

Lecture Notes in Artificial Intelligence

746

Subseries of Lecture Notes in Computer Science

Edited by J. Siekmann

Lecture Notes in Computer Science

Edited by G. Goos and J. Hartmanis



Andranick S. Tanguiane

Artificial Perception and Music Recognition

Springer-Verlag

Berlin Heidelberg New York

London Paris Tokyo

Hong Kong Barcelona

Budapest

Series Editor

Jörg Siekmann

University of Saarland

German Research Center for Artificial Intelligence (DFKI)

Stuhlsatzenhausweg 3

D-66123 Saarbrücken, Germany

Author

Andranick Tanguiane

Fernuniversität Hagen

Postfach 940, D-58048 Hagen, Germany

CR Subject Classification (1991): I.2, I.5, E.2

ISBN 3-540-57394-1 Springer-Verlag Berlin Heidelberg New York

ISBN 0-387-57394-1 Springer-Verlag New York Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 1993

Printed in Germany

Typesetting: Camera ready by author

Printing and binding: Druckhaus Beltz, Hemsbach/Bergstr.

45/3140-543210 - Printed on acid-free paper

Foreword

In this book I summarize my studies in music recognition aimed at developing a computer system for automatic notation of performed music. The performance of such a system is supposed to be similar to that of speech recognition systems: acoustical data at the input and music score printing at the output.

In this essay I develop an approach to pattern recognition which is entitled *artificial perception*. It is based on self-organizing input data in order to segregate patterns before their identification by artificial intelligence methods. The performance of the related model is similar to distinguishing objects in abstract painting without their explicit recognition.

In this approach I try to follow nature rather than to invent a new technical device. The model incorporates the *correlativity of perception*, based on two fundamental perception principles, the grouping principle and the simplicity principle, in a very tight interaction.

The grouping principle is understood as the capacity to discover similar configurations of stimuli and to form high-level configurations from them. This is equivalent to describing information in terms of generative elements and their transformations.

The simplicity principle is modeled by finding the least complex representations of data that are possible. The complexity of data is understood in the sense of Kolmogorov, i.e., as the amount of memory storage required for the data representation.

The tight interdependence between these two principles corresponds to finding generative elements and their transformations with regard to the complexity of the total representation of data. This interdependence justifies the term “correlativity”, which is more than relativity of perception.

The model of correlative perception is applied to voice separation (chord recognition) and rhythm/tempo tracking.

Chord spectra are described in terms of generative spectra and their transformations. The generative spectrum corresponds to a tone spectrum which is repeated several times in the chord spectrum. The transformations of the generative spectrum are its translations along the \log_2 -scaled frequency axis. These translations correspond to intervals between the chord tones. Therefo-

re, a chord is understood as an acoustical contour drawn by a tone spectral pattern in the frequency domain.

Time events are also described in terms of generative rhythmic patterns. A series of time events is represented as a repetition of a few rhythmic patterns which are distorted by music elaboration and tempo fluctuations associated with the tempo curve. The interdependence between tempo and rhythm is overcome by minimizing the total complexity of representation, e.g., the total amount of memory needed for storing rhythmic patterns and the tempo curve.

The model also explains the function of interval hearing, certain statements of music theory, and some phenomena in rhythm perception.

Generally speaking, I investigate hierarchical representations of data. In particular, I pose the following questions:

- (a) Why a hierarchy?
- (b) Which hierarchy? and
- (c) How does the hierarchy correspond to the reality?

From the standpoint of the model, the answers to these questions are, respectively:

- (a) A hierarchy makes a data representation compact, which is desirable in most cases;
- (b) consequently, a better hierarchy is one which requires less memory for the related data representation; and
- (c) under certain assumptions such a hierarchy reveals perception patterns and causal relationships in their generation, making the first step towards a semantical description of the data.

One can see that the main distinction of this approach is finding optimal representations of data instead of directly recognizing patterns. In a sense, analysis of patterns is replaced by synthesis of data representations. Since self-organization is used instead of learning, the threshold criteria used in most pattern recognition models are avoided.

The correspondence between music perception and the performance of the model, together with the diversity of its applications, can hardly be regarded as simply a coincidence. It makes an impression that the model really simulates certain perception mechanisms. Probably, the related model can be applied to speech recognition, computer vision, and even simulation of abstract thinking. All of this is a subject for discussion.

This book has been written during my work in Grenoble at the ACROE-LIFIA (Association pour la Création et la Recherche sur les Outils d'Expression-Laboratoire d'Informatique Fondamentale et d'Intelligence Artificielle).

I acknowledge Professor Philippe Jorrand, the director of the LIFIA, and Dr. Claude Cadoz and Dr. Annie Luciani, the heads of the ACROE, for inviting me to Grenoble and the stimulating working conditions; my colleagues, especially Martial Barraco, for their friendly attitude and help in various domains; and Remi Ozoux, a student of ENSIMAG (Ecole Nationale Supérieure d'Informatique et de Mathématiques Appliquées de Grenoble), who has fruitfully collaborated on the project and who has programmed the latest version of the algorithm for chord recognition.

Hagen, 14 September 1993

Andranik Tangian (Tanguiane)

Contents

1	Introduction	1
1.1	Formulation of the Problem	1
1.2	Brief Survey of Music Recognition	2
1.3	Brief Survey of Artificial Perception	6
1.4	Development of Correlativity Principle	15
1.5	Contribution to Music Recognition	23
1.6	Summary of the Book	24
2	Correlativity of Perception	27
2.1	Introductory Remarks	27
2.2	Principle of Correlativity of Perception	28
2.3	Model of Correlative Perception	35
2.4	Method of Variable Resolution	38
2.5	Complexity of Transformation as a Distance	40
2.6	Distinctions of the Model	41
2.7	Summary of the Chapter	43
3	Substantiating the Model	45
3.1	On the Adequacy of the Model	45
3.2	Tone Spectra	47
3.3	Representation of Tones	52
3.4	Generation of Chord Spectra	53
3.5	Unique Deconvolution of Chord Spectra	56
3.6	Causality and Optimal Data Representation	70
3.7	Interpretation of the Results	72
3.8	Main Items of the Chapter	75
4	Implementing the Model	77
4.1	Reduction of the Model	77

4.2	Properties of Boolean Spectra	79
4.3	Necessary Condition for Generative Patterns	86
4.4	Algorithm for Finding Generative Patterns	89
4.5	Summary of Reduced Model	94
5	Experiments on Chord Recognition	95
5.1	Goals of Computer Experiments	95
5.2	Example of Chord Recognition	98
5.3	Testing the Simple Correlation Approach	105
5.4	Recognition Mistakes	109
5.5	Efficiency and Stability of Recognition	112
5.6	Testing the Decision Making Approach	121
5.7	Testing the Structural Approach	124
5.8	Judging Computer Experiments	129
6	Applications to Rhythm Recognition	131
6.1	Problem of Rhythm Recognition	131
6.2	Rhythm and Correlative Perception	134
6.3	Correlativity and Recognition of Periodicity	137
6.4	Timing Accentuation	140
6.5	Rhythmic Segmentation	142
6.6	Operations on Rhythmic Patterns	144
6.7	Definition of Time and Rhythm Complexity	147
6.8	Example of Analysis	148
6.9	Summary of Rhythm Perception Modeling	152
7	Applications to Music Theory	153
7.1	Perception Correlativity and Music Theory	153
7.2	Logarithmic Scale in Pitch Perception	155
7.3	Definition of Musical Interval	156
7.4	Function of Relative Hearing	160
7.5	Counterpoint and Orchestration	163
7.6	Harmony	168
7.7	Rhythm, Tempo, and Time	174
7.8	Summary of the Chapter	176
8	General Discussion	177

<i>CONTENTS</i>	<i>xi</i>
-----------------	-----------

9 Conclusions	181
----------------------	------------

References	185
-------------------	------------

Index	201
--------------	------------

Name Index	201
----------------------	-----

Subject Index	204
-------------------------	-----

List of Figures

1.1	Parallel primes, fifths, and octaves prohibited in counterpoint . .	15
1.2	The use of parallel voices in <i>Bolero</i> by M.Ravel	17
1.3	Duality of chord contours and melody trajectories	19
2.1	High-level pattern of <i>B</i> composed by low-level patterns of <i>A</i> . .	29
2.2	Pattern of <i>B</i> composed by unknown symbols	29
2.3	A succession of time events	31
2.4	Four representations of the same succession of time events . . .	31
2.5	Theme in augmentation and diminution from Bach's <i>The Art of Fugue</i>	33
2.6	Contextual similarity of rhythmic patterns	35
2.7	Flying ball in a cinema sequence	37
2.8	Illustration to the method of variable resolution	39
3.1	The overtone series	49
3.2	Example of tone spectra	51
3.3	Example of chord spectrum	55
4.1	Example of Boolean spectra of tones	80
4.2	Example of Boolean spectrum of chord	81
4.3	No unique deconvolution of Boolean spectra	85
5.1	Spectral representations of chords	99
5.2	Correlation analysis of chord spectra	101
5.3	Two hardly recognizable chords	111
5.4	The 130th chorale from J.S.Bach's <i>371 Four-Part Chorales</i> . . .	113
5.5	Maximal values of the sufficient number of patterns and maximal number of accidentals in the sufficient set of patterns . . .	117
5.6	Average value of the sufficient number of patterns and average number of accidentals in the sufficient set of patterns	118

5.7	Maximal and average values of the limits of the number of successive standard partials inherent in true patterns from the sufficient set	119
5.8	Chords used for testing the recognition algorithm	125
6.1	Multilevel repetition structure of rhythm from M.Ravel's <i>Bolero</i>	135
6.2	Representation of time events with variable resolution	139
6.3	Ambiguous segmentation of a periodical sequence of time events	141
6.4	Accentuation by timing cues	143
6.5	Rhythmic segmentation by timing cues	143
6.6	Syllable as an indecomposable unit	143
6.7	The elaboration of a crotchet rhythmic pattern	145
6.8	Two different junctions of the same rhythmic syllables	147
6.9	A rhythm with the indicator of complexity 2	149
6.10	Determination of time by recognizing rhythmic syllables	149
7.1	Ambiguity in defining coordinates of a complex object	157
7.2	Distance between dissimilar objects	157
7.3	Distance between similar objects	157
7.4	Thorough bass notation based on interval relationships	161
7.5	Transposition in notation but not in sound	162
7.6	Transposed chord and melody in standard notation	162
7.7	Invariance of guitar fingering with respect to chord transpositions	162
7.8	Sonorous effect from a complex harmony in <i>Bolero</i> by M.Ravel	165
7.9	Beginning of <i>Ricercar</i> from J.S.Bach's <i>Musical Offering</i>	167
7.10	A.Webern's orchestral arrangement of J.S.Bach's <i>Ricercar</i>	167
7.11	Spacing between voices more than an octave prohibited in counterpoint	168
7.12	Frequency ratios for the scale of just intonation in <i>do-major</i>	169
7.13	The voices parallel in notation but not in sound	169
7.14	Voice crossing prohibited in counterpoint	171
7.15	Voice overlapping prohibited in counterpoint	171
7.16	Filling skips in strict counterpoint	171
7.17	Cluster chords with resolutions	173
7.18	Sequence of cluster chords	173
7.19	Major triads in root position in harmonic and melodic junctions	173
7.20	Minor triads in root position in harmonic and melodic junctions	173
7.21	Instability of optima	175

List of Tables

2.1	The complexity of representation of time events	31
2.2	Artificial intelligence and artificial perception in pattern recognition	43
3.1	Correspondence between visual and audio data	45
5.1	Most salient melodic intervals between two chords	103
5.2	Most salient harmonic intervals in the first chord	103
5.3	Most salient harmonic intervals in the second chord	103
5.4	Correlations of chord spectra	107
5.5	Specifications of recognition procedure	109
5.6	Specifications of recognition procedure restricted to 12 semitones	110
5.7	Specifications of recognition procedure not restricted	110
5.8	Specifications of recognition procedure for one experiment on chord recognition	114
5.9	Preliminary and final results of chord recognition	122
5.10	Recognition of chord $C_7 = (c; e; g; b)$	125
5.11	Recognition of chord $C_{7M} = (c; e; g; h)$	126
5.12	Recognition of chord $C_{7/9} = (c; e; g; b; d_1)$	126
5.13	Recognition of chord $C_{7M/9} = (c; e; g; h; d_1)$	127
5.14	Recognition of chord $C_{6/9} = (c; e; g; a; d_1)$	127
6.1	Correspondence between visual and time data	133
6.2	Autocorrelation $R(p)$ of time events	139