60% of arrangements are improved for the criteria brightness or darkness. In general, if an arranger is less than proficient, the results tend to be evaluated as unstable or unfavorable. Therefore, it is more likely that a piece will be arranged and evaluated as unstable or unfavorable than as stable or favorable.

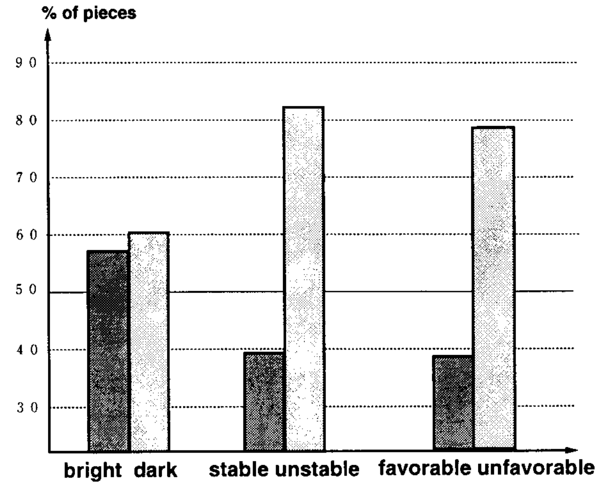

Figure 8: Improvement in impression

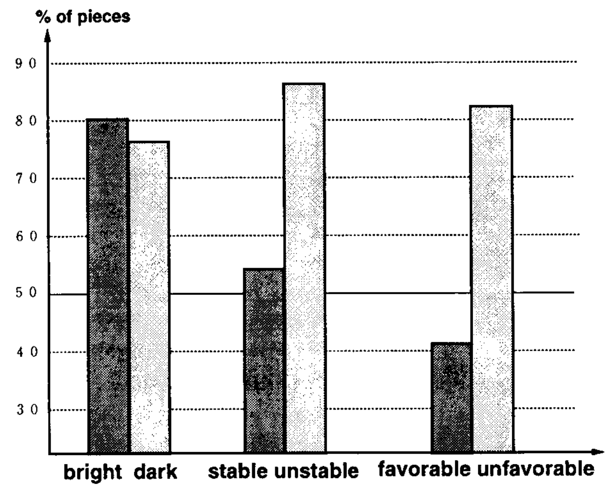

Figure 9: Impression modified (opposing criteria)

According to the figure, some arrangements are evaluated as stable or favorable, although these make up less than half of the total. This suggests that an arranger based on learning personal feelings is effective.

If an original arrangement was evaluated as very bright, it would be difficult to rearrange the piece for a higher level of brightness. It would be easier to rearrange the piece to less bright. Figure 9 shows the average percentage of pieces successfully rearranged to cause the opposite impression. According to the figure, 80% of arrangements are rearranged in this fashion.

6 Related Works

Widmer [1994] proposed a method of accomplishing explanation-based learning by attaching harmonies — chord symbols to the notes of a melody. This task is closely related to that described in Section 3. The present paper further discusses a means of controlling the process based on learned feelings.

Katayose, Imai and Inokuchi [1988] approached the understanding of music based on the following rules:

Melody:
Both the fourth and the seventh do not exist
→ Oriental mood

Our system not only uses this kind of rule but also creates it and applies it through a process that considers its weight.

Most applications of machine learning related to music investigate approaches to interpretation of a music score for playing, for example, measuring the loudness of each note in a score [Widmer, 1993; 1994], or acquiring playing viola [Furukawa, 1997; 1996].

Neural Networks have been used for dealings with problems in music [Todd and Loy, 1991]. The present authors believe that inductive logic programming offers a better means of describing music scores, in which structures are very important. To quantify feelings in music, we introduce a weight W_j to each clause. A better solution is to combine logic programming and neural networks by weighting links in dynamic logical networks, as described in [Numao *et al.*, 1997].

7 Conclusion

We present an approach for utilizing human feeling in listening to and composing music. This approach to feelings is an interesting test of the application of machine learning techniques, and should be developed further due to the lack of practical tests to date. If background knowledge of the learner incorporates some important theories in musicology and psychology [Meyer, 1956; Hiraga, 1987; Longuet-Higgins, 1987], we may obtain a more powerful tool for composing music and analyzing human feelings.

Acknowledgments

An earlier version of the system was implemented by Takashi Shirai (Nintendo), Koji Yamaguchi (Matsushita Electric) and Masatake Saito (SONY).

References

[DeRaedt, 1996] L. DeRaedt, editor. *Advances in Inductive Logic Programming.* IOS Press / Ohmsha, Amsterdam, Oxford, Tokyo, Washington, DC, 1996.

[Furukawa, 1996] K. Furukawa. Towards verbalization of tacit knowledge by inductive logic programming. In *Keio International Workshop on Verbalization of Tacit Knowledge based on Inductive Inference.* 1996.

[Furukawa, 1997] K. Furukawa. A framework for verbalizing unconscious knowledge based on inductive logic programming. In *Machine Intelligence 15.* Oxford University Press, 1997.

[Hatano, 1987] G. Hatano, editor. *Music and Cognition (in Japanese).* University of Tokyo Press, Tokyo, 1987.

[Hiraga, 1987] Y. Hiraga. A knowledge representation for recognition of music (in Japanese). In *In [Hatano, 1987]*, chapter 4, pages 97–130. 1987.

[Katayose *et al.*, 1988] H. Katayose, M. Imai, and S. Inokuchi. An approach to extract sentiments in music (in Japanese). *Journal of Japanese Society for Artificial Intelligence*, 3(6):748–754, 1988.

[Longuet-Higgins, 1987] H. C. Longuet-Higgins. Music. In *Mental Processes*, chapter II, pages 57–188. The MIT Press, Cambridge, MA, 1987.

[Meyer, 1956] L. B. Meyer. *Emotion and meaning in music.* University of Chicago Press, 1956.

[Michalski and Tecuci, 1994] R. S. Michalski and G. Tecuci, editors. *Machine Learning: A Multistrategy Approach (Vol. IV).* Morgan Kaufmann, San Francisco, CA, 1994.

[Mooney and Ourston, 1994] R. J. Mooney and D. Ourston. A multistrategy approach to theory refinement. In *In [Michalski and Tecuci, 1994]*, chapter 5, pages 141–164. 1994.

[Numao and Shimura, 1989] M. Numao and M. Shimura. Explanation-based acceleration of similarity-based learning. In *Proceedings of the Sixth International Workshop on Machine Learning*, pages 58–60, Palo Alto, CA, 1989. Morgan Kaufmann.

[Numao *et al.*, 1997] M. Numao, S. Morita, and K. Karaki. A learning mechanism for logic programs using dynamically shared substructures. In *Machine Intelligence 15.* Oxford University Press, 1997.

[Quinlan, 1990] J. R. Quinlan. Learning logical definitions from relations. *Machine Learning*, 5:239–266, 1990.

[Tangkitvanich and Shimura, 1992] S. Tangkitvanich and M. Shimura. Refining a relational theory with multiple faults in the concept and subconcept. In *Machine Learning: Proc. 9th International Workshop*, pages 436–444, 1992.

[Todd and Loy, 1991] P. M. Todd and D. G. Loy, editors. *Music and Connectionism.* The MIT Press, Cambridge, MA, 1991.

[Tsuji, 1995] S. Tsuji, editor. *KANSEI Information Processing (in Japanese).* Grant in Aid for Scientific Research of Ministry of Education, Science, Culture, and Sports of Japan under Grant-in-Aid for Special Project Research 04236107, 1995.

[Widmer, 1993] Gerhard Widmer. Understanding and learning musical expression. In *Proc. International Computer Music Conference*, pages 268–275, 1993.

[Widmer, 1994] G. Widmer. Learning with a qualitative domain theory by means of plausible explanations. In *In [Michalski and Tecuci, 1994]*, chapter 25, pages 635–655. 1994.