Profiles of Pitch Classes Circularity of Relative Pitch and Key – Experiments, Models, Computational Music Analysis, and Perspectives

vorgelegt von
Diplom-Mathematiker
Hendrik Purwins
aus Münster

von der Fakultät IV – Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
– Dr. rer. nat. –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. B. Mahr
Berichter: Prof. Dr. K. Obermayer
Berichter: Prof. Dr. W. Auhagen

Tag der wissenschaftlichen Aussprache: 15. August 2005

Berlin 2005
D 83

Abstract

The doubly-circular inter-relation of the major and minor keys on all twelve pitch classes can be depicted in toroidal models of inter-key relations (TOMIR). We demonstrate convergence of derivations on the explanatory levels of a) an experiment in music psychology, b) geometrical considerations in music theory, and c) computer implementation of musical listening scenarios.

Generalizing Shepard [1964] to full overtone spectra, circular perception of relative pitch is experimentally verified and mathematically modeled as the spectrum of pitch differences, derived from virtual pitch [Terhardt, 1998]. Musical examples of circular pitch, tempo, and loudness are analyzed.

For each pitch class calculating the intensity in a musical recording, our constant quotient (CQ-) profile method is a) consistent with psychological probe tone ratings, b) highly efficient, c) computable in real-time, d) stable with respect to sound quality, e) applicable to transposition, f) free of musical presupposition, except approximately equal temperament, and g) sensitive to substantial musical features (style, composer, tendency towards chromaticism, and major/minor alteration) in a highly compact reduction. In Bach, Chopin, Alkan, Scriabin, Hindemith, and Shostakovich, the latter features are extracted from overall CQ-profiles by classification (support vector machine [SVM], regularized discriminant analysis) and clustering. Their inter-relations are visualized by a technique called Isomap.

Leman [1995] models acquisition of inter-key relations. Artificial cadences are preprocessed through modeling the auditory periphery. Then, in the toroidal self organizing feature map (SOM) a TOMIR is arrived at. We extend this approach by a) using a great many actual performances of music as input, and/or b) not presupposing toroidal topology a priori. Visualizing CQ-profiles from Bach's WTC by correspondence analysis (CA) and Isomap reveals the circle of fifths. TOMIRs evolve from a) average CQ-profiles of Chopin's Préludes *and toroidal SOMs, b) overall annotated duration profiles of Bach's WTC I from score and CA, and c) formalization of music theoretical derivation from Weber [1817]'s key chart by the topographic ordering map introduced here. These results are consistent with Krumhansl and Kessler [1982]'s visualization of music psychology ratings.*

As auxiliary results, we discuss fuzzy distance, spatial and systematized synesthetic visualization in conjunction with beat weight detection on multiple time scales and suggested applications to automatic tone center tracking. Furthermore, statistics on the key preference of various composers are collected. Based on the latter, CA visualizes composer inter-relations.

This thesis substantially contributes to content retrieval (MPEG-7), automated analysis, interactive audio-based computer music, and musicology.

Acknowledgments

I am very grateful towards my supervisor Klaus Obermayer for his support. His vivid Neural Information Processing (NI) Group provided an effective infrastructure and an open and pleasant atmosphere of scientific adventure.

I want to thank some people that made substantial contributions to this research and that have been especially inspiring and encouraging. First of all I want to express my gratitude towards Benjamin Blankertz. We worked together on auditory models, the constant Q profil technique, toroidal models of inter-key relations, and machine learning. I collaborated with Immanuel Normann on experiments and models of circular perception of relative pitch in harmonic complex tones. Thore Graepel and I worked on correspondence analysis.

Always helpful, Christian Piepenbrock set up the initial software infrastructure for the NI group. Thanks to Wilhelm Köhler and Uli Pralle for maintaining a huge network of workstations. I am grateful to a couple of computer wizards, mostly NI group members, who supported me: Robert Anniés, Hauke Bartsch, Anca Dima, Thomas Knop, Roman Krepki, Nabil Lashin, Johannes Mohr, André Przywara, Stephan Schmitt, Holger Schöner, Susanne Schönknecht, Lars Schwabe, Frank Segtrop, Michael Sibila, Ilya Souslov, and Peter Wiesing. I have to thank the audio experts Kamil Adiloglu, Michael Natterer, and Marcus Verwiebe. Cornelius Weber and Oliver Beck were helpful "room mates" at the office. I owe a lot to Thomas Noll. Several times he pointed out to me new fruitful directions in my research. In addition, he always let me use the equipment of the research group Mathematical Music Theory. Jörg Garbers also was helpful. I learned much from Ulrich Kockelkorn's well-founded statistical insight. Thanks for his support in solving statistical problems. I want to express my gratitude towards the people from the studio for electroacoustic music at TU Berlin for letting me use the equipment and for support: Gerhard Behles, Onnen Bock, Folkmar Hein, Manfred Krause†, Axel Roebel, and anonymous subjects of the listening tests.

At Stanford I was warmly welcomed by Jonathan Berger, Stefan Bilbao, Fabien Gouyon, Stefan Harmeli, Heidi Kugler, Fernando Lopez-Lezcano, Max Matthews, Eleanor Selfridge-Field, Julius Orion Smith. I had a very fruitful exchange with Craig Sapp. I had inspiring discussions with Marina Bosi, Takuya Fujishima, Dan Gang, David Huron, Dan Levitin, Bob Sturm, and Harvey Thornburg.

At McGill Auditory Research Laboratory, I have to thank Al Bregman for thorough discussions. Pierre Ahad was very helpful with setting up my working environment. I had an inspiring discussion with Jim Ramsay.

Antonie Budde, Christopher W. Jones, Hans Peter Reutter, and Kotoka Suzuki advised me in music analysis. Mark Lindley shared his insight into temperament with me. Young-Woo Lee played many piano samples. Joo-Kyong Kim recorded for me. Thanks to Wolfgang Auhagen for evaluating this thesis. I am indebted to Stefan Bilbao for proof reading the major part of the thesis and to Thomas Gonsior, Christopher W. Jones, Timour Klouche, Ingo Lepper, Dennis Livingston, Thomas Noll, and Joshua Young for proof reading sections thereof. In addition I have to thank Matthias

Bode, Joachim Buhmann, Matthias Burger, Gunnar Eisenberg, Georg Hajdu, Sepp Hochreiter, Petr Janata, Carsten Kruse, Marc Leman, Markus Lepper, Fabien Lévy, Guerino Mazzola, Helga de la Motte-Haber, Hans Neuhoff, Miller Puckette, Ingo Schießl, Sambu Seo, Manfred Stahnke, Baltasar Trancon-y-Widemann, Gregor Wenning, Daniel Werts, and Ioannis Zannos. Since so many people supported me, I would like to ask pardon for forgetting someone.

I am obliged to my parents who supported me in the most comprehensive sense of the word but also very specially related to this thesis: my father, who triggered my interest in physics, and my mother, who provided me with music education from early on. Last not least I have to thank my friends.

I am grateful for financial support by the German National Academic Foundation (Studienstiftung des deutschen Volkes) and Axel Springer Stiftung.

Contents

Is Music Reducible to Numbers, Mechanics, and Brain States? 13
A Machinist Intruding Musicology . . . 13
Profiles from Theory, Score, and Sound . . . 14
Inter-Key Relations . . . 16
Infinity – Construction of "Paradoxes" . . . 20
Interdisciplinary Perspectives on Music . . . 21
Surface Level or Underlying Deep Structure? . . . 21
Sister Sciences Mathematics and Music . . . 22
The Brain Comprehended? . . . 24
Machine "learning" . . . 26
Ground Truth through Inquiry of the Listener . . . 27
Organization of Thesis . . . 28

1 From Pitch Criticism to Tone Center Geometry 31
1.1 Fuzzy Foundations of Pitch . . . 31
1.1.1 The *Psycho*acoustic Nature of Pitch . . . 31
1.1.2 Sums of Sine Tones – Natural Sounds . . . 32
1.1.3 Frequency Scaling . . . 33
1.1.4 Loudness Summation of Frequency Components . . . 34
1.1.5 Singling Out Frequency Components . . . 36
1.1.6 Few Frequencies – Several Tones . . . 37
Beating, Roughness, and Combination Tones . . . 37
Virtual Pitch . . . 38
1.1.7 Several Frequencies – One Tone . . . 40
1.2 Musical Landscapes . . . 43
1.2.1 Tonnetz, Temperaments, and Torus . . . 45
Criticism of the Notion of Tone . . . 47
1.2.2 Pitch Classes . . . 48
1.2.3 Pitch Decomposition into Pitch Class and Brightness . . . 48
1.2.4 Chords and Harmony . . . 51
The Key to Tonality . . . 51
Categorization of Synchronous Pitch Classes . . . 52
Syntax and Music . . . 53
1.2.5 Key Proximity . . . 54

1.2.6 Chart of Tone Centers . . . 54
1.2.7 Torus by Synchromatic and Enharmonic Identification . . . 56
1.2.8 Key Character . . . 59
1.2.9 Key – Reduction to an Empirical Profile . . . 61
Probe Tone Ratings . . . 61
A TOMIR from Probe Tone Ratings . . . 63
Limitations as a Psychological Reference . . . 64
1.2.10 Spatial and Planar Views of Tones, Triads, and Keys . . . 64
1.2.11 Sequences of Tone Centers . . . 67
Temporal Granularity . . . 67
Modulation . . . 67

2 Signals, Ear Models, Computer Learning 69
2.1 Constant Q Transform . . . 69
2.1.1 Deriving Filter Parameters . . . 69
2.1.2 An Efficient Algorithm . . . 70
2.2 Auditory Models and Autocorrelation . . . 71
2.2.1 Auditory Models . . . 71
Non-linear Transduction . . . 72
Information Coding Hypotheses . . . 74
Hebbian Learning and Tonotopy . . . 74
Meddis Auditory Model . . . 74
2.2.2 Pitch Detection Algorithms . . . 75
Frequency Differences Between Partials . . . 75
Calculation of Virtual Pitch . . . 76
Virtual Pitch and the Proximity Principle . . . 78
Pitch and Timbre . . . 78
Autocorrelation with and without an Auditory Model . . . 79
Parameters . . . 79
Missing Fundamental . . . 79
Missing Fundamental of Non-Resolved Partials . . . 81
Ambiguous Pitch . . . 81
Shifted Residual Series . . . 82
Limited Impact of Auditory Model . . . 83
2.3 Classification, Clustering, Visualization . . . 83
Supervised versus Unsupervised Learning . . . 83
Evaluating Classification Algorithms . . . 83
Cross-Validation . . . 84
Receiver Operating Characteristic . . . 84
2.3.1 Supervised Learning . . . 84
K-Nearest-Neighbor Classifier . . . 84
Regularized Discriminant Analysis . . . 85
Support Vector Machines . . . 86
2.3.2 Analysis of Co-occurrence . . . 86

2.3.3 K-Means Clustering . . . 89
2.3.4 Embedding and Visualization . . . 89
Correspondence Analysis . . . 89
Principle . . . 89
Historical Note . . . 91
Isomap . . . 91
Toroidal Self-Organizing Feature Map (SOM) . . . 91

3 Results **93**
3.1 The Sawtooth Culmination Pattern . . . 93
3.2 Circular Pitch in Harmonic Complex Tones . . . 95
3.2.1 Octave Ambiguity and Octave "Paradox" . . . 95
3.2.2 Experiment: Judgment on Pitch Comparison . . . 97
Results . . . 98
Statistical Formulation and Evaluation . . . 101
The Subjects´ Comments . . . 101
3.2.3 Dimensionality, Proximity, Quantization, and Competition . . . 102
3.2.4 Model: Spectrum of Pitch Differences . . . 105
Applications . . . 106
3.2.5 General Construction Principle . . . 108
3.2.6 Circularity in Music Literature . . . 110
Circular Pitch . . . 110
Circular Tempo and Loudness . . . 117
3.2.7 Discussion . . . 119
3.3 Pitch Class Content of Musical Audio . . . 120
3.3.1 Constant Q Profiles . . . 120
Calculation of CQ-Profiles . . . 121
The CQ-Reference Set . . . 123
3.3.2 Fuzzy Distance and Key-Finding . . . 123
3.3.3 CQ-Profiles and Probe Tone Ratings . . . 124
3.3.4 Musical Corpus . . . 125
3.3.5 Composers and Pieces Portrayed in CQ-Profiles . . . 126
3.3.6 Discussion . . . 130
3.4 Mode, Style, and Composer Information in Audio . . . 130
3.4.1 Composer Classification . . . 130
3.4.2 Which Scale Degree Does Signify Mode? . . . 131
3.4.3 Mode as Salient Feature in Clustering . . . 132
3.4.4 Visualizing Modes, Composers, and Pieces . . . 135
3.4.5 Discussion . . . 139
3.5 Circle of Fifths . . . 139
3.5.1 Equal Temperament for Infinite Tonal Progression . . . 140
3.5.2 Circle of Fifths Emergence in Correspondence Analysis . . . 141
Circle of Fifths from Score . . . 141
Circle of Fifths from Performance . . . 144

3.5.3 Circle of Fifths Evolving in Isomap . . . 146
3.5.4 Discussion . . . 146
3.6 Composers Characterized by Key Preference . . . 147
3.6.1 Key Preference Statistics . . . 149
3.6.2 Key Character and Key Statistics . . . 150
3.6.3 Analysis . . . 151
3.6.4 Discussion . . . 151
3.7 Visualization of Inter-Key Relations . . . 152
3.7.1 Key-Color Maps . . . 152
3.7.2 Topographic Ordering Map . . . 154
3.7.3 Discussion . . . 156
3.8 Toroidal Models of Inter-Key Relations . . . 156
Topology of a Toroidal Surface . . . 157
3.8.1 TOMIR Evolving in Topographic Ordering Map . . . 157
A Note on our Visualization of SOMs . . . 157
Simulations . . . 158
3.8.2 TOMIR Evolving in a Neuromimetic Environment . . . 159
3.8.3 TOMIR Evolving in a Cognitive Model from Performance . . . 161
Tone Center Dissimilarities: Theory – Model . . . 162
3.8.4 TOMIR Emergence in Correspondence Analysis . . . 164
Consistency with Chew's Geometric Model. . . . 164
3.8.5 Discussion . . . 166
3.9 Outlook – Tone Center Transitions . . . 168
3.9.1 Timing and Tone Centers . . . 168
Onset Detection . . . 168
Beat Strength Extraction . . . 170
3.9.2 Chopin's c–Minor Prélude . . . 171
3.9.3 Moving Average Analysis . . . 173
3.9.4 Multiple Time Scale Analysis . . . 174
3.9.5 Ambiguity and Poly Tone Center Analysis . . . 176
3.9.6 Tone Center Path on the Torus . . . 177
3.9.7 Error Function . . . 178
3.9.8 Moving Average Analysis of Simple Modulations . . . 179
3.9.9 Discussion . . . 180

Summary **183**

A The Binding Problem **187**
A.1 Implementations of Acoustical Source Separation . . . 188

B More, Explicit CQ-Profiles **189**
B.1 Profile Decomposition . . . 191

C Technical Details **195**
C.1 Data and Metrices 195
C.2 Support Vector Machine 197
C.3 Visualization 198
C.3.1 Correspondence Analysis 198
C.3.2 Isomap 199
C.3.3 Chew's Model with Choice of Parameters 200

Bibliography **201**

Introduction – Is Music Reducible to Numbers, Brain States, and Mechanics?

Inspired by the Pythagorean identification of musical harmony and geometrical beauty, musical notions like pitch and key are depicted by geometrical objects such as lines, circles, and toroids. Circular structures are motivated by the desire of the musician to ascend in one musical parameter beyond any limit. This is approximated in the sawtooth pattern of slow rise, sudden fall, slow rise and so forth. The helix [Drobisch, 1855] is a stark description of the most striking musical "paradox", a scale endlessly ascending in pitch [Shepard, 1964]. The helix decomposes pitch into the linear component of brightness and the circular component of pitch class. It proves to be fruitful to represent music in the highly compressed form of a profile of pitch classes. Such a profile is a 12-dimensional vector, each component referring to a pitch class. For each component the frequency of occurrence or the accumulated intensity of that pitch class is given. This pitch class representation not only reveals much information about style and composer, but also about mode and key. In a natural way the circle of fifths and a toroidal model of inter-key relations (TOMIR) emerge from pieces of music given in the pitch class profile representation.

A Machinist Intruding Musicology As expressed in Die Automate [Hoffmann, 1821], it is common to distinguish between, on the one hand, science and engineering and arts and humanities, especially music:

> Der größte Vorwurf, den man dem Musiker macht, ist, daß er ohne Ausdruck spiele, da er dadurch eben dem eigentlichen Wesen der Musik schadet, oder vielmehr in der Musik die Musik vernichtet, und doch wird der geist- und empfindungsloseste Spieler noch immer mehr leisten als die vollkommenste Maschine, da es nicht denkbar ist, daß nicht irgend einmal eine augenblickliche Anregung aus dem Innern auf sein Spiel wirken sollte, welches natürlicherweise bei der Maschine nie der Fall sein kann. [1]

[1] The harshest criticism against a musician is that he plays without feeling, since thereby he harms the very nature of music, or rather destroys music by playing the music, and even the most

Abandoning Hoffman for now, let us show that the machine *can* play a role in music.

In the traditional study of literature, problems like authorship, dating, and interrelations between literary works require high scholarship encompassing not only a very refined sense for the aura of each word and the nuances of the language, but also knowledge of the writer's biography, their historical circumstances, and philosophy. However, a paper appears, entitled: A Mechanical Solution of a Literary Problem (Mendenhall [1901], cf. also Farey [2004]). Mendenhall merely counts how often each word length occurs in a literary work. For each work he obtains a histogram displaying the average frequencies of occurrence of all found word lengths. He then discovers that the word length profiles for Christopher Marlowe and William Shakespeare are almost identical. The similarity of these numerical profiles supports the – heavily debated – hypothesis that Christopher Marlowe and William Shakespeare are the same person. Even though the Shakespeare identity problem cannot be answered definitively, it views literature and linguistics from an – at that time – unorthodox perspective. It indicates the effectiveness of, taking a minimum of domain specific knowledge into consideration, mere counting and calculation in the analysis of a literary work.

Profiles of Pitch Classes from Theory, Score, and Sound In the simplest case, binary twelve-dimensional vectors, like in Figure 0.1, are employed to represent chords and keys, e.g. a C–major scale is represented by a "1" in the 1st (c), 3rd (d), 5th (e), 6th (f), 8th (g), 10th (a), and 12th (b) component and a "0" for all other pitch classes. Such twelve-dimensional binary vectors can be considered pitch class sets, interpreting each entry in the twelve-dimensional vector as a membership function, "1" indicating that the specified pitch class is contained in the pitch class set and "0" otherwise. So far, we make the same abstraction as in musical set theory [Forte, 1973], before branching out. Instead of only using "0" and "1" for each pitch class, we can further refine the resolution of the profile, e.g. by using integer or real numbers. Instead of merely indicating whether a pitch class is contained in a particular set or not, we can assign a degree of prominence to it. We could say, for example, that the tonic (1st component) is most prominent, followed by the dominant (8th component), followed by the major third, and so forth. In the context of a given key, Krumhansl and Kessler [1982] derive such a prominence profile from a psychophysical protocol, the probe tone experiment, yielding a so called probe tone rating. For a piece in that key, mere counting of the occurrences of pitch classes results in the *profile of frequency of occurrence*, highly consistent with the probe tone ratings. In place of mere occurrences we can also add up the annotated durations for each pitch class throughout the entire score, yielding the *profile of overall annotated durations*. It would be worthwhile to retrieve pitch class information directly from

soulless and insensitive player will still perform better than the most perfect machine, since it is unthinkable that even a transient passion from inside would never affect his play, which, naturally, can never be the case for the machine.

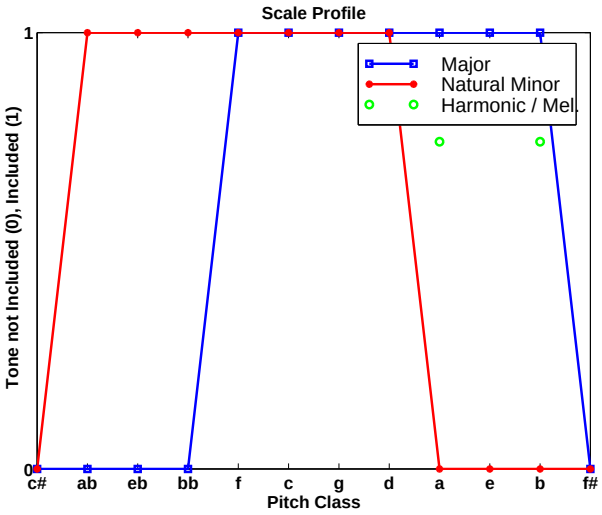


FIGURE 0.1: For each key a binary pitch class vector indicates if a pitch class is included in the scale (> 0) or not ($= 0$). Pitch class profiles are shown for C–major and for melodic and harmonic C–minor. The green circles highlight the substituted tones for harmonic (b for $b^\flat$) and melodic minor (b for $b^\flat$ and a for $a^\flat$).

a recorded performance without the necessity of having a symbolic representation, e.g. a score, on-hand. For several tasks this could usefully work around the generally unsolved, hard problem of automatic polyphonic pitch tracking and transcription. A straight forward approach is to use a more or less neuromimetic pitch analysis technique, e.g. some spectral representation, preferably one with equal resolution in the logarithmic frequency domain. It would be desirable to have a method available that is highly efficient, robust, real-time operable, and validated by a psychological experiment as a reference. A fast method would enable the implementation in low cost sound chips. High robustness especially could cope with historical recordings of poor sound quality. Operability of the algorithm with no look-ahead would make the method suitable for responding to a performance in real-time, e.g. as an analysis module for monitoring and further (semi-)automatic sound transformation such as sound control and machine accompaniment. Reference to psychological data would emphasize the perceptual relevance of the method. For this reason, in Section 3.3.1, we develop the constant Q (CQ-) profile method that exactly meets all of these requirements. A constant Q profile is calculated with the highly efficient, robust, real-time constant Q transform [Brown and Puckette, 1992; Izmirli and Bilgen, 1996]. CQ-profiles closely match with probe tone ratings if spectral information of the playing instrument is taken into consideration. Fujishima [1999] calculates pitch class profiles [2] to represent chords, without using the efficient constant Q transform. Sheh and Ellis [2003] extend his work. We can also consider more special pitch

[2]Sometimes they are also called "chroma profiles".

class profiles, e.g. by counting only pitch classes occurring on certain beat weight positions. It is also possible to consider seven-dimensional diatonic profiles, only indicating the scale degree intensities. Norming the pitch class profile with respect to the keynote yields the normalized constant Q (NCQ-) profile. In the NCQ-profile, the first component always refers to the keynote.

Which features of the musical content of a piece can be retrieved, based on the noisy and very compressed information contained in the normalized constant Q profile? Can the composer be determined, with nothing but a normalized constant Q profile at hand? What can it tell us about the stylistic relations between pieces? How does a composer discriminate between major and minor? These are the questions we scrutinize in *Section 3.4*, covering results presented in Purwins et al. [2004a]. Regularized discriminant analysis and support vector machines with radial basis functions can discriminate between composers quite well using no other feature than the normalized constant Q profile. These methods outperform (regularized) linear discriminant analysis and k-nearest neighbors. By means of t-test and the sparse linear programming machine, in *Section 3.4.2*, we derive how composers use pitch class intensities to discriminate between major and minor. *Section 3.4.3* discloses how the mode of a piece is the most salient property, emerging in the data using k-means clustering. The position of the piece relative to the cluster centers reveals its "majorness". Positions in between the cluster centers for major and minor can indicate three phenomena: 1) chromaticism, 2) a modulation to the parallel major/minor key in the middle section, or 3) a wide range of tone centers on the circle of fifths that are visited during the piece. In *Section 3.4.4*, from visualization by Isomap it can be seen how certain pieces either group together or distinguish from each other due to their chromaticism, switch to parallel major/minor, range of modulations, and ambiguous or extended tonality. Statistical analysis of inter style relations based on counting pitch class occurrences in the score are done by Fucks and Lauter [1965] and Beran [2003], the latter using principal component analysis, linear discriminant analysis, hierarchical clustering, and various other visualization techniques.

How can we group or separate composers solely based on their preference for a particular key and avoidance of another one? Statistics on key preference in composers yields a composer & key co-occurrence matrix. A *stylescape* is a planar or spatial arrangement of composers and schools of composition, displaying their interrelations. In *Section 3.6*, stylescapes are generated by correspondence analysis. Overlaying the map of composers and the map of keys, the biplotting technique links stylistic characteristics to favored keys. Interdependence of composers and schools is meaningfully visualized according to their key preferences.

Inter-Key Relations How can we arrange keys on a manifold so that walking on the surface of this manifold corresponds to a smooth transition from one key to a musically adjacent other key?

We will see that from key kinships and charts of tone centers the circle of fifths and TOMIRs emerge in a couple of different setups under varying epistemologi-

cal assumptions: [3] 1) in a purely speculative music theory framework that can be modeled by an algorithm, 2) in hand-made analysis of music literature, 3) in music psychology experiments, and 4) in various cognitive models operating on score as well as audio data. A cognitive model can process a piece of music in audio or symbolic representation. Such a model consists of preprocessing, modeling the auditory periphery, and a schema,[4] implemented as a machine learning technique with some allusion to cognitive learning mechanisms. Several cognitive approaches share two properties: First, analysis is based on text-book like simple cadences composed of Shepard tones. Second, toroidal SOMs are used. To use musical examples synthesized with Shepard tones is, in general, a drastic simplification of the acoustical reality. The claim of some authors to use the acoustical signal instead of a symbolic representation, as e.g. given in a score [Leman, 1995], cannot be fully met by using textbook like cadences made of Shepard tones, since in Shepard tones all components have the same pitch class, with no noise or transient components. Therefore, they are so simple that they can almost be considered explicitly annotated pitch classes. We can refer to Shepard tones as a quasi symbolic representation.

In Western music of the major-minor tonality period, roughly from Bach's birth (1685) to Wagner's Tristan (1857), inter-key relations are determined by the tonic-dominant relations and the major-minor duality, i.e. parallel and relative relations. In search for an explanation, Weber [1817] remarks that for dominant, subdominant, and relative keys the scales differ from the tonic scale only in one tone. The tonic and its parallel scale share the same principal scale degrees, i.e. I, IV, and V. Werts [1983]'s analysis of modulations in a large corpus of music reveals that modulations to dominant, subdominant, parallel and relative keys are most common (*Section 1.2.7*). The outcome of iterative application of these key kinships can be visualized by Weber [1817]'s grid-like chart of tone centers, later echoed by Schönberg [1969]. Within this chart, the identification of enharmonically different spellings yields the circle of fifths.

The circle of fifths traces back to the circular display of keys in Figure 1.13 by Heinichen [1728]. [5] Fed with musical data sets, the circle of fifths emerges in various setups of cognitive models. Petroni and Tricarico [1997] apply a simplified version of Terhardt [1998]'s virtual pitch model (*Section 2.2.2*) in conjunction with a non-toroidal SOM. The data consists of major and minor triads, and dominant seventh chords, composed of Shepard tones. After training on the SOM, the major triads arrange along the circle of fifths. When training the SOM with probe tone ratings of all keys, instead of chords, the SOM projects the major triads, not used in training, along the circle of fifths as well. [6] In a different experiment, correspondence analysis

[3] TOMIR related research in this thesis has been also published in Blankertz et al. [1999a]; Purwins et al. [2000a, 2001b, 2004b, 2006]

[4] Definition on p. 42.

[5] For a detailed account on the history of the circle of fifths cf. Werts [1983].

[6] Petroni and Tricarico [1997], p. 171. Using the same data fed to a *toroidal* SOM (Leman [1995], p. 90–92) fragments of the circle of fifths emerge. In Leman and Carreras [1997], preprocessing by an auditory model [Immerseel and Martens, 1992] in conjunction with a toroidal SOM is applied

serves as a schema. This schema can be used to process the profiles of frequency of occurrence of pitch classes. Specific results are shown for profiles extracted from digitized scores of fugues in Bach's WTC (*Section 3.8.4*). There is exactly one profile for each key. Correspondence analysis then projects keys, expressed by the frequency profile, on a planar *keyscape*. [7] Vice versa, pitch classes can be viewed as profiles of frequency of occurrence in a specific key, indicating how often this pitch class is used in pieces of the various keys. Vice versa to the process applied to keys, pitch classes can be projected to a space (*pitchscapes*) spanned by the keys. In both scenarios, a homogeneous circle of fifths emerges in the scapes. (*Section 3.5.2*) In the correspondence analysis of CQ-profiles of Bach's WTC II (fugues) the circle of fifths evolves also. If an Isomap is used as a schema and the Isomap is provided with all CQ-profiles in WTC the circle of fifths evolves as well. The two latter results from correspondence analysis and from Isomap are more far-reaching than the above mentioned experiments, since they are based on digitized music recordings, do not require averaging or normalization of the data, and circularity is not included as an inherent feature of the model.

In Weber [1817]'s chart of tone centers, we can identify multiple occurrences of the same or enharmonically equivalent tone. In addition to the circle of fifths, we observe a circular parallel and relative relation, resulting in a torus, referred to specifically as a TOMIR. [8] This genesis can be mathematized using a relative of the SOM referred to as the topographic ordering map. The input is a set of key pairs that are assumed to be closely related. The algorithm aims at finding a spatial configuration which respects stipulated constraints. Presupposing fifth, parallel, and relative key relationships as "close", a simulation with the topographic ordering map yields a TOMIR, very similar to the familiar one derived from Weber [1817]. Furthermore, this doubly circular representation is consistent with music psychology experiments. The intrinsic doubly circular inter-relation of the probe tone ratings becomes apparent after down-scaling them to a four-dimensional Euclidean space by multidimensional scaling. [9] Closer inspection reveals that the scaled points which correspond to the major and minor keys lie approximately on a sub-manifold formed by the crossproduct of two cycles (cf. Figure 1.20). Similar to processes previously described, this sub-manifold can be mapped to a TOMIR.

Werts [1983]'s analysis of modulations in Schubert supports the significance of the toroidal representation of inter-key relations. (cf. *Section 3.5.1*) In cognitive models, provided with musical data, it is studied whether, in addition to the circle of fifths, a

---

to the preludes of Bach's WTC I. Then the preprocessed cadential Shepard chord progressions in all keys are mapped by the trained SOM. The configuration of mapped major keys is, to a low degree, reminiscent of the circle of fifths. In Leman [1995], the display of Figure 6 on p. 156 apparently aims at resembling images produced in optical recording in brain imaging.

[7] By "keyscape" and "pitchscape" we mean a low, usually two- or three-, dimensional arrangement of keys or pitch classes respectively. The notion keyscape has been introduced by Sapp [2001].

[8] Cf. Werts [1983], p. 18, Figure 47. For a detailed history of TOMIRs cf. Hyer [1995]; Cohn [1997]; Douthett and Steinbach [1998].

[9] Cf. Krumhansl and Kessler [1982]; *Section 1.2.9* and Figure 1.19.

TOMIR might emerge as well. A toroidal SOM, preceded by a model of the early auditory pathway, is fed by cadential chord progressions played on a piano: a TOMIR clearly emerges.[10] (*Section 3.8.2*) In another experiment, instead of a very realistic auditory model, average constant Q profiles are calculated in the preprocessing stage, preceding the toroidal SOM. A recording of Chopin's PRÉLUDES, formated as audio data, constitutes the analyzed corpus. Fed by the profiles a TOMIR evolves in the SOM. [11] (*Section 3.8.3*) In this setup, only octave equivalence and the chromatic scale – a minimal amount of knowledge of music in the tonal period – is employed. Again we apply correspondence analysis to the pitch class profiles of frequency of occurrence of Bach's WTC I. We employ biplots to embed keys and pitch classes in the keyscape to visualize their interdependence. After a change of co-ordinates, the four-dimensional biplots can be interpreted as a configuration on a torus (*Section 3.8.4*), closely resembling results from music theory and experiments in listener models. In summary, the eminent role of the TOMIR is emphasized in various experiments and epistemological settings. Also colors may be employed to depict inter-key relations. In *Section 3.7.1*, synesthetic key-color codes are developed and classified.

Listeners are able to trace modulations from key to key, even if they cannot explicitly name them. Can a machine capture modulations as well? In the last four decades, several approaches for tone center tracking have been published. Early attempts mostly explicitly implement rule systems suggested by harmonics textbooks. [Simon, 1968; Winograd, 1968; Longuet-Higgins and Steedman, 1970; Holtzmann, 1977] Many approaches [Krumhansl, 1990; Huron and Parncutt, 1993] are based on representing key as a pitch class profile of frequency of occurrence [Krumhansl and Kessler, 1982], or as a matrix of pitch class interval occurrences [Butler, 1989]. Bharucha and Todd [1989] and Leman [1990] model music cognition. For Leman [1990], the tonal reference for tone center analysis is provided by idiomatic Shepard tone cadences, preprocessed by an auditory model, subsequently treated by a multi time scale attractor model, as an allusion to retrospective tonal re-evaluation. Gang and Berger [1999]'s system accepts music in MIDI format. Their recurrent net learns to predict harmonic development, by learning metrical-harmonic inter dependence. Griffith [1994] conducts tone center tracking on binary pitch class profiles and second-order pitch class transition tables [Browne, 1981], employing various models of memory and a modular architectures of artificial neural nets (ART-2, SOM, and feed-forward network; Haykin [1999]). Izmirli and Bilgen [1996] use the constant Q transformation (Brown [1991], cf. *Section 2.1*) for tone center analysis in combination with a refined frequency estimation, observing phase changes [Brown and Puckette, 1993]. Context is integrated adaptively based on chord changes. But in their model, by canceling out harmonics of a detected fundamental, fundamentals of other tones are possibly canceled out also. Sapp [2001] suggests multi-time scale visualizations of tone center analysis for MIDI. Temperley [2001] uses Bayesian inference for automatic tone center analysis of MIDI. Pardo and Birmingham [2002] delineate a har-

[10] The experimental setup is adapted from Leman [1994].

[11] The material has been presented also in Purwins et al. [2000b] and Purwins et al. [2001a].

monic chord-by-chord analysis of MIDI. Sheh and Ellis [2003] use hidden Markov models based on previous chord segmentation. In *Section 3.9* an outlook of applications of CQ-profiles in the domain of tone-center tracking is presented [Purwins et al., 2002], including new features, an appropriate dissimilarity measure, various visualization methods, and multi time scale analysis.

Infinity – Construction of Acoustical "Paradoxes" Is it possible to perceive pitch as (monotonically) increasing, while returning the physical parameter to its starting point? With respect to other acoustical parameter domains the question would be: Is it possible to infinitely accelerate the beat of a piece? How about loudness? Can we modulate to remote keys keeping one direction on the circle of fifths and nonetheless reach the initial key? We will present examples in the domain of pitch and tempo and similar phenomena for loudness and harmonic modulation. The idea of these "paradoxes" is the following. The everyday experience of this perception obeys a linear order (greater/smaller relation), not a circular one. That implies, if we proceed in one perceptual quality in one direction (higher, faster, louder, farther away from tonal center), it appears "paradoxical" to us, if we reach the starting point again at some point. *Sections 3.1 and 3.2* contain results published in Normann et al. [2001a,b] and Purwins et al. [2005]. From the sawtooth culmination pattern we derive the pitch and tempo "paradoxes". In an experiment, we demonstrate that the phenomenon of circular pitch perception also holds for sequences of a much larger class of harmonic complex tones [Normann, 2000; Normann et al., 2001b]. Calling a phenomenon an illusion or a paradox usually means that it is not consistent with particular common assumptions. In the case of circular pitch perception such a common implicit assumption is the equalization of psychoacoustic and basic physical entities. We can dissolve the "paradox" of circular pitch by discarding this inadequate premise. Then circular culmination, e.g. in pitch, does not appear as a paradox anymore. Due to just noticeable differences, perception quantizes analog signals. Thereby small differences turn into equalities, eventually producing artifacts. At some point, the accumulation of small differences exceeds the perceptual threshold, whereas their quantized percepts add up to zero. In addition, in the pitch "paradox" we can see the gestalt principle of proximity at work, searching for the shortest path on the pitch class circle. As a mathematization of these principles we will introduce an equation, the spectrum of pitch differences, that extends the virtual pitch algorithm [Terhardt, 1998] to pitch intervals. Compared to artificial sounding Shepard tones, the spectra of the investigated harmonic complex tones are in general much closer to the stationary part of sounds of acoustical instruments. This explains the potential of employing circular pitch perception in pieces of music for acoustical instruments (e.g. by Berg, Ligeti, Stahnke, Reutter, cf. *Section 3.2.6*), even though their spectra do not merely consist of octave components.

Interdisciplinary Perspectives on Music

We will now discuss our approach with respect to various musicological traditions, mathematics, brain science, machine learning, psychoacoustics, and music psychology.

Surface Level or Underlying Deep Structure? There are two main views on music: first the description and analysis of musical works on the *surface level* and second the investigation of the musical mind, the *deep structure* that comprises the cognitive mechanisms at work in the creation and perception of music. According to the latter concept, a piece of music is considered merely an epiphenomenon of the deep structure. Looking at the deep structure in music means searching for the universal basis of musical intelligence, beyond a particular musical style. Let us investigate how far various theories proceed in this direction.

Traditional music theory [Rameau, 1722; Riemann, 1877; Schenker, 1935] tries to find universal laws beyond the surface level. But when applied to various styles of music the limitations of these theories become apparent. Each theoretic approach is more or less adapted to the musical style it describes best. Generalization to music not belonging to that particular style creates a lot of theoretical problems. E.g. Salzer [1962]'s extension of Schenkerian theory to twelve-tone music is not as smooth as, e.g., Schenker's own analysis of Beethoven's Piano Sonatas, since Schenkerian theory implies tonality. In some respects, functional harmony [Riemann, 1877] is inconsistent. It cannot adequately cope with music that is based on patterns like stereotype fifth fall sequences. It assigns a harmonic function to each chord, rather than treating the sequence as an idiomatic mechanism. The assignment of chords to functions is ambiguous, since e.g. the mediant can have tonic or dominant function. (cf. p. 52) Functional harmony reaches its limits in extended tonality (Richard Wagner's Tristan and later). However, traditional music theory is useful as a reference, since for complex phenomena in music often no other approach, e.g. neuroscience or music psychology yield results that could serve as a reference.

Yet other approaches that do not generally lead to the deep structure are the scrutinizing of the historical genesis of concepts and the etymologic derivation of the terminology, demanding authority. However, for showing convergence of distinct levels of explanation, we follow this understanding in the exposure of the historical development of the circle of fifths and the TOMIR in *Section 1.2*, although the exploration of a musical notion by etymologic derivation, common in traditional humanities, is more misleading than clarifying if the actual usage and meaning of a word deviates too much from its historical roots. Although descriptive on the surface level, another direction of research gives insight into the deep structure.

Ethnomusicology provides a therapy for Eurocentric attempts towards universalism by delivering counter-examples contradicting music theories that are mainly developed from the historically biased perspective of classical Western European composition styles especially from 17th to 19th century. On the other hand, there is such a diversity of music cultures on the world that for an arbitrary surface level hypothe-

sis on music almost always two music examples can be found, one to support and another one to contradict the hypothesis.

Other music theories consider *acoustics* to be a stable ground to build upon. The overtone series is thoroughly studied by Helmholtz [1863], who relates it to consonance and dissonance. [12] Upon that Oettingen [1866] builds his major/minor duality theory. But Dahlhaus [1967] points out that this theory cannot consistently build on the symmetry of major (major third - minor third) and minor (minor third - major third).

As a consequence to these problems, Lévy [2004] critically discusses and questions science methods in musicology and the possibility of universal laws in general. But thereby he questions his own scientific effort. However, some theorists abandon the search for deep structure and universal laws. Instead they restrict themselves to the mere description of a style in a particular musical epoch, composer, or score [de la Motte, 1980; Schneider, 1987].

How can mathematics help to investigate the deep structure of music?

Sister Sciences Mathematics and Music The earliest surviving fragments of original writings of a philosopher later assigned to the Pythagorean school are by Philolaus (470 – 385 B. C.). He emphasizes the eminent epistemological role of the number:

> καὶ πάντα γα μὰν τὰ γιγνωσκόμενα ἀριθμὸν ἔχοντι. [13]

As an example, he gives the intervals of the so-called Pythagorean diatonic scale, defined by the intervals 8:9, 8:9, 8:9, 243:256, 8:9, 8:9, and 243:256. Plato adopts this scale in the construction of the world soul (Timaeus: 36a-b, cited in Huffman [2005]). In his famous image of the cave, Plato divides between the world of thought and the world of sense-perception (Republic: 514a). Giving the world of thought higher priority, "Plato is calling for the study of numbers in itself" (Huffman [2005] *Section* 2.2). Plato (Republic: 531c) criticizes Archytas for not ascending to generalized problems [Huffman, 2005]. But Archytas uses numbers just to describe the musical scales in use in his days. He gives "an account of individual things in the phenomenal world in terms of numbers" (Huffman [2005], 3.1):

> περί τε δὴ τᾶς τῶν ἄστρων ταχυτᾶτος καὶ ἐπιτολᾶν
> καὶ δυσίων παρέδωκαν ἁμῖν σαφῆ διάγνωσιν καὶ περὶ γαμετρίας
> καὶ ἀριθμῶν καὶ σφαιρικᾶς καί οὐχ ἥκιστα περι μωσικᾶς.
> ταῦτα γὰρ τὰ μαθήματα δοκοῦντι ἡμεν ἀδελφεά. [14]

[12] Due to the constrained length of this thesis, we cannot address this popular topic here. It is frequently and thoroughly discussed elsewhere (cf. Palisca and Moore [2004]).

[13] Fragment 4 in Diels [1903], p. 408: "And indeed all things that can be known have number."

[14] Fragment 1 in Diels [1903], p. 432: "They [the mathematicians] reported to us a clear insight into the speed of celestial bodies, their rise and set and about geometry, numbers, and the (astronomic) study of spheres, and at any rate into music, because these sciences seem to be sisters."

For Archytas, music and numbers are united by the superior role of mathematics. "In reviewing the accounts of music that have characterized musical and intellectual history, it is clear that the Pythagoreans are reborn from age to age." [Epperson, 2003] Twenty centuries later, the "harmony of the spheres" reappears in Kepler's Harmonice Munde (1619). For Leibniz [1734] "musica est exercitium arithmeticae occultum nescientis se numerare animi". [15] The cosmological concept is still preserved, e.g. in the New Age work Nada Brahma: The World is Sound [Behrendt, 1983] and Karlheinz Stockhausen: "I know ... that I have come from Sirius, myself." [Felder, 1977] Maintaining close ties between mathematics and music, *mathematical music theory* [Mazzola, 1990, 2002] [16] reformulates concepts in musicology in more precise mathematical language. Thereby the vagueness of natural language is removed, thereby providing the basis for scientific work in musicology. Distinct music theory schools can be related to each other and possibly unified or extended to new theories. With this formalism at hand, (digitized) scores can be precisely and objectively investigated. This is one step towards operationalization[17] called for by Laske [1987] and Seifert [1993]. So far, mathematical music theory has delivered some strong results in music theory, e.g. for counterpoint (Mazzola [1990], p. 241-260). However, complying with Plato, it focuses on beautiful mathematics beyond the music more than its manifestation in actual pieces of music. Aspects of the related phenomenon are considered even less, such as the sensitive appearance of music, that is the acoustical reality of the sound signal, its auditory perception, and its particular sociocultural context. As for any applied mathematics, the relevance of mathematical music theory for music understanding depends on two critical transitions: the transfer of a musical event into a mathematical formalism and – vice versa – the retranslation of the mathematical result into musical meaning. When translating music into mathematics sometimes inadequately complex formalism is used, not complying with one of the basic principles of mathematics, the precedence of simplicity. [18] E.g. algebraic geometry [Mazzola, 1990] is used to explain the significance of the diminished seventh chord in the Hammerklavier-Sonate (cf. p. 68), pseudo-Riemannian geometry to discuss harmony and the problem of enharmonic spelling [Noll, 2002], and group theory to derive the twelve pitch classes (cf. p. 55). From music theory, Mazzola ascends further to a meta theory, he calls "encyclospace", mathematical music theory itself appearing as a special case thereof. With respect to the translation from music into mathematics, for chord classification by means of affine-linear mappings the psychoacoustic relevance of the theory is not clear (cf. p. 53). As a consequence, for subsequent mathematical conclusions the retranslation into music is not clear either. Motivic similarity is formulated mathematically in

[15] Music is the hidden mathematical endeavor of a soul unconscious it is calculating.

[16] Independently, Lewin [1987] and his school develop a similar approach.

[17] Operationalism is "the attempt to define all scientific concepts in terms of specifically described operations of measurement and observation." [Britannica, 2004]

[18] Ockham: "Pluralitas non est ponenda sine necessitate; 'Plurality should not be posited without necessity. The principle gives precedence to simplicity; of two competing theories, the simplest explanation of an entity is to be preferred."' [Britannica, 2003b]

terms of subsets of tones [Buteau, 1997] or the counterpoint group [Nestke, 2004]. But it is equivocal whether their definitions of melodic similarity are consistent with psychological findings. [Muellensiefen and Frieler, 2004] As a consequence, the musical relevance of the mathematical conclusions is equivocal as well. The translation of the mathematical conclusions back into general musical terms in Buteau [1997] and Nestke [2004], if meaningful at all, is not given yet. In the analysis of modulations, a test-bed has been developed [Noll and Brand, 2004], but, so far, no work has been devoted to numerically handle the problem of harmonic path finding. In metric analysis (cf. p. 170, Fleischer [2002]), mathematics is retranslated to Brahms' music in score format. However, it could be further extended to music in which the distinct simultaneous meters are handled in a more elaborate way, such as the music of the Aka pygmies, classical music of North India, and Nancarrow. Additional future work could comprise the translation into music in audio format. In general, mathematical music theory settles in a terrain that is difficult of access to music theorists not trained as mathematicians. Therefore for them it is very hard to adequately evaluate mathematical music theory. A pitfall – committed also by much writing in traditional musicology – is the relating of theories to special cases, singular hand-picked musical examples, instead of a representative sample with well-defined error estimates.

In our research we follow Archytas rather than Plato. In contrast to mathematical music theory, we employ only very simple mathematical description, such as pitch class profiles, which we derive from empirical data, various recordings in audio data format. We use models in auditory neuroscience, cognition and learning, to the extent they are available. For analysis we employ statistics and machine learning with quantifiable significance. In addition, we refer to psychological experiments. However, in some instances we have to refer to traditional music theory (cf. *Section 3.9.2*), especially for complex music phenomena, since for the latter, music psychology often does not provide enough results yet.

The Brain Comprehended? Cognitive science is an "interdisciplinary study that attempts to explain the cognitive processes of humans and some higher animals in terms of the manipulation of symbols using computational rules" [Britannica, 2004]. "*Cognitive musicology* is that branch of cognitive science which investigates musical thinking." [19] Lerdahl and Jackendoff [1983] [20] emphasize the criterion of operationalism in cognitive musicology. This implies stating hypotheses, subject to experimental validation, based on statistical significance, rather than on few particularly chosen examples. However, early accounts on cognitive musicology [Laske, 1987; Seifert, 1993] point out the necessity of operationalism rather than actually develop operational theories. "Apart from the cultural background and the individual experience of the beholder", Leman [1995] points out the role of the percept and its context. This is in contrast to considering music as a construction of building

[19] Boldface by the author, Lischka [1987], p. 191, cited in Seifert [1993], p. 39.
[20] p. 332, cited in Seifert [1993], p. 39.

blocks in the world of thought, e.g. notes, chords, and bars in traditional Western European music notation. Instead of dividing between the world of thought and the world of sense perception, cognitive musicology assumes a "correspondence between mental representations of perceived phenomena and brain states or cortical representations." [Leman and Carreras, 1997] Seifert [1993] (p. 39) thinks that in cognitive musicology psychology, auditory neuroscience, and artificial neural networks "converge" and lead to the crucial representations and processes at work in music cognition. Auditory images and schemata (Leman [1995], cf. *Section 1.1.7*) are considered more appropriate building blocks than notes, chords, etc. . An *auditory image*

> is assumed to contain a state or snapshot of the neural activity in a brain region during a short time interval. It is modeled as an ordered array of numbers (a vector), where each number represents the probability of neural firing of a group of neurons during a short time interval. (Leman [1995], p. 38)

But to what extent can neural correlates of mental processes be empirically identified? Auditory processing in the auditory periphery including the hair cell is well known and successfully applied in music compression, sound analysis, and music cognition modeling [Leman, 1995; Rosenthal and Okuno, 1998]. Hair cells are thoroughly investigated. [Meddis, 1988] Therefore the auditory image of the hair cell is biologically plausible. However, the knowledge about music processing in the auditory pathway subsequent to the hair cell, especially in the cortex, is sparse at this stage of research (Leman [1995], p. 172). All subsequent auditory images used in Leman [1995]'s model, including correlogram, tone context image, and semantic image, lack biological evidence at the current stage of auditory neuroscience. The correlogram is the autocorrelation vector (cf. Section 2.2.2, Equation 2.21) calculated from the hair cell output. But biological evidence for a process like autocorrelation is doubted (cf. *Section 2.2* and citation p. 83). In consequence, the biological plausibility of the correlogram (Leman [1995], p. 56) is not clear. A semantic image assumes characteristic groups of neurons that respond most strongly to particular keys. (Leman [1995], p. 18-19) However this image is not supported by neurobiological evidence either. [21]

Euphemistically called "invasive" animal experiments are the principle source of knowledge about neural correlates of hearing. But can they be employed for the investigation of music cognition as well? The relevance of animal experiments for music cognition research is limited by the extent, up to which we think of an animal having musical abilities as well. That leads to ethical problems. For the investigation of certain questions, species of a high degree of organization and close kinship to humans are required. But according to a generally accepted compromise between

[21] Cf. the discussion of Janata et al. [2002] on p. 168 of this thesis.

anthropocentrism[22] and biocentrism[23], those species require relatively high protection. For animal experiments in an ethical evaluation often the solidarity principle is cited in wide interpretation as a counterbalance to the value of animal life. The solidarity principle comprises the obligation to help others that suffer from illness or diseases, e.g. by conducting medical research involving animal experiments in order to develop effective treatment. We would like to assume that the main motivation in research is curiosity, and the solidarity principle mostly is just used as a vehicle in the train of arguments. However, in music it cannot be easily employed convincingly.

How far can brain imaging in humans help the understanding of music cognition? Sometimes diagnosis of epilepsy in humans indicates the need of electrode implantation in the patient for localizing the center of the epileptic seizure. Sometimes this gives the opportunity of undertaking delimited invasive music cognition experiments. We can imagine the brain as a clew of approximately 10^{11} neurons (Kandel et al. [1991], p. 18), each often having a complex topology and synaptically connecting to up to as much as 150 000 other neurons. [24] Hence, in general it is impossible to localize electric current flow in individual neurons by measuring electricity in electrodes positioned around the skull, like in the electroencephalogram (EEG). Other non-invasive methods, such as functional magnetic resonance imaging (fMRI) or positron emission tomography (PET), have better spacial but worse temporal resolution. Different musical activities and the processing of different musical styles can be vaguely localized in the cortex [Petsche, 1994; Marin and Perry, 1999]. The fMRI study of Janata et al. [2002] gives some evidence to the localization of tone center tracking activity in the brain (cf. p. 168).

Machine "learning" *Artificial neural networks* are a formalization of neural activity based on a computational model called the connectionist neuron. From the current viewpoint of research, the connectionist neuron presents a simplified, not entirely accurate view on the neuron, since it excludes certain ways of information coding like time and inter spike interval coding (cf. *Section 2.2.1*). In the mid 1980s, artificial neural networks demand great attention, being presented as an adaptive, flexible "learning" method opposed to rule based artificial intelligence, immoderately overwhelmingly suggesting a quick solution to hard problems like automatic translation before a disappointing exposure revealed a déjà vu of similar optimism praising artificial intelligence some twenty-five years before. Furthermore, in artificial neural networks the employment of a terminology borrowed from brain science supports

[22] "The defining feature of anthropocentrism is that it considers the moral obligations humans have ... to each other and, less crucially, to future generations of humans rather than from any obligation to other living things or to the environment as a whole." [Elliott, 2003]

[23] "In contrast to anthropocentrism, biocentrism claims that nature has an intrinsic moral worth that does not depend on its usefulness to human beings, and it is this intrinsic worth that gives rise directly to obligations to the environment. Humans are therefore morally bound to protect the environment, as well as individual creatures and species, for their own sake." [Elliott, 2003]

[24] Purkinje cell in the cerebellum (Kandel et al. [1991], p. 22).

this hype, suggesting that such an algorithm, a black box, would autonomously "learn" complex behavior without explicit provision of domain specific knowledge, just like humans do. In practice, domain specific knowledge is often required but hidden in preprocessing and laborious tuning of various parameters. Nonetheless, artificial neural networks are often even outperformed by classical methods based on heuristics. But this is not surprising, since a closer look reveals that many of the seductive terms denote rediscoveries of methods known for a long time in numerical mathematics and statistics (cf. Ripley [1996]). The first popular "neural net", the "multi layer perceptron" is nothing but a concatenation of linear and sigmoid functions. The term "learning", in effect, means "finding zeros of the first derivative of such a function". The methods are now rather labeled *machine learning* and treated as special cases in mathematics and statistics (e.g. Vapnik [1998]).

The self-organizing feature map (SOM,Kohonen [1982]) is frequently presented as a model of cortical organization [Obermayer et al., 1990]. Its backbone is the winner-takes-all step (Equation 2.30), requiring a global comparison between all neurons. Due to the high number of neurons in the cortex, such a step is biologically implausible. Although we know little about the cortex, the SOM is used as a rough, yet inaccurate implementation of instances of the schema concept [Leman, 1995]. Leman describes how musical meaning of pitch, e.g. its functional role as a leading tone, tonic etc., is partly defined by its context, e.g. the key, whereas vice versa this context, the key, is defined by the pitches. We will see that the employment of correspondence analysis (*Sections 2.3.2, 2.3.4, and 3.8.4*) as a schema is especially suited to model this duality.

Due to the incompleteness of knowledge in auditory neuroscience, the potential of insights from experiments with neuromimetic models is limited. In general, psychoacoustics and music psychology experiments give a far more precise and reliable reference. Machine learning methods should be thought of as heuristics with occasional allusions to neurobiological phenomena. Therefore, usually a compromise has to be made between a neuromimetic model of high explanatory value and an efficient practical application that is rather based on heuristics. Even though the bottom-up approach from auditory neuroscience does not lead so far, by machine learning we can tentatively fill the gap up to the top down line of attack reached by psychoacoustics and music psychology.

Ground Truth through Inquiry of the Listener In general it can be observed that musical qualities are *psychoacoustic* qualities more than physical ones (*Section 1.1*). Referring to Chomsky [1988], the deep structure of music cannot be a formalized rule system extracted solely from a corpus of music. The major reference is the musician or naive music listener, who can be accessed in a psychological experiment, such as the one on circular perception of pitch in *Section 3.2.2* and the probe tone experiment in *Section 1.2.9*. In addition, various dissimilarities, as an implementation of the gestalt principle of proximity, provide theories and models of circular perception of pitch (cf. *Section 3.2.3*) and of inter-key relations (cf. *Sections 1.2.5 and*

1.2.7). But also in music psychology the surface level could be falsely taken for the deep structure: 1) The "role that anthropological field-workers [musicologists in our context] play in shaping, rather than merely describing, the situations they report on." [Britannica, 2003a] E.g. music students are coined by musicologist's concepts that dominate the curriculum. In some cases, such as the music psychology experiments with probe tones, *Section 1.2.9*, or circular pitch, *Section 3.2.2*, the judgments of the musically educated subjects reflect identified explicit music theory concepts rather than some deep structure. Likewise, awareness of a regularity or a pattern in music provokes composers to break it, in order to eschew a cliché. 2) In our global musical monoculture, it is difficult to find unbiased naive subjects that have never heard main stream music before. However, so far music psychology has predominantly studied simple musical attributes such as pitch, tempo, or rhythm. Lacking music psychology data, for more complex musical attributes we need to refer to traditional music theory.

We do not claim that the methods we apply comprehensively explain music. On the contrary, by sorting out the minority of explicable aspects, we wish to focus the attention to the residuum, the (yet) inexplicable marrow of striking music experience.

Organization of Thesis *Chapter 1* covers the background. We view music from two contrary perspectives. We refer to the more Platonic perspective including mathematical music theory and geometrical models. The other, reductionist, viewpoint is the discussion of neural correlates of musical notions. *Section 1.1* prepares the ground for the definition of pitch as a psychoacoustic notion, introducing gestalt principles, especially for grouping sound features, virtual pitch, and the circular perception of relative pitch [Shepard, 1964]. In *Section 1.2*, we develop the TOMIR from geometrical concepts in music theory. Starting from the tonnetz [Euler, 1774], we show how equal temperament supports the identification of tones that are separated by commata [25]. Enharmonic identification (e.g. d♯ ∼ e♭) and abstraction from brightness leads to pitch classes. Chords, harmony, cadences, and modulations are introduced defining the sense of key. Key transposability is discussed as a matter of key character. Finally, geometrical models of inter-relations between tones, chords, and keys are presented.

Chapter 2 introduces methods relevant to this thesis. We explain the constant Q transform derived from the fast Fourier transform [Gauss, 1866] [26]. In *Section 2.2*, we see to what extent we can use knowledge from auditory neuroscience to construct a functional neuromimetic model. As an example, we discuss autocorrelation as an implementation of virtual fundamental detection in *Section 2.2.2*. We demonstrate

[25] To give an example, the Pythagorean comma is the difference between seven just octaves and twelve just fifths.

[26] Gauss used the algorithm 1805 to interpolate the trajectories of the asteroids Pallas and Juno, published posthumously. [Heideman et al., 1984]

that for the purpose of pitch detection in our context, harmonic complex tones and pitch class intensities accumulated across time, the impact of most of the features of the neuromimetic model are irrelevant. Finally, machine learning (supervised and unsupervised learning, clustering, and embedding/visualization algorithms) provides an implementation of schemata with loose neurophysiological allusions. Correspondence analysis especially is suited to capture inter-relations between pairs of musical features that mutually constitute each other's meaning by defining the context. In particular this is true for key and pitch. A key is coined by the frequency of occurrence of pitch classes. Vice versa, pitch class can be defined by its prominence in various keys.

Chapter 3 presents the results on circular pitch, pitch class profiles from audio, the circle of fifths, the toroidal models of inter-key relations, and tone center tracking. A comprehensive summary follows. In the end, appendices include a discussion of the binding problem, auxiliary pitch class profile data, and similarities as the basis for classification and visualization.

1 Background – From Pitch Criticism to Geometry of Tone Center Inter-Relations

In this chapter, we challenge the linear notion of pitch through a psychoacoustic investigation of its (traditional) definitions. Furthermore, at the level of geometric considerations in music theory, various visualizations of tone, pitch class, triad, key, and their inter-relations are described.

Parts of this chapter are adopted from other publications: Sections 1.1.1, 1.1.2, 1.2.3 [Purwins et al., 2005], Section 1.1.6 [Normann, 2000], Section 1.1.7 [Purwins et al., 2000a], and Section 1.2.9 [Purwins et al., 2006].

1.1 Fuzzy Foundations of Pitch

Pitch is an attribute which is distinct from, and more general than physical frequency. Two sinusoids can evoke the perception of several pitches, combination tones, and virtual pitch. From another perspective, due to grouping effects, several sinusoids can fuse to a tone of a single pitch.

1.1.1 The *Psycho*acoustic Nature of Pitch

Even though for an acoustical signal, frequency and pitch are usually closely related, it is not generally possible to reduce pitch perception to the physically measurable attribute of frequency. Neglecting transient and noise components, if they exist, we can treat natural tones as approximately periodic signals. These tones are not exactly periodic in the mathematical sense of the definition, since, e.g. for vibrato, the cyclic duration varies and the wave form slightly changes. This evokes questions addressed to the psychology of perception. What is the minimal duration of an ideal acoustical signal, so that a pitch still can be assigned to it? To what extent may a signal deviate from the exact repetition of its wave form? (Terhardt [1998], p. 245) Already at this stage, it is apparent that the mere physical analysis of the signal does not suffice. Furthermore, we rely on psychoacoustics to adequately describe pitch. But even with a roughly periodic stimulus, a simple example of generally aperiodic signals, at hand we will have to employ the psychology of perception to seek an explanation for pitch perception. American National Standards Institute (ANSI,1973) defines pitch as

> ... that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from high to low.

According to this definition, pitch is not clearly distinguished from loudness. Furthermore, we will see in the sequel, e.g. in Section 1.2.2, that the assumption made in this definition, pitches can be ordered, is not generally valid.

1.1.2 Sums of Sine Tones – Natural Sounds

The sine tone plays a special role in the investigation of pitch perception:

> On it we can base a reasonable operational definition of pitch. We can assume a specific pitch existing in an arbitrary sound signal, if it is possible to adjust the frequency of a sine tone with sufficient reproducibility in a listening test so that both pitches coincide. (translated from Terhardt [1998])

Exploring pitch cannot be restricted to a single sine tone or to narrow-band noise. Leaving aside any discussion of transients and noise and focusing on the sustain phase of an acoustic signal, in psychoacoustic investigations we often consider periodic signals, as a rough approximation of natural sounds like those of musical instruments. *Complex tones* are finite sums $s(t) = \sum_{k=1}^{N} s_k(t)$ of sine tones $s_k(t) = a_k \sin(2\pi f_k t + \theta_k)$ of different amplitudes a_k. If for a frequency $f_0 : f_k = k * f_0$ for $1 \leq k \leq N$, $s(t)$ is called *harmonic complex tone*[1]. f_0 is its *fundamental*, f_k is the k-th *partial*[2]. The preeminent role of harmonic complex tones is evidenced by several arguments that we can only briefly summarize here:

1. Internal physical model of vocalized speech signals: These signals are harmonic. An internal model of these sounds greatly helps human communication.

2. The activation pattern on the basilar membrane evoked by a sine tone displays a spot with maximal excitation and, in addition, a couple of narrower local maxima, aural harmonics, at membrane locations relating to integer multiples of the sine tone frequency [Stevens and Newman, 1936].

3. Coincidence neurons: For lower frequencies the hair cells synchronize with the positive phase of the bandpass filtered sound signal (phase locking). Special coincidence neurons detect synchronous maxima of spike frequency, referring to signals with cycle durations expressible by integer fractions [Langner, 1997].

[1] Also called "harmonic overtone series".

[2] Also called "k-th harmonic" or "$k-1$st overtone". Please note that in our terminology the fundamental f_0 equals the first partial f_1.

It is not astonishing, therefore, that many music theorists use the overtone series as a reference to ground their theories. A few examples are Euler [1739]'s *gradus suavitatis*, Helmholtz [1863]'s dissonance model based on beating, and Barlow [1995]'s harmonicity function. The emphasis on overtone series, on the other hand, propels the composer's interest in *in*harmonic complex tones (cf. Section 3.2.1).

$f'_k = f_0/k$ is named the k-th *subharmonic* of f_0. A *residual series* is a harmonic complex tone so that the first n partials vanish ($\alpha_k = 0$ for $1 \leq k \leq n$). The *missing fundamental* is then called the *residual*. Sometimes we may wish to *shift* the overtone series by a frequency difference Δ: $f_k \longrightarrow f_k + \Delta$. We can also *stretch* the overtone series by a factor s: $f_k \longrightarrow s f_k$.

1.1.3 Frequency Scaling

The *Weber-Fechner law* quantifies a perceptual entity B, expressed by a physical entity A, relative to a reference value A_0, with scaling factor s:

$$B = s \log(\frac{A}{A_0}). \tag{1.1}$$

In the inner ear, incoming sound waves cause a traveling wave on the *basilar membrane*, to which are attached hair cells. The latter transduce mechanical vibrations into electrical activity. (Cf. Figure 2.1 and p. 72 for more details.) Due to the varying stiffness of the basilar membrane and the adjusted stiffness, size, and electrical resonance of the hair cells, spots along the basilar membrane are tuned to specific frequencies (Kandel et al. [1991], p. 487 f.). Frequencies below 500 Hz are mapped approximately linearly. In the range of 500 Hz to 8 kHz the mapping is approximately logarithmic (cf. Figure 3.49). This is in agreement with the Weber-Fechner perceptual law, Equation 1.1.

This particular mapping has strong implications on pitch perception. Lower 500 Hz and higher than 8 kHz, relative pitch (measured in mel) deviates from strictly logarithmic scaling. The musical understanding of pitch class (cf. p. 48) as octave equivalence is based on uniform spacing of the corresponding resonance frequencies on the basilar membrane. The common time/log-frequency representation in stave notation can be linked to the logarithmic resonating range, since this range contains a significant portion of overtones of common musical instruments. Apart from correlograms (see Section 2.2.2), the time log-frequency representation is widely used to start with higher level analysis, e.g. such as receptive fields for trajectories in the time frequency domain [Todd, 1999; Weber, 2000].

A model of the basilar membrane may be implemented as a filter bank. The spacing and the shape of the filters have to be determined. A first approach is the discrete Fourier transform, which is very quick, but gives equal resolution in the linear (i.e. non-log) frequency domain. For analysis tasks that do not rely on the detection of very low frequencies the constant Q transform (Brown [1991]; Brown and Puckette [1992], cf. Section 2.1) is an appropriate method. Equal logarithmic resolution can be achieved also by the continuous wavelet transformation (CWT,

Strang and Nguyen [1997]). For frequency f (in kHz), a more exact modeling of the basilar membrane is supplied by a filter spacing according to the equivalent rectangular bandwidth (ERB, Moore and Glasberg [1983], p. 751)

$$\Delta_{\text{ERB}} = 6.23f^2 + 93.39f + 28.52 \tag{1.2}$$

or the critical bandwidth units (Bark, Zwicker and Fastl [1990], p. 147)

$$\Delta_{\text{Bark}} = 25 + 75(1 + 1.4f^2)^{0.69}. \tag{1.3}$$

A linear approximation for ERB is given in Glasberg and Moore [1990], p. 114. Especially in the range of center frequencies below 500 Hz, ERB is more suitable to describe the result of the notched noise experiment. (Moore and Glasberg [1983], cf. Figure 3.49 for a comparison of filter banks.) Depending on its center frequency, the *critical bandwidth* is associated with a qualitative change in perception, e.g. of accumulation of loudness, sensation of roughness, frequency masking, tone fusion, and just noticeable frequency difference. [3] The *critical band-rate* is a frequency scaling based on critical bandwidth units (Bark). Twenty-five Bark cover the audible frequency range.

Establishing the ground for calculating pitch in Equation 3.2, we need to investigate the loudness of frequency components and of complex tones.

1.1.4 Loudness Summation of Frequency Components

Loudness perception obeys the Weber-Fechner law. The *sound pressure level (SPL) L* is measured in dB:

$$L = 20 * \log_{10}(\frac{P_{\text{rms}}}{P_0}) \tag{1.4}$$

with P_{rms} being the root-mean-square pressure level (Unit: N/m^2) and $P_0 = 0.00002\ \text{N/m}^2$, e.g. $P_{\text{rms}} \cong \frac{1}{\sqrt{2}}$ for a sinusoid with amplitude 1.

The *hearing threshold* (in quiet) is the sound pressure level of a sine tone that can be just heard. The hearing threshold depends on the frequency (cf. Figure 1.1). We observe that it undergoes a qualitative change at approximately 500 Hz. Below this value in the log-frequency / dB domain the curve shows rather linear characteristic, whereas above it has a polynomial shape. From Terhardt [1998] (p. 243 Equation (9.15)) we cite an analytical expression, describing the hearing threshold L_T (in dB) for frequency f (in kH):

$$L_T = 3.64f^{-0.8} - 6.5e^{-0.6(f-3.3)^2} + 10^{-3}f^4. \tag{1.5}$$

Other mathematical formulations rather focus on numerical efficiency and treat the regions below and above 500 Hz separately.

[3] Zwicker and Fastl [1990], p. 134–141, Terhardt [1998], p. 261.

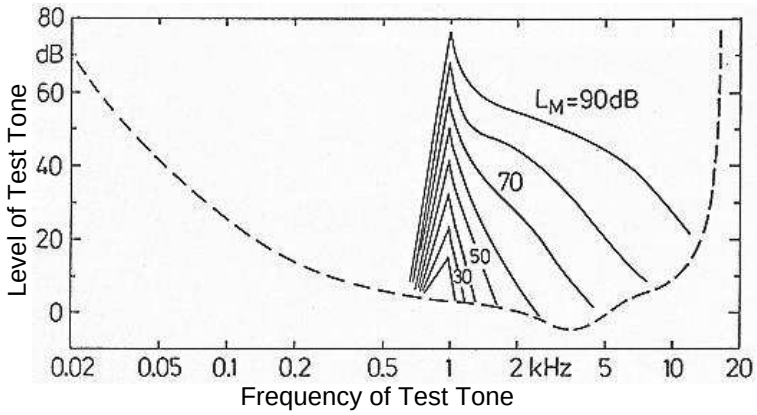


FIGURE 1.1: Hearing threshold in quiet (dashed line) depending on frequency. The ear is most sensitive in the frequency region between 2 kHz and 5 kHz. Masked thresholds (solid lines) induced by a masker, a 1 kHz sinusoid of different, indicated, sound pressure levels. (from Zwicker and Fastl [1990], p. 63, Figure 4.7.)

Fletcher and Munson [1933] derive *equal loudness curves*. For sine tones of different frequencies these curves give the sound pressure levels that are necessary to produce tones of equal perceived loudness. An approximation for these curves can be derived from the hearing threshold (Equation 1.5). Within the frequency range from 500 Hz to the upper hearing limit and the dB range from 0 dB to 100 dB the equal loudness curves can be roughly approximated by a log-linear shift of the polynomial describing the hearing threshold. In Section 3.2.4, we will linearly interpolate the *specific loudness* N_ν from the equal loudness curves. [4]

The presence of a tone (*masker*) may change the hearing threshold of another tone heard at the same time. "The *masked threshold* is the sound pressure level of a test sound ... necessary to be just audible in the presence of a masker." The masked threshold mostly lies above threshold in quiet. "If the masker is increased steadily, there is a continuous transition between an audible (unmasked) test tone and one that is totally masked. This means that besides total masking, partial masking also occurs. Partial masking reduces the loudness of a test tone but does not mask the test tone completely" (Zwicker and Fastl [1990], p. 56, cf. Figure 1.1).

The perceived loudness of superposed sine tones with intensities I_i and random phases crucially depends on their frequency difference. We distinguish the following cases:

1. If all frequencies lie within a critical bandwidth, from the respective intensities I_i, the perceived loudness L of the resulting complex tone yields (Roederer

[4] Note that we use a different approach than Zwicker and Fastl [1990], Section 8.7.1.

[1995], p. 92 Equation (3.18)):

$$L = c \sqrt[3]{\sum_i I_i}, \tag{1.6}$$

where c is a constant.

2. If the frequency spread of the components exceeds a critical bandwidth but is not too large the resulting subjective loudness tends towards the sum of the individual loudnesses L_i (Roederer [1995], p. 93 Equation (3.19)):

$$L = \sum_i L_i. \tag{1.7}$$

If the L_i differ considerably among each other masking occurs. The louder tone deforms the hearing threshold of a softer tone by lifting the threshold up in a sort of triangular shape centered at the frequency of the masker (cf. Figure 1.1 and Equations 1.8-1.12).

3. If the frequency difference of the signal tones is large the listeners "tend to focus on only one of the component tones (e.g. the loudest, of that of the highest pitch) and assign the sensation of total loudness to just that single component" (Roederer [1995], p. 93 Equation (3.20)).

1.1.5 Singling Out Frequency Components

The frequency differences Δf of adjacent k partials be greater than the critical bandwidth. N_ν be their specific loudnesses and z_ν their frequencies. Then for the μ-th partial *incisiveness* [5] $L_{X\mu}$ is defined as:

$$L_{X\mu} = N_\mu - 20 \cdot \log_{10}(A_\mu)\ \text{dB} \tag{1.8}$$

with

$$A_\mu = \sum_{\nu=1}^{\mu-1} 10^{\frac{L_{\nu r}}{20\ \text{dB}}} + \sum_{\nu=\mu+1}^{k} 10^{\frac{L_{\nu l}}{20\ \text{dB}}}. \tag{1.9}$$

Sinusoidal masking is displayed in the Bark-Phon-plane:

$$L_{\nu r} = N_\nu - R(z_\mu - z_\nu), \tag{1.10}$$

$$L_{\nu l} = N_\nu - 27(z_\nu - z_\mu)\frac{\text{dB}}{\text{Bark}}, \tag{1.11}$$

$$R = (24 + \frac{230}{f_\nu} - 0.2\frac{N_\nu}{\text{dB}})\frac{\text{dB}}{\text{Bark}}. \tag{1.12}$$

Equations 1.8-1.12 are adapted from (11.7)-(11.12) in Terhardt [1998] using specific loudness instead of sound pressure level.

[5] "Prägnanz" in German.

1.1.6 Few Frequencies – Several Tones

Beating, Roughness, and Combination Tones Depending on the frequency difference $\Delta f = f_2 - f_1$ of two sine tones, the phenomena of roughness, beating, difference, or combination tones are perceived (Figure 1.2). At high intensity levels and sufficiently large Δf, pitches can be heard with frequencies not present in the sound stimulus, the two sine tones in this example. *Combination tones* (Plomp [1965],

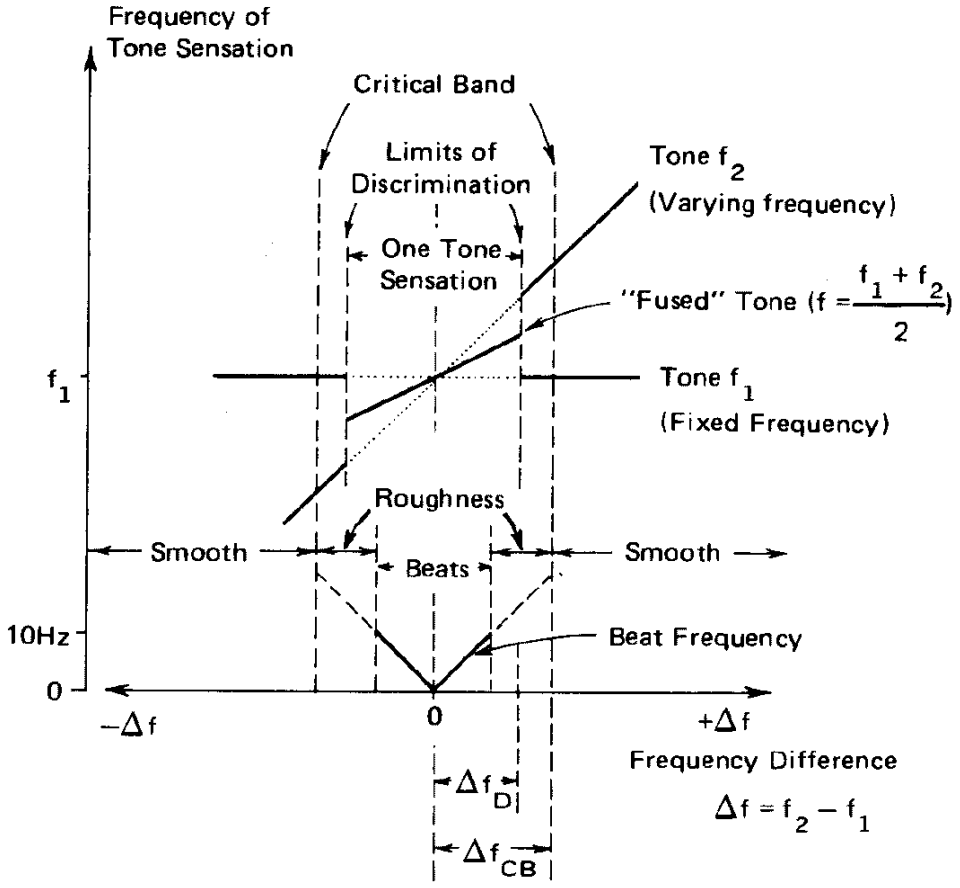


FIGURE 1.2: Schematic representation of the frequency (heavy lines) corresponding to the tone sensations evoked by the superposition of two sine tones of nearby frequencies f_1 and $f_2 = f_1 + \Delta f$. A beating with beat frequency Δf is heard for small Δf. For higher Δf still smaller than the critical bandwidth Δf_{CB} we hear roughness. The transition from beating to roughness to the absence of roughness is smooth [Terhardt, 1998; Zwicker and Fastl, 1990]. The just noticeable frequency difference Δf_D lies within the range of roughness. For $\Delta f < \Delta f_D$ both tones are perceived as one, having a pitch equal to the average frequency $\frac{f_1+f_2}{2}$ (from Roederer [1995], p. 33 Figure 2.12). For $\Delta f > \Delta f_{\mathrm{CB}}$, cf. Figure 1.3.

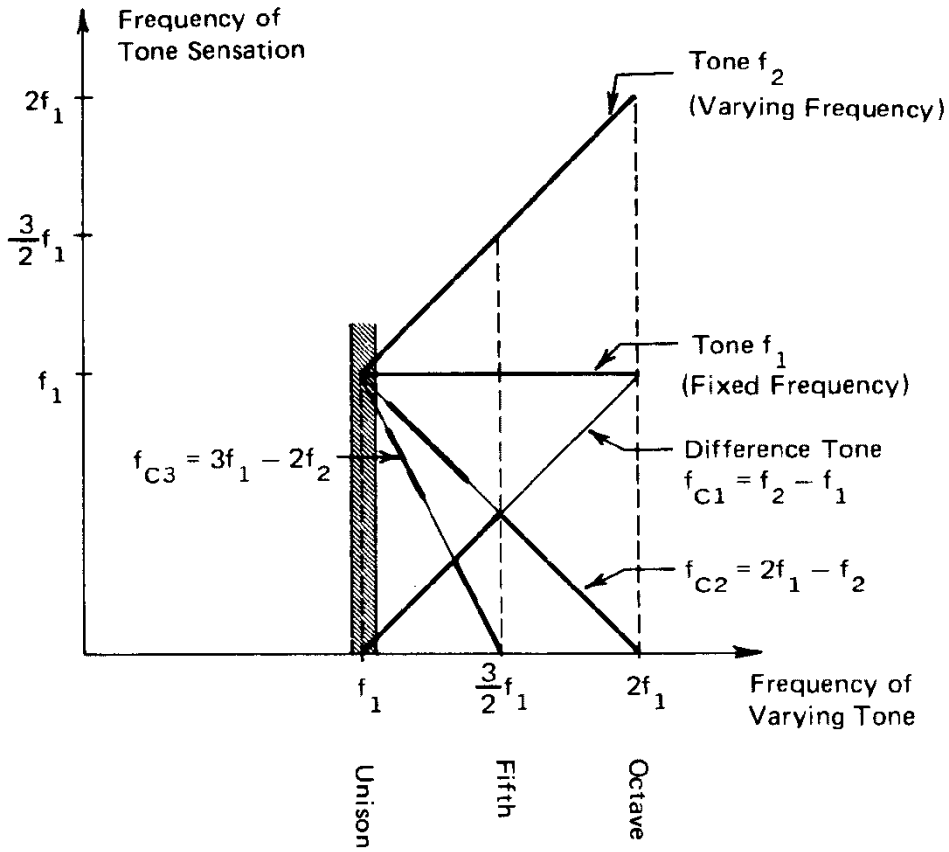


FIGURE 1.3: Combination tones evoked by two frequencies f_1 and f_2. At high intensity levels the combination tones f_{C1}, f_{C2}, f_{C3} can be heard out, especially at the ranges marked by heavy lines. For f_1 being very near to f_2, the hatched area is shown as a close-up in Figure 1.2. (from Roederer [1995], p. 38 Figure 2.14)

cf. Figure 1.3) can be detected at frequencies $f_C = mf_1 - nf_2$ with $m, n = 1, 2, \ldots$. The combination tone at frequency $f_{C1} = f_1 - f_2$ is easily detectable. Its strength is almost independent from f_2/f_1. Even at low intensity levels the combination tones corresponding to the frequencies $f_{C2} = 2f_1 - f_2$, the cubic difference tone, and $f_{C3} = 3f_1 - 2f_2$ can be perceived. Mathematically, difference frequencies result from non-linearities [Hartmann, 1997].

Virtual Pitch Under certain conditions (cf. Figure 1.4), residual series can evoke a perceived pitch (*virtual pitch*[6]) that is lower than the frequency of the lowest partial. Cf. Section 2.2.2, including Figure 2.3, for more detail. The phenomenon of virtual pitch is also used in the design of musical instruments:

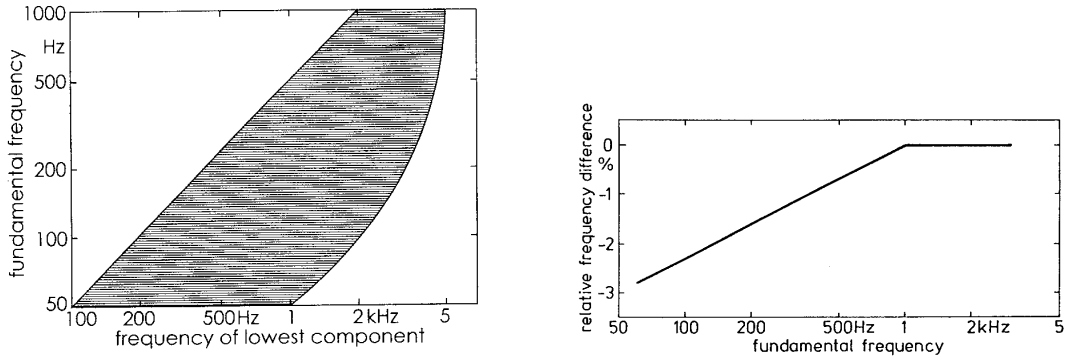
[6]Virtual pitch is also called "periodicity pitch", "subjective pitch", or "residue tone".

> Actually, periodicity pitch has been used for many centuries. For instance, since the end of the sixteenth century, many organs include a stop composed of pipes sounding a pitch higher than the pitch of the written note actually played. The purpose is to stimulate or reinforce the bass one octave below the pitch of the written note. E.g. for the *resultant bass* a pipe at an upper fifth is added to a pipe with a basic frequency. [Roederer, 1995]

This results in a perceived pitch one octave lower than the basic frequency.

Virtual pitch is different from the phenomenon of combination tones. Both effects can be observed in residual series, but the causes are different. Combination tones are produced by non-linearities in the inner ear. [7] Virtual pitch cannot be explained by non-linearities [Licklider, 1954]. An experiment [Houtsma and Goldstein, 1972] indicates that virtual pitch is a neural process: virtual pitch can be perceived, when neighboring partials are presented to separate ears.

[7] "Most scientists believe that the source of the cubic difference tone is in the inner ear; it probably results either from nonlinear motion of the basilar membrane or from some nonlinearities that exist when the hair cells stimulate the 8th nerve fibers" [Yost and Nielsen, 1989]. For a more thorough investigation of the physiological reasons of the non-linear processes cf. Dallos [1992].



(a) (b)

FIGURE 1.4: Existence region of virtual pitch: Virtual pitch comes into play if both fundamental frequency and the frequency of the lowest partial are in the existence region (shaded area on the *left*). Virtual pitch is not observed, if the frequency of the lowest partial is higher than 5 kHz or lower than 100 Hz. The fundamental frequency has to be in a range between 50 Hz and 1 kHz (*left*, Schouten et al. [1962]; Ritsma [1962]; Plomp [1967]; Moore [1973], from Zwicker and Fastl [1999], p. 121 Figure 5.10). For fundamental frequencies lower than 1 kHz, the virtual pitch of the harmonic residual series lies slightly below the fundamental. If the fundamental has a frequency higher than 1 kHz it equals the virtual pitch in frequency (*right*, from Zwicker and Fastl [1999], p. 119 Figure 5.8).

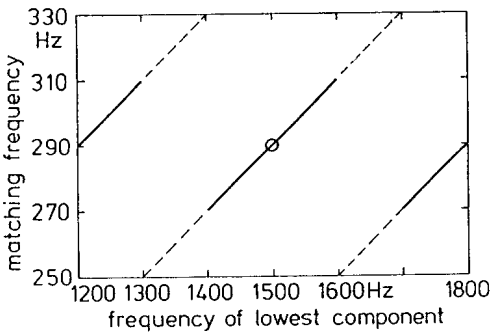


FIGURE 1.5: Virtual pitch of shifted residual series: harmonic residual series are shifted, while maintaining the frequency distances between partials. The matching virtual pitch increases slightly until the residual series is shifted by half of a partial distance. For an even higher shift the matching virtual pitch is smoothly transformed into a pitch slightly lower than the matching pitch of the original residual series. (from Zwicker and Fastl [1999], p. 122 Figure 5.11)

A requirement for the perception of virtual pitch is that both the fundamental and the lowest partial be in the *existence region* (cf. Figure 1.4). Not all partials are equally significant in their contribution to virtual pitch. *Dominance region* for virtual pitch is described by Ritsma [1970]: "...if pitch information is available along a large part of the basilar membrane the ear uses only the information from a narrow band. This band is positioned at 3-5 times the pitch value. Its precise position depends somewhat on the subject" (also cf. [Dai, 2000]). The virtual pitch of shifted residual series slightly increases until the shift amounts to half of the virtual pitch initially perceived (cf. Figure 1.5).

1.1.7 Several Frequencies – One Tone

Certain sensory features, such as detected partials, may be grouped, forming a *gestalt*, such as a tone, according to particular criteria. [8] The gestalt concept originates from Ehrenfels [1890] and Mach [1886]. They initially present musical examples. Subsequently visual perception is investigated. From the 1970s onward, computer-supported sound synthesis and analysis enforce the application of gestalt theory to auditory perception, exhaustively reviewed in Bregman [1990].

In the following, principles are introduced which aid binding in auditory perception (Figure 1.6, Bregman [1990]): The principle of *proximity* refers to distances between auditory features, e.g. onsets, pitch, and loudness. Features that are grouped together have a small distance between each other, and a long distance to elements

[8] For a treatment of more general theoretical and practical aspects of the closely related *binding problem* cf. Appendix A.

of another group. Temporal and pitch proximity are competitive criteria. E.g. the slow sequence of notes A-B-A-B... (Figure 1.6 A 1) which contains large pitch jumps, is perceived as one stream. The same sequence of notes played very fast (Figure 1.6 A 2) produces two perceptual streams, one stream consisting of the As and another one consisting of the Bs.

Similarity is very similar to proximity, but refers to properties of a sound, which cannot be easily identified along a single dimension (Bregman [1990], p. 198). For example, we speak of similar rather than of proximate timbres.

The principle of *good continuation* denotes smoothly varying frequency, loudness, or spectra with a changing sound source. Abrupt changes indicate the appearance of a new source. In Bregman and Dannenbring [1973] (Figure 1.6 B) high (H) and low (L) tones alternate. If the notes are connected by frequency glides (Figure 1.6 B 1) both tones are grouped to a single stream. If high and low notes remain unconnected (Figure 1.6 B 2) Hs and Ts each group to a separate stream. "Good continuation" is the continuous limit of "proximity".

The principle of *closure* completes fragmented features, which already have a

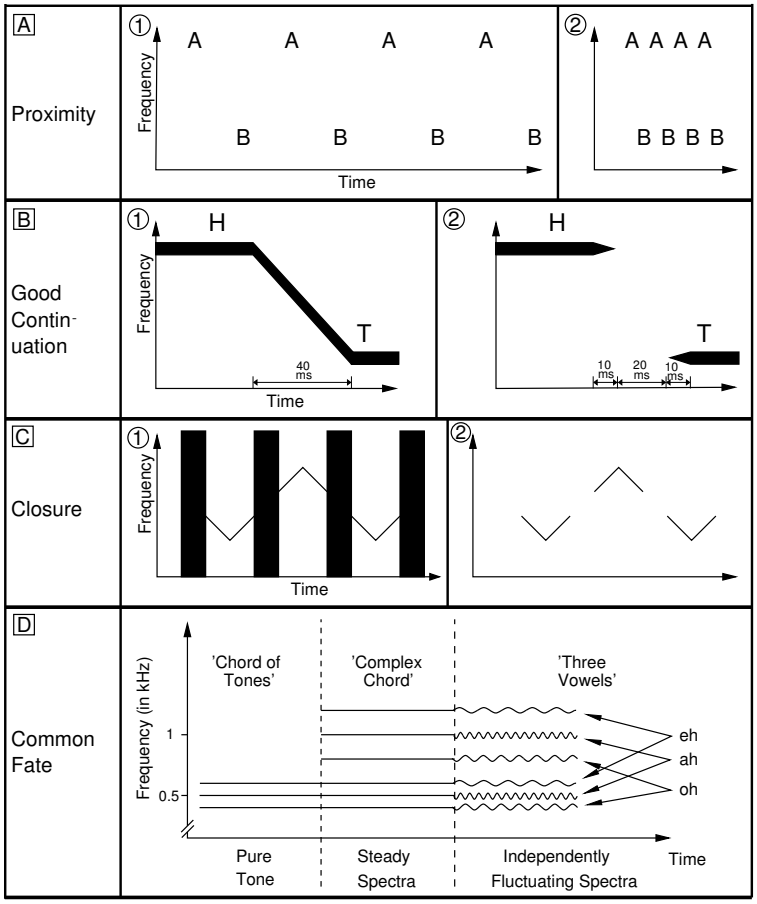


FIGURE 1.6: Psychoacoustic experiments demonstrating grouping principles (cf. Section 1.1.7, Figure according to Bregman [1990]).

"good gestalt", e.g. ascending and descending frequency glides that are interrupted by rests in a way depicted in Figure 1.6 C 2. Separated by rests, three frequency glides are heard one after the other. Then noise is added during the rests, as shown in Figure 1.6 C 1. This noise is so loud that it would mask the glide, even if it were to continue without interruption. Surprisingly, the interrupted glides are perceived as being continuous. They have "good gestalt": They are proximate in frequency and glide direction before and after the rests. So they can easily be completed by a perceived good continuation. This completion can be understood as an auditory compensation for masking (cf. p. 36).

The principle *common fate* groups frequency components together, when similar changes occur synchronously, e.g. onsets, glides, or vibrato. Chowning [1980] (Figure 1.6 D) made the following experiment: First three sine tones are played. A chord is heard, containing the three pitches. Then the full set of harmonics for three vowels ("oh", "ah", and "eh") is added, with the given frequencies as fundamental frequencies, but without frequency fluctuations. This is not heard as a mixture of voices but as a complex tone in which the three pitches are not clear. Finally, the three sets of harmonics are differentiated from one another by their patterns of fluctuation. We then hear three vocal sounds being sung at three different pitches.

Other important topics in auditory perception are *attention* and *learning*. In a cocktail party environment, we can focus on one speaker. Our attention selects this stream. Also, whenever some aspect of a sound changes, while the rest remains relatively steady, then that aspect is drawn to the listener's attention ("figure ground phenomenon"). Let us give an example for learning: The perceived illusory continuity (cf. Figure 1.6 C) of a tune through an interrupting noise is even stronger, when the tune is more familiar (Bregman [1990]: p. 401). Whereas auditory grouping principles perform a sort of low level processing, a *schema* is

> an organization (in a person's brain) of information pertaining to some regularity in his or her environment. Sometimes it is conceptualized as an active structure analogous to a computer program, sometimes as similar to a complex organization of "memory" records in a computer. In all cases it is abstract enough to be able to fit a range of environmental situations. Schemas are conceived of as being at different levels of generality. (Bregman [1990], p. 734)

Attention and learning favor either of two opposing auditory strategies. In *analytic listening* particular components, e.g. partials, are singled out. *Synthetic listening* gives rise to holistic perception, e.g. of a tone. In musical composition often both strategies are supported. E.g. in Webern's orchestration of the Ricercar in Bach's Das Musikalische Opfer a competition of gestalt principles takes place in the domains pitch and timbre. Focusing on pitch as a cue, the melody is heard. Focusing on timbre the main line decomposes into small phrases. Bregman [1990], p. 470 points out the visual analog in Giuseppe Arcimboldo's paintings (cf. Figure 1.7).

In *computational auditory scene analysis* gestalt principles are translated into numerical algorithms. Two questions are of particular importance: Which is the domain in

FIGURE 1.7: In Giuseppe Arcimboldo's L'AUTOMNE, an analytical perspective shows various fruits. By the help of gestalt principles, in the synthetic perspective the face becomes apparent.

which the principles are applied (e.g. spectral domain or contour space, cf. p. 103)? Provided an appropriate representation is given, how can we translate the principles into numerical equations? E.g. for the principle of proximity the question reduces to: What kind of dissimilarity or (more specific) metric shall we employ to measure proximity? (cf. Appendix C) In Sections 3.5 and 3.8 we implement a schema of interkey relations by visualization algorithms, such as the self-organizing feature map, correspondence analysis, and Isomap.

In the evolution of the auditory system, gestalt principles serve as a means of identification of sound sources. But in music there is not a one-to-one correspondence between tones and sound sources. A single instrument can play several tones simultaneously. On the other hand one tone can be produced by several instruments in chordal streams.

1.2 Musical Landscapes of Tones, Triads, and Keys

> Es sei demnach eine gewisse Klasse von einfachen Vorstellungen so beschaffen, dass, wenn viele derselben zugleich im Bewusstsein sind, alsdann aus ihrer Qualität bestimmte Abstufungen ihres Verschmelzens erfolgen müssen: so ordnen sich unfehlbar diese Vorstellungen dergestalt

> *neben* und *zwischen* einander, dass man sie nicht anders als auf räumliche Weise zusammenfassen, und sich darüber nicht anders als in solchen Worten ausdrücken kann, welche dem Schein nach vom Raume entlehnt, eigentlich aber eben so ursprünglich der Sache angemessen sind, als wenn man sie auf den Raum bezieht. (Herbart [1850], p. 298, $ 139)[9]

Space is a natural medium for imagining inter-relations between percepts. Similarity of percepts can be identified with spatial proximity. Mathematically, spatial distances can be represented by a similarity measure. Some integrated percepts can be decomposed into principle perceptual attributes that correspond to axes or their spatial projections. To discover spatial equivalents of inter-relations between percepts visualization and clustering algorithms are often used, such as multidimensional scaling (MDS, [Shepard, 1962]), principal component analysis, independent component analysis (Comon [1994], cf. Appendix A.1), SOM (cf. Section 2.3.4), or correspondence analysis (Section 2.3.4). There are spatial perceptual models of emotions, timbre, pitch, and keys. Some perceptual qualities are circular, e.g. pitch class. For such structures, the Euclidean space is not an appropriate model. Manifolds like circles, helices, and toroids are more suitable. Camurri et al. [2000] use a spatial arrangement of emotional states combined with a mapping of emotions to rhythmic and melodic features. In their installation, autonomous robots react to visitors by humming sad or happy music and by moving in a particular manner. Since the perception of timbre is complex, it is suitable to project timbres into a two- or three-dimensional timbre space, in which the dimensions are related to physical quantities [Grey, 1977; McAdams et al., 1995; Lakatos, 2000]. For example, in the three-dimensional timbre space of McAdams et al. [1995] the first dimension is identified with the log attack time, the second with the harmonic spectral centroid (Equation 2.19), and the third with a combination of harmonic spectral spread and harmonic spectral variation. Decomposing pitch into pitch class and brightness, Drobisch [1855] supplies a spatial model of pitch in Figure 1.9. In one model of inter-key relations, in Weber [1817]'s key chart, a key can be located on the axes of fifths and major/minor kinships. Another plausible model of inter-key relations is provided by the MDS analysis of the outcome of the psychoacoustic probe tone experiment [Krumhansl and Kessler, 1982]: all major and minor keys are arranged on a torus, preserving dominant, relative and parallel major/minor relations. This structure also emerges from a learning procedure with real music pieces as training data (cf. Section 3.8). Walking on the surface of the tone center torus results in "maximally smooth transitions between keys" (Werts [1983], p. 72).

[9] Accordingly a particular class of simple imaginations be of a kind so that if many of them are in consciousness simultaneously, derived from their quality certain degrees of fusion have to occur: Then these imaginations organize certainly *next to* and *between* each other in a way so that one cannot integrate them in another but a spatial manner and one cannot express oneself but in those words that are seemingly borrowed from space, but actually are equally naturally appropriate, as if one relates them to space. (cf. also Noll [2005])

1.2.1 Tonnetz, Temperaments, and Torus

The tonnetz (Figure 1.8; Euler [1774] [10]) is a spatial representation of just intervals. In the tonnetz the tone space is spanned by just major thirds and just perfect fifths (and just perfect octaves). On the horizontal axis tones progress in just perfect fifths. Along the diagonal from lower left to upper right, tones progress in just major thirds. We can extend the planar grid spanned by integer multiples of just perfect fifths and just major thirds, as in Figure 1.8, to a three-dimensional grid by adjoining the just perfect octave as a third basis vector spanning the third dimension, the Z-axis. Other intervals can be constructed as combinations of major thirds, perfect fifths and perfect octaves. The tonnetz is the origin of many spatial arrangements of tones. Several music theorists, among them Oettingen [1866], Riemann [1877], Lewin [1987], and Chew [2000], base their account of harmony on the tonnetz.

In music making a compromise is made between just and adjusted intervals. Tuning schemes provide a receipt which intervals to tune just and which to adjust.

[10] Euler [1774], p. 584 = Paragraph 29., p. 350-351 of the original.

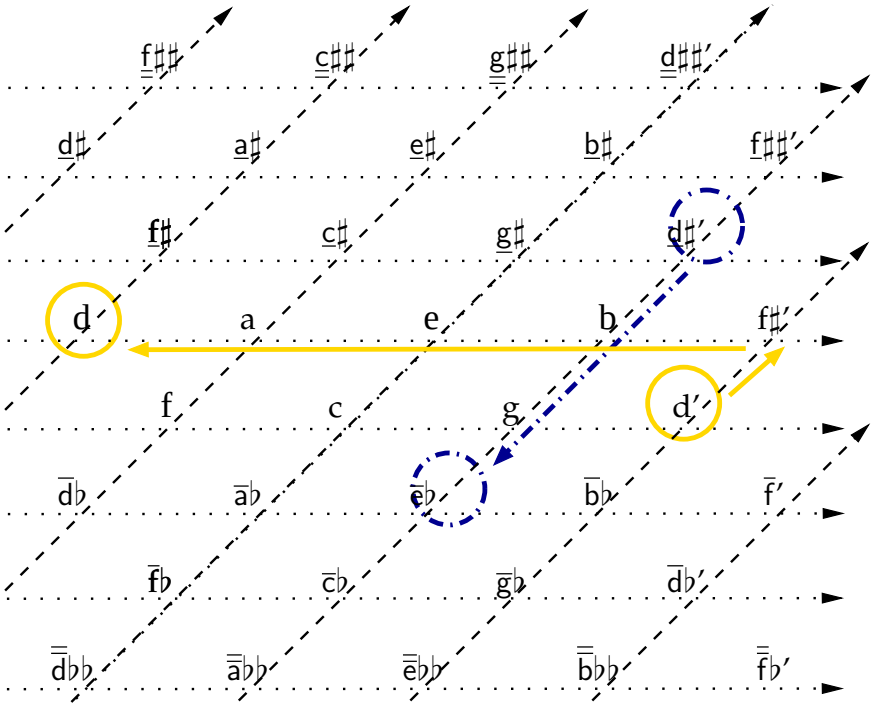


FIGURE 1.8: Euler's tonnetz (cf. Section 1.2.1) in Riemann's notation. On the horizontal axis tones ascend in just perfect fifths from left to right. On the diagonal tones ascend in just major thirds from the lower left to the upper right. Assuming enharmonic identification tones ascend chromatically in vertical direction from bottom to top. The solid arrow indicates synchromatic identification. The dashed dotted arrow signifies enharmonic identification (cf. text).

Historically, temperament developed from Pythagorean, mean-tone, well-tempered to equal temperament. The old Greeks build a scale from two perfect fourths, so called tetrachords, each of them divided into four notes. There are many different concepts how to fill the second and third note in a tetrachord (Philolaus, cf. p. 22, Archytas, cf. p. 22, Aristoxenus [330]; al Farabi [950]). The *Pythagorean temperament* is constructed from pure fifths. To yield the twelve tones of the chromatic scale, Pythagorean temperament consists of eleven pure perfect fifths. The sour twelfth fifth, the wolf, has to compensate for the other just fifths. It is smaller than a pure fifth by 23.46 Cent, the *Pythagorean comma*. Historically, the wolf is assumed to be positioned at $g\sharp$–$e\flat$ by most authors.Another tuning – *mean-tone temperament* – emphasizes the pure major thirds (frequency ratio $\frac{5}{4}$). They are divided into two equal whole tones.[11] As a consequence, in mean-tone temperament all fifths except the wolf are tuned a $\frac{1}{4}$ syntonic comma ($\frac{21.05}{4}$ Cent) smaller than the pure fifth. The wolf fifth is most commonly positioned at $g\sharp$–$e\flat$, or distributed between $c\sharp$–$a\flat$, $g\sharp$–$e\flat$, or $d\sharp$–$b\flat$. Werckmeister [1697]'s recommendation for *well-tempered tuning* requires that it "makes a pleasing variety". Similarly Kirnberger [1776] says it should not "injure the variegation of the keys." [12] *Equal temperament* of the chromatic scale is defined by uniformly tuned minor seconds with frequency ratio $\sqrt[12]{2}$. It is widely established around 1800. We will see in Section 3.6.2 how key preference is interdependent from development of temperaments.

It is convenient to identify tones in the tonnetz, yielding a toroidal configuration. Some tones appear twice in the tonnetz, e.g. d and d'. Every two points, represented by knots in the cubic grid, can be derived from each other by linear combination of horizontal steps, integer multiples, of just perfect fifths, diagonal steps of just major thirds, and steps of just perfect octaves in Z-dimension. Starting from d' we step up a just major third to $f\sharp'$ and then step down four just perfect fifths and reach d. By stepping up in Z-dimension two just octaves from d, we arrive at a tone d^{**}, close in frequency to tone d'. But the frequencies of d' and d^{**} differ by the *syntonic comma* that is 21.51 Cent. In Western music it is prevalent to identify them as being the same tone and to assign the same pitch class to d and d'.

In contrast, another identification, the *enharmonic equivalence*, e.g. $d\sharp/e\flat$ identifying $\overline{e\flat}$ and $\underline{d\sharp'}$ (abstracting from octave position), is more debated. For example, the archicembalo, described by Nicola Vicentino in 1555, has separate keys for $d\sharp$ and $e\flat$ (Lindley [1987], p. 153). But on a contemporary piano the two enharmonically differently spelled notes $d\sharp$ and $e\flat$ are represented by the same key, i.e. $d\sharp$ and $e\flat$ have the same acoustical correlate. However, enharmonic spellings have different musical meaning (Dahlhaus [1967], p. 41). According to Leman [1995] (p. 3) and Kurth [1913], the meaning of a tone cannot be fully provided by the isolated tone itself. Instead, the context substantially contributes to its meaning. To give another reference to linguistics, the meaning of a word depends on the situational context of the conversation ("speech acts", Austin [1962]). If the key of a piece of

[11] Sadie [2001], Vol. XVI, p. 205-6 by Mark Lindley.

[12] Both cited in Lindley [2004]. For more details on well-tempered tuning cf. Section 1.2.8 on page 60.

Pitch Class	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Major	*i*	*i♯*	*ii*	*iii♭*	*iii*	*iv*	*iv♯*	*v*	*v♯*	*vi*	*vii♭*	*vii*
Minor	*i*	*ii♭*	*ii*	*iii*	*iii♯*	*iv*	*v♭*	*v*	*vi*	*vi♯*	*vii*	*vii♯*

TABLE 1.1: Heuristics to assign enharmonic spelling of the non-diatonic pitch class as downward alteration of the higher scale degree or upward alteration of the lower scale degree if the key is the only knowledge provided a priori.

major/minor tonality is the only information known beforehand a straightforward heuristics to determine the enharmonic spelling of a non-diatonic note would be the following (cf. Table 1.1). For major, we introduce accidentals according to their appearance in the circle of fifths, and under the assumption that the tendency towards upward modulation in the direction of the dominant is more frequent than downward modulation. So we add the first three upward alterations (sharps) and the first two downward alterations (flats). Instead, we also could have introduced *ii♭* accounting for a frequent use of Neapolitan chords. For minor, in the same way we add tones that do not belong to Aeolian minor. The choice of *vi♯* and *vii♯* is clear due to the melodic and harmonic variant of the minor scale. A more precise mapping of pitch class to enharmonic spelling can be yielded if statistics of enharmonic spelling are collected for particular styles. However, sometimes the a priori knowledge provided by the key does not suffice. The melodic and harmonic development, previous and subsequent to a pitch class, have to be considered. In addition, the actual spelling in the score is of minor importance. Let us give an example. In an e–minor context a $d\sharp/e\flat$ pitch class naturally means $d\sharp$ as a leading note resolving upwards to e. Nonetheless, Chopin effectively misleads this expectation in bar 2 of his PRÉLUDE op. 28 No. 4 in e–minor. $d\sharp/e\flat$ resolves downwards to d. So in this situation the key of the piece does not provide enough information to clarify the enharmonic spelling. The knowledge of the resolving note and the general descending voice leading reveals the spelling as $e\flat$.

By synchromatic ($d \cong d'$) and enharmonic ($\overline{e\flat} \cong \underline{d\sharp'}$) identification, the tonnetz turns into a $\mathbb{Z}_4 \times \mathbb{Z}_3$ torus. The circle of fifths wraps around the torus three times. Cf. the torus analogously derived from the chart of tone centers in Section 1.2.7.

Criticism of the Notion of Tone Tuning sometimes varies within a piece. E.g. violinists adjust their intonation reacting to their fellow musicians. In the music of the Aka pygmies, Central Africa, investigated by Arom [2003], throughout a particular performance of a piece pitches of a well-defined set are precisely repeated. In contrast, the set of employed pitches varies drastically across various performances of the very same piece. These findings suggest that the delineated basis of tones, temperaments, and pitch makes cultural presuppositions prevalent in Western European music.

1.2.2 Pitch Classes

A subspace in the tonnetz defines a set of tones, such as the pentatonic, diatonic, or chromatic scale. Temperaments distort the just inter-relations of tones provided in the tonnetz in order to achieve scales for practical music making. By enharmonic and octave equivalence we yield the twelve *pitch classes*. Pitches that are related to each other by intervals of integer multiples of an octave are treated as one equivalence class. It is not always appropriate to consider pitch classes instead of pitches: E.g. the pitches used by the Aka pygmies [Arom, 2003] in a single piece do not repeat in other octaves. Also in harmony of Western European music and Jazz the octave position of tones matters. Chord inversions can only be interchanged according to specified rules. However, there are several arguments emphasizing the eminent role of octave equivalence: 1) biological reference by the uniformly spaced pattern of resonance frequencies on the basilar membrane (cf. p. 33), 2) concept to explain psychoacoustic phenomena (Figure 1.1.3 on page 33), 3) strength of the second partial, the octave, in the singing voice and in Western musical instruments, and 4) important role in Western harmony.

Are all twelve pitch classes equally important with respect to key? Sometimes a few notes suffice to indicate the key. Harrison [1994] (p. 73) names this concept *position finding*. E.g. for a major scale a tritone (b–f) reduces the potential keys in question to two (C–major and G♭–major), thereby giving rise to the tritone substitution, a harmonization technique in jazz [Levine, 1996]. Provided an established tonal context and depending on the keys perceived as most kindred (cf. Section 1.2.5) in that musical style the number of key candidates is further reduced. E.g. for Bach in C–major a transition to G–major is very likely. If there is an $f\sharp$ appearing following a passage in C–major, this most likely indicates a tone center (cf. p. 54) of G–major. Provided a keynote, in Section 3.4.2 we will show that for major/minor mode discrimination, the major/minor third is most important. For Bach all pitch classes are significant. For Chopin only thirds, sixths, and sevenths are discriminating, for Shostakovich only thirds. The selected pitch class material, e.g. the major, minor, or whole tone scale, has a great impact on the style and the character of a piece of music. In Section 1.2.9 on page 61, we will describe how pitch class profiles of frequency of occurrence signify keys.

1.2.3 Pitch Decomposition into Pitch Class and Brightness

Pitches of tones can be thought of as aligned on a straight line, such as the chromatic scale on the vertical axis in the tonnetz (Figure 1.8). Naturally pitch is related to brightness. Normally a tone of higher pitch is brighter as well. On the other hand, by introducing pitch class, it becomes apparent that pitch also has a circular attribute. Integrating both aspects results in a *helix* (Drobisch [1855], cf. Figure 1.9 on the facing page and p. 36 ff.). The helix arrangement decomposes pitch into the two dimensions: pitch class, defined by the position on the circle, and *brightness*, given by the vertical position. Pitch class represents the top view on the helix. Brightness

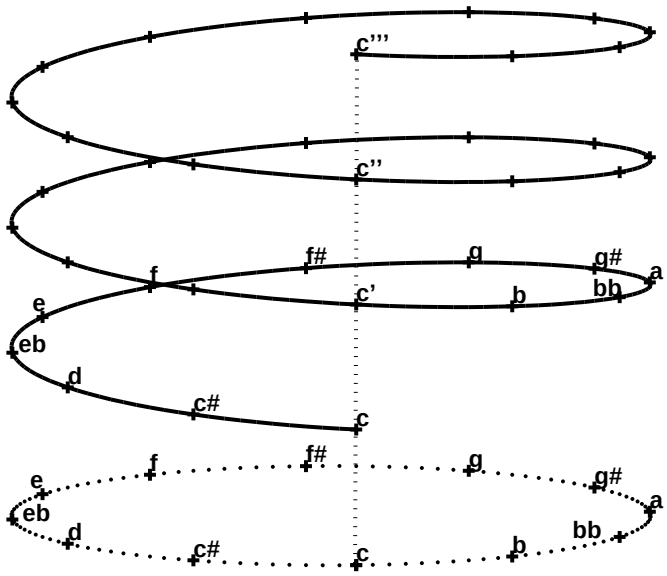
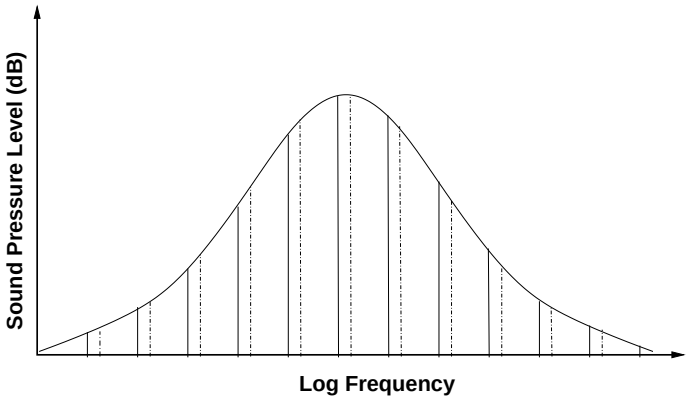


FIGURE 1.9: Escalating chromatic scale depicted in a helix. Drobisch [1855] decomposes pitch into brightness (vertical axis) and pitch class (angle in the pitch class circle). Projecting the tones of the helix on the top view results in the dotted pitch class circle below.

is ambivalent, since besides being a component of pitch, it is also a quality of timbre. If brightness and pitch class are independent dimensions of pitch, there should be tones that differ only in either attribute. E.g. a brightness value can be assigned to band-limited noise. But the latter has no pitch class.

It is Shepard [1964]'s intention to build tones that differ in pitch class but not in brightness. In the geometric pitch model of a helix (Figure 1.9), a tone sequence made up of such tones does not move vertically, but on the pitch class circle that is the projection of the helix on the top view. In this projection, pitches in octave interval distance are represented by exactly one point. This is the ground for Shepard's construction of complex tones that are used to build tone sequences that are perceived circularly in pitch. A *Shepard tone* is a complex tone (Figure 1.10 on the next page) with all partials in octave distance (double frequency) to their adjacent partial. Figure 1.10 plots logarithmic frequency against loudness. The figure depicts the partials of two Shepard tones, marked in solid and in dashed-dotted lines. A spectral envelope is fixed for all Shepard tones. It defines the loudness of all partials. For now, we assume a spectral envelope with a bell shape in the loudness domain covering the entire hearing range. By using a bell shape, partials uniformly decrease in intensity towards low and high frequencies, so that the hearing threshold is reached evenly. In contrast to the frequency relations, the exact shape of the



(a) Spectral structure of Shepard tones

FIGURE 1.10: Shepard [1964] (p. 2347 Figure 1) designs an intransitive tone sequence. The complex tones only consist of partials in octave distance. A fixed bell shape controls the sound pressure levels of the partials. When the complex tone is transposed to a higher frequency (dashed line), higher partials are muted until they vanish and lower partials fade in. Transposing the complex tone by one octave leaves the tone unchanged.

spectral envelope is not crucial for the realization of the idea. In our experience, a constant intensity envelope for the intensities yields almost the same sound quality, since the hearing threshold then induces a bell shape in the perceived loudness. If we listen to a continuous sequence of Shepard tones we let the grid of partials slide from left to right underneath the fixed spectral envelope. A low partial inaudibly moves in, so that it does not ruin the synthetic perception of the tone as a whole. Another intention of choosing a fixed and continuous spectral envelope is the constant loudness of all Shepard tones. In contrast to harmonic complex tones, all partials are eliminated that do not stem from frequency doubling. Another property signifies Shepard tones. In Shepard tones the notion of fundamental frequency becomes fuzzy, even though the maximum of the brightness envelope highlights a particular octave component. Relative to this pseudo fundamental a Shepard tone does not only contain partials that are generated by frequency doubling. A Shepard tone also contains the frequencies generated from bisection of the frequencies of its other partials. A Shepard tone has a pitch class, but cannot be assigned to an octave position. The sound of a Shepard tone resembles the sound of an organ.

The pitch "paradox" is depicted in Escher's lithograph ASCENDING AND DESCENDING (1960), Figure 1.11: A closed chain of people ascends a stair. The "lowest" person is above the "highest" person. One can infinitely ascend this staircase. After one round the stairs reach their starting point again. The stairs remain in the same level of the building, the roof, as if the constant height in level would depict constant brightness.

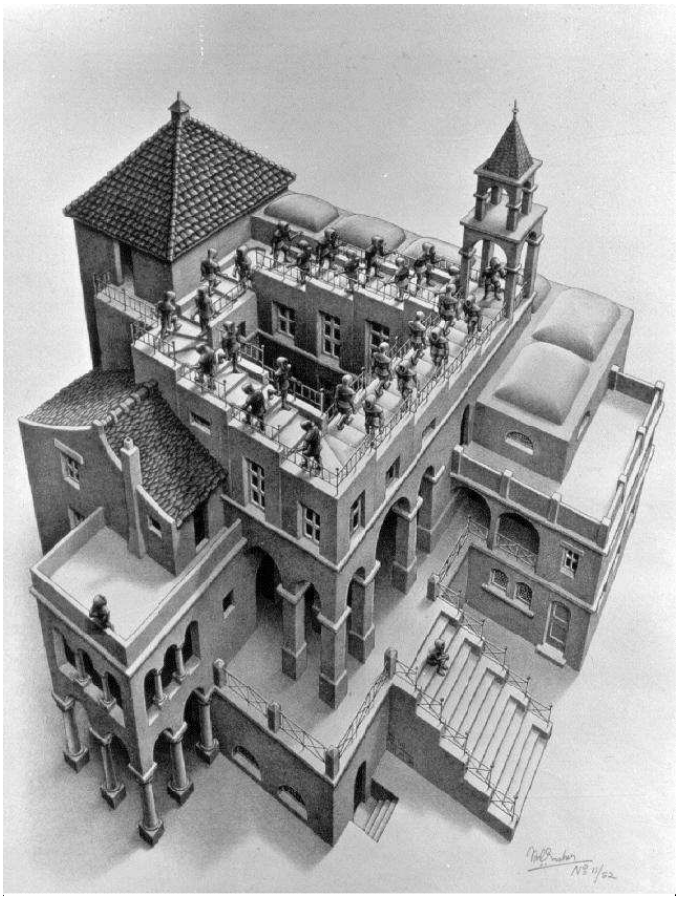


FIGURE 1.11: Escher's ASCENDING AND DESCENDING. The circular staircase depicts the pitch class circle, the stairs remain in the same level, analogously to the constant brightness in the pitch "paradox".

1.2.4 Chords and Harmony

How can we categorize sets of simultaneous pitch classes, such as chords? Scale degrees, harmonic function, and chord morphology are possible criteria. As stereotype sequences, chords coin cadence and key.

The Key to Tonality Let us first delineate a cadence as a constituting feature of key. A *cadence* is

> the conclusion to a phrase, movement or piece based on a recognizable melodic formula, harmonic progression or dissonance resolution ..... The cadence is the most effective way of establishing or affirming the tonality ... of an entire work or the smallest section therof. [Rockstro et al., 2004]

We can distinguish cadences on the basis of their varying degree of "finality". Either the end of the cadence deviates from the $V - I$ pattern (perfect versus plagal, imper-

fect, and interrupted cadence) or at least one chord is not in root position (perfect versus authentic and medial cadence).

For a profile of frequency of occurrence of pitch classes (cf. Section on page 14) accumulated across an entire piece, the sequentiality of notes and their positions relative to beat weights are neglected, thereby missing major defining features of a cadence. Likewise, Mazzola [1990] reduces a cadence to the minimal set of triads on scale degrees completing the pitch class set of a particular major scale or its minor relative, thereby ignoring sequentiality aspects as well. However, ignoring sequentiality we will see that profiles of pitch classes contain much information about key and composer.

For training in Leman [1995] and in Section 3.8.2 cadences such as the following are used: $I - IV - V^7 - I$, $I - ii - V^7 - I$, $I - vi - V^7 - I$. Koelsch et al. [2000] also use $I - II\flat - V^7 - I$, including the Neapolitan $II\flat$.

Categorization of Synchronous Pitch Classes Rameau [1722]'s *scale degrees* associate a chord class with each tone of the diatonic scale. From the viewpoint of *functional harmony* [Riemann, 1877], a cadence is the interplay between the tonic (T), dominant (D), and subdominant (S) function. Typical representatives are the IV^6_5 for the subdominant and dominant seventh chord V^7 for the dominant (cf. Figure 1.12).

Dahlhaus [1967] understands function theory [Riemann, 1913] to be a further abstraction of the seven scale degrees to the three tonal functions. But the scale degrees cannot be mapped to the tonal functions in a unique mapping. The third scale degree, e.g. e–minor in C–major, is ambiguous in its double role, as tonic mediant and dominant relative. With dominant functionality, the third scale degree could constitute part of a sequence, such as $I - V/vi - iii$, but not a cadence, i.e. $I - IV - iii - I$ (Dahlhaus [1967], p. 146). Facing ambiguity problems of such kind, we could replace the unique mapping to harmonic functions by fuzzy ones, also interpretable

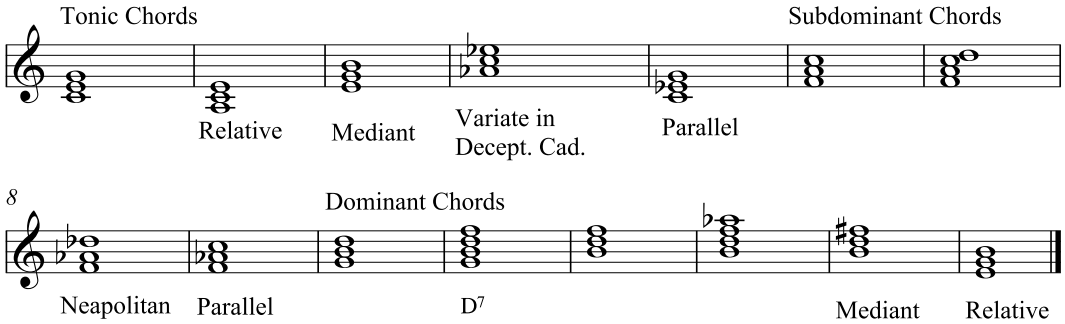


FIGURE 1.12: In functional harmony chords are related to three functions: tonic, subdominant, and dominant. For C–major typical representatives for these functions are given. Terminology: The relative function is also called sub-mediant. The mediant is also called leading-tone change, or "Leittonwechselklang" in German. Confusingly, relative is "parallel" in German.

as probabilities of harmonic functions, given in a Riemann matrix [Mazzola, 2002].

Oettingen [1866] relates the major/minor duality to the duality of the harmonic and the subharmonic series. The latter plays an important role in virtual pitch (cf. Terhardt [1998] and Section 2.2.2). Noll [1995] suggests a classification of chords based on their group theoretic *morphology*. Affine linear mappings, concatenations of a transposition and a multiplication, operate on pitch class sets. These mappings are employed to classify chords. Morphological chord analysis by means of group theory is also performed by Cohn [1997].

Instead of classifying chords into a small number of classes, chords can be evaluated on a refined scale with respect to how well they match with the tonal context. Provided a keynote as a focal point, the tonic triad fits best in the tonal context, followed by triads built on the fifth, fourth, sixth, and second scale tone, according to a psychological experiment by Bharucha and Krumhansl [1983]. Besides the appropriateness of a chord with respect to a given tonal context, also the sequentiality of chords is important. Thereby syntactical aspects of music come into play.

Syntax and Music In an electroencephalogram (EEG) study, Koelsch et al. [2000] demonstrate that parsing harmonies is part of music cognition. Event related potentials significantly reflect violation of music expectancy, built up from preceding musical context, especially for the final chord of a cadence. Various cadential chord progressions consisting of five chords are presented. The replacement of the final tonic by a Neapolitan chord, $\mathrm{II}_3^\flat$, evokes both an early right-anterior negativity and a late bilateral negative frontal deflection. Koelsch et al. [2000] interpret this result as a hint that language and music are closely related, since both require a grammatical parsing process. Some kind of parsing, similar to language processing seems to be involved in music listeners.

Lerdahl and Jackendoff [1983] implement Schenkerian analysis[13] by context free grammars that nowadays appear a plain mechanization of Schenker's in-depth analysis that is enriched by his musical intuition. Bod [2002] introduces a data-driven method to automatically and recursively construct grammars from previously derived subtrees according to principles of likelihood and simplicity. Whereas syntactic investigations might be appropriate to understand some features in music, other aspects may be rather described as a continuous dynamic process[14] or as rigid idiomatic phrases, such as stereotype cadential formulas, fifth fall sequences, and other mechanical musical elements. For elucidation let us refer to linguistics once more. In his controversy to Fodor, Smolensky [1991] suggests that language processing consists of complex processes of connectionist information flow, whereas on a high level it is coarsely describable as a hierarchical grammar not parallel to its syntactical neural implementation. Likewise, traditional music theory based on the idea that musical knowledge can be reduced to building blocks, such as tones,

[13] Schenker [1935] reduces a score to a sequence of principle notes. The latter are in turn reduced to the Urlinie, "the conceptual upper voice of a piece in its simplest terms." [Drabkin, 2004]

[14] Lidov [1999] tracing back to Schopenhauer [1859], Nietzsche, and Herbart.

chords, and keys, can meet with cognitive musicology that treats objects of musical knowledge as results of a connectionist processing of sensorial information (Leman [1995], p. 182), describable as a complex dynamic system.

1.2.5 Key Proximity

Like tones, keys are coined by their inter-relations. What makes keys adjacent? A criterion frequently mentioned is the *number of tones shared by the corresponding scales* (e.g. Weber [1817]). Considering the harmonic scale for minor, according to this measure dominant, subdominant, and relative key (cf. Figure 1.12) are close to the tonic. Their scales differ from the tonic scale only in one tone ($f\sharp, b\flat, g\sharp$ in C–major). Mainly based on the number of common scale notes, Schönberg [1969] (p. 67 f., p. 73) divides up tone center inter-relations into five classes. The first two classes contain keys having five to six, respectively three to four scale tones in common with the reference major key.

Another viewpoint reduces a key to its tonic triad. As a consequence, the resulting criterion for key proximity is the *number of common-tones shared by their tonic triads*. There are only three non-identical pairs of major/minor triads sharing two common-tones, as pointed out by Heinichen [1728]. [15] The three chords can be identified with the relative, parallel, and mediant. Yet another criterion for key kinship is the *identity of the scale degree meaning of shared pitch classes*. It is employed by Weber [1817] to account for the close parallel relationship between C–major and c–minor [Werts, 1983]. Both keys share the same principal scale degrees, i.e. I, IV, and V. Weber [1817] in this case especially emphasizes that both keys share the tonic note and the fifths scale degree.

These key kinships also appear in the visualization of the psychological probe tone experiment [Krumhansl, 1990] in Figure 1.21 on page 63. Cf. Appendix C for an overview of various dissimilarities and distance measures that can be employed to mathematically express proximities, e.g. between keys.

1.2.6 Chart of Tone Centers

Measures for key proximity imply a spatial display of keys.[16] Illustrated by Figure 1.13 on the next page, the head of Chapter V in Heinichen [1728], p. 837, reads:

> Von einem Musicalischen Circul, aus welchen man die natürliche Ordnung, Verwandschafft, und Ausschweiffung aller Modorum Musicorum gründlich erkennen, und sich dessen so wohl im Clavier als Composition mit vortrefflichen Nutz bedienen kann. [17]

[15] cited in Werts [1983] and revisited in Cohn [1997].

[16] A *tone center*, also called "key area" or "key region" [Schönberg, 1969], is a predominating key within a given time frame. Throughout this thesis we will, for the sake of simplicity, occasionally not differentiate between keys and tone centers.

[17] About a musical circle, by which one can thoroughly comprehend the natural order, relationship,

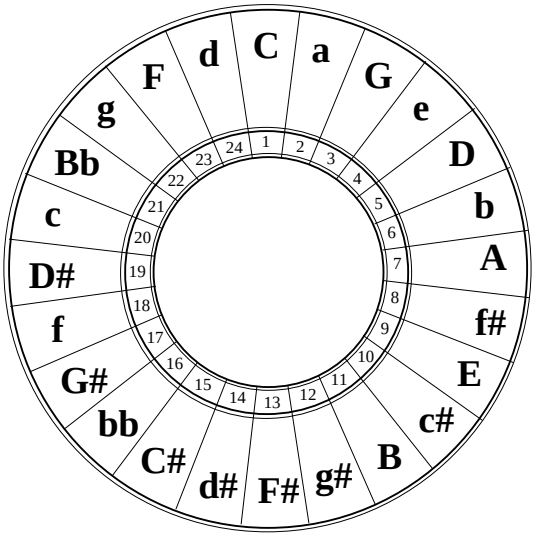


FIGURE 1.13: Circle of fifths adopted from Heinichen [1728], p. 837. Major keys (capital letters) alternate with minor keys (small letters).

"The organization principle of Heinichen's scheme appears to be the maximization of common-tones between adjacent scales," (Werts [1983], p. 71) if we refer to minor by the Aeolian scale. In the sequel, alterations of this graph are proposed by Mattheson [1735] and Kellner [1737], the latter suggesting two nested circles, an outer circle of major keys and an inner circle of minor keys, adopting the adjacency of a major and its relative minor key. Thereby Kellner's arrangement reflects the inherent two-dimensionality, spanned by the circle of fifths and the major/minor kinship. Heinichen's circle of fifths can be linked to the tonnetz, Figure 1.8, in the following way: Take the extended horizontal axis of just fifths and replace each tone by the major key with that keynote and its relative minor key. Then enharmonically identify $g\sharp$ and $a\flat$. The circle of fifths can be interpreted mathematically. Noll [2003] relates the circle of fifths to the twelve elemented factor group $Sp_1(\mathbb{Z})/Sp_1(\mathbb{Z})^c \cong \mathbb{Z}_{12}$ of the modular group[18] and its commutator subgroup.[19]

In the chart of tone center inter-relations, tone centers are arranged according to their proximity to the prevailing key. Figure 1.14 on the following page depicts this for C–major. The organization principles of the chart are the maximization of common-tones between the referring scales and the identity of scale degree meaning of the shared pitch classes. Dominant, subdominant, and relative relations are given reason by corresponding scales sharing all but one tone. For the parallel relationship,

and modulations of all musical keys, and of which one can make good use of at the keyboard as well as in composition.

[18]The modular group $Sp_1(\mathbb{Z})$ is the group of 2×2 matrices M in $\mathbb{Z}$ with $M^t J M = J$ and $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.

[19]For a group $\mathcal{G}$ the commutator subgroup $\mathcal{G}^c$ is the subgroup generated by all elements of the form $aba^{-1}b^{-1}$ with $a, b \in \mathcal{G}$.

d♯	F♯	f♯	A	a	C	c
g♯	B	b	D	d	F	f
c♯	E	e	G	g	B♭	b♭
f♯	A	a	C	c	E♭	e♭
b	D	d	F	f	A♭	a♭
e	G	g	B♭	b♭	D♭	d♭
a	C	c	E♭	e♭	G♭	g♭

FIGURE 1.14: Weber [1817]'s chart of tone centers arranges all major (capital letters) and minor keys (small letters) according to their relation to a prevailing key, in this case C–major. The chart is structured so that fifth relationships lie along the vertical axis, and in the horizontal axis relative and parallel relations alternate.

C-major – c–minor, the tonic note and the harmonies on the fifths scale degree are the same. Schönberg [1969] adopts Weber [1817]'s chart in Figure 1.15 on the next page. The remote tone centers are somewhat compressed. We also witness the Neapolitan D♭ [20] distorting the previous symmetry of the chart. In contrast to Weber's chart, in Schönberg's chart distances between tone centers are relative to the key. For instance, the distance between tone centers G♭–major and g♭–minor is smaller in the key of C–major than in G♭–major.

1.2.7 Torus by Synchromatic and Enharmonic Identification

We will see that the toroidal structure of the inter-key relations is implicitly contained in Weber [1817]'s chart of tone centers (Figure 1.14). [21] This is plausible according to geometric considerations. We give reasons why the surface of a torus supplies a suitable structure for a configuration of the major and minor keys. In addition, we investigate how such an arrangement can be derived using consider-

[20] Here, we use the term Neapolitan in a broader sense, although it usually refers to the II♭ triad only if the third is in root position and if all tones are resolved appropriately.

[21] Barely noticed, Werts [1983] (p. 18, Figure 47) presents the same train of thoughts.

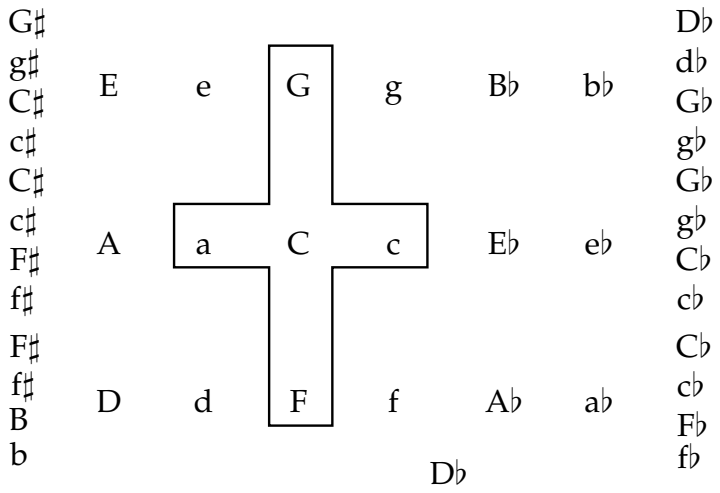


FIGURE 1.15: Chart of tone centers (Schönberg [1969], p. 20), going back to Weber [1817]'s chart, Figure 1.14. Note the asymmetry of the Neapolitan D♭–major and the squeezed remote keys on the left and right side.

ations based solely on fundamental information about musical inter-key relations. Let us first consider a strip, Figure 1.16, cut out from Weber's chart of tone centers, Figure 1.14, then rotated from vertical to horizontal orientation.

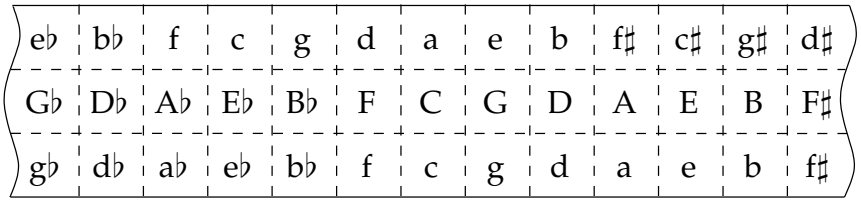
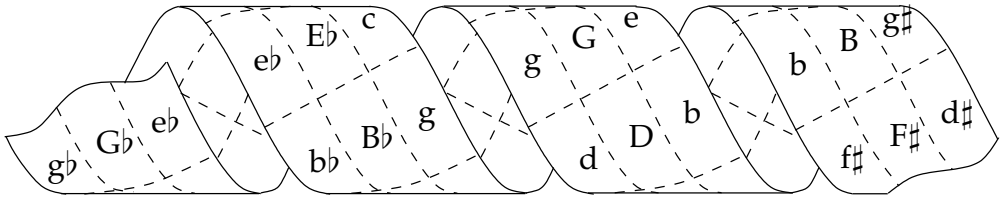


FIGURE 1.16: Strip of keys cut out from Weber's chart of tone centers, Figure 1.14, and rotated to horizontal orientation.

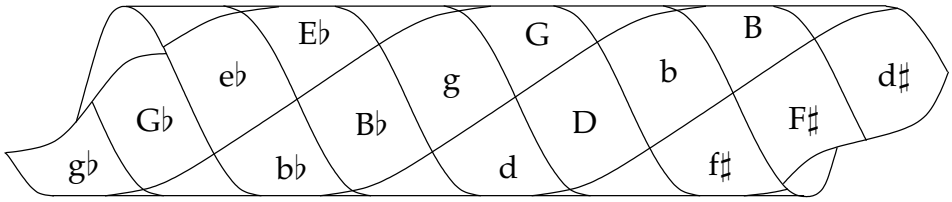
In this scheme, some keys and their enharmonic equivalents appear in multiple places. For example, g–minor appears in the upper row in Figure 1.16 as well as three positions to the right in the lowest row. In order to arrange a unique position for each key, we have to apply a spatial transformation. One curls up the strip, forming a tube (Figure 1.17 on the next page) in order to unite the minor keys of the upper row with those of the lower one. The redundant occurrences of g–minor in Figure 1.17 (a) coincide in Figure 1.17 (b). This is the synchromatic identification we apply to tones in Figure 1.8 on page 45.

The ends of the tube in Figure 1.17 (b) show keys that are enharmonically equivalent. These are united by bending the tube and joining its ends. This creates a toroidal surface, a *toroidal model of inter-key relations (TOMIR)*, around which the circles of fifths for major and minor keys wrap three times before joining themselves. If

we unite major keys with their parallel minor keys, the circle of minor thirds evolves on the diagonal from the lower left to the upper right (g♭, e♭, ...). The circle of major thirds occurs horizontally (g♭, b♭, d, ...).



(a) Curling of the strip (The strip is shown somewhat stretched for better display)



(b) Double appearances of minor keys are brought to accordance

FIGURE 1.17: The two locations of g–minor are identified by synchromatic identification: (a) The strip of keys is curled up. (b) Double occurrences of minor keys join. The same procedure is applied to tones in Figure 1.8 on page 45. At both ends of the tube enharmonically equivalent keys appear. Their identification by joining both ends of the tube yields a torus.

In Section 3.7.2 on page 154, the same arrangement is established by a computer program that is designed to derive an appropriate configuration from a set of stipulated close relationships.

The curling of Weber's chart to a torus can be subtly related to tone center modulation path and key character symbolizing the underlying idea in Richard Wagner's PARSIVAL: spiritual and physical journey.

> Gurnemanz: Du siehst mein Sohn, zum Raum wird hier die Zeit [22] (PARSIVAL, Act I)

Temperley [2001], p. 119 ff, deducts character and semantics of tone centers from such a journey encoded by the leitmotifs of mystery, wound, spear, and sorrow. We expand and reorient Weber's chart of tone centers, Figure 1.14, to the 12×9 grid shown in Figure 1.18 on the next page, centered at D-major as an intersection point (the wound) of a cross formed by the vertical axis of the circle of fifths from north, A♭–major for "heaven", to south, A♭–major for "earth", and the horizontal axis of minor thirds from west, A♭–major for "evil", to east, A♭–major for "good". In this light, e.g. in the Act I Prelude from bar 80 to bar 100, the tone center trajectory from

[22] "You see, my son, here time is turned into space"

					Heaven					
	D	b	B	g♯	A♭	f	F	d	D	
	G	e	E	c♯	D♭	b♭	B♭	g	G	
	C	a	A	f♯	F♯	e♭	E♭	c	C	
	F	d	D	b	B	g♯	A♭	f	F	
	B♭	g	G	e	E	c♯	D♭	b♭	B♭	
	E♭	c	C	a	A	f♯	F♯	e♭	E♭	
Evil	A♭	f	F	d	D	b	B	g♯	A♭	Good
	D♭	b♭	B♭	g	G	e	E	c♯	D♭	
	G♭	e♭	E♭	c	C	a	A	f♯	G♭	
	B	g♯	A♭	f	F	d	D	b	B	
	E	c♯	D♭	b♭	B♭	g	G	e	E	
	A	f♯	F♯	e♭	E♭	c	C	a	A	
	D	b	B	g♯	A♭	f	F	d	D	
					Earth					

FIGURE 1.18: In PARSIVAL, tonal space, in Weber's extended chart, depicts the cross of spiritual journey. The north-south passage from northern A♭–major, symbolizing "heaven", southboud to another A♭–major, "earth", meeting with the east-west passage from a third A♭–major, "evil", westward to yet another A♭–major, "good". Both paths cross at D–major, signifying the wound. (Graphics and description following Temperley [2001], p. 119 ff.)

A♭–major to A♭–major, along the horizontal axis in minor third steps, symbolizes the spiritual journey from "good" (west) to opposite "evil" and "sin" (east). In the projection of the trajectory onto the torus the opposite positions of A♭–major, west and east, are identified, signifying identity of the physical location of Amfortas' home as both departure and destination place of his journey. Physically, A♭–major has a unique location on the TOMIR. Spiritually, A♭–major has different meanings located in four separate spots on Weber's chart.

1.2.8 Key Character

Weber [1817]'s chart of tone centers (Figure 1.14) and Schönberg [1969]'s adaptation thereof (Figure 1.15) are isomorphic for any key of the same mode. Structurally they look the same for C–major or for D♭–major. However, in PARSIVAL we have seen that every key can be associated with a specific meaning or character. This *key character* is in particular connected to the history of temperaments.

Based on his analysis of Bach's sacral vocal works Mayer [1947] [23] gives an overview of keys and their associated character in Table 1.2 on the following page. Certain rarely used keys, such as B–major, C♯–major, e♭–minor, and g♯–minor, are not indistinguishably characterized. In his WTC, Bach transposed those keys from

[23] Cited in Auhagen [1983].

Key	*Character*
E–major	ethereal, transcendental, consecration
A–major	overwhelming joy, transfigured blessedness
D–major	joy
G–major	merry, charming, content, pastoral
C–major	solemn, majestic, bright
F–major	blessed happiness
B♭–major	confident, solemn
E♭–major	love
f♯–minor	agonizing
h–minor	God
e–minor	graceful, majestic, desperation
a–minor	lamenting, consoling
d–minor	desperate, sad, dreary
g–minor	grievous
c–minor	grievous, consoling
f–minor	intense sorrow

TABLE 1.2: Key character for the most prominent keys in Bach according to Mayer [1947], as cited in Auhagen [1983], p. 240.

pieces originally written in a key one half-tone below or above [Keller, 1965]. Auhagen [1983] points out that the affect of one key partly holds also for the affect of the keys next to it on the circle of fifths. He adds that the intensity of affects increases with the number of accidentals. From investigating Bach's vocal music, Auhagen [1983] argues that Bach uses particular tone centers with threefold intention: First, he associates certain affects with a particular key. Second, he uses tone centers with many sharps to symbolize the cross. Third, the sequel of tone centers is arranged to portrait the affective content of the text. Before being eroded, key character undergoes changes depending on composer and musical epoche. According to Keller [1965], by Beethoven the character of some keys is changed, e.g. c–minor is associated with passion, c♯–minor characterized by the MOONLIGHT SONATA, F♯–major is considered emotional. For Chopin (sonatas) C–major is passionate and F♯–major is associated with "fragile damped light" [Keller, 1965]. In contrast to such a characterization, Auhagen [1983] relates keys in Chopin to genre, meter, and tempo. He finds that A♭–major, C–major, a–minor, and c♯–minor are preferredly used for mazurkas (A♭–major also for waltzes) and therefore are connected with triple meter.

Lindley [2003] points out the relation between keyboard tuning and key character. In well-tempered tuning before equal temperament, with increasing numbers of accidentals, sharps and flats, the third on the major tonic is stretched and the minor second from the leading note to the tonic is diminished. Thereby with keys with more accidentals there is more beating and roughness. Therefore these keys sound more vibrant, nervous, and less calm. By the establishment of equal temperament

during the decades after ∼1800, nuances of key character slowly vanish. Whereas in Chopin key character can be still identified, this is not evident for Scriabin. [24] In spite of the far-reaching implications of Lindley's demonstrations for performance practice, this hypothesis has not yet been tested based on systematic psychological experiments.

Key character is also influenced by the heritage of the corresponding ecclesiastical modes, the pitch range available to the performing singer or instrument, particularities of certain instruments, e.g. open strings of the violin etc., and playability to wind or keyboard instruments [Lindley, 2003].

In Section 3.6.2 on page 150, we will reveal a link between key preferences of particular composers and temperaments. In the discussion of normalized constant Q profiles (Section 3.4.5 on page 139) we will suggest how this representation can contribute to testing for significance of key character. We account for the differences of key character in the experiments that reveal the circle of fifths based on performance data with correspondence analysis (Section 3.5.2 on page 144) or Isomap (Section 3.5.3 on page 146), and in the evolving toroidal model of inter-key relations based on score data (Section 3.8 on page 156). However, in the late 19th century's perspective of equal temperament and eroded key character we will use transpositions and average profiles in several computer simulations in Chapter 3.

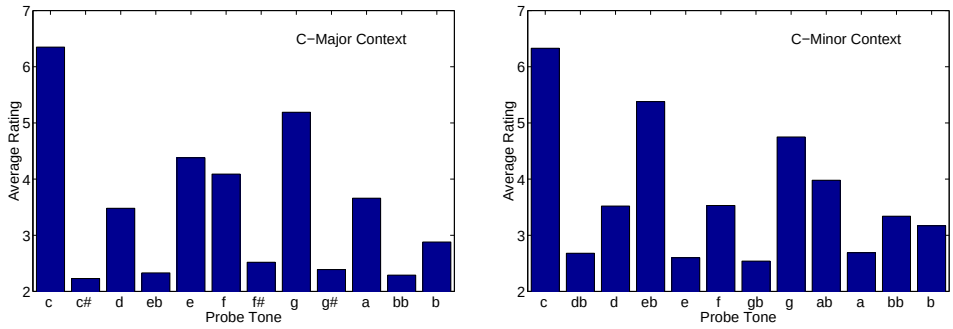
1.2.9 Key – Reduction to an Empirical Profile

It would be worthwhile to have a simple quantitative description of the notion of key that makes it more accessible to statistics and computation in music analysis. One constituting factor of a key is its scale that can be displayed as a binary pitch class vector (Figure 0.1 on page 15). Instead of providing binary information on whether a pitch class is contained in a piece of music or not, we can count the occurrences of pitch classes yielding a profile of frequency of occurrence (cf. Introduction on page 14). Despite the complex inter-relation between major/minor tonality and various musical features, like voice leading, form, beat strength, tonal rhetoric[25], and gestalt, the mere frequencies of occurrences of pitch classes are a major cue for the percept of a tonality [Krumhansl, 1990]. Entirely ignoring the sequential structure of a piece of music, in this reductionist view, a major/minor key is a prominence profile of pitch classes. E.g. for C–major the tone c is most important. The other tones g, e of the tonic triad follow. The non-diatonic notes are least important. The relevance of this representation of a key is supported by an experiment.

Probe Tone Ratings The probe tone experiment is pursued by Krumhansl and Shepard [1979]. *Probe tone ratings* are a quantitative description of a key that creates the possibility of relating statistical or computational analyses of music to cognitive

[24] Personal communication with Mark Lindley.

[25] The initial and the final chord of a piece or a section thereof particularly enforce the perception of the referring key, a phenomenon named *tonal rhetoric* in Harrison [1994].



(a) (b)

FIGURE 1.19: Probe tone ratings supply a pattern of pitch class prominence for each musical key. A tonal context C–major (c–minor) is established by playing cadential chord progressions. Then the subject is asked to rate how well different pitch classes fit into the previous context. The figure shows the averaged answers for subjects with an average of 7.4 years of music instruction [Krumhansl and Kessler, 1982].

psychology. The probe tone experiment consists of two stages: establishment of a tonal context, and rating of the relation of a probe tone to that context. The tonal context is provided by examples, which are unambiguously written in a certain key. In our case the subjects listen to simple cadential chord progressions composed of Shepard tones [Krumhansl and Kessler, 1982], such as $\mathrm{IV} - \mathrm{V} - \mathrm{I}$, $\mathrm{vi} - \mathrm{V} - \mathrm{I}$, $\mathrm{ii} - \mathrm{V} - \mathrm{I}$ for major and analogous cadences for minor. Subsequently the probe tone, a Shepard tone chosen randomly from the chromatic scale, is played. The subject is asked to judge how well the note fits with the tonal context provided by the cadential chord progression. The test subjects rate by a number from 1 ("fits poorly") to 7 ("fits well"). After this procedure is repeated several times, with different chromatic notes, the average rating for each pitch class is calculated (Figure 1.19). The 12-dimensional vector containing the averaged answers for each pitch class is called the probe tone rating. There are two types of rating vectors, one for major and one for minor – depending on the mode of the context. Rating vectors of keys in the same mode but with different tonic keynotes are assumed to be related by a shift that compensates for the transposition interval. Abstracting from key character for the sake of simplicity is in accordance with empirical results (see Krumhansl and Kessler [1982], p. 342). One observes that the first scale degree is rated highest. The third and fifth scale degrees are also rated high. Diatonic notes are rated higher than non-diatonic notes. Thereby pitch classes are ordered by their prominence by a ">" relation, defining a so called *pitch class hierarchy*. According to an observation reported in Krumhansl [1990] (p. 66–76), each component in the probe tone rating vector corresponds to the frequency and the overall duration of occurrence of the corresponding pitch class at metrically prominent positions in a tonal piece that is written in a given key. The meaning of one key develops by referencing the other

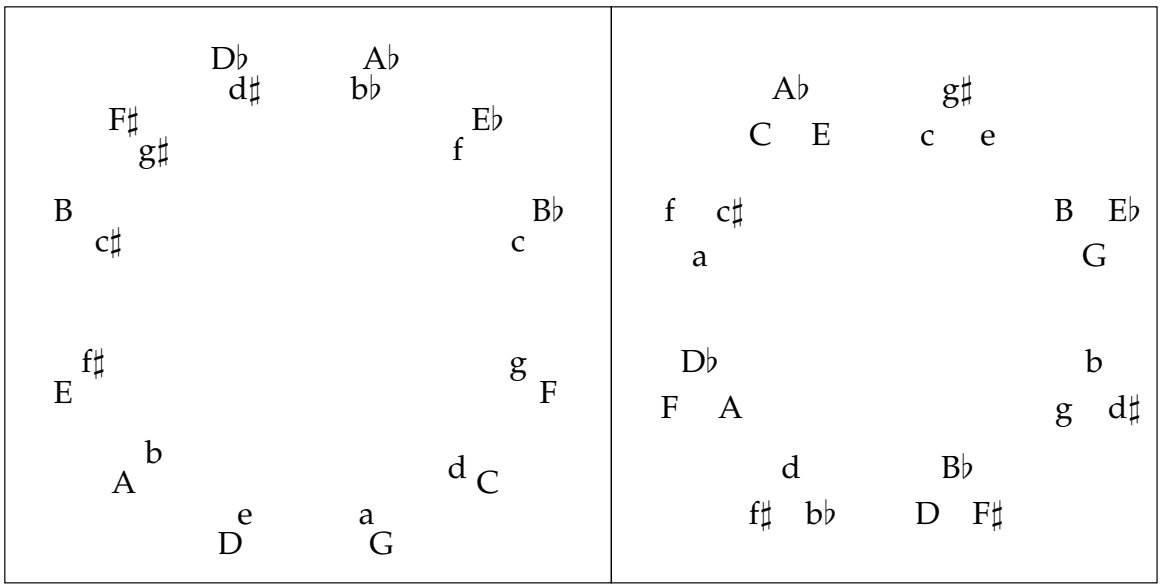


FIGURE 1.20: The probe tone ratings are scaled down to a four dimensional Euclidean space by multidimensional scaling. The scaled points lie approximately on a sub-manifold formed by the crossproduct of two cycles (left graph: 1st+2nd dimension, right: 3rd+4th dimension).

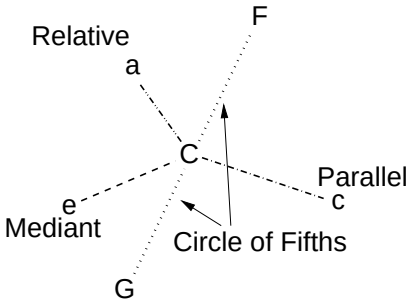


FIGURE 1.21: The implicit two-dimensionality in the key visualization of Krumhansl and Kessler [1982]: One axis is the circle of fifths, the other the inhomogeneous axis of parallel and relative third major/minor relationships. In addition, the mediant, e–minor, is also adjacent to the tonic.

keys, resembling the role of the context for meaning formation of a tone, described on p. 46. Accordingly, key distances are calculated by comparing the corresponding probe tone ratings by correlation, Euclidean distance, or other dissimilarities described in Appendix C on page 195.

A TOMIR from Probe Tone Ratings With a dissimilarity at hand, keys represented as probe tone ratings can be visualized. Krumhansl and Kessler [1982] obtain a TOMIR by scaling down the probe tone ratings to a four dimensional Euclidean

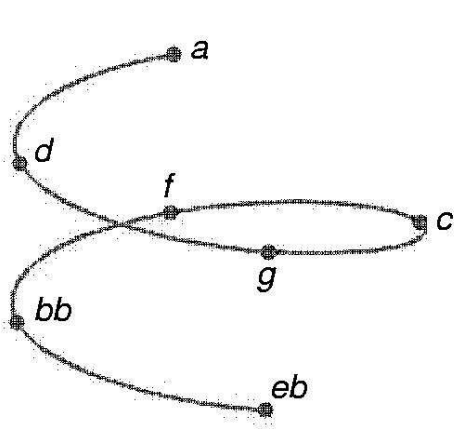
space by multidimensional scaling. Closer inspection reveals that the scaled points, which correspond to the major and minor keys, lie approximately on a sub-manifold formed by the crossproduct of two cycles (cf. Figure 1.20 on the preceding page). This sub-manifold can be mapped to a torus, with local adjacencies depicted in Figure 1.21. The topology of this configuration and its relations to the torus will be discussed in more detail in Section 3.8 on page 157.

Limitations of the Probe Tone Experiment as a Psychological Reference In the psychological experiments (Krumhansl [1990], pp. 21), probe tone ratings are derived from group A, subjects with strong musical background (on average 7.4 years of music instruction), differ significantly from the profiles derived from group B, subjects with less musical education (5.5 years of musical instruction) and even more from group C, subjects with very little education (0.7 years of musical instruction, Krumhansl [1990], p. 22–24). Whereas subjects of group A give high ratings to the diatonic notes, especially the fifth, the ratings of diatonic notes given by naive listeners, the subjects of group C, are inversely related to the distance to the keynote measured in chromatic steps on the pitch class circle. We suspect that subjects of group A partly recognize intervals, apply their musical knowledge, and judge the probe based on their explicit knowledge of the prominence of the scale tone. Nevertheless, the probe tone rating method seems to be a general concept for the description of tonal structures, applicable to non-Western music as well. Probe tone ratings can be employed successfully to describe pitch organization in Indian music [Castellano et al., 1984]. Another way to investigate pitch use with psychological methods is based on a betting paradigm: The listener has to guess, how a musical segment is going to be continued, i.e. what notes are going to be played. In contrast to the mere grading in the probe tone experiment, it involves more complex considerations on the part of the listener. The listener is encouraged to employ all his musical skills to solve the task. Coupled with a prize to win it can further motivate the subject in a playful manner. [26] Beyond the intensities of pitch classes, research of Auhagen [1994] and Huovinen [2002] indicates the impact of sequentiality, beat weight, final intervals, border notes, chords as frames of orientation for tonality perception. However, in this work we limit ourselves to the aspect of accumulated pitch class intensities.

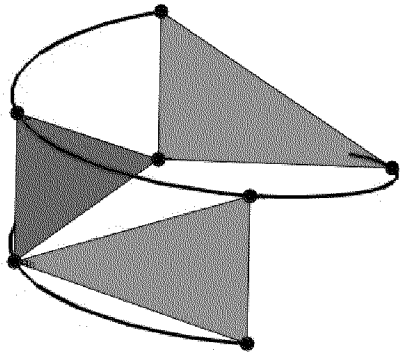
1.2.10 Spatial and Planar Views of Tones, Triads, and Keys

After separately visualizing tones and tone centers in various spatial and planar models, we would like to embed distinct musical entities, i.e. tones, triads, and keys, in the same plane or space, thereby viewing their inter-relations. In particular, we will discuss the models by Chew [2000] and Lehrdahl [2001]. Concerning the notation, not entirely consistently throughout this thesis, we will mark major keys by capital letters, minor keys by small letters (or capital letters with an appended "m"), and tones by small italics.

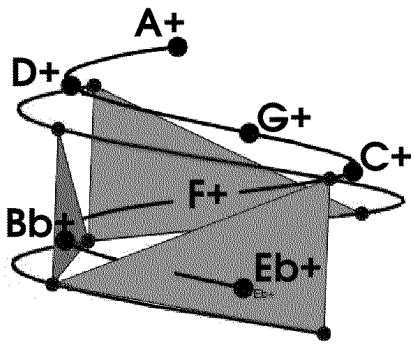
[26] Personal communication with David Huron.



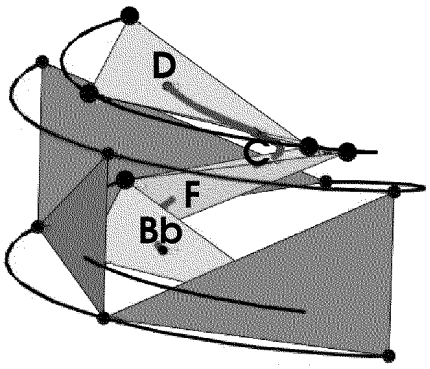
(a) Tones



(b) Triad Triangles



(c) Triad Spiral



(d) Key Spiral

FIGURE 1.22: Chew [2000]'s model of inter-dependence of tone, chord and key: (a) Tones spiral up along the circle of fifths. (b) The three tones contained in the basic triad of each harmonic function, tonic, mediant, and dominant, are connected, thereby forming a triangle for each function. (c) The center of mass of the triangle defines the locus of the triads (note names with extension "+", forming another spiralling trajectory in the 3-dimensional space. (d) The centers of gravity of the three basic functional triads form additional triangles. The centers of gravity of these triangles again constitute an ascending helix, the trajectory of major keys. (The figure is adopted from an illustration by Thomas Noll.)

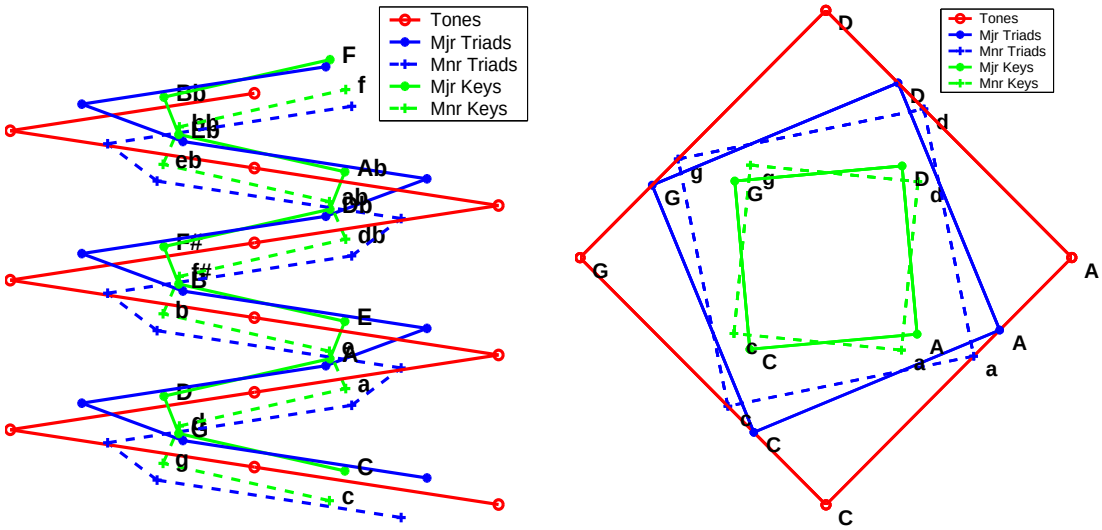


FIGURE 1.23: Side (a) and top (b) view of Chew [2000]'s helix. In contrast to Figure 1.22, points are connected by staight lines instead of smooth ones. In (a), the helix is related to a torus by gluing upper and lower side together. In (b), the first four tones, triads, and keys are depicted in the top view, the X-Y-projection. Major and minor keys are closer together than in Figure 3.51 on page 161, a configuration derived from self-organizing Chopin profiles. Compare this picture to the alternative biplot of pitches and keys in Figure 3.37 on page 142, resulting from correspondence analysis.

In Figure 1.22 on the preceding page, we depict the construction principle of Chew [2000]'s three-dimensional model. It is derived from the tonnetz (Section 1.2.1 on page 45, Figure 1.8, cf. Euler [1739]; Lewin [1987]). Tones are lined up on a helix along the circle of fifths, circular in the X-Y-plane and elevating in the Z-direction. For a triad composed of three tones, construct the triangle whose vertices are provided by the tones constituting the triad. Then the triad is represented by the center of gravity of the triangle. In the same way a key is represented as the center of gravity of the triangle whose vertices are the points identified with the three main triads (tonic, dominant, and subdominant) of the key. Finally we observe three nested spirals, of tones, triads, and (major) keys, escalating in fifths. The circle of fifths, curling around the tube in Figure 1.17 (a), can be identified with the innermost spiral of major or minor keys in Chew's model. We also show side and top view in Figure 1.23. For more detail see the comparison with data-driven geometric models in Section 3.8.4 on page 164. For technical details cf. Appendix C.3.3.

Lehrdahl [2001] presents five nested circles in a plane. They represent a hierarchy of tone sets. From the outermost to the innermost circle the referred musical entities are: first, the chromatic scale tones, second, the diatonic scale tones, third, the basic triad, fourth, the basic fifth, and fifth, the keynote identified with the center of the

circles. By adding up the instances of a pitch class in all circles he arrives at a profile similar to the probe tone rating (cf. Figure 1.19 on page 62).

We will consider embeddings of keys and pitch classes displayed in biplots in Figures 3.37 on page 142 and 3.55 on page 165.

1.2.11 Sequences of Tone Centers

Sequentiality and temporal development are crucial aspects of music. Up to now, we have limited our discussion to planar and spatial views on static profiles of pitch classes and of tone centers, neglecting the temporal dimension. What are reasonable time scales for calculating pitch class profiles? What are musically meaningful sequences of tone centers?

Temporal Granularity We can investigate harmonic progression on different time scales, on the level of chords, beats, bars, phrases, sections, or the entire piece. Harrison [1994] (p. 128) refers to varying temporal resolution of harmonic function by using the term "segmental analysis with varying thickness". According to the Schenkerian school pieces are viewed at different reduction stages. After a preliminary reduction, further reduction stages comprise foreground, middle ground and background level. Also Leman [1995]'s tone center attractor dynamics includes different levels of temporal resolution.[27] The consideration of long time windows enables the reinterpretation of tone center progression from a looking-back perspective. This implies viewing the musical structure on different levels of temporal resolution. In the analysis of harmonic content, the temporal granularity ranges between two extremes: harmonic analysis of single chords without considering the context and finding the key of the entire piece. Harrison [1994] (p. 129) cites an analysis of Hermann Grabner in which a harmonic function is assigned to every chord. This kind of analysis granularity level is appropriate for passages that contain very short tonicizations. (cf. next paragraph) On the other hand, according to the principle of *monotonality*, every tonal piece has a key as a whole, no matter how far modulations lead away from that key. [Schönberg, 1969] For a piece of music this is the coarsest temporal resolution. In Section 3.9.4 on page 174 we will provide tone center analysis on multiple levels of temporal resolution.

Modulation In Western major/minor tonality tone centers normally do not follow one after the other in an arbitrary fashion. Changes from one tone center to another require particular musical preparation and continuation. A *modulation* is

> a firmly established change of key, as opposed to a passing reference to another key, known as a *tonicization*. The scale or pitch collection and characteristic harmonic progressions of the new key must be present. (Saslaw [2004], emphasized by the author)

[27] E.g. Leman [1995], Equation 9.8, p. 130, and Equation 9.9, p. 131.

Thereby the significance of the pitch class set is emphasized.

> [In a modulation] there will usually be at least one cadence to the new tonic. Firm establishment also depends on these features having a certain duration. [Saslaw, 2004]

There are the following variants of modulations:

> "incidental" or "arbitrary", involving brief chromaticism ..., "passing", involving longer references to the new key but no cadence, and "formal", requiring a cadence in the new key ... Techniques of modulation often involve *pivot chords*, that is, chords common to the original key and the new key which can provide a transition between the two ... The choice of a pivot chord or chords depends on the range of pitches and chords held in common between the original key and the new key. ... The closer two keys are on the circle of fifths ... the more pitches they have in common, and the larger the repertoire of available pivot chords. [Saslaw, 2004]

For more distant keys even a single pitch can be used as a *pivot note*. A *sequential* or straightforward restatement of a phrase in a different key can be executed without a pivot chord.

What types of modulations are there? In *diatonic modulation* all main chords (I, IV, V) are frequently used as pivot chords, although their suitability is debated. [Saslaw, 2004] In *chromatic modulation* by upward alteration of the keynote or by downward alteration of the fifth of the tonic we can reach the tonic, dominant, and subdominant function chord of the objective key. Chromatic modulation of such kind allows the composer to reach any key within no more than two consecutive modulations. If we allow downward / upward alteration on prime, third, and fifth we can reach any of the 24 major/minor keys immediately. In addition, *tone central modulation*, *mediants* with no tone in common, *tonal shift*, and *sequences* (Schneider [1987], p. 171–80) are used to reach remote keys.

Schönberg [1966] describes a modulation as a three part process, consisting of neutral, fundamental, and cadential scale degrees. Mazzola [1990] (p. 202, 206) formalizes this idea in the modulation theorem. He applies the theorem to Beethoven's Hammerklavier-Sonate. The automorphism group built on the tones of the diminished seventh chord on $c\sharp$ separates the tone centers of the piece into two orbits. These two groups of tone centers describe the parts of the piece that are circumscribed by the world/anti-world metaphor [Ratz, 1973]. However, the diminished seventh chord is very popular during long phases in music history, so it is not so significant that it plays a special role in the Hammerklavier-Sonate.

Although of great interest, the sequential and temporal aspects of music are beyond the main focus of this thesis. However, covering some auxiliary results, Section 3.9 is devoted to temporal aspects such as tone center transitions.

2 Methods – Signals, Ear Models, Computer Learning

Some issues introduced and discussed in the Introduction and in Section 1 will reappear in this section, covering mathematical and algorithmic details of auditory signal processing, including an in-depth evaluation of autocorrelation for pitch detection, and machine learning.

Most of this chapter is compiled and revised from previous publications: Section 2.1 [Purwins et al., 2000b], 2.2.1 [Purwins et al., 2000a], 2.2.2 [Normann, 2000], 2.3.1 and 2.3.3 [Purwins et al., 2004a], 2.3.2 and 2.3.4 [Purwins et al., 2004b].

2.1 Constant Q Transform

After constructing the constant Q transform from Fourier filters we will describe its fast calculation exploiting sparseness.

2.1.1 Deriving Filter Parameters

To derive the calculation of the *constant Q (CQ-) transform* of some sequence $\mathbf{x}$, we begin with an inspection of the familiar equation of a Fourier filter at z

$$\sum_{n<N} x[n]\, e^{-2\pi i n z/N}. \tag{2.1}$$

Each component of the constant Q transform, in the sequel called *CQ-bin*, will be calculated as such a filter, but suitable values for z and window length N have to be found in order to match the properties discussed above. The bandwidth of the filter (2.1) is $\Delta_z^{\mathrm{ft}} = f_s/N$ (f_s denotes the sampling rate) independently of z. Thus the desired bandwidth $\Delta_k^{\mathrm{cq}} = f_k/Q$ can be realized by choosing a window of length $N_k = \frac{f_s}{\Delta_k^{\mathrm{cq}}} = Q\frac{f_s}{f_k}$. The frequency to resolution ratio of the filter in Equation 2.1 is $\frac{f_z}{\Delta_z^{\mathrm{ft}}} = z$. To achieve a constant value Q for the frequency to resolution ratio of each CQ-bin one has to set $z := Q$. Thus for integer values Q the k-th CQ-bin is the Q-th DFT-bin with window length $Q\frac{f_s}{f_k}$.

Summarizing, we get the following recipe for the calculation of a constant Q transform: First choose minimal frequency f_0 and the number of bins per octave b according to the requirements of the application. The maximal frequency $f_{\max}$ only affects

the number of CQ-bins to be calculated [1]:

$$K := \ulcorner b \cdot \log_2(\frac{f_{\max}}{f_0}) \urcorner, \tag{2.2}$$

$$Q := (2^{\frac{1}{b}} - 1)^{-1}, \tag{2.3}$$

and for $k < K$ set

$$N_k := \ulcorner Q\frac{f_s}{f_k} \urcorner, \tag{2.4}$$

$$x^{\mathbf{cq}}[k] := \frac{1}{N_k} \sum_{n<N_k} x[n]\ w_{N_k}[n]\ e^{-2\pi i n Q/N_k}. \tag{2.5}$$

To reduce spectral leakage (cf. Harris [1978]), it is advisable to use the filter in conjunction with some window function: $\langle w_N[n] \ : \ n < N \rangle$ is some analysis window of length N. Following Brown and Puckette [1992], we use Hamming windows.

2.1.2 An Efficient Algorithm

Since the calculation of the constant Q transform, according to Equation 2.5, is very time consuming, an efficient algorithm is highly desirable. Using matrix multiplication, the constant Q transform of a row vector $\mathbf{x}$ of length N ($N \geq N_k$ for all $k < K$) is

$$\mathbf{x}^{\mathrm{cq}} = \mathbf{x} \cdot \mathbf{T}^* \tag{2.6}$$

where $\mathbf{T}^*$ is the complex conjugate of the temporal kernel[2] $\mathbf{T} = (t_{nk})_{\substack{n<N \\ k<K}}$

$$t_{nk} := \begin{cases} \frac{1}{N_k}\, w_{N_k}[n]\ e^{2\pi i n Q/N_k} & \text{if } n < N_k \\ 0 & \text{otherwise.} \end{cases} \tag{2.7}$$

Since the temporal kernel is independent of the input signal $\mathbf{x}$ one can speed up successive constant Q transforms by precalculating $\mathbf{T}^*$. But this is very memory consuming and since there are many non vanishing elements in $\mathbf{T}$ the calculation of the matrix product $\mathbf{x} \cdot \mathbf{T}^*$ still takes quite a while.

For improving the calculation, Brown and Puckette [1992] carry out the matrix multiplication in the spectral domain. Since the windowed complex exponentials of the temporal kernel have a DFT that vanishes almost everywhere except for the immediate vicinity of the corresponding frequency the spectral kernel

$$\mathbf{S} = (s_{nk})_{\substack{n<N \\ k<K}} := \mathrm{DFT}(\mathbf{T}) \quad \text{(1-dim. DFTs applied columnwise)} \tag{2.8}$$

[1] $\ulcorner x \urcorner$ denotes the least integer greater than or equal to x.

[2] In (2.7) we center the filter domains on the left for the ease of notation. Right-centering is more appropriate for real-time applications. Middle-centering has the advantage of making the spectral kernel (2.8) real.

is a sparse matrix (after eliminating components below some threshold value). This fact can be exploited for the calculation of $\mathbf{x}^{\mathrm{cq}}$ owing to the identity

$$\sum_{n<N} x[n]\ y[n]^* \ = \ \frac{1}{N} \sum_{n<N} x^{\mathrm{ft}}[n]\ y^{\mathrm{ft}}[n]^* \tag{2.9}$$

where $\mathbf{x}$ and $\mathbf{y}$ are sequences of length N and $\mathbf{x}^{\mathrm{ft}}$, $\mathbf{y}^{\mathrm{ft}}$ denote their unnormalized discrete Fourier transform. Applying this identity to the equation of the constant Q transform (2.5), using definitions (2.7) and (2.8) yields

$$x^{\mathrm{cq}}[k] = \frac{1}{N} \sum_{n<N} x^{\mathrm{ft}}[n]\ s^*_{nk}$$

or, equivalently, in matrix notation

$$\mathbf{x}^{\mathrm{cq}} = \frac{1}{N}\ \mathbf{x}^{\mathrm{ft}} \cdot \mathbf{S}^*. \tag{2.10}$$

Due to the sparseness of $\mathbf{S}$ the calculation of the product $\mathbf{x}^{\mathrm{ft}} \cdot \mathbf{S}^*$ involves essentially less multiplications than $\mathbf{x} \cdot \mathbf{T}^*$.

The Fourier transforms that arise in the efficient algorithm should of course be calculated using FFTs. To this end, N is chosen as the lowest power of 2 greater than or equal to N_0 (which is the maximum of all N_k). The calculation of the spectral kernel is quite expensive, but having done this once all succeeding constant Q transforms are performed much faster.

2.2 Auditory Models and Autocorrelation

2.2.1 Auditory Models

Lyon and Shamma [1996], p. 227, give a general characterization of auditory models:

> All models, however, can be reduced to three stages: analysis, transduction, and reduction. In the first stage, the unidimensional sound signal is transformed in a distributed representation along the length of the cochlea. This representation is then converted in the second stage into a pattern of electrical activity on thousands of auditory nerve fibers. Finally, perceptual representations of timbre and pitch are extracted from these patterns in the third stage.

In this three-level description, analysis refers to linear processing. Transduction and reduction are non-linear processes.

With reference to biology, the linear part concerns the outer and middle ear, and the basilar membrane. Transduction refers to the non-linear transformation by the inner hair cell, and in a broader sense, the stochastic transmission at the auditory

nerve fiber. Finally, the "mechanical vibrations along the basilar membrane are transduced into electrical activity along a dense, topographically ordered array of auditory nerve fibers." [Lyon and Shamma, 1996]

The physiological correlate of the reduction stage is understood very poorly. Auditory models are quite speculative at that stage. In using embedding and clustering algorithms (cf. Section 2.3.4 and Appendix C.3), we can claim biological plausibility only within narrow bounds, e.g. we can mimic the auditory principle of tonotopy.

Both outer and middle ear combined show a response curve maximal at between 1 and 3 kHz and decreasing towards lower and higher frequencies. [3] For imitating various pitch phenomena the effect can be implemented as a band pass filter. [4] The inner ear contains the basilar membrane, with hair cells adjoined to it.

Non-linear Transduction

Within a neuron, voltage pulses passively propagate along a dendrite until they reach the axon hillock. In the axon hillock all incoming voltage pulses accumulate until they cross a threshold. As a consequence, a voltage pulse (*spike*) is generated. The spike propagates across the axon until it reaches a synapse. Neurotransmitters fill the vesicles of the presynaptic axon (Figure 2.1, left). The vesicles release the neurotransmitters into the synaptic cleft, when triggered by the incoming spikes, and the emitted neurotransmitters then enter the receptor channels of the postsynaptic dendrite of the receiving cell.

The major properties of the *hair cell* in the inner ear are temporal coding and adaptation behavior. According to Meddis and Hewitt [1991], the synapse of the hair cell can be described as a dynamic system consisting of four elements (Figure 2.1, right). In this model, the activity transmitted by the hair cell to the auditory nerve is assumed proportional to the number of neurotransmitters $c[t]$ in the synaptic cleft. $c[t]$ depends on the number of transmitters $q[t]$ in the hair cell multiplied by a non-linear function

$$k[t] = \begin{cases} \frac{gd[t](x[t]+A)}{x[t]+A+B} & \text{for } x[t]+A > 0 \\ 0 & \text{for } x[t]+A < 0 \end{cases} \tag{2.11}$$

where g, A, B are suitable constants (cf. Table 3.8.1). $k[t]$ describes the permeability of the presynaptic hair cell membrane and is triggered by the presynaptic membrane potential. The closer $q[t]$ is to the maximum capacity m, the less neurotransmitter is produced. A portion of the transmitters (factor r) returns from the synaptic cleft into the presynaptic hair cell. The temporal behavior of the system is described by a non-linear first order differential equation. The change in parameter is calculated from the balance of the incoming and outgoing quantities:

[3] Meddis and Hewitt [1991], p. 2868, Figure 2. Similar filter curves are motivated by different psychoacoustic phenomena, such as hearing threshold [Yost and Hill, 1979] or dominance region [Terhardt et al., 1982].

[4] IIR filter, Oppenheim and Schafer [1989], of second order, Meddis and Hewitt [1991], Footnote p. 2881.

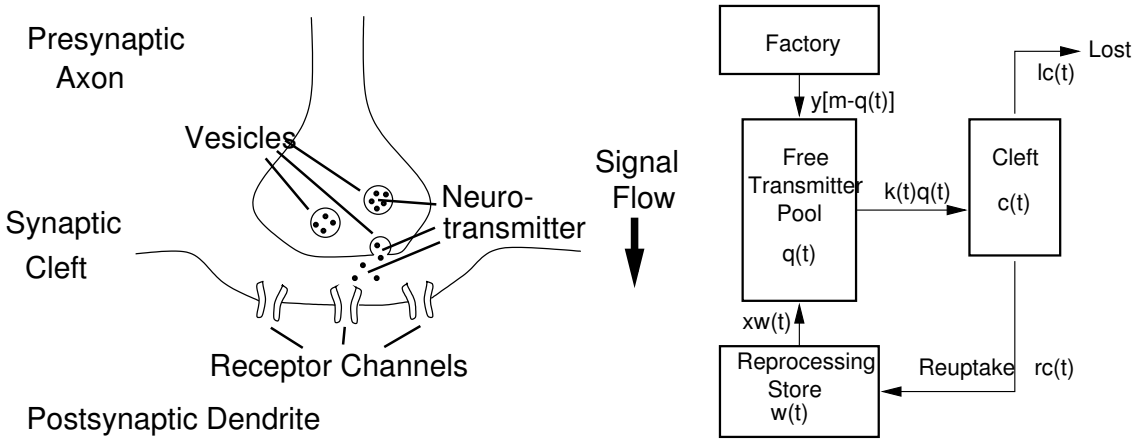


FIGURE 2.1: Information is transmitted from the presynaptic to the postsynaptic cell (left). The action potential in the presynaptic cell forces the vesicles to empty neurotransmitters into the synaptic cleft. The emitted neurotransmitters enter the receptor channels and change the potential of the postsynaptic cell. The hair cell synapse is modeled by a dynamic system consisting of four stages (right, after Figure 4, Meddis and Hewitt [1991], p. 2869). The non-linear function $k[t]$ includes the presynaptic action potential and controls the permeability of the presynaptic membrane. The evoked potential in the auditory nerve scales with $c[t]$. (cf. Section 2.2.1)

$$\frac{\Delta q}{\Delta t} = y(m - q[t]) + mw[t] - k[t]q[t], \tag{2.12}$$

$$\frac{\Delta c}{\Delta t} = k[t]q[t] - lc[t] - rc[t], \tag{2.13}$$

$$\frac{\Delta w}{\Delta t} = rc[t] - xw[t]. \tag{2.14}$$

This hair cell model reproduces some experimental results from hair cells of gerbils, in particular: (i) Frequent spikes occur with the onset of a tone. The spike rate decreases to a constant value, when the tone continues, thereby revealing adaptation behavior. After the offset of the tone, it decreases to about zero. (ii) Below 4-5 kHz spikes occur in the hair cell nearly exclusively during the positive phase of the signal (*phase locking*). In this range, thus, frequency is coded both by the position of the responding hair cell on the basilar membrane and by temporal spike behavior. For frequencies above 5 kHz spikes occur about equally often during the positive and the negative phase of the signal. Therefore above 5 kHz frequency is only coded by the place information on the basilar membrane.

After mapping frequency on a critical band rate (Bark) scale, masking curves can be approximated by triangular functions. Gammatone filters realize a fairly good approximation of the filter shape. More complex masking templates are suggested in Fielder et al. [1995]. Additive properties of simultaneous maskers is an ongoing research topic.

Information Coding Hypotheses

There are three major hypotheses which attempt to explain how a sequence of spikes encodes information: (i) by exact neuron firing times, (ii) by the time interval between proceeding spikes, the inter spike interval, and (iii) the spike rate, the inverse of the mean inter spike interval. To precisely model (i) and (ii) we could solve a system of non-linear differential equations [Hodgkin and Huxley, 1952] describing current flow in the axon. As a simplification, we can make use of the "integrate and fire" model [Maass, 1997]. Voltage is integrated until threshold is reached. After a refractory period, integration starts again from rest potential. (iii) is a rough simplification of spike behavior. Nonetheless, (iii) is the basis of the auditory image (cf. p. 25) and the connectionist neuron used in multilayer feedforward networks in artificial neural networks (cf. e.g. Haykin [1999], p. 156 f.).

Hebbian Learning and Tonotopy

If presynaptic and postsynaptic electrical activity occur synchronously, the postsynaptic receptor channels become more permeable so that a presynaptic activity evokes stronger activity at the postsynaptic dendrite. This principle is called *Hebbian learning*.

According to the principle of *tonotopy* proximate hair cells on the basilar membrane project to proximate neurons in the central nervous system. In computer science the tonotopic principle can be implemented by algorithms such as the self-organizing feature map (SOM, cf. Section 2.3.4, Kohonen [1982]).

Meddis Auditory Model

In Meddis [1988] the temporal signal undergoes the following transformations:

1. bandpass filter (outer and middle ear)
2. gammatone filter bank scaled according to ERB (basilar membrane)
3. per channel: non-linear transduction (inner hair cell, auditory nerve)
4. per channel: autocorrelation
5. summation of autocorrelation

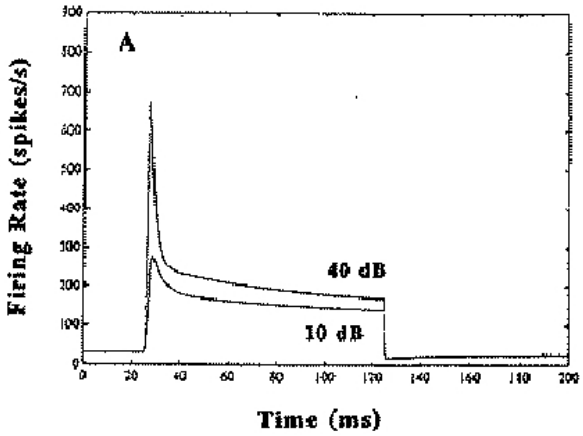


FIGURE 2.2: A hair cell model (Meddis [1988], cf. Section 2.2.1) responds with an impulse to stationary sine wave signals with sharp onset and the indicated sound pressure levels. Shortly after the onset, the firing rate decreases to a lower level. Throughout the duration of the signal, the firing rate remains at that level [O'Mard et al., 1997].

An onset of a sound signal evokes a sharp increase in firing rate. If the signal persists, the firing rate relaxes to a constant level. This spike pattern is reproduced in Meddis [1988]'s hair cell model, as shown in Figure 2.2. The final sum vector, the *pitch vector*, encodes the perceived pitch.

2.2.2 Pitch Detection Algorithms

We have discussed the phenomenon of virtual pitch in Section 1.1.6. How is it possible to design an algorithm that can capture the perceived pitch, e.g. of a residual series? We will give an account of research related to Terhardt et al. [1982]'s algorithm, before presenting the latter. Finally for complex tones, we will compare autocorrelation with and without preprocessing by an auditory model.

Frequency Differences Between Partials

Whitfield [1970] considers pitch to be a process of pattern recognition. Activation patterns are learned. Usually, natural sounds contain the fundamental. The activation pattern of a residual series differs only slightly from the activation patterns of "complete" overtone series. Due to the pattern similarity the same pitch is heard.

Consider a residual series composed of partials with frequencies 1800, 2000 and 2200 Hz. A pitch of 200 Hz is perceived. The components of the residual series are recognized as being the 9th, 10th, and 11th partial of 200 Hz. These models cannot account for the pitch of a shifted residual series (cf. Figure 1.5). Let us look at a residual series shifted 40 Hz up. According to this model we would determine the pitch to be 40 Hz. 1840, 2040, and 2240 Hz would be considered the 46th, 51st, and

56th partial of 40 Hz. In contrast, the experiment shows that we actually hear 204 Hz (in some instances 184 Hz or 227 Hz are perceived).

Walliser [1968] overcomes this difficulty by emphasizing the frequency distance between partials. He considers the frequency of the lowest partial as an additional factor. According to Walliser [1968], the virtual pitch is the subharmonic of the lowest partial that is closest to the frequency distance between partials. To be more precise, let f_l be the frequency of the lowest partial and Δf the frequency distance of neighboring partials. Pitch is then calculated according to $f = \frac{f_l}{m}$ $(m = 1, 2, \ldots)$ so that $|f - \Delta f|$ is minimal. In our example this is 204.4 Hz. 1840 Hz is the 9th harmonic of 204.4 Hz. Terhardt [1992] modified this principle further by limiting the partials that can be considered (cf. Figure 1.4). Terhardt [1974] considers subharmonics of all partials within the existence region. In Terhardt [1974], virtual pitch is the pitch that can be considered a subharmonic of most partials. In shifted residual series, we do not yield exactly the actually perceived pitch. But the subharmonics of the partials cluster around the perceived pitch.

For a given residual series, Goldstein [1973]'s "optimum processor" determines the best matching overtone series in the sense of minimizing mean squared error. Virtual pitch is the fundamental of the overtone series.

A similar approach is the "harmonic sieve" [Duifhuis, 1982; Scheffers, 1983]. To each pitch we assign a sieve. Frequencies can only pass the sieve, when they are proximate to the harmonics of the fundamental of that sieve. The pitch is determined by the sieve that lets pass most of the frequencies.

A problem occurs in all these models: To each overtone series with fundamental frequency f there is another one with fundamental frequency $\frac{f}{2}$. Let us look at the harmonic sieve. All frequencies that pass the first sieve also pass the second one. Different approaches are given to uniquely determine virtual pitch: limiting the considered pitch range [Goldstein, 1973], limiting the number of partials considered, and "harmonic summation" [Cohen et al., 1995]. In the latter the weighting of subharmonics is inspired by Terhardt [1974].

Calculation of Virtual Pitch

According to Terhardt [1998] (p. 362-366), we calculate virtual pitch from spectral weights and subharmonic coincidence detection, as shown in Figure 2.3. A *spectral weight* S_i for frequency component f_i with incisiveness L_{X_i} (cf. Equations 1.8-1.12) is calculated as

$$S_i = \begin{cases} \frac{1 - e^{\frac{-L_{X_i}}{15\,\mathrm{dB}}}}{\sqrt{1 + 0.07(\frac{f_i}{700\,\mathrm{Hz}} - \frac{700\,\mathrm{Hz}}{f_i})^2}} & \text{for } L_{X_i} \geq 0 \\ 0 & \text{for } L_{X_i} < 0 \end{cases} . \tag{2.15}$$

The denominator models the phenomenon of dominance regions (Section 1.1.6, Figure 1.4) in virtual pitch. The coincidence C_{ij}^{mn} of the m-th subharmonic of the i-th

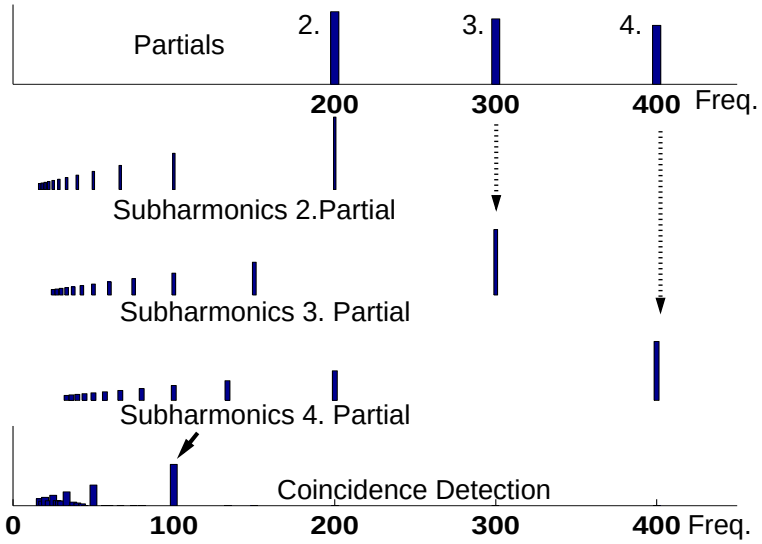


FIGURE 2.3: Virtual pitch as recognition of regularities in the frequency pattern of the partials. A harmonic complex tone with missing fundamental frequency $f_0 = 100$ Hz consists of partials with frequencies $2f_0, 3f_0, 4f_0$. For each partial f_i and spectral weight S_i a subharmonic series H_i with frequencies $\frac{f_i}{m}$ is constructed. In the further step of coincidence detection, the accumulation of the subharmonic series leads to the spectrum of virtual pitch. Its maximum (arrow) presents the (most prominent) virtual pitch.

spectral pitch with the n-th subharmonic of the j-th spectral pitch is determined as follows:

$$C_{ij}^{mn} = \begin{cases} \sqrt{\frac{S_i S_j}{mn}}(1 - \frac{\gamma}{\delta}) & \text{for } \gamma \leq \delta \\ 0 & \text{for } \gamma > \delta \end{cases} \tag{2.16}$$

with [5]

$$\gamma = |\frac{nf_i}{mf_j} - 1| \ , \quad n = \lfloor \frac{mf_j}{f_i} + 0.5 \rfloor. \tag{2.17}$$

The weight W_{im} of the virtual pitch referring to the m-th subharmonic of the i-th spectral pitch for a total of k spectral pitches holds:

$$W_{im} = \sum_{j=1, j \neq i}^{k} C_{ij}^{mn}. \tag{2.18}$$

[5] $\lfloor x \rfloor$ is the highest natural number smaller than x

Virtual Pitch and the Proximity Principle

The spectrum of virtual pitch provides a much more appropriate domain to determine pitch proximity than the ordinary spectrum. Let us compare two tones: a harmonic complex tone with a missing fundamental and a sine tone of the same frequency. We then compare both tones by applying a common dissimilarity (e.g. Euclidean distance, cf. Section C) to their spectrum. This yields a high degree of dissimilarity. This opposes the perception that the pitches of both tones are identical (cf. quote on p. 32). But if we apply the distance to the spectrum of virtual pitch, the dissimilarity between both tones is low. This is consistent with the perceived pitch. Thereby the spectrum of virtual pitch can be seen as an implementation of the gestalt principle of proximity with respect to pitch perception.

Pitch and Timbre

There are several approaches to formally describing timbral features. The helix model (Section 1.2.3, Figure 1.9) decomposes pitch into pitch class and brightness. We would like to have some measures for the latter.

For a complex tone, the *harmonic spectral centroid* [Peeters et al., 2000] for frequency components f_i and intensities I_i is given by:

$$\mu_s = \sum_i f_i I_i. \tag{2.19}$$

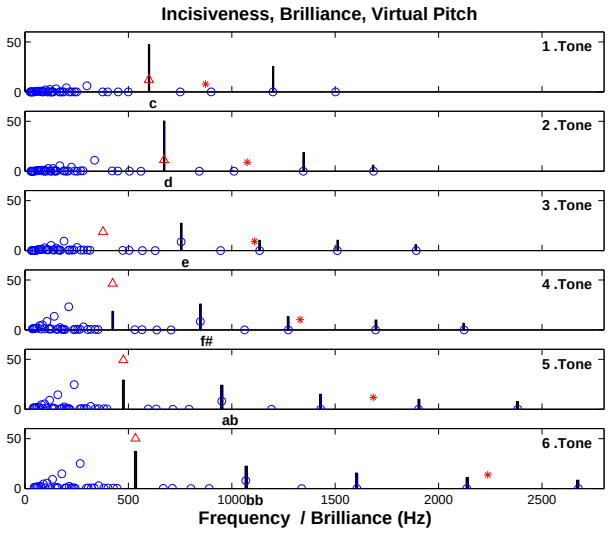


FIGURE 2.4: Virtual pitch of the stimuli used in the experiment in Section 3.2.2: The incisiveness of the partials is displayed by vertical bars. The spectral weights in the calculation of virtual pitch are indicated by "○" (secondary pitch) and "△" (maximum: principal pitch). Brilliance is indicated also by "∗".

In this equation frequency can be measured in Bark or in Hz; instead of the intensity also the loudness or the spectral weight can be considered. We modify Aures [1985] [6] to calculate the more refined concept of *brilliance*:

$$\mu_s' \sim \frac{2.7782}{\ln(1 + 0.05\frac{N}{\text{sone}})} \int_0^{24\ \text{Bark}} L_X(z)\, 0.0165 \cdot e^{\frac{0.171z}{\text{Bark}}}\, dz \tag{2.20}$$

where z is the frequency in Bark, $L_X(z)$ the specific loudness (cf. p. 35), and N the overall loudness integrated across a time window. On the basis of harmonic spectral centroid or brilliance, also spectral spread can be calculated [Aures, 1985; Peeters et al., 2000].

Autocorrelation with and without an Auditory Model

Meddis and Hewitt [1991] review several results on pitch perception. They then try to reproduce these results with an auditory model. We will focus on a couple of psychoacoustic effects that are more relevant with respect to pitch perception in musical instruments. We compare the auditory model to direct autocorrelation with respect to the capability of reproducing selected psychoacoustic effects.

Since in Meddis and Hewitt [1991] only specific cases are considered, we cannot give a statistically significant benchmark for the auditory model and pure autocorrelation. We will demonstrate the capacity of autocorrelation to reproduce the results from a set of remarkable psychoacoustic experiments. [7] Autocorrelation for time lag l at time t_0 within a window of w samples is given by

$$x^{\text{acf}}[l] = \sum_{t=t_0-w+1}^{t_0-l} x[t]x[t+l]. \tag{2.21}$$

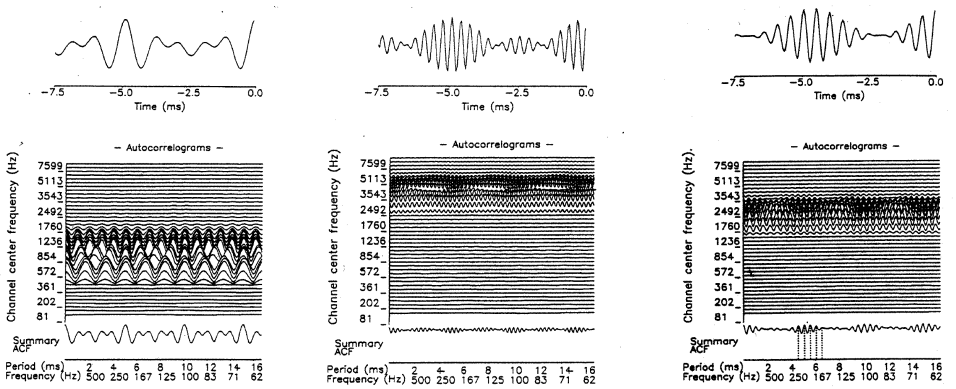
Parameters For determination of the pitch vector of the pure autocorrelation, the input signal is sampled at 22050 Hz. The window length is 512 samples. In contrast to figures from Meddis [1988] there are discontinuities in Figures 2.5 (e) and (f). They are due to short window length. But this window length proves to be sufficient, since the relevant maxima already appear in this setup.

For comparison of pure autocorrelation to the auditory model we use residual series with partials of equal intensity as test signals.

Missing Fundamental of Separately Resolved Partials The signal consists of a residual series of frequency components at 600, 800 and 1000 Hz. So the missing fundamental (cf. Section 2.2.2) has frequency 200 Hz. All partials lie in the existence region (cf. Figure 1.4).

[6] Cited in Terhardt [1998], p. 303, Equations (10.11), (10.18), and (10.19).

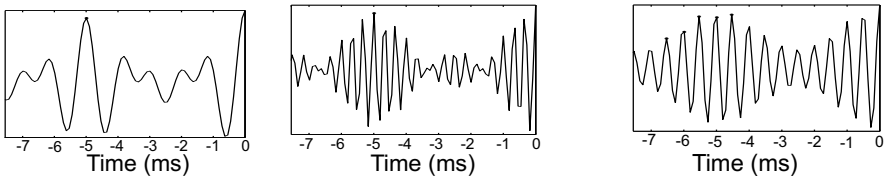
[7] Since all transformation steps are explicitly stated in equations, we could have discussed the reproducibility of the data analytically. But the non-linear transduction presents a difficult obstacle. We thus restrict ourselves to some numerical demonstrations.



(a) Auditory model: resolved harmonics

(b) Aud. mod.: harmonics not res.

(c) Auditory model: ambiguity



(d) Direct ACF: resolved harmonics

(e) Direct ACF: harmonics not resolved

(f) Direct ACF: ambiguity

FIGURE 2.5: Comparison of pitch detection by an auditory model (a, b, c from Meddis and Hewitt [1991], p. 2871 f., Figure 6, 7, 8; d, e, f adopted from Normann [2000]) and by direct autocorrelation (d, e, f). In (a, d) the stimulus is a residual series of frequency components 600, 800, and 1000 Hz (top in a). The partials are resolved into different channels (middle in a). Both the auditory model (summary ACF in a) and the direct autocorrelation (d) have a maximum at 5 ms, indicating a missing fundamental (200 Hz). In (b,e) the stimulus is a complex tone consisting of frequencies 2800, 3000, and 3200 Hz (top in a). Not all partials are resolved into different channels (middle in a). Again both the auditory model (summary ACF in b) and the direct autocorrelation (e) detect the missing fundamental at 200 Hz by having a maximum at 5 ms. The signal consists of of 2189, 2388, 2587 Hz (top in c). It is perceived ambiguously [Schouten et al., 1962]. We consider local maxima close to the maximum in the auditory model (bottom in c) and direct autocorrelation to be candidates for the various perceived pitches in the experiment [Schouten et al., 1962]. The perceived pitches do not coincide with the 5 highest local maxima, since the local maximum at 4 ms in (f) is higher than the one at 6.5 ms (cf. Table 2.1).

Detection Method	*Frequencies (in* Hz*)*				
Perception	145	160	178	198	220
Auditory Model	153	166	181	199	221
Direct ACF	153	167	182	200	218

TABLE 2.1: Five different pitches (first row) are perceived in the pitch ambiguity experiment by Schouten et al. [1962]. These pitches are matched to local maxima closest to the absolute maximum in the pitch vector of the auditory model (second row) and direct ACF (third row). Cf. Figure 2.5 (c) and (f).

Figure 2.5 (a) shows the autocorrelogram generated by the auditory model. The partials are resolved in different channels. The sum vector displays a clear maximum at 5 ms. This is exactly the frequency of the missing fundamental.

As we would expect based on the theoretical results, the pure autocorrelation of the signal delivers the same result. Around the maximum, the pitch vector from direct autocorrelation (cf. Figure 2.5 (d)) does not significantly differ from the auditory model.

Missing Fundamental of Non-Resolved Partials In this example the signal consists of partials with frequencies 2800, 3000, 3200 Hz. The missing fundamental has a frequency of 200 Hz. Not all frequencies are still in the existence region (cf. Figure 1.4). Partials are not resolved into separate channels, as the autocorrelation shows in Figure 2.5 (b). Here again, the pitch vector does not differ significantly from the one produced by pure autocorrelation in Figure 2.5 (e). The maximum again occurs at 5 ms.

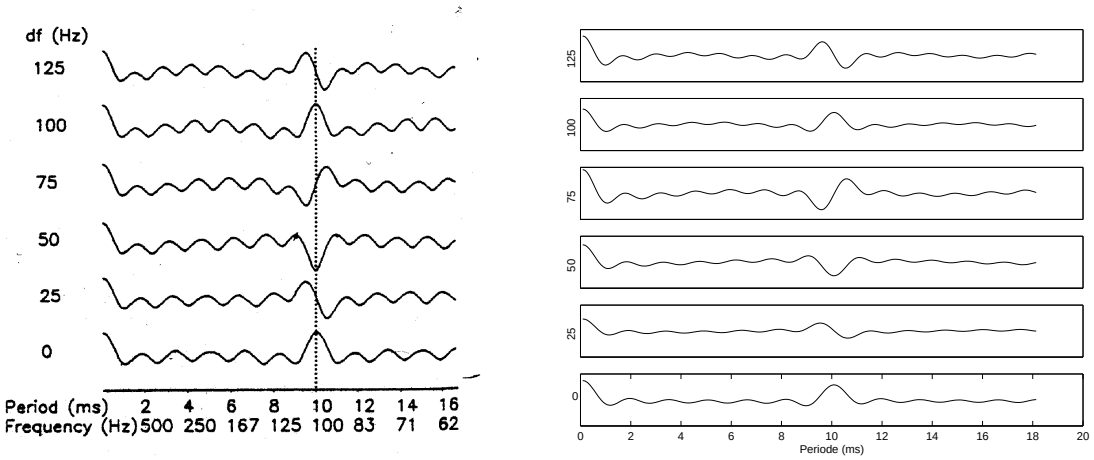
Ambiguous Pitch Schouten et al. [1962] conduct a psychoacoustic experiment in order to explore the ambiguity of pitches in a complex tone that consists of the 11th, 12th and 13th component of a missing fundamental of frequency 199 Hz. The subjects are asked to adjust a reference tone so that both have the same pitch. The results are listed in Table 2.1

The pitch vectors of this tone are displayed in Figure 2.5 (c) (auditory model) and Figure 2.5 (f) (direct autocorrelation). Local maxima proximate to the absolute maximum are considered candidates for a possible pitch perception (we will call them pitch of the pitch vector in the sequel). In Meddis [1988] the figure and its caption do not give information on whether the five maxima are used for determination of different pitches. In direct autocorrelation this is not the case: the maximum at 4 ms is obviously higher than the one in 6.5 ms. In Table 2.1 we list the pitches closest to the one observed in the experiment.

The deviation between auditory model and pure autocorrelation are always below

the just noticeable difference ($\sim$ 3.6 Hz).[8] But the perceived pitch in Schouten et al. [1962] deviates from both the other values more than the just noticeable frequency difference. Meddis [1988] links these effects to non-linearities.

Shifted Residual Series Schouten et al. [1962] explore different shifted overtone series (cf. Figure 1.5). The dependence of virtual pitch is determined by a shift of the overtone series that consists of six harmonics. The fundamental frequency of the original series is 100 Hz. Figure 2.6 (a) (auditory model) and Figure 2.6 (b) (direct autocorrelation) display the pitch vectors of these overtone series. Both the auditory model and the direct ACF give similar results, but the maxima vary excessively in frequency, whereas in Figure 1.5 virtual pitch varies only slightly.



(a) Auditory model and ACF (b) Direct ACF

FIGURE 2.6: Pitch determination of shifted residual scales by autocorrelation with (a, from Meddis and Hewitt [1991], p. 2872, Figure 9; b adopted from Normann [2000]) and without an auditory model (b). The input signal is an overtone series with frequency components 100, 200, 300, 400, 500, 600 Hz. In the lowest row in (a) and (b) the ACF of the original series gives a maximum at 100 Hz corresponding to the perceived pitch of 100 Hz. The upper rows give the ACF of the original series shifted by 25 (second lowest row), 50, 75, 100, and 125 Hz (top row). The position of the maximum is shifted towards higher frequencies. Also it decreases. For a shift of 50 Hz (4th row) the situation is ambiguous. The local maxima right and left from 100 Hz have the same height. For a shift of 75 Hz the maximum below 100 Hz (right from the dotted vertical line) dominates the one above 100 Hz. The process repeats cyclically.

[8] These deviations could be due to rounding errors.

Limited Impact of Auditory Model We have compared auditory models to direct autocorrelation. This gives rise to the conjecture that processing of filter bank and the non-linear transduction in the auditory model do not have a significant impact on the resulting pitch vector, since it varies only slightly from pitch vector generated by direct autocorrelation. As long as we work with signals in the form of overtone series, we can omit the preprocessing and can restrict the processing to direct autocorrelation. The temporal properties of the non-linear transduction are not considered here.

The auditory model's strength is that it can reproduce several pitch phenomena within the framework of physiological constraints.

> The model has the merit of being able to deal with pitch percepts arising from both resolved and unresolved signal components while offering an explanation for the dominance of low frequency harmonics in giving rise to the sensation of low pitch. In so doing, it combines insights from two rival schools, championing, respectively, place and temporal theory. [Meddis and Hewitt, 1991]

Meddis and Hewitt [1991] do not give unequivocal biological evidence of the central component in their model, autocorrelation:

> The biggest conceptual difficulty with the time interval detection approach is knowing how it is achieved physiologically. There is a considerable gap between the knowledge that timing information is available and identifying a nucleus in the nervous system that does the job. [Meddis and Hewitt, 1991]

2.3 Classification, Clustering, Visualization

In this section the theoretical background for classification is outlined. Then *k*-means clustering (Section 2.3.3) and visualization tools, such as correspondence analysis (p. 89), Isomap (p. 91), and the self-organizing feature map (p. 91) are presented.

Supervised versus Unsupervised Learning Supervised learning can be used to classify data according to certain labels (e.g. key, mode, composer, performer). Unsupervised learning lets salient structural features emerge without requiring any assumption or focusing on a specific question like a classification task according to predefined categories. While the performance of supervised methods can be quantified by – more or less – objective measures (cf. subsequent paragraphs) the evaluation of unsupervised analyses is more delicate due to their exploratory nature and the lack of a specific goal that has been fixed beforehand.

Evaluating Classification Algorithms We will discuss how performance of supervised learning can be measured. The concept of cross-validation and the receiver operating characteristic are discussed.

Cross-Validation Machine learning algorithms for classification work in two steps. First the learning algorithm is fed with some labeled data (*training set*) from which regularities are extracted. After that the algorithm can classify new, previously unseen data. In order to validate how well a classification algorithm learns to generalize from given data, e.g., a technique called k-fold cross-validation is applied: the data set is randomly split into a partition of k equally sized subsets. Then the classifier is trained on the data of $k-1$ sets and evaluated on the hold-out set (*test set*). This procedure is done until each of the k sets have been used as test set, and all k error ratios on those test sets are averaged to get an estimation of the *generalization error*, i.e., the error which the classification algorithm is expected to make when generalizing to new data from given training data on which the classifier has been trained before. Of course this quantity greatly depends on the structure and complexity of the data and the size of the training set. To get a more reliable estimate one can also do n-many partitions of the data set and do k-fold cross-validation of each partitioning (*n times k-fold cross-validation*).

Technical note: With our music corpus some care has to be taken when doing cross-validation in order to avoid underestimating the generalization error. Since some pieces exist in versions from various performers, all pieces are grouped in equivalence classes, each holding all versions of one specific piece. But most pieces exist only in one interpretation and form a singleton equivalence class. In cross-validation all pieces of one equivalence class are either assigned to the training or to the test set.

Receiver Operating Characteristic *Receiver operating characteristic (ROC)* analysis is a framework to evaluate the performance of classification algorithms independent of class priors (relative frequency of samples in each class, cf. Provost et al. [1998]). In a ROC curve the false positive rate of a classifier is plotted on the X-axis against the true positive rate on the Y-axis by adjusting the classifier's threshold. In a plot all ROC curves from n times k-fold cross-validation procedure are averaged. To subsume the classification performance in one value the *area under the curve (AUC)* is calculated. A perfect classifier would attain an AUC of 1, while guessing has an expected AUC of 0.5, since false positive rate equals true positive rate in the latter case.

2.3.1 Supervised Learning

Two classifiers, regularized discriminant analysis and support vector machines, prove successful in Section 3.4.1 for determining the composer based on normalized constant Q profiles.

K-Nearest-Neighbor Classifier Be $\mathbf{X}$ an $n \times l$-matrix, with l vectors of dimension n. Be $d(\mathbf{x}, \mathbf{x}')$ a dissimilarity between columns $\mathbf{x}, \mathbf{x}'$ of $\mathbf{X}$ and $\mathrm{rk}_d(\mathbf{x}, \mathbf{x}')$ their respective rank order (Equation C.8). $\mathcal{G}_{\mathbf{x}}(k)$ be the set of k-nearest neighbors, that means, the k

points closest to $\mathbf{x}$ according to dissimilarity d:

$$\mathcal{G}_{\mathbf{x}}(k) = \{\mathbf{x}' \text{ column of } \mathbf{X} : \mathrm{rk}_d(\mathbf{x}, \mathbf{x}') \leq k\}. \tag{2.22}$$

Then we define the number of points $g_{\mathbf{x},i}$ in $\mathcal{G}_{\mathbf{x}}(k)$ that belong to class $\mathcal{C}_i$:

$$g_{\mathbf{x},i} = |\{\mathbf{x}' \in \mathcal{G}_{\mathbf{x}}(k) : \mathbf{x}' \in \mathcal{C}_i\}|. \tag{2.23}$$

In the *k-nearest-neighbor classification rule*, a data point $\mathbf{x}$ is assigned to a class $\mathcal{C}_{i'}$ by

$$i' = \arg\max_i g_{\mathbf{x},i}. \tag{2.24}$$

We assign $\mathbf{x}$ to the class $\mathcal{C}_{i'}$ that has the highest number $g_{\mathbf{x},i}$ of representatives among the k-nearest neighbors $\mathcal{G}_{\mathbf{x}}(k)$ of $\mathbf{x}$.

Regularized Discriminant Analysis The *quadratic discriminant analysis (QDA)* is a classification method which assumes that each class $\mathcal{C}_i$ has a Gaussian distribution $\mathcal{N}(\mu_i, \Sigma_i)$ with mean μ_i and covariance matrix Σ_i. Under this assumption with known parameters μ_i, Σ_i it is possible to derive a classification rule which has the minimum misclassification risk. The regions of the classes in input space are separated by a quadratic function. A related but simpler classifier is the *linear discriminant analysis (LDA)* which makes the further assumption that the covariance matrices of all classes are equal ($\Sigma = \Sigma_i$ for all i). In this case, the rule that minimizes the misclassification risk leads to a linear separation. Whether it is better to take QDA or LDA depends on the structure of the data. The covariance matrix of a Gaussian distribution describes in what way individual samples deviate from the mean. In classification problems where this deviation is class independent LDA should be preferred, otherwise QDA is based on the more appropriate model. So far the theory. Apart from that, in real-world problems one is faced with additional issues, even if we suppose that the Gaussian assumption is valid. The true distribution parameters μ_i and Σ_i are not known and thus have to be estimated by $(\hat{\mu}_i, \hat{\Sigma}_i)$ from given training data. If the number of training samples is small compared to the dimensionality n of the data this estimation is prone to error and degrades the classification performance. This has two consequences. Even when the true covariance matrices are not equal, LDA might give better results than QDA, because for LDA less parameters have to be estimated and it is less sensitive to violations of the basic assumptions. We modify the estimated covariance matrices according to

$$\hat{\Sigma}_i \mapsto (1 - \lambda)\hat{\Sigma}_i + \frac{\lambda}{\#\text{classes}} \sum \hat{\Sigma}_j. \tag{2.25}$$

Then one can mediate between QDA (for $\lambda = 0$) and LDA (for $\lambda = 1$). This strategy is called *regularization*. On the other hand, the estimation of covariance matrices from too little samples holds an inherent bias causing the ellipsoid that is described by the matrix deviating too much from a sphere: Large eigenvalues are estimated too large and small eigenvalues are estimated too small. To counterbalance this

bias a so called *shrinkage* of the covariance matrices towards the identity matrix I is introduced:

$$\hat{\Sigma}_i \mapsto (1-\gamma)\hat{\Sigma}_i + \gamma I \cdot \operatorname{trace}(\hat{\Sigma}_i)/n. \tag{2.26}$$

Of course regularization and shrinkage can also be combined which gives *regularized discriminant analysis (RDA)*, cf. Friedman [1989], while LDA with shrinkage is called *RLDA*. The choice of parameters λ and γ is made in a model selection, e.g., by choosing that pair of parameters that results in the minimum cross-validation error on the training set.

Support Vector Machines The *support vector machine* (*SVM*, Vapnik [1998]) is a popular classification tool. The method is based on the idea to use large margins of hyperplanes to separate the data space into several classes. Also non-linear functions, e.g. radial basis functions, can be used to obtain more complex separations by applying the kernel trick, cf. Appendix C.2 and the review Müller et al. [2001].

2.3.2 Analysis of Co-occurrence

Co-occurrence data frequently arise in various fields ranging from the co-occurrences of words in documents (information retrieval) to the co-occurrence of goods in shopping baskets (data mining). In the more general case, we consider the co-occurrence of two different features. One feature $\mathcal{K}$ is described by a vector that contains the frequencies how often it co-occurs with each specification of the other feature $\mathcal{P}$ and vice versa. Correspondence analysis aims at embedding the features $\mathcal{K}$ in a lower-dimensional space such that the spatial relations in that space display the similarity of the features $\mathcal{K}$ as reflected by their co-occurrences together with feature $\mathcal{P}$.

	c	...	...	b	$\mathbf{h}^{\mathcal{K}}$
C	$h^{\mathcal{K},\mathcal{P}}_{\mathrm{C},c}$	...		$h^{\mathcal{K},\mathcal{P}}_{\mathrm{C},b}$	$h^{\mathcal{K}}_{\mathrm{C}}$
...	...				...
B	...				$h^{\mathcal{K}}_{\mathrm{B}}$
Cm	...				$h^{\mathcal{K}}_{\mathrm{Cm}}$
...	...				...
Bm	$h^{\mathcal{K},\mathcal{P}}_{\mathrm{Bm},c}$	...		$h^{\mathcal{K},\mathcal{P}}_{\mathrm{Bm},b}$	$h^{\mathcal{K}}_{\mathrm{Bm}}$
$\mathbf{h}^{\mathcal{P}}$	$h^{\mathcal{P}}_{c}$	...		$h^{\mathcal{P}}_{b}$	n

TABLE 2.2: Co-occurrence table $\mathbf{H}^{\mathcal{K},\mathcal{P}}$ of keys $\mathcal{K}$ and pitch classes $\mathcal{P}$. Cf. text for details.

Consider, as our running example, the *co-occurrence table*

$$\mathbf{H}^{\mathcal{K},\mathcal{P}} = (h_{ij}^{\mathcal{K},\mathcal{P}})_{\substack{1\leq i\leq 24\\ 1\leq j\leq 12}} \tag{2.27}$$

for keys ($\mathcal{K}$) and pitch classes ($\mathcal{P}$). In Table 2.2, $\mathbf{H}^{\mathcal{K},\mathcal{P}}$ reflects the relation between two sets $\mathcal{K}$ and $\mathcal{P}$ of features or events (cf. Greenacre [1984]), in our case $\mathcal{K} = \{\mathrm{C},\ldots,\mathrm{B},\mathrm{Cm},\ldots,\mathrm{Bm}\}$ being the set of major and minor keys, and $\mathcal{P} = \{c,\ldots,b\}$ being the set of different pitch classes. Then an entry $h_{ij}^{\mathcal{K},\mathcal{P}}$ in the co-occurrence table would just be the number of occurrences of a particular pitch class $j \in \mathcal{P}$ in musical pieces of key $i \in \mathcal{K}$. The frequency $h_i^{\mathcal{K}}$ is the summation of occurrences of key i across all pitch classes. The frequency of pitch class j accumulated across all keys is denoted by $h_j^{\mathcal{P}}$. The sum of the occurrences of all pitch classes in all keys is denoted by n.

From a co-occurrence table one can expect to gain information about both sets of features, $\mathcal{K}$ and $\mathcal{P}$, and about the relation between features in $\mathcal{K}$ and $\mathcal{P}$, i.e., between keys and pitch classes in the example above. The *relative frequency* of the entries is denoted by

$$f_{ij}^{\mathcal{K},\mathcal{P}} = \frac{1}{n} h_{ij}^{\mathcal{K},\mathcal{P}} \,. \tag{2.28}$$

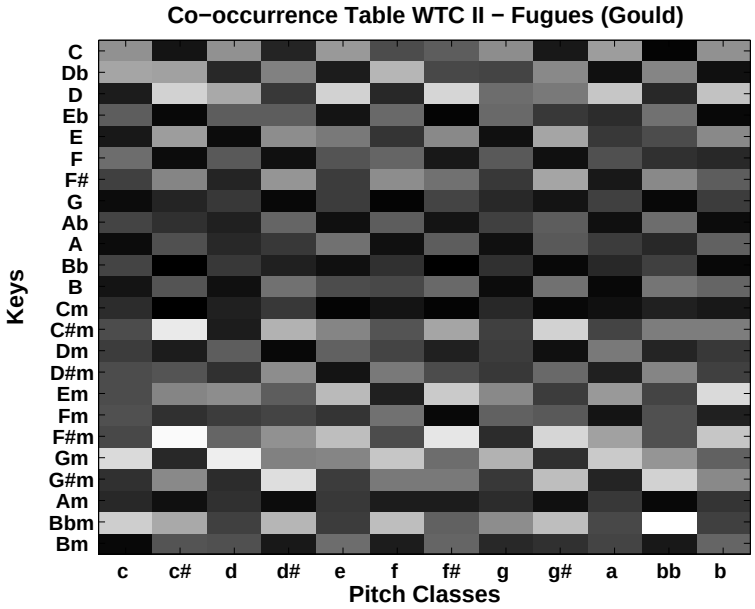


FIGURE 2.7: Co-occurrence table (cf. Table 2.2) of Bach's WELL-TEMPERED CLAVIER, fugues of Book II, recorded by Glenn Gould. The keys of the 24 fugues are labeled on the vertical axis. For each fugue the intensities are accumulated for each pitch class, calculating CQ-profiles (cf. Section 3.3.1 and Purwins et al. [2000b]). Light color indicates high intensity. Dark color indicates low intensity. This table is analyzed in Section 3.5.2.

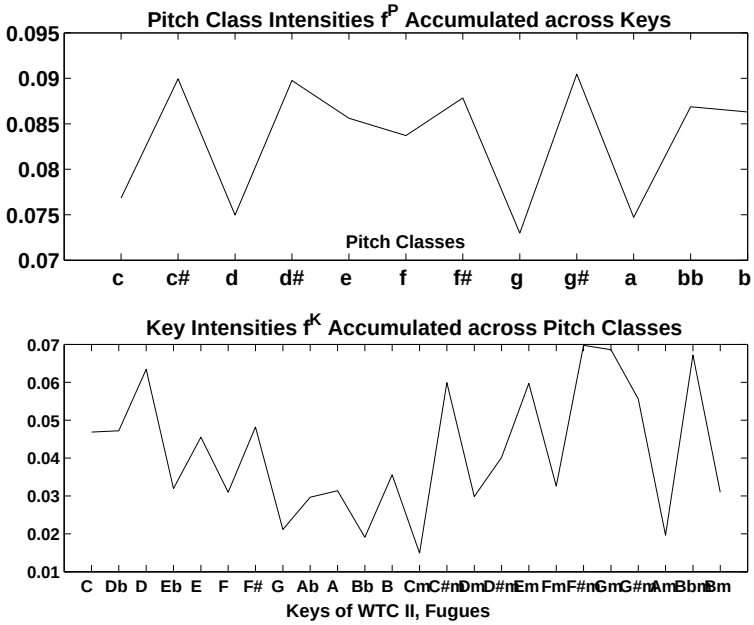


FIGURE 2.8: Relative frequency of pitch classes $\mathbf{f}^{\mathcal{P}}$ and keys $\mathbf{f}^{\mathcal{K}}$ of performed WTC II, fugues, accumulated from the co-occurrence table (Figure 2.7). It is remarkable that the non-diatonic notes in C–Major are the most prominent notes, as if Bach (and/or the performer Gould) wanted to oppose to the emphasis of C–Major in pre-well-tempered tuning. *Upper:* $\mathbf{f}^{\mathcal{P}}$ is the normalized vector of pitch class intensities accumulated across all fugues in WTC II. *Lower:* $\mathbf{f}^{\mathcal{K}}$ is the normalized vector of accumulated intensity of each fugue, i.e. each key.

It is the joint distribution of $\mathcal{K} \times \mathcal{P}$. The relative frequency of column j is $f_j^{\mathcal{P}} = \frac{1}{n} h_j^{\mathcal{P}}$. It is the marginal distribution of $f_{ij}^{\mathcal{K},\mathcal{P}}$. The diagonal matrix with $\mathbf{f}^{\mathcal{P}} = (f_1^{\mathcal{P}}, \ldots, f_{12}^{\mathcal{P}})$ on the diagonal is denoted $\mathbf{F}^{\mathcal{P},\mathcal{P}}$. The *conditional relative frequency* is denoted by

$$f_j^{\mathcal{P}|\mathcal{K}=i} = \frac{h_{ij}^{\mathcal{K},\mathcal{P}}}{h_i^{\mathcal{K}}}, \tag{2.29}$$

in matrix notation: $\mathbf{F}^{\mathcal{P}|\mathcal{K}} = (f_j^{\mathcal{P}|\mathcal{K}=\mathrm{i}})_{ji}$.

Instead of co-occurrence tables $\mathbf{H}^{\mathcal{K},\mathcal{P}}$ of frequencies of occurrences, in the sequel, we will also consider co-occurrence tables of overall annotated durations (cf. Section 3.5.2) as well as co-occurrence tables of accumulated intensities (cf. Figures 2.7, 2.7 and Section 3.5.2). The rows and columns of a co-occurrence table can be analyzed by clustering, correspondence analysis, Isomap, and the self-organizing feature map.

2.3.3 K-Means Clustering

Cluster analysis [Jain and Dubes, 1988] is a technique of exploratory data analysis that organizes data as groups (*clusters*) of individual samples. The *k-means clustering* method employs the *expectation maximization (EM) technique* in the following way: The data points are randomly separated into k clusters. First the mean of each cluster is calculated (E-step) and then each point is re-assigned to the mean to which it has the least Euclidean distance (M-step). The E- and the M-step are iterated until convergence, i.e., the M-step does not change the points-to-cluster assignment. It can be proven that this algorithm always converges. But runs with different initial random assignments could lead to different final configurations. In our experiment in Section 3.4.3 repetitions led to the same result.

2.3.4 Embedding and Visualization

For the visualization of high dimensional data that effectively lie on a low dimensional manifold of the data space we will introduce a linear method (correspondence analysis), a linear method with previous non-linear transformation such as graph distance calculation (Isomap), and a non-linear approach (self-organizing feature map).

Correspondence Analysis

In this section we extend this general idea of embedding musical structure in two-dimensional space by considering the Euclidean embedding of musical entities whose relation is given in terms of a co-occurrence table, such as Table 2.2. This general approach enables us not only to analyze the relation between keys and pitch classes (Sections 3.5.2 and 3.8.4), but also of other musical entities including aspects of the style of composers (Section 3.6.3). We can, for instance, exploit the fact that composers show strong preferences towards particular keys. This provides the basis for arranging the composers by correspondence analysis reflecting their stylistic relations.

We will first introduce the technique of correspondence analysis with a focus on the analysis of co-occurrences of keys and pitch classes in Section 2.3.2.

Principle Given a co-occurrence table $\mathbf{H}^{\mathcal{K},\mathcal{P}}$, for visualization purposes we aim at displaying the features $\mathcal{K}$ and $\mathcal{P}$ in a two-dimensional space, such that aspects of their tonal relation are reflected by their spatial configuration. In particular, correspondence analysis can be thought of as a method that aims at finding a new co-ordinate system that optimally preserves the χ^2-distance between the frequency $\mathcal{K}$-, and $\mathcal{P}$-profiles, i.e., of columns and rows. For 12-dimensional pitch class frequency vectors we consider the χ^2-distance, Equation C.5. The χ^2-distance is equal to the Euclidean distance in this example if all pitch classes appear equally often. The χ^2-distance weights the components by the overall frequency of occurrence of

pitch classes, i.e., rare pitch classes have a lower weight than more frequent pitch classes. The χ^2-distance satisfies the natural requirement that pooling subsets of columns into a single column, respectively, does not distort the overall embedding because the new column carries the combined weights of its constituents. The same holds for rows.

We can explain correspondence analysis by a comparison to principal component analysis. In principal component analysis eigenvalue decomposition is used to rotate the co-ordinate system to a new one with the axes given by the eigenvectors. The eigenvalue associated with each eigenvector quantifies the prominence of the contribution of this particular co-ordinate for explaining the variance of the data. The eigenvector with highest eigenvalue indicates the most important axis in the data space: the axis with highest projected variance. Visualization in this framework amounts to projecting the high-dimensional data (the 12-dimensional pitch class frequency space or, respectively, the 24-dimensional key frequency space) onto a small number of (typically two or three) eigenvectors with high eigenvalues. Hereby only insignificant dimensions of the data space are discarded, leading, effectively, to a plot of high dimensional data in two- or three-dimensional space.

In principal component analysis by rotating the co-ordinate system, the Euclidean distances between data points are preserved. Correspondence analysis is a generalization of principal component analysis: The χ^2 distance (a generalization of the Euclidean distance) between data points is preserved.

If the data matrix is not singular and not even symmetric, generalized singular value decomposition instead of eigenvalue decomposition yields two sets of *factors* $\mathbf{u}_1, \ldots, \mathbf{u}_d$ and $\mathbf{v}_1, \ldots, \mathbf{v}_d$, instead of one set of eigenvectors. So either for the m-dimensional column vectors of the data matrix the co-ordinate system can be rotated yielding a new co-ordinate system given by the column factors $\mathbf{u}_1, \ldots, \mathbf{u}_d$, or the n-dimensional row vectors of the data matrix are expressed in terms of co-ordinates in the new co-ordinate system of row factors $\mathbf{v}_1, \ldots, \mathbf{v}_d$. In principal component analysis each eigenvector is associated with an eigenvalue. In the same sense for each pair of column and row vectors $\mathbf{u}_k$ and $\mathbf{v}_k$, an associated singular value δ_{kk} quantifies the amount of variance explained by these factors (cf. Appendix C.3.1 for technical details). Consider the conditional relative frequency of pitch classes $\mathbf{F}^{\mathcal{P}|\mathcal{K}}$ being the data matrix. If we project the 12-dimensional pitch class profile $\mathbf{f}^{\mathcal{P}|\mathcal{K}=i}$ into the space spanned by all d vectors $\mathbf{v}_1, \ldots, \mathbf{v}_d$ and represent each profile $\mathbf{f}^{\mathcal{P}|\mathcal{K}=i}$ by its d-dimensional co-ordinate vector $\mathbf{s}_i$, then the χ^2-distance between $\mathbf{f}^{\mathcal{P}|\mathcal{K}=i}$ and $\mathbf{f}^{\mathcal{P}|\mathcal{K}=l}$ equals the Euclidean distance between the co-ordinate vectors $\mathbf{s}_i$ and $\mathbf{s}_l$ of their projections. But if we only use the two co-ordinates with highest singular value, instead of all d co-ordinates, then all distances are contracted and more or less distorted, depending on the singular values.

A *biplot* provides a simultaneous projection of features $\mathcal{K}$ and $\mathcal{P}$ into the same space. Both the co-ordinates of a $\mathcal{K}$-profile in the co-ordinate system of the $\mathbf{u}_k$'s and the co-ordinates of a $\mathcal{P}$-profile in the co-ordinate system of the $\mathbf{v}_k$'s are displayed in the same co-ordinate system. Such a biplot may reveal the inter-set relationships.

Using the multiv library in R ported by Friedrich Leisch from the original S package by Fionn Murtagh, in Sections 3.5.2 and 3.8.4 we will present the results of our correspondence analysis of inter-key relations in scores and recorded performances that leads to the emergence of the circle of fifths and to a toroidal model of inter-key relations. We show how these results relate to a similar model from music theory [Chew, 2000] and to earlier experiments with a different cognitive model [Purwins et al., 2000a]. In Section 3.6 we apply correspondence analysis to the problem of stylistic discrimination of composers based on their key preference. Please note that we provide a more technical perspective on correspondence analysis in Appendix C.3.1.

Historical Note According to Greenacre [1984], the interest in studying co-occurrence tables emerged independently in different fields such as algebra [Hirschfeld, 1935], psychometrics [Horst, 1935; Guttman, 1941], biometrics [Fisher, 1940], and linguistics [Benzécri, 1977]. Correspondence analysis was discovered not only in distinct research areas but also in different schools, namely the pragmatic Anglo-American statistical schools as well as the geometric and algebraic French schools. Therefore, various techniques closely related to correspondence analysis have been discussed under various names, e.g., "reciprocal averaging", "optimal (or dual) scaling", "canonical correlation analysis of contingency tables", "simultaneous linear regressions".

Isomap

Often high dimensional data can be approximately described by a curved manifold of a lower dimension. For Isomap [Tenenbaum et al., 2000] data points in high dimensional space are given. Each of them is connected to its k nearest neighbors (cf. Section 2.3.1 and Ripley [1996]), with respect to a given dissimilarity, e.g. Euclidean distance. The idea is that from one point another point cannot be reached directly, only by taking a route via data points that are connected to each other. The graph distance assigned to a pair of data points is the length of the shortest path via connected points. Multidimensional scaling [Shepard, 1962] is performed on the graph distance matrix, so that projecting the points onto the eigenvectors with highest eigenvalues shows the configuration of points in a low dimensional Euclidean space that optimally preserves the graph distances. E.g. a two dimensional manifold hidden in the high dimensional data is visualized by Isomap as a planar display that appropriately reflects the inter point distances on the surface of the manifold. Cf. Appendix C.3.2 on page 199 for technical details.

Toroidal Self-Organizing Feature Map (SOM)

The weight vectors $\mathbf{w}_i \in \mathbb{R}^n$ of the SOM have the same dimensionality as an analyzed input datum $\mathbf{x} \in \mathbb{R}^n$. During training, the weight vectors adjust to the input vectors. The indices of the weight vectors form a grid structure defined by some

neighborhood function. In our application, a two-dimensional toroidal grid is defined by the toroidal distance $d_{\mathcal{T}}(l, l')$ (Equation C.7), serving as a neighborhood function. During training, the grid is smoothly embedded in the high dimensional input space. The algorithm works as follows, d_n being the Euclidean distance in $\mathbb{R}^n$:

1. random initialization of weights $\mathbf{w}_i \in \mathbb{R}^n$,
2. random choice of input vector $\mathbf{x} \in \mathbb{R}^n$,
3. $$i_{\mathbf{x}} = \arg\min_i(d_n(\mathbf{x}, \mathbf{w}_i)), \quad (\text{winner} - \text{takes} - \text{all step}) \tag{2.30}$$
4. update:
$$\mathbf{w}_i := \begin{cases} \mathbf{w}_i + \eta(\mathbf{x} - \mathbf{w}_i) & \text{if } d_{\mathcal{T}}(i, i_{\mathbf{x}}) < N \\ \mathbf{w}_i & \text{otherwise} \end{cases}, \tag{2.31}$$
5. continue at 2. with learning parameters η and N (neighborhood range) gradually decreasing.

Graepel et al. [1997] further develop a more robust version of the SOM by introducing stimulated annealing into the algorithm. An energy function can be defined for infinitely many simulations by viewing the evolution of the $\mathbf{w}_i$ as a stochastic process [Ritter and Schulten, 1988]. The SOM preserves the topology, i.e. adjacent (defined by d_n) vectors in the input space map to neighboring (defined by $d_{\mathcal{T}}$) positions on the grid. The topology preserving property of the SOM is also optimal in the sense of the information theoretic principle of maximal information preservation [Linsker, 1989]. Topology preservation is a property that can be observed in the brain also, e.g. in tonotopic or retinotopic maps. Therefore, the SOM serves as a model of the cortical organization, e.g. of orientation preference and ocular dominance patterns. [Obermayer et al., 1991] However, the biological relevance of the SOM is limited, since the "winner-takes-all step" (Equation 2.30) would assume a global comparison with *all* other "neurons" (= weights) in the brain. Lacking biologically more plausible models, we will use the SOM as a coarse cognitive model of a schema for the perception of key and tone centers.

3 Results – Sawtooth, Pitch Class Profiles, Circles, and Toroids

We will first discuss the pitch "paradox", then the constant Q profiles and their applications.

The material presented in this section is mostly a revised version from previous publications: Sections 3.1, 3.2.1, 3.2.3, 3.2.5, 3.2.6, and 3.5.1 [Purwins et al., 2005], 3.2.2, 3.2.2, and 3.2.4 [Normann et al., 2001a,b], Subsection "The subject's comments" of 3.2.2 [Normann, 2000], 3.3.1 – 3.3.5, 3.9.3 [Blankertz et al., 1999b,a; Purwins et al., 2000b, 2001a], 3.4 [Purwins et al., 2004a], 3.5.2, 3.6, and 3.8.4 [Purwins et al., 2004b], 3.7.2 and Subsection "Simulations" of 3.8.1 [Purwins et al., 2006], 3.8.2 and Subsection "Simulations" of 3.8.3 [Purwins et al., 2000a], 3.9.4 [Purwins et al., 2002].

3.1 The Sawtooth Culmination Pattern

The range of musical parameters, like pitch, tempo, and loudness is limited by the capability of the auditory system. This is reflected by design constraints of musical instruments and in playing technique. Every instrument has an ideal pitch range. But in addition, the hearing range limits the perceived pitch by an upper and a lower bound. Also tempo is limited to a specified range. An accelerating pulse sequence is perceived as a beat. But by exceeding a certain tempo a pitch is heard. Also loudness is limited by the threshold of pain and the hearing threshold. For ultimately extending the period of suspense before reaching the climax, composers use a trick. The parameter in question (e.g. pitch, tempo, or loudness) slowly increases. Then the parameter is suddenly turned down. After that, the slow increase repeats, possibly in a varied form, abruptly collapses again and so forth. Inspired by Schneider [2005] we observe a sawtooth pattern.

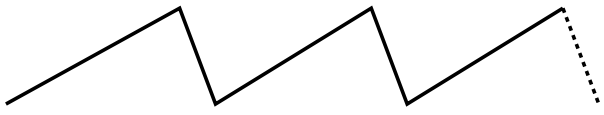


FIGURE 3.1: The sawtooth culmination pattern: a repeating cycle of slow rise and quick fall.

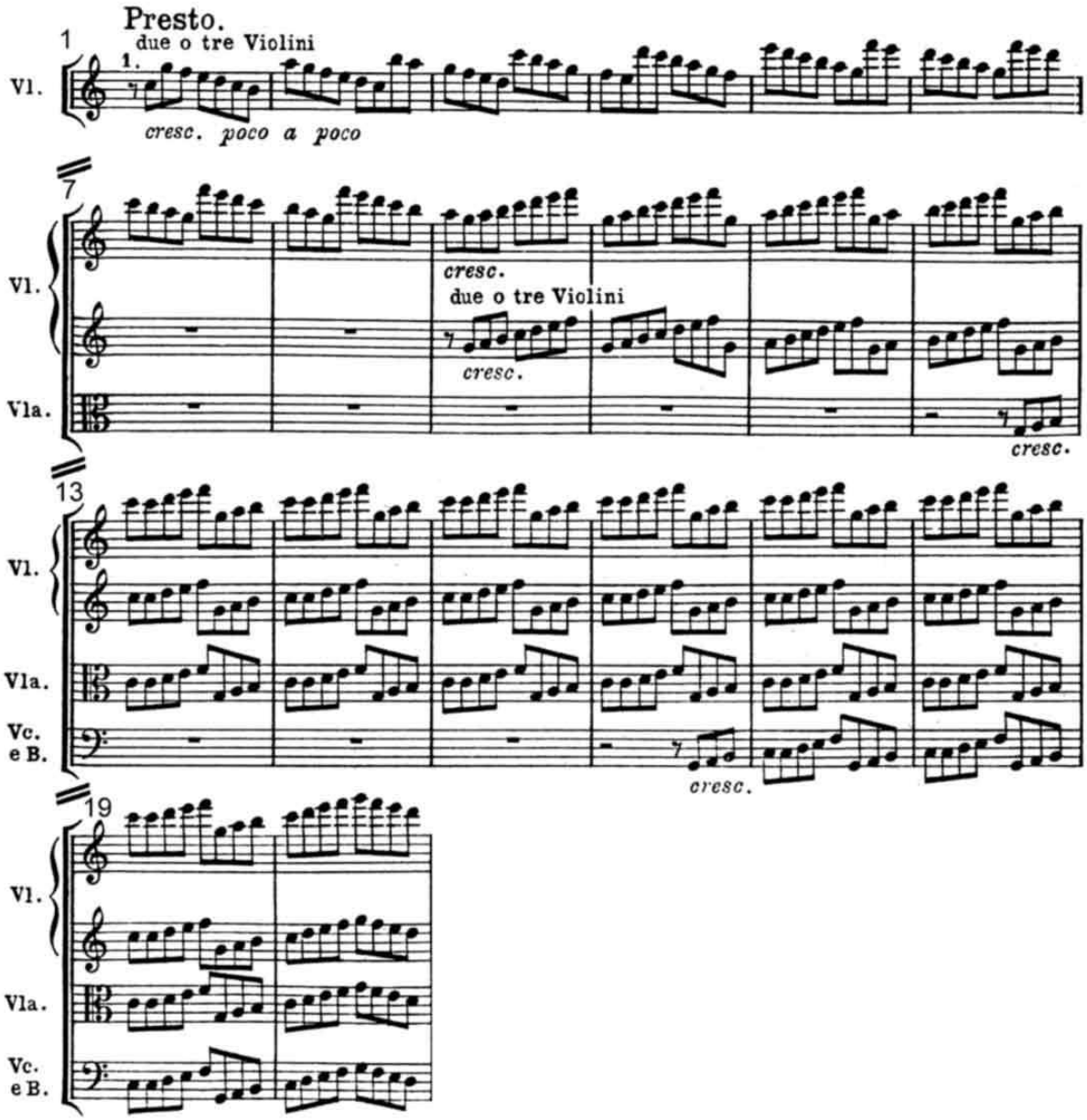


FIGURE 3.2: Final PRESTO of Beethoven's overture LEONORE III.

Let us take a closer look at the first eight bars from the final PRESTO in Beethoven's overture LEONORE III (Figure 3.2). At a crude level we can assume a quick long falling scale. But this idea cannot be realized directly. A sawtooth-like pitch sequence (Figure 3.1) is a good approximation: The solo violins jump up a seventh after every six descending diatonic scale notes. They then continue with the falling scale playing up an octave. The jump is audible though. Shepard [1964] (cf. p. 48) aims at camouflaging the jump. The next falling scale already inaudibly fades in before the last scale fades out.

Just to name one of numerous examples for a sawtooth-like cumulation in the loudness domain, we may look more closely at Danse Générale from Maurice Ravel's Daphnis et Chloé, Suite No. 2. Danse Générale consists of one long steady cumulation of a driving repetitive motive in triplets. Throughout this section the dynamics slowly build up from *p* or *pp* to *mf* or *ff*, before collapsing, and slowly dynamically rising again (new cycles at No. 210, 2nd bar after 212, 2nd bar after 214, 2nd bar after 215, 216, 217, 218, 4th bar after 220).

3.2 Circular Perception of Relative Pitch in Harmonic Complex Tones

> die Idee von etwas, was immer weiter wächst, immer nach oben wächst, aber nie ... ankommt [1] (György Ligeti referring to his etudes L'escalier du diable and Columna infinită)

3.2.1 Octave Ambiguity and Octave "Paradox"

Tritone intervals of Shepard tones are either perceived as going up or going down. In one listener the perception is consistent. But it varies from one listener to another. [Deutsch, 1986] In addition there is the phenomenon of octave ambiguity: Even trained singers sometimes miss a tone by one octave, especially when a female voice wants to meet the tone of a man's voice or vice versa.

Bells are special among the instruments of classical Western music. Bell spectra can only insufficiently be approximated by harmonic complex tones. [2] Due to this feature, composers often feel challenged to imitate bell spectra by suitable chords, thereby opening new harmonic perspectives: Franz Liszt's Funérailles, the crowning scene in Modest Mussorgsky's Boris Godunov, Arnold Schönberg's Sechs kleine Klavierst cke op. 19 No. 6, the highlights in György Ligeti's L'escalier du diable, and Manfred Stahnke's Bellringer's Song, about 1200. The pitch of bells is often ambiguous with respect to the octave position [Terhardt and Seewann, 1984].

How can we explain this phenomenon? Let us consider Shepard tones, introduced in Section 1.2.3. Due to their spectrum, octave intervals cannot be generated with Shepard tones. This is so because transposing a Shepard tone up an octave yields the identical tone. Let us put it this way. Shepard's trick is as follows: On the pitch helix (Figure 1.9) tones in octave interval distance can be distinguished due to their vertical position. Shepard [1964] lets the helix shrink to a circle in the X-Y-plane, so that octaves unify in one point. Let us imagine a variation on Shepard [1964]'s experiment, which works the other way around: we let pitch class stay

[1] "the idea of something that is continuously growing, always growing upwards, but never arrives" Ligeti in his introduction to the first performance of Columna infinită (1993) in Münster.

[2] Terhardt [1998], p. 220. Definition of a "harmonic complex tone" on p. 32 of this thesis.

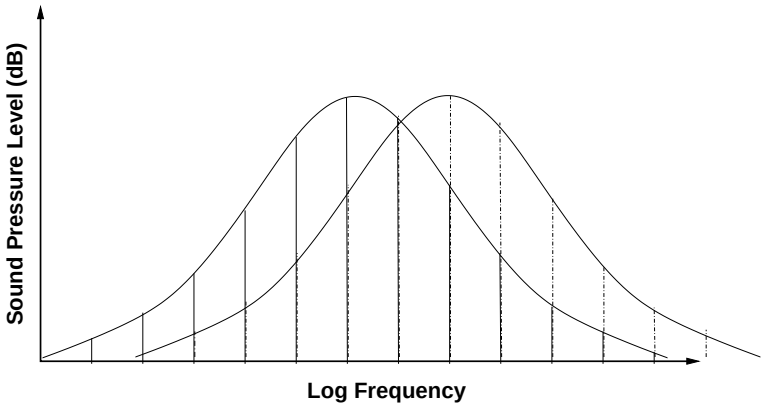


FIGURE 3.3: Generation of octave ambiguity. The spectral envelope of a Shepard tone is varied, while the partials remain the same.

the same and vary the spectral envelope, as seen in Figure 3.3. In other respects the structure of the complex tone remains the same. It is constructed from partials in octave distance. In this experimental design we explore brightness in isolation, referring to the vertical axis of the pitch helix. Without actually having conducted the experiment, we can expect that a complex tone with a spectral envelope centered in the lower register is perceived lower than a complex tone with an envelope in the higher pitch register. The spectral envelope can be shifted continuously from low to high frequencies. Octave ambiguity occurs if the centroid of the spectral envelope is in a certain position, so that octave position of that tone is judged differently by different subjects. The experimental design could be as follows: A complex tone is presented initially. Two partials in octave distance, one lower and the other one higher than the spectral centroid, are then presented in isolation. The subject is asked: Which sine tone of the latter pair resembles the initially presented tone more closely? We would expect that the answers vary among the subjects.

From octave ambiguity we can construct the octave "paradox", using the gestalt principle of similarity. A small shift of the spectral envelope would not change the perceived octave position of the tone. But an accumulation of consecutive small shifts causes the final note to have another octave position from that of the starting note.

The Shepard scale is still perceived "paradoxically" if we combine the pitch class movement with a smooth temporal variation of the spectral envelope. Brightness can fluctuate or can move parallel to or opposite to pitch class. We can double the pitch "paradox" (Figure 3.4) by designing a sound sequence that is circularly perceived as ascending and descending in pitch at the same time. Pitch class and brightness both change cyclically with different cycle durations. Pitch class infinitely rises. A sinusoidal spectral envelope cyclically descends and thereby lets brightness infinitely get duller. We will see in Section 3.2.6 how spectral centroid and spectral spread are tuned for usage of circular pitch in particular pieces of music.

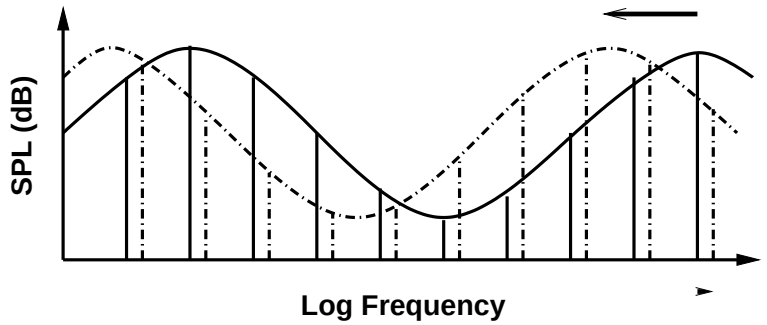


FIGURE 3.4: Doubled "paradox". Instead of a fixed bell shape such as in Figure 1.10, we employ a sinusoidal spectral envelope moving opposite to pitch class. When the pitch class moves up (small arrow on the bottom from solid to dashed-dotted lines) the spectral envelope moves down (long arrow on top). Pitch class ascends, whereas brightness gets darker at the same time.

To give an example in music for acoustical instruments, the beginning of the PRESTO at the end of Beethoven's overture LEONORE III refers to the octave "paradox" with falling spectral envelope (Figure 3.2). The strings play the same pitch class *unisono* but in different octave positions, starting with the violins, later joined by other instruments in increasingly lower registers. First only the violins play fast runs. One after another, the viola, cello and double-bass join in. By bar 16 the passages run in three parallel octaves. Techno music sometimes employs a comparable effect, technically realized by the steady fall of the boundary frequency of a low pass filter.

A problem arises here. If the ear hears analytically it may decompose a single tone into several components or even its partials. For a tone sequence with circular pitch we may hear how particular components become more prominent and others vanish. In the description of the experiment in Section 3.2.2, we will explain how we support synthetic perception of single tones.

3.2.2 Experiment: Judgment on Pitch Comparison

Usually an order is considered to be transitive. A relation "$\prec$" is termed *circular*, if there are elements $a_0, a_1, \ldots, a_n$ with $a_0 \prec a_1 \prec \ldots \prec a_0$. The "paradox" of our experiments is based on circular and intransitive pitch relation. According to common notion we would assume that any tone determines a unique pitch on a linear scale. But such a linear and transitive order falls short in explaining our experiments.

For our experiment, stimuli are designed in the following way. Starting from the first tone (top Figure 3.5), step by step, partials of a fundamental with half of the lowest frequency are added (2nd to 6th row in Figure 3.5). Six tone pairs are combined from these six consecutive harmonic complex tones: 1st with 2nd; 2nd with

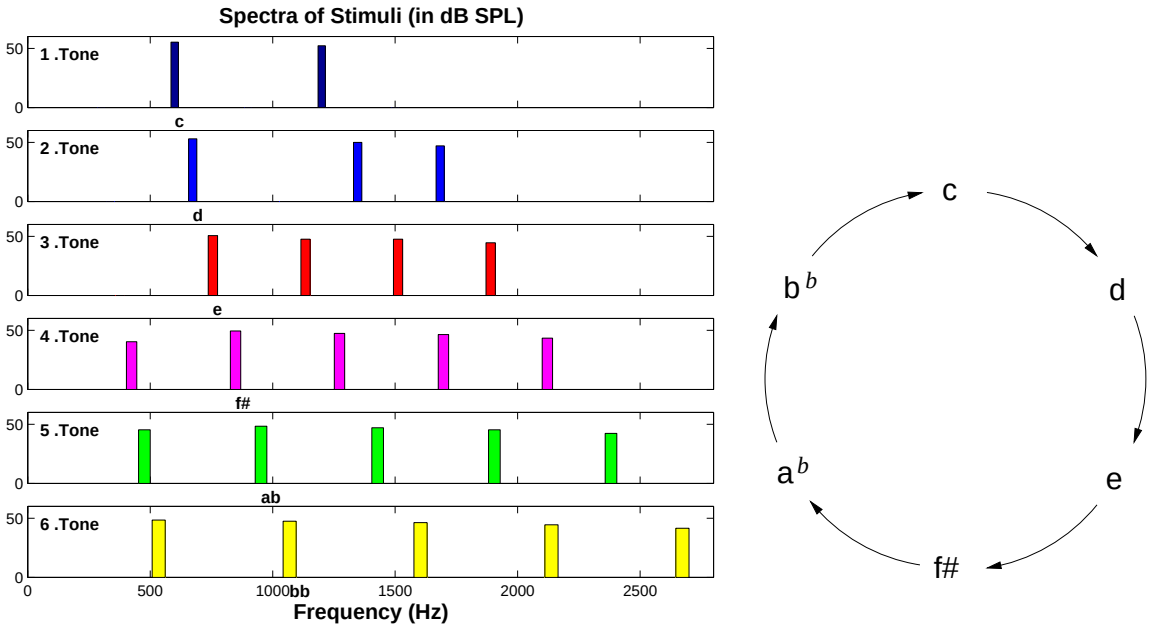


FIGURE 3.5: The harmonic complex tones of a group differ in amplitude of the even and the uneven partials (*left*). The timbre manipulation leads to an octave uncertainty that enables paradoxical, namely circular, pitch perception (*right*). c → d denotes "perceived pitch of d higher than perceived pitch of c".

3rd; ... ; 5th with 6th; and 6th with 1st. This group of tone pairs is considered in nine transpositions. A second block of stimuli is derived by reversing the order of all $6 * 9 = 54$ tone pairs. Each of the 18 subjects has to give one of the following alternative judgments on pitch comparison to each of the $6 * 9 * 2 = 108$ tone pairs: 1) "first tone higher", 2) "second tone higher", 3) "both tones equal", or 4) "undecidable". The sequence of tone pairs is presented in random order. [3]

A statistical investigation of the stimuli reveals that the brilliance (Equation 2.20 in Section 2.2.2) remains between 5.5 to 7 Bark with the exception of the first tone (> 4 Bark).

Results The results are shown in Figure 3.6 and they are summarized in Figure 3.7. Each column of each block in Figure 3.6 reflects 54 judgments of one subject. In the lower block, tone pairs are presented in reversed order. Each block is divided into nine rows, one for each transposed group of tone pairs. For each subject and each transposed group, one rectangle contains the judgments on the six tone pairs of that group. Each individual pitch judgment is not interesting in isolation. The group of six pitch judgments on the whole, the contents of one rectangle, is of interest. As shown in Figure 3.5, the tones of a group are compared with each other in a circular

[3] For a more detailed description of the experiments cf. Normann [2000].

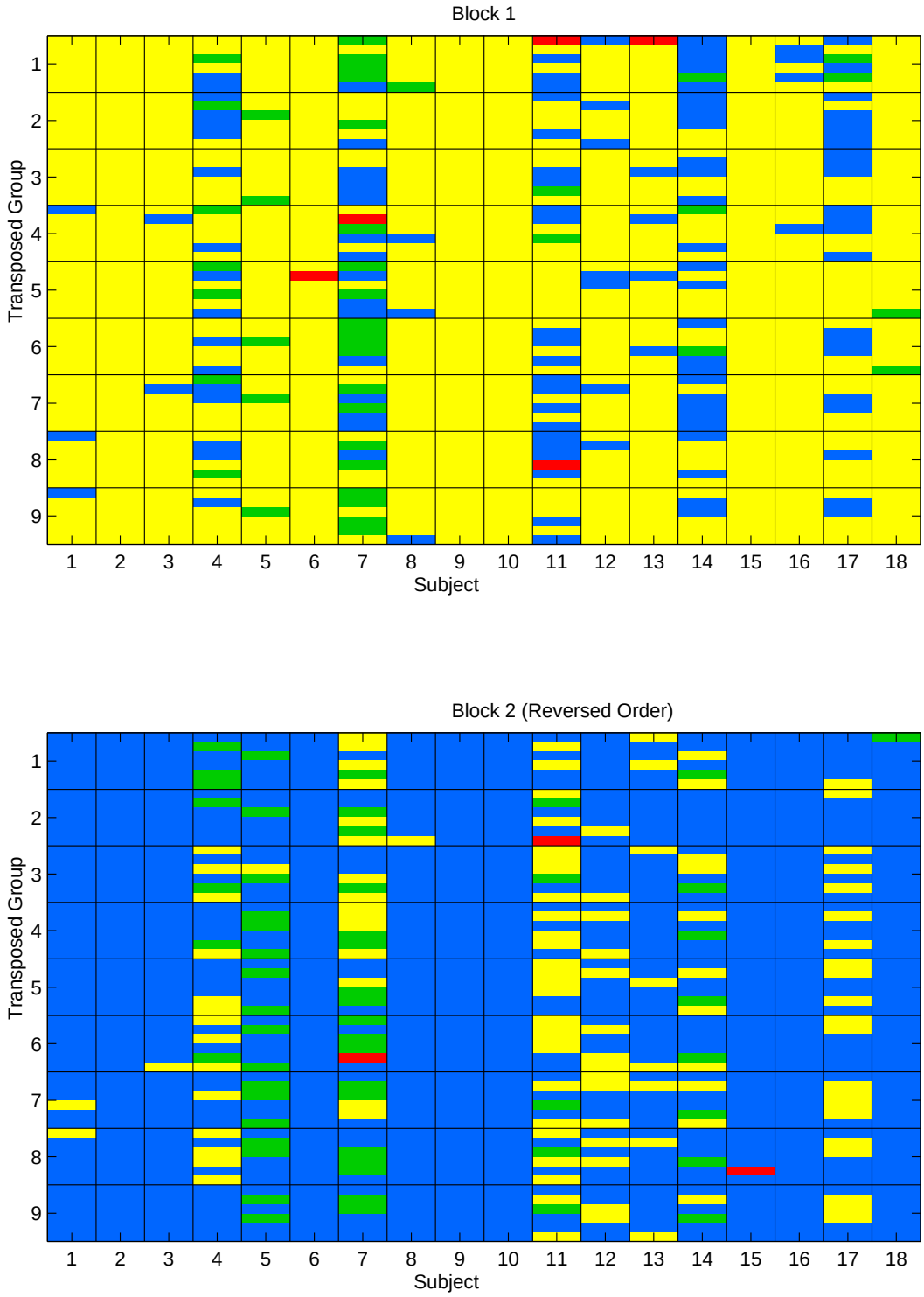


FIGURE 3.6: Pitch judgments on pairs of consecutive tones from Figure 3.5. Each of the nine rows of groups indicates a different transposition of the group of tone pairs. Each column shows the judgments of a different subject. The lower block refers to the same tone pairs as the upper block, but in reversed order. The pitch judgments are encoded by colors as follows: "first tone higher" (blue), 2) "second tone higher" (yellow), 3) "both tones equal" (green), and 4) "undecidable" (red). (Confer text for details. From Normann [2000], p. 29.)

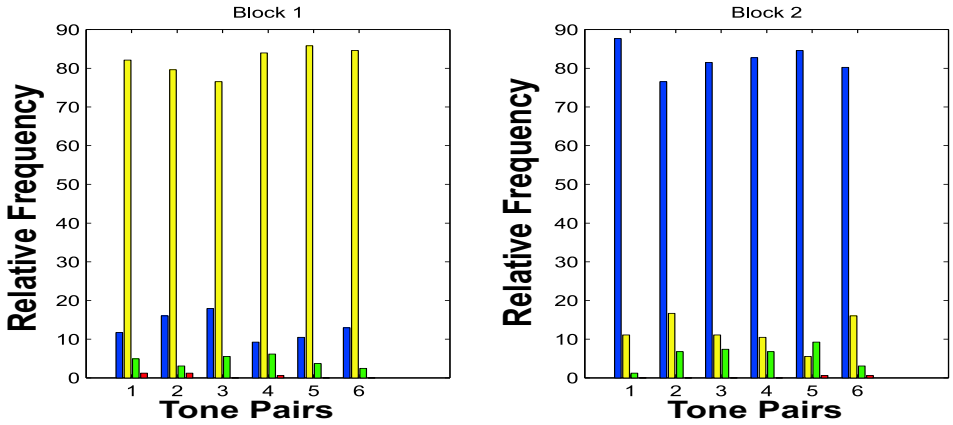


FIGURE 3.7: Percentage of answers for the six tone pairs. Tone pair no. 1 refers to tone no. 1 compared with tone no. 2 in Figure 3.5 ... tone pair 6 to tone no. 6 compared with tone no. 1. In the right figure tones are presented in reverse order. The colors encode the judgments on pitch difference: yellow - ascending; blue - descending; green - equal; red - not comparable in pitch. The mean is taken across the 9 transpositions of the tone pairs and the 18 subjects. We can see that the judgments across tone pairs are quite homogeneous with relatively low probability for the comparison of the ascending third (left no. 3) and the descending second tone pair (right no. 2). A more advanced design of stimuli could smooth out the differences (cf. Section 3.2.5 and Figure 3.13).

order. A subject has perceived six tones of a group circularly if there is at least one ascending tone pair ("second tone higher") and all others are perceived ascending as well or equal in pitch. An analogous formulation holds for circularly descending tones of one group. If we expect that tones should arrange in a linear order such a sequence of tones appears paradoxically. In Figure 3.6, circularly perceived tone sequences can be identified as rectangles that contain only the color(s) yellow (and green, first block) or blue (and green, second block). On the whole in the first block, 96 of the 162 rectangles are circular, in the second 102. That indicates that the constructed tone sequence is perceived circularly in more than half of the instances. There are four subjects that confirm the circularity in all groups in both blocks.

The histogram of different degrees of circular perception (right Figure 3.8) reveals two blobs, one on the right, another one on the left side. We interpret this as a bimodal distribution related to two groups of listeners. An informal screening of the background of the subjects suggests that the musically experienced listeners tend to "paradoxical" perception (right blob). Naive listeners (left blob), on the other hand, guess more than actually hear the pitch attribute and distinguish less between pitch and brilliance. How significant is the "paradoxical" perception of the subjects in the right blob?

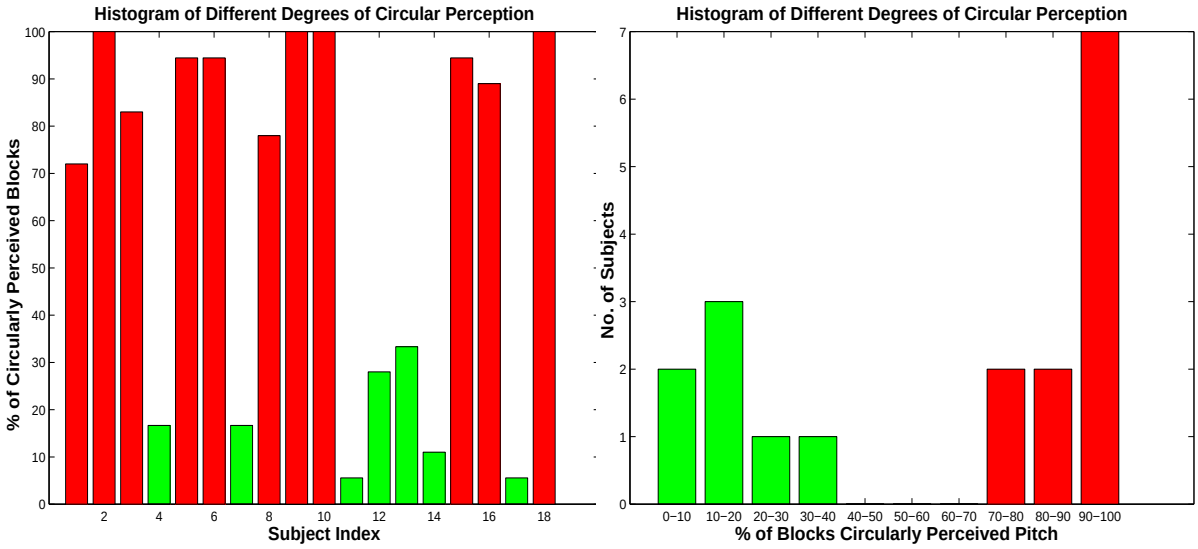


FIGURE 3.8: 11 out of 18 subjects perceive at least 70% of all blocks circularly. For all the subjects equivalent statements hold for mostly 30% of all blocks (*left*). The result is a bimodal distribution that indicates the different types of listeners (*right*).

Statistical Formulation and Evaluation For the i-th block, define a binary random variable X_i: $X_i = 1$, if the i-th block is perceived circularly descending, otherwise $X_i = 0$. Be $P\{X_i = 1\} = \pi_k, P\{X_i = 0\} = 1 - \pi_k$, for subject k. Under the simplifying assumption of stochastic independence $Y_k = \sum_{i=1}^{18} X_i$ is binomially B_{18,π_k}-distributed.

The uniformly best, randomized Neymann-Pearson test for two-sided test problems to error level $\alpha = 0.05$ yields that at least 39 % of the subjects have a circular pitch perception ("$\pi_k > 0.5$"; for comparison "wild guess" would yield $\pi = 0.094$).

The Subjects´ Comments Before the experiment started, most of the subjects were convinced that for any tone a pitch could be determined unequivocally. So they thought their judgments were either right or wrong. The experimenter explained to the subjects that pitch is a perceptual quality. So the matter of investigation would be rather "how" than "how well" the subjects would judge pitch, he said. In the end the subjects were asked: When would you think two tones were not comparable in pitch? Various responses to this question were given:

1. "The tones sound too different with respect to timbre"
2. "At the same time, a tone sounds brighter and lower than another tone"
3. "A tone is musically lower but physically higher then another one"
4. "In one tone one could hear two pitches simultaneously"
5. "One could hear two or even more tones simultaneously"

6. "The fundamental was missing"

These answers reflect different ways of hearing in music: Musically experienced participants have another way of listening. Some listeners tend to listen more synthetically, others more analytically. Some of the subjects have in-depth experience with using sound synthesis techniques on the computer, e.g. additive synthesis. The answers also reveal the competition of timbre and pitch cues.

Only rarely the subject chose the judgment "tones are not comparable in pitch". Most subjects said they mostly heard only one pitch, given one tone. Or at least one pitch would have dominated other pitches, they explained.

We should also report how the subjects helped themselves to give a judgment on pitch difference. Often they tried to reproduce the tone sequence by singing, humming, or whistling. Obviously the pitch comparison referred to the reproduced tone pair rather than to the original tone pair itself. It looked as if the underlying strategy was to translate two tones with distinct timbre into tones with familiar timbre. So the subjects could solve the task in a timbre domain they were more familiar with.

3.2.3 Dimensionality, Proximity, Perceptual Quantization, and Competition

We would like to discuss the "paradoxes" with regard to the following aspects:

1. the multidimensionality of the adequate psychoacoustic model for pitch and tempo,

2. the gestalt principle of proximity at work in the contour space of pitch classes,

3. the perceptual quantization accumulating inaudibly small changes to a noticeable qualitative change, and

4. the contradictory arrangement of competitive cues that are usually redundant in a natural environment.

According to Drobisch [1855] (see Section 1.2.3 and Figure 1.9 on page 49) pitch is an inherently two-dimensional feature that can be decomposed into pitch class and brightness. On a straight line one cannot walk in circles. But if a feature is defined by several degrees of freedom it is not surprising to embed a circle in the multi dimensional feature space.

As we have seen above, the pitch of Shepard tones is solely represented by pitch class given by the angle that is a point on the pitch class circle (Figure 1.9). But in the experiment we only ask for the relation between two pitches. After this preparation, the concept of Shepard pitch offers a convincing explanation of experimental results. In essence, the subjects involved in the experiment were asked to establish a relation between two points on the circle by using their auditory perception. The subject's

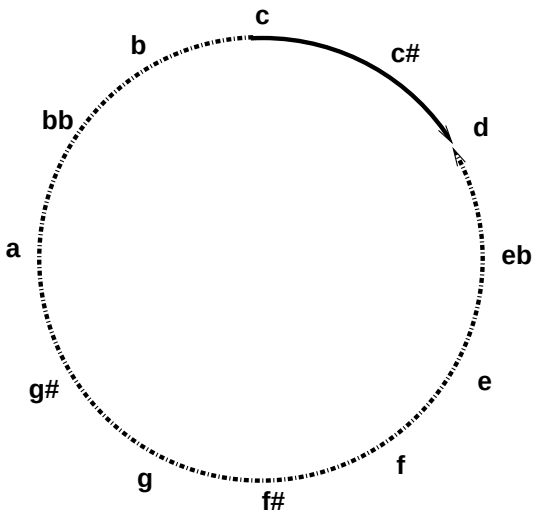


FIGURE 3.9: Gestalt principle of proximity on the circle of pitch classes. d is higher than c, because the shortest path between c and d is clockwise. Two ascending halftone steps (solid line) are shorter than ten descending half tones (dotted line).

answer "a is lower than b" can be visualized as a clockwise walk on the circle from a to b. "a is higher than b" implies a counterclockwise walk on the circle from a to b. The experimental results show significantly that the subjects hear pitch comparisons according to the shortest path on the pitch class circle (Figure 3.9).

The subjects choose a perspective, so that a and b are closer to each other. This can be interpreted as a result of the gestalt principle of proximity, if viewed in the appropriate formulation: For pitch classes a and b, let us consider the directed interval

$$d = d(a, b) = (b - a + 6) \bmod 12 - 6 \tag{3.1}$$

The signature $\sigma(d)$ of d is positive, if the shortest path on the pitch class circle runs clockwise, otherwise $\sigma(d)$ is negative. For a melody the *contour* is the sequence of ups ("1"), downs ("-1"), and repetitions ("0") in pitch. Therefore, in the contour space $C = \{-1, 0, 1\}^n$, the gestalt of a heard contour is determined by the closest path connection ($\sigma(d) = \pm 1$). In this formulation, the phenomenon of circular pitch perception obeys the gestalt principle of proximity.

In the experiment we make another observation: the more the directed interval between a and b approaches six halftones (a tritone), the more the variation of perceived pitch relation increases across different subjects. For this interval, the variability of the subject's judgments reaches its maximum. This result is consistent with psychological experiments in general, in that the task requires the subject to give a yes/no answer depending on whether he perceives a stimulus to be above or below a certain threshold. Then the averaged answers can be described by a sigmoid with its symmetry point at the threshold.

Due to the just noticeable differences and perceptual thresholds the perceptual system quantizes the analog stimulus, thereby generating a quantization error. A

similar value in the analog signal can result in an *equal* value in the quantized signal. The accumulation of inaudible differences finally results in a change of perceptual quality. In the design of circularly perceived sequences, consecutive stimuli have to be perceived similarly according to the related gestalt principle. We can also read this in light of the motto from gestalt psychology: "The whole is more than the sum of its parts". Frequency components with octave relationships fuse into one perceived tone. For a Shepard scale, this is a way to camouflage the entry of single components. If one could analytically extract them, their sudden entry would disrupt the sequence. Shepard [1964] describes how spectra of superimposed thirds, instead of octaves, are not suited for constructing a gradually changing sequence, since they are perceived as chords instead of fusing to a single tone perception. Furthermore in synthesis experiments we observe that for circular sequences, of both octave spectra and harmonic complex tones, the cycle duration of the rise should not be too short. In that case, changes (e.g. fade-out of high, fade-in of low components) of the tone are camouflaged better because they are closer to the perceptual threshold. Thus we can return to the initial sound imperceptibly.

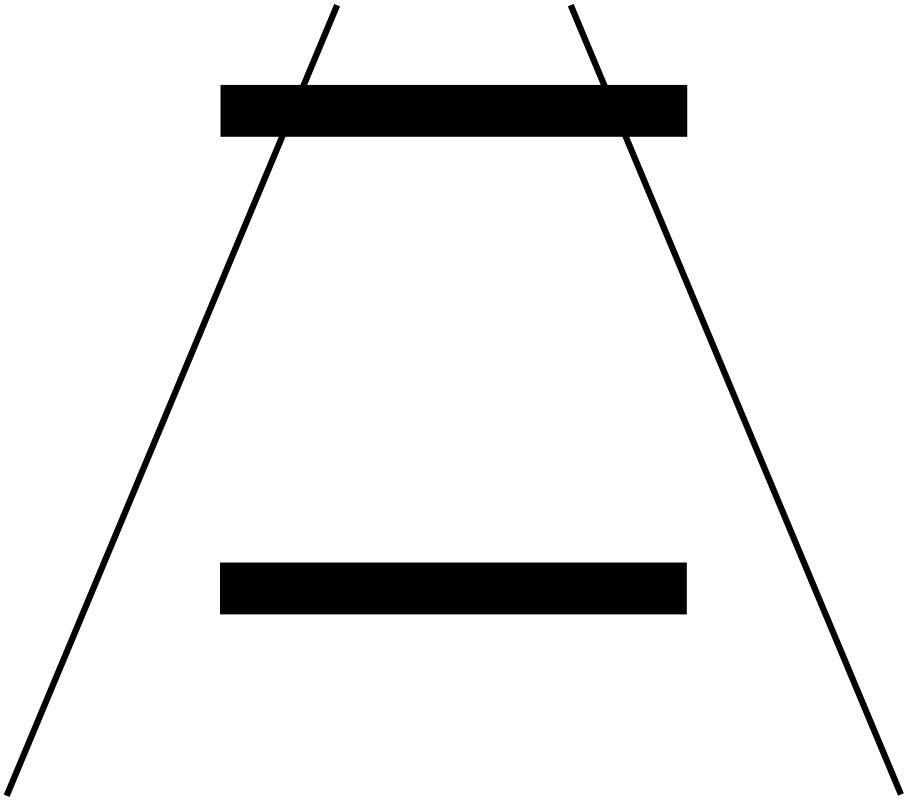


FIGURE 3.10: Ponzo scalability paradox: Two lines are running towards the vanishing point. There are two parallel bars of equal size on top. The bar "behind" apears bigger.

Starting from Risset [1985], we would like to present another hypothesis that will guide us to further formalization in Section 3.2.4. In the perception of our environment, certain information is encoded by several redundant features. For illustration of the problem, let us look at Ponzo's scalability paradox (Figure 3.10). Two lines converge towards the vanishing point, thereby creating a perspective. Two bars of identical shape and size lie on top of the lines. The upper bar appears bigger. What could be the explanation? On one hand, due to their equal shape, both bars should be perceived as being the same. On the other hand, the item is closer to the vanishing point. In this optical illusion we synthesize a combination of features that does not appear in this way in our environment. Normally these features are coherent. If an object of the same shape is further away in perspective, it is scaled down. But in our setup we present an incoherent combination of features. The size of the bar remains the same. But it is stationed further toward the vanishing point. Two features compete with each other: "same shape" versus "position closer to vanishing point". The feature "position closer to vanishing point" dominates.

We can apply this principle to the Shepard tones (Section 1.2.3 on page 48): The pitch class changes, the brightness stays the same. For natural sounds in general, it applies that a higher tone is brighter. Refined adjustment of parameters results in the circularly perceived Shepard scales that maintain brightness while ascending in pitch class. For circular sequences of harmonic complex tones, the following cues for the determination of the fundamental are competitive. In the spectrum of pitch difference (Equation 3.2) the two main peaks compete with each other. Different cues are incorporated into this equation: On the one hand, the difference of the frequencies of adjacent partials is the fundamental frequency. On the other hand the fundamental of a complex sound is determined by the lowest partial. We are now ready to formalize the competition of cues for pitch determining.

3.2.4 Model: Spectrum of Pitch Differences

We want to design a psychoacoustic model for higher/lower pitch comparison between tones, such as in the experiment in Section 3.2.2 on page 97. Ambiguity in pitch (e.g. bell sounds) and in relative pitch (e.g. the tritone "paradox") motivates the idea to model relative pitch as a spectrum rather than a single scalar value, in the same way as virtual pitch expands the idea of a definite unique pitch to a spectrum of pitch candidates with various prominence. Within the limited scope of this thesis we have to restrict ourselves to the pitch comparison of complex tones. Like in virtual pitch, for each partial of each complex tone its prominence (spectral weight), given by the existence and dominance region in virtual pitch, is calculated (Section 1.1.6 on page 38). For each partial the spectral weight is expanded to its subharmonic series that is then decreasingly weighted. For a given frequency difference of subharmonics their spectral weights are multiplied. These products of subharmonic weights are summed up across all subharmonic pairs of the same frequency interval. Finally we yield a spectrum of pitch difference indicating various degrees of prominence of the

intervals.

In our approach, to each compared pair of complex tones, we assign a *spectrum of pitch difference (SPD)*. In this mathematical framework, pitch comparison is an operator that maps two functions – the spectra of the compared complex tones – to another function Γ of pitch differences δp. $\Gamma(\delta p)$ denotes the intensity of the perceived pitch difference δp. A pair of complex tones with an ambiguous pitch relation should lead to a SPD with two or more peaks at the same level, whereas distinct global pitch differences should show a single significant peak in the SPD (e.g. when pure tones are compared).

To give the explicit equation, the equation for calculating virtual pitch in Terhardt [1998], p. 362-364 (Section 2.2.2 on page 76) is adapted to pitch intervals:

$$\Gamma(\delta p) := \sum_{\substack{1\le j\le N \\ 1\le k\le M \\ \delta p=\log_2 \frac{f_j m}{f'_k n}}} \sum_{1\le n,m\le 12} \frac{S_j S'_k}{nm} \qquad \text{(SPD)}, \tag{3.2}$$

where $f = (f_1, \ldots, f_N)$ and $f' = (f'_1, \ldots, f'_M)$ are the spectral frequencies of the two complex tones. $S = (S_1, \ldots, S_N)$ resp. $S' = (S'_1, \ldots, S'_M)$ are their associated spectral weights (Equation 2.15 on page 76). For each partial f_j of the first tone f, the spectral weight S_j/m of the subharmonic f_j/m is compared to the spectral weight S'_k/n of the subharmonic f'_k/n of each partial f'_k of the second tone f'. [4] Products of spectral weights are added, if their corresponding subharmonics have the same logarithmic pitch dissimilarity $\delta p = \log_2 \frac{f_j m}{f'_k n}$. A reversed order of pitch comparison leads to a mirrored SPD ($\Gamma(\delta p) \to \Gamma(-\delta p)$).

Applications To give application examples, we will first calculate pairwise SPD between the harmonic complex tones used in our experiment. Then we will determine the SPD between Shepard tones. We calculate the pitch difference between successive tones in the scale of harmonic complex tones and in the Shepard scale. If the maximum of the spectrum of pitch difference is either always positive or always negative circular pitch perception is indicated.

Let us look at the SPD of circularly perceived tone series, of harmonic complex tones in Figure 3.11 on the next page and of Shepard tones in Figure 3.12 on page 108. For a qualitative description of the experiments by the SPD, the SPD must be interpreted. If the maximum of the SPD Γ lies left from the vertical zero line ($\delta p < 0$) it means that the subjects predominantly judge "the first tone is lower then the second one". Respectively $\delta p > 0$ means "the first tone is higher then the second one". Both tone sequences are circularly ascending meaning that most subjects judged the tone

[4] This idea is inspired by Terhardt's principle of subharmonic coincidence for calculation of virtual pitch (Terhardt [1998], p. 363-364, cf. Section 2.2.2). Though we do not determine virtual pitch explicitly, the SPD equation can be considered to some extent a comparison of virtual pitches.

of the k-th row lower then that of the $k+1$-th row and the last tone lower then the first one. These results are consequently reflected by the corresponding SPD: Each SPD has its absolute maximum for a $\delta p < 0$. Thus the SPD describes the experimentally found circular structure of pitch relations. Amplitudes of two subsequent tones according to Figure 3.5 yield an octave ambiguous interval perception. The octave ambiguity is apparent in the two most prominent peaks in the right column of Figure 3.11.

In Figure 3.12 and 3.11, virtual pitch is calculated according to Equations 2.15, 2.16, and 2.18 on page 77. For plausible reasons but with no experimental psychoacoustic verification so far, in contrast to Terhardt [1998], virtual pitch calculation starts from inciciveness (Equations 1.8,1.9, and 1.12 on page 36).

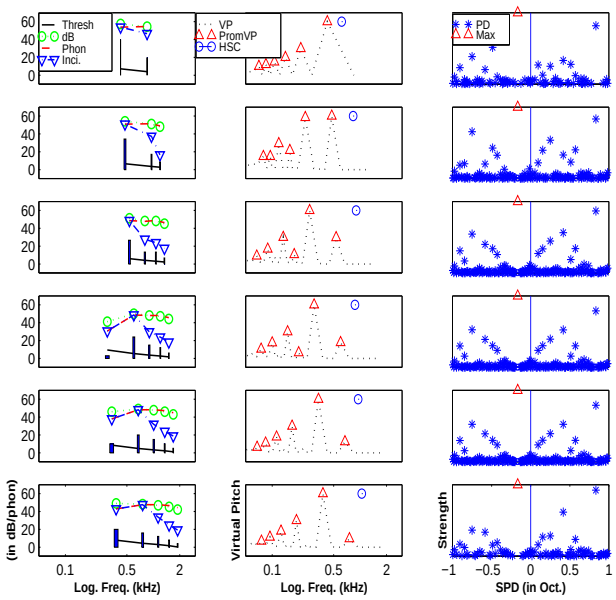


FIGURE 3.11: Pitch determination of the harmonic complex tones in the experiment (cf. Figure 3.5). The equation models the perception of a circular sequence of harmonic complex tones. In the *left* column we see scaled intensities (bars), loudness (○), incisiveness (dashed, ∇) of the harmonic complex tones, and the hearing threshold (solid line). In the *middle* column virtual pitches (dotted, cf. Section 2.2.2) are indicated including the six most prominent virtual pitches (△). The *right* column displays the spectrum of pitch differences (SPD) of complex tones belonging to this and the subsequent row. The most prominent value of SPD is indicated by △. The SPD in the last row belongs to the complex tones of the last and the first row. The △ is always on the same side of the vertical zero line throughout the right column. That indicates circular pitch perception.

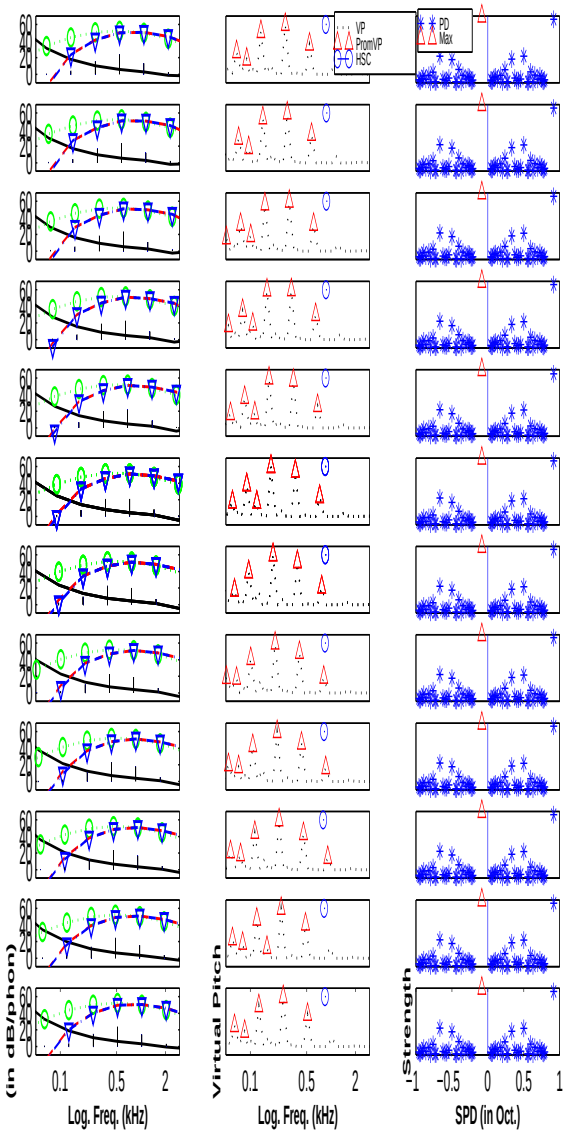


FIGURE 3.12: Pitch determination of the chromatic Shepard scale. (Cf. Figure 3.11 for explanation.)

3.2.5 General Construction Principle

We will outline a general construction algorithm for circular pitch. The algorithm is capable of generating Shepard tones as well as the stimuli used in the experiment in Section 3.2.2. To provide a more general framework, we describe the continuous version of circular pitch, the circular *glissando*. From that, stepwise circular pitch sequences, such as the ones discussed above, can be deduced. We superimpose

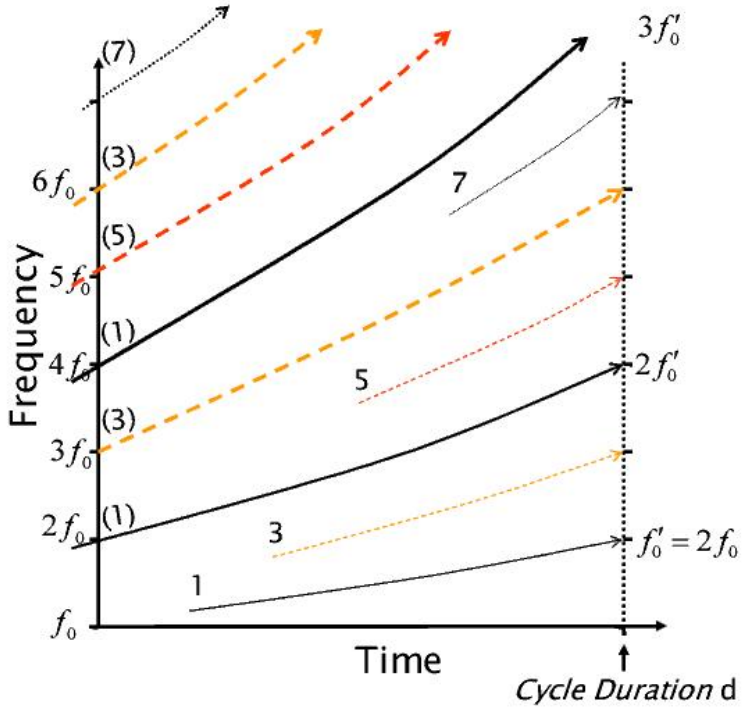


FIGURE 3.13: Generalized Shepard *glissando*: The spectrum of the sound reaches the same state (right dotted vertical line) after cycle duration d that it used to have in the beginning (left vertical axis). The *glissando* repeats without disruption. The solid lines sketch out the continuous Shepard *glissando*, consisting of "sinusoidal" components in octave distance, with a spectral envelope above the basic start frequency f_0. The partials between octave components are generated by *glissandi* from uneven multiples of the basic start frequency. In order to avoid discontinuity, *glissandi* starting from lower basic frequency fade in earlier and more slowly.

glissandi with bell shaped envelopes. Each *glissando* starts at a different time and with a different "partial". We can only coarsely outline the concept here. It is beyond of the scope of this thesis to prove and evaluate the construction recipe in detail. This task has to be left for future work.

To start with, we choose a Shepard tone with a spectral envelope that is squeezed so that all octave components below the basic start frequency f_0 become inaudible (Figure 3.13). Shepard tones only consist of partials in octave-multiple relationships to each other. For harmonic complex tones, partials in between the octaves have to be added. To generate N partials and their upper octaves within a circle duration d, we have to superimpose the initial Shepard *glissando* G_1 with other Shepard *glissandi* G_n that are shifted in frequency and enter at a time offset. The additional Shepard *glissandi* start with uneven multiples $(2n-1)f_0$ $(n = 1, 2, \ldots, \lceil \frac{N}{2} \rceil)$ of the start basic frequency f_0. To avoid disruption in the progression of the sound, all Shepard

glissandi enter at a distinct time $d_n = \frac{(2n-1)}{N}d$. The lower components start first. An abrupt entrance of these components would not be masked by lower frequencies, and hence would be more obvious. The additional Shepard *glissandi* G_n differ from the initial Shepard *glissando* G_1 in two aspects: the shifted frequency position and the shrunk spectral envelope.

3.2.6 Circularity in Music Literature

How is circularity of musical attributes referred to in music? First we will discuss examples of circular pitch. Then we will shortly describe the role of circular tempo and loudness (p. 117) in selected musical examples.

Circular Pitch For circular pitch, the following properties are constituent: first, a steady movement up or down respectively, second, brightness decoupled from pitch class, and third, the perceived similarity of consecutive tones. For selected music samples, we will set out to what degree these criteria are or are not met. Furthermore, the role of circular pitch in the overall context of the piece of music will be discussed. The ascension of pitch class is realized as scales. Due to its uniformity the chromatic scale is especially well-suited to support circular pitch. Other scales are used also, sometimes in conjunction with special tuning systems. Sophisticated variations on the chromatic scale occur. Brightness is characterized by the mean (*spectral centroid*) and the width (*spectral spread*) of the spectral envelope integrated over a small section of the piece (cf. Section 2.2.2 on page 78). Do the varied scales move across several octaves or do they remain within the range of a high or low register? As long as the variation of brightness is not directly synchronized with the pitch class trend, brightness does not need to be kept constant. The *brightness trend* can go in a direction opposite to the trend of pitch classes, can fluctuate, or can imitate the trend of pitch classes in an alleviated manner. A homogeneous circular pitch class trend is perceived, when consecutive sounds *smoothly* morph into each other, due to their proximity in the spectrum of virtual pitch. The question rises, whether the drive occurs synchronously in several voices that fuse together. For ascending pitch class trend this is realized in the following way: Before an ascending line fades out, a new ascending line fades in almost imperceptibly. Sometimes varied scales ascend in parallel. Are the tones then grouped as chords or chordal streams? Is a *crescendo* used in order to camouflage the onsets of low pitch, a *decrescendo* for the off-sets of high pitch? What does circular pitch mean in the particular musical context? Does it serve as a metaphor for the central musical idea? Does it take the part of a background foil, behind another musical action? Or is circular pitch just the starting point for a musical development that leads further away from it? In the sequel we will discuss these issues based on scores and recordings. It is beyond the scope of this thesis to explicitly calculate the SPD and other features based on mathematical formalism, such as spectral centroid and spectral spread, brightness trend, and smoothness, even though the equations of Aures [1985] and Peeters et al.

[2000] can be employed to do so (cf. Schneider [2005]).

Alban Berg anticipates Shepard scales in Wozzeck (1925, Act III, Scene 4, Lévy [2004]). Wozzeck is drowning in the pond. The vanishing rising air bubbles and the decreasing concentric wave motion are onomatopoetically illustrated. Chromatic scales rise. The scales cross-fade. Before one scale decreases in loudness in the high register, another one starts in the low register, increasing in loudness. Remarkably, Berg explicitly notes *crescendo* and *decrescendo*. Electro-acoustically generated sounds in Ernst Krenek's Pentecost Oratorio Spiritus Intelligentiae Sanctus (1956) are given by Stuckenschmidt [1969] as another anticipation of circular pitch. The slow descending *glissando* in the beginning of the fourth region of Hymnen (1966-67) by Karlheinz Stockhausen can be linked to circular pitch as well [Lévy, 2004]. Jean-Claude Risset uses descending Shepard *glissando* in the movement Fall of the computer-generated tape Little Boy (Computersuite 1968). First he uses fast cycles of descending Shepard *glissandi* in its pure form, adding other descending *glissando* tones. Later he playfully employs interference of various discrete and continuous ascending and descending cyclic pitch class trends. He uses different cycle durations and sound textures. In the movement Contra-Apotheosis such sections alternate with wavering pitch class trends of Shepard tones. In one part brightness changes cyclically. Circular pitch stands for the trauma of the pilot, who dropped the atomic bomb on Hiroshima. In his hallucination the bomb falls endlessly without ever reaching the ground. Circularly ascending and descending pitch sequences cross-fade as an onomatopoeia of a schizophrenic mind. A sparse Shepard scale can be found in Ligeti's Doppelkonzert f r Fl te, Oboe und Orchester (1972), second movement bars 69-70 (Sabbe [1987], p. 11).

In the beginning of the movement Bellringer's Song, about 1200 in Manfred Stahnke's Partota (1983-85, Figure 3.14 on the next page) the pitch class trend is performed by the descending diatonic scale $b\sharp, a\sharp, g\sharp, f\sharp, e\sharp, d\sharp, c\sharp$. Mixtures of the following kind are used: The right and left hands play tones out of triads that are transposed relative to each other by a major tenth. As in Hofstadter [1979] [5], small and big notes graphically indicate that chord components subtly fade in and out. However, while listening to an interpretation (Hofer) an endlessly descending scale cannot be perceived. Between the second and the third chord stroke of the left hand brightness constantly decreases along the direction of the pitch class trend. Thereby particular tones in the mixture are highlighted, forming a new melody. The circular pitch motion builds the background against the chords in *fff* in the foreground. Opposing the descending pitch class trend, the accelerating chords ascend.

Ligeti's L'escalier du diable (Floros [1996]; Schneider [2005], Figure 3.15 on page 113) from Études pour piano – deuxième livre (1988-94) starts with a variation on a chromatic run. The left hand plays the chromatic scale. At the same time the right hand escalates in a pendulum movement consisting of alternating major seconds and major thirds, rhythmically grouped in sections of nine, seven, nine,

[5] p. 766 of the German edition.

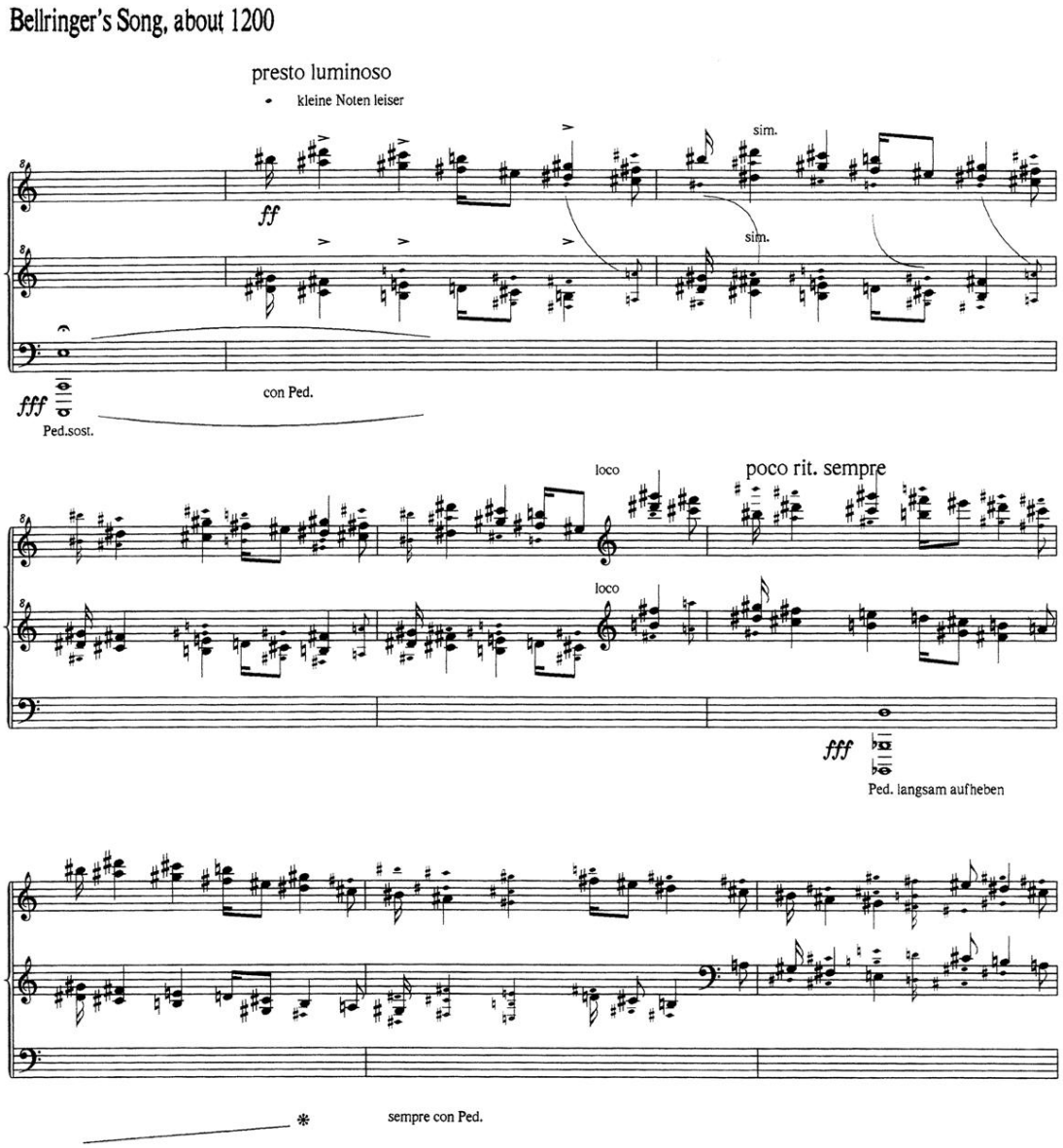


FIGURE 3.14: The beginning of the fourth movement BELLRINGER'S SONG, ABOUT 1200 of Manfred Stahnke's PARTOTA. The cycle of mixture sounds repeats barwise. From the *fff* chord in the left hand in bar 6, the spectral envelope falls slowly. By this envelope shift the emphasis of particular, thereby highlighted, tones changes. Out of the, up to then, synthetically perceived mixtures, a melody consisting of these emphasized tones can be analytically singled out.

and eleven eighth notes. Consisting of thirty-six eighth notes, this rhythmic pattern repeats across the first nine bars, comparable to the medieval *talea* concept [Gann, 1995; Purwins et al., 2005]. In addition, the piece offers other implementations of the idea of pitch class rise. Three- and four-tone chords, often constructed in layers of thirds, tie over and slowly push chromatically upward (bars 37 middle – bar 43). Chromatically ornamented melodic phrases (bars 11-12, 15-17), and chromatic chord shifts (bars 7-9) contribute to the sense of a rise throughout the piece. On a coarser time scale, the ascension is echoed by a sawtooth pattern. The ambitus is shifted upward slowly and smoothly, interrupted by harsh crashes (bars 10, 17, 26, and 43) that articulate the formal structure. Twice the piece breaks through the ascending spiral, at a culmination in bars 29-32 ("wild ringing of bells") and in bar 46 until the end, an onomatopoeia of bells ("like bells, gongs, tamtams"). Circular pitch impressively works for conveying the central idea of the piece: the Sisyphean labor, "the fruitless

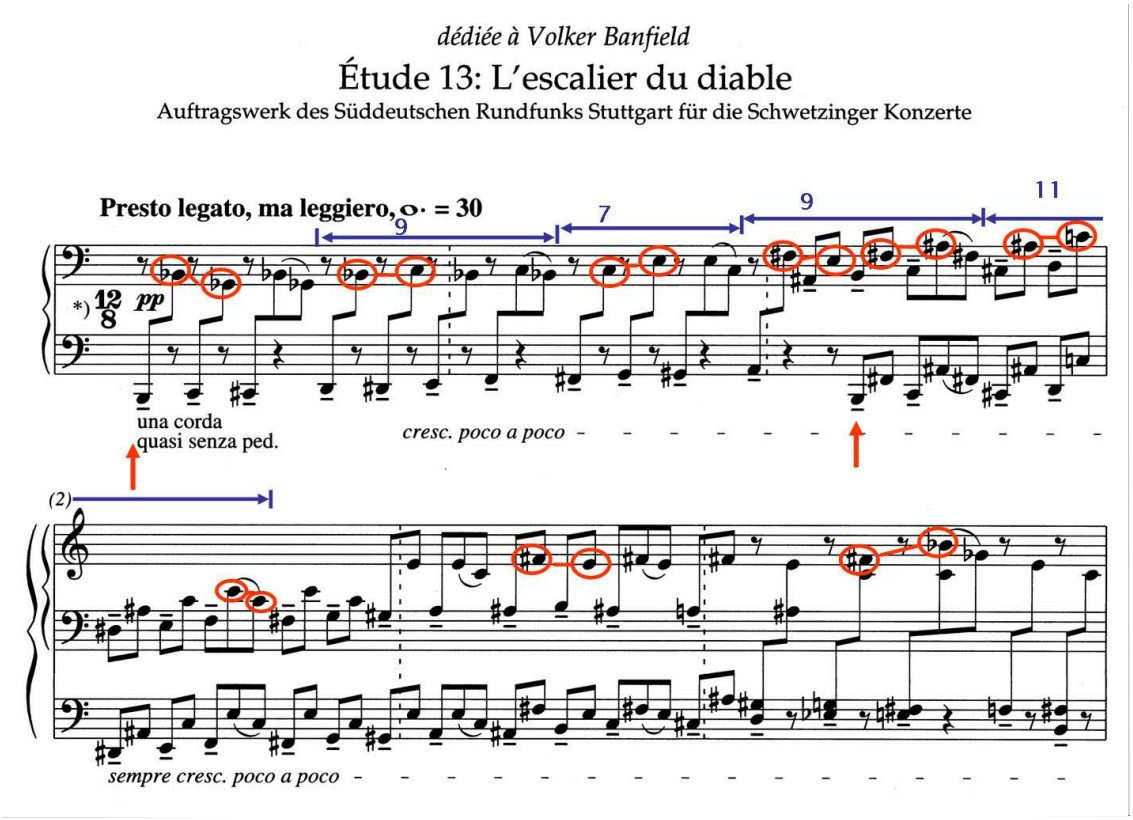


FIGURE 3.15: Ascending passages with shifted onsets in the beginning of L'ESCALIER DU DIABLE from Ligeti's ÉTUDES POUR PIANO – DEUXIÈME LIVRE. The left hand plays a chromatic scale. The right hand escalates in major thirds and seconds (encircled notes). This pitch organization interferes with a rhythmical cycle consisting of nine, seven, nine, and eleven eighth notes. (reprinted with friendly permission from SCHOTT MUSIK INTERNATIONAL, Mainz)

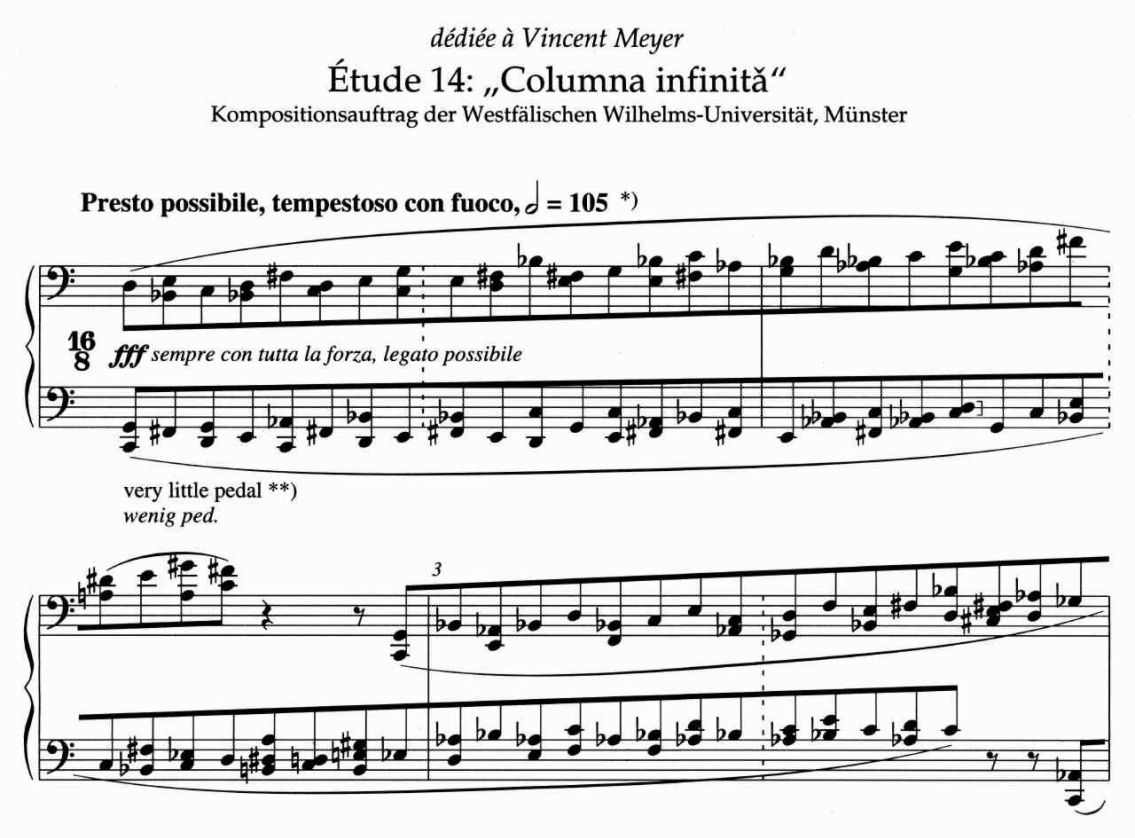


FIGURE 3.16: The beginning of COLUMNA INFINITĂ from Ligeti's ÉTUDES POUR PIANO – DEUXIÈME LIVRE. (reprinted with friendly permission from SCHOTT MUSIK INTERNATIONAL, Mainz)

effort to ascend" (Floros [1996], p. 193).

Ligeti's etude VERTIGE reveals flexible employment of spectral envelopes. The principal layer of the piece is built from a very fast descending chromatic scale, consisting of sixteenth notes. First, the ambitus remains the same until bar 4, then it expands by semitone steps. In the general development the spectral envelope describing brightness first remains at about the same, then rises, fluctuates, and descends again toward the end [Toop, 1999; Schneider, 2005].

Ligeti's etude COLUMNA INFINITĂ (1988-94, Figure 3.16) is based on ongoingly rising eighth notes. The spectral envelope remains more or less constant [Schneider, 2005]. In several aspects this etude resembles L'ESCALIER DU DIABLE, except, COLUMNA INFINITĂ appears more static, in repeated blocks of rising figuration.

The idea of circular pitch is also the principal ingredient of the third movement INTERMEZZO (Figure 3.17 on the next page) of Ligeti's VIOLIN CONCERTO (1990/92, cf. Floros [1996], p. 215 and 221-2). Closely resembling Shepard [1964]'s original sequence, throughout the entire movement, chromatic runs with offset entrances are played by the strings. They are perceived as a curtain moving from background, at

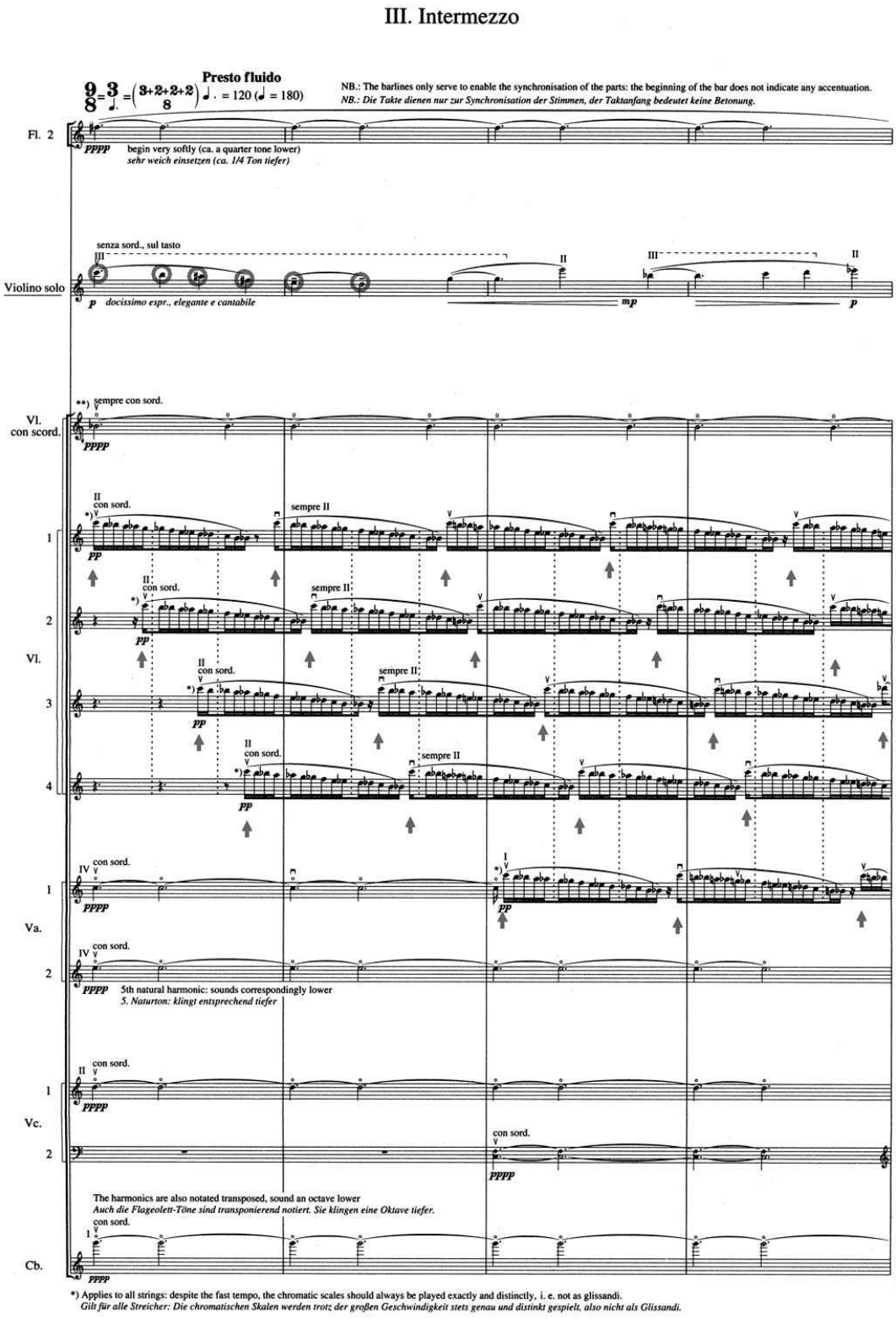


FIGURE 3.17: The beginning of the third movement INTERMEZZO from Ligeti's VIOLIN CONCERTO. (reprinted with friendly permission from SCHOTT MUSIK INTERNATIONAL, Mainz)

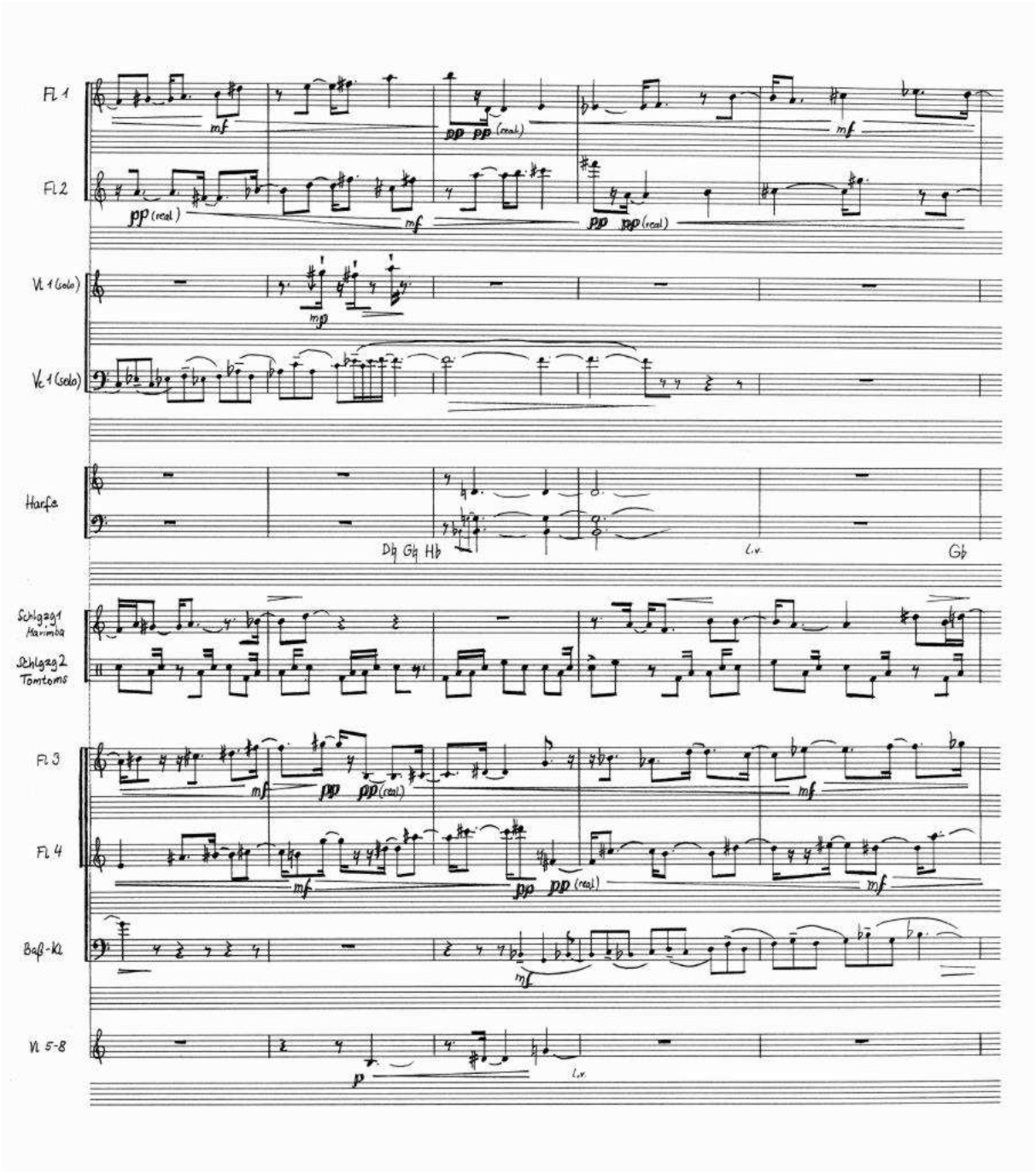


FIGURE 3.18: The second page of Hans Peter Reutter's THE SHEPHARD'S FLUTE from ORCHESTRAL SET NO. 1. The four flutes play freely rising melodies with shifted entries. The melodies tend to break the pattern of strict ascension. First cello and bass clarinet escalate in triplets.

an almost inaudible loudness level (*pp*), to foreground, played at *ffffff* from bar 72. An offspring of this pitch class movement is heard in the beginning of the principal theme of the solo violin: six descending notes, composed of three notes of a chromatic scale and three notes of a whole tone scale, connected by a major third. For accompaniment, the chromatic runs increase in range from fifteen to twenty-five tones by bar 28. The range of this pitch material starts with $b\flat^1 - c^3$ and finally reaches b^3 by bar 42. Then the spectral envelope separates into a high and a low frequency region. While the lower frequency region falls, the upper region narrows, limited by $b\flat^3$ until it merely consists of two tones, $b\flat^3$ and a^3. Playing *cantabile*, the solo violin fills in the omitted middle range of the spectrum. Circular pitch perception is supported by two factors. First, the selection of instruments (strings, later flutes and clarinets) supports fusion and thereby synthetic hearing. Second, the irregular entries of the falling runs follow closely one after another. In this piece, circular pitch appears airily in character. Finally, the sixteenth-note runs reach the range $B\flat_1 - b^3$. This momentum takes over the entire orchestra, including the solo violin.

Four flutes play freely rising melodies in The Shephard's Flutes by Hans Peter Reutter's Orchestral Set No. 1 (1993, Figure 3.18 on the facing page), composed of microtonal pitch material. The design of the ascending melody is very alive, often breaking the pattern of strict ascension. The flutes are accompanied by the spiraling triplets of the first cello and the bass clarinet, continued from the previous movement. The spiraling momentum resembles the right hand part in the beginning of Ligeti's L'escalier du diable (cf. Figure 3.15 on page 113). The brightness does not change significantly during this movement. Despite the free design of the rising movement, a sense of the circularity of pitch still remains for two reasons. First, the ascending melodies enter with a *crescendo* and fade out with a *decrescendo*. Second, all melodies enter offset to each other. Propelled by percussion, circular pitch acts as a *perpetuum mobile*, continually generating new rising melodies.

Other examples include: James Tenney's For Ann (rising) (1969) and Georg Hajdu's Fingerprints for Piano (1992). Lévy [2004] further mentions Manfred Trojahn's Enrico (premiered in 1991) and Was ihr wollt (premiered in 1998), Marc-André Dalbavie's Concerto pour Violon (1997), and Fabien Lévy's Coincidences (1999).

Circular Tempo and Loudness Meter can be obliterated. Different meters can exist simultaneously. The Aka pygmies make extensive and elaborate use of the interference of simultaneous meters. In Western music, the latter play an important role in the work of Colon Nancarrow [Gann, 1995], tracing back to Chopin and Brahms (cf. Purwins et al. [2005] for more details). Metrical confusion fertilizes the ground to cultivate another "paradox" upon.

Downscaling frequency is the means for adapting circular pitch to the domain of tempo. The spectrum of partials of a complex tone is interpreted as a spectrum of beat frequencies in a slower time scale. But, in contrast to pitch, phase has a more

important impact for rhythm.

We can reinterpret a sine tone of frequency 440 Hz to a pulse of 440 beats per second. Slowing down by a factor of 100 results in 4.4 beats per second, equivalent to a tempo of 264 MM. We apply the analogy straight forward to a single Shepard tone, shown in Figure 1.10 on page 50. After suitable rescaling and assigning each partial to a different percussive timbre, the Shepard tone turns into overlaid pulses of frequencies $\ldots, \frac{f}{4}, \frac{f}{2}, f, 2f, 4f \ldots$. To enforce the tempo "paradox", ideally, the percussive timbres assigned to each tempo should continuously morph into each other. While speeding up to double tempo, the timbre of a pulse should steadily transform into the timbre associated with double tempo. For such a *Shepard rhythm*, we let the beats slide under the envelope that in this framework defines the intensity of the beats. Then a steadily accelerating tempo is perceived. This train of thoughts leads to circular tempo by Risset and K. Knowlton [Risset, 1985]. If we replace frequency with pulse speed, all considerations for pitch and Shepard tones apply for tempo and Shepard rhythm as well. The equivalent of octave ambiguity for pitch is the ambiguity in perceiving a given tempo or the tempo of doubled speed. The theoretical explanation exploiting the gestalt principle of proximity for relative pitch (Section 3.2.3 on page 103) is applicable to tempo as well.

There is a rhythmical pattern in music literature that resembles the Shepard rhythm (Figure 3.19). The pulse sequence speeds up, while fewer beats are emphasized. First every beat is emphasized, then every second, every third, and so forth. At the same time, the volume of the the unstressed beats is reduced. This variant is realized in Arthur Honegger's Pacific 231 (1923). A smoother version of this pattern

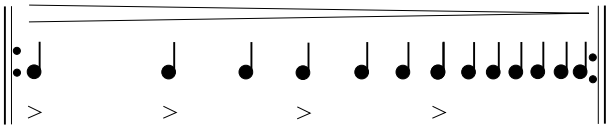


Figure 3.19: The ancestor of the Shepard rhythm. A beat sequence accelerates while decreasing in loudness. At the same time the emphasized beats become more rare. After the fade out of the fast beats the emphasized beats constitute the main beat. The cycle repeats again.

can be generated in sound post-production. The accents are then replaced by manipulated temporal volume envelopes. Given an accelerating pulse sequence, the peaks of the volume envelopes are gradually shifted from the attack phase of the tone towards the sustain phase and furthermore to the attack phase of the following tone. Similar ideas are applied in techno music. Risset presents a decelerating sequence, consisting of only one beat pair, trickily manipulating phase and degree of fusion to support the intended circular tempo perception (Figure 3.20). The sequence starts with a pulse with regular beat. Then a second pulse enters in a low voice shortly after the first pulse. At that stage both pulses almost fuse. Gradually the entry of the second pulse is shifted towards the position exactly in the middle between two beats of the main pulses, increasing volume until the second pulse reaches the volume of

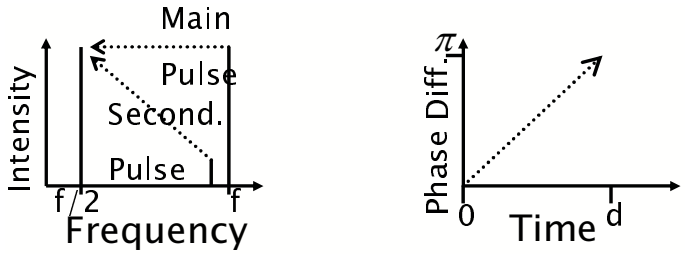


FIGURE 3.20: Idea of Risset's circularly accelerating sequence of a pair of pulses. Circular pitch is adapted to the tempo domain. In addition, phase and degree of fusion are controlled. Initially the pulse of frequency f almost fuses with the slightly offset secondary pulse. During a cycle duration d the main pulse decelerates down to tempo $\frac{f}{2}$ (*left*) and the secondary pulse reaches loudness of the main pulse (*left*). Gradually the entry of the secondary pulse is shifted in phase relative to the main pulse (*right*). After a time d it enters exactly in between two main beats. Then we hear a single pulse of frequency f and the cycle starts again.

the main pulse. By then the two pulses cannot be distinguished any longer. They form one homogeneous pulse identical to the initial pulse. The cycle starts again.

Adapting circular pitch to loudness, how would a culmination sound? The specific loudness of a single component rises slowly, then falls abruptly, while in another frequency band a new tone slowly increases in loudness, and so forth... Richard Wagner uses this technique in the endings of his operas. For the final culmination, string *tremoli* are played in a sawtooth pattern of loudness. A *tremolo* on one tone slowly rises in loudness from *piano* to *forte*, before abruptly falling back to *piano*. A *tremolo* on another pitch slowly rises again. Despite the abrupt withdrawal of loudness such passages are perceived as continuously stepping up.

In Section 3.5.1 we find a more abstract adaptation of the idea of circularity in the domain of tone centers.

3.2.7 Discussion

The SPD (Equation 3.2 on page 106) has been mainly designed to express the circular perception of "paradox" sound sequences, such as the stimuli of our experiment as well as Shepard tones. For the analysis of a greater variety of sounds, the SPD needs to be modified further. Subharmonics of two partials do not always match exactly. An analyzed tone may be out of tune or its partials may be approximately, but not exactly, harmonic. To cope with small deviations of subharmonics, it would be useful to adopt the degree of inharmonicity γ_{jk} from Equation 2.17 on page 77 for calculating coincidences. The corpus of musical data to be analyzed could include the pieces discussed in Section 3.2.6 on page 110 and Pierre Schaeffer's "paradox". The SPD is not limited to describe "lower/higher" pitch relations. It provides more

detailed information. It can be used to calculate ambiguous relative pitch also, for instance, when comparing the pitch of two bell sounds. An appropriate experiment for detailed testing of the SPD and the general construction principle of circular pitch (Section 3.2.5 on page 108) is not designed so far.

3.3 Pitch Class Content of Musical Audio

We have indicated how to represent pitch by a vector of virtual pitches.[6] Another pitch representation is the correlogram. (Cf. Section 2.2.2 on page 79 and Figure 3.50 on page 160 a.) In contrast to virtual pitch and correlogram, a profile of pitch classes (Section 1.2.2 on page 48) can be calculated more effectively and more reliably and it can be applied to music theory more easily. In the simplest case (Figure 0.1 on page 15), a binary pitch class profile indicates which pitch class is included ("1") and which is not ("0"). (Introduction p. 14) Taking statistics and durations of pitch classes into account, we yield the pitch class profile of frequency of occurrence and the profile of overall durations (p. 14). These profiles are similar to the psychological probe tone ratings (Section 1.2.9). In a harmonic complex tone,[7] the fundamental, the 2nd, and the 4th partial have the same pitch class. But the 3rd, 5th, and 6th partial have other pitch classes. For this and other reasons, such as temperament, inharmonic partials, spectral leakage, onsets, and noise, in a pitch class profile from audio one tone yields intensities for several pitch classes.

First we describe how to calculate pitch class profiles from audio, how to compare them, and how to generate key prototypes. Then we compare these profiles with the probe tone ratings, calculate, and discuss the profiles of the analyzed musical corpus.

3.3.1 Constant Q Profiles

In this section we develop a new concept of key profiles (cf. Introduction, p. 15), called *constant quotient (CQ-)profiles*. One of our aims is to build a system for automatic key-finding in audio recordings. It is desirable that the tonal hierarchy on which the system relies is derived from music literature itself and not merely from music theory (cf. Introduction). This is a more direct way and ensures the flexibility of the architecture. Such a system can be adapted to different musical styles simply by using a reference set (cf. p. 123) that is built from material of that style. In Section 3.9 we present several experiments which give strong indications that the CQ-profiles are a suitable tool for the tracking of tone center transitions (Section 3.9 on page 168).

The CQ-profiles unite features of Krumhansl and Kessler [1982]'s probe tone ratings and Leman [1995]'s correlograms (Fig. 1.19 on page 62 and 3.50 on page 160).

[6] Cf. Section 2.2.2 on page 76. Leman [1995], p. 49, uses a simplification of the virtual pitch model, the subharmonic sum spectrum.

[7] Definition on p. 32.

Advantages include:

- Each CQ-profile has a simple interpretation, since it is a 12–dimensional vector like a probe tone rating. The value of each component corresponds to a pitch class.
- A CQ-profile can easily be calculated from an audio recording. Since no complex auditory model or other time consuming methods are used, the calculation is very quick, and can be done in real-time.
- The calculation of the CQ-profiles is very stable with respect to sound quality. In Sections 3.8.3 and 3.9 we pursuit an analysis based on Alfred Cortot's recording of the Chopin Préludes from 1933/34.
- The CQ-profiles are transposable. That means a CQ-profile calculated from a transposed piece of music is the transposition of the CQ-profile of the original piece. In Section 3.4 this property will prove to be very useful.

Normalized constant Q (NCQ-) profiles [8] are derived from constant Q profiles by transposition to the same reference key. That means that the first entry in the profile corresponds to the keynote and the last entry corresponds to the major seventh.

Calculation of CQ-Profiles

In this section, we will describe how a short-time CQ-profile is calculated from some audio sequence and how successive profiles are summarized in a long-term CQ-profile. First the digitized pieces are converted to 11025 Hz mono, in WAV format. As underlying DSP method for further processing serves the constant Q transform which has been presented in Section 2.1 on page 69. The Matlab implementation by Benjamin Blankertz [Purwins et al., 2000b] is employed. First a reasonable choice for the frequency range and the resolution of the constant Q transform has to be made. To cope with spectral leakage effects, we decide to use a resolution of $b = 36$ bins per octaves instead of 12. To cover a wide range of tones we chose the minimum frequency f_0 about[9] 131.25 Hz (corresponding to c) and a maximum frequency five octaves above. Tones below the minimum frequency are also captured, by their harmonics. So there will be 12 CQ-bin triples per octave. In approximately equal temperament, only CQ-bins in the middle of such triples correspond to tones of the chromatic scale and they will be protected from spectral leakage due to the high frequency resolution. The outer bins are used to check for note onset and other inharmonic noise.

To concentrate the so found frequency information in a 12-dimensional vector $\mathbf{x}^{\mathrm{cqp}}$ – which we call CQ-profile – all the values that belong to one pitch class are

[8] In Purwins et al. [2004a] NCQ-profiles are named scale degree profiles.

[9] To be exact: f_0 is chosen "one CQ-bin below" such that the frequency of the first scale tone resides in the middle of the first CQ-bin triple. The value should be trimmed to the tuning of the recorded instrument(-s). This can be done by a heuristics (cf. p. 123).

summed up from the constant Q transform $\mathbf{x}^{\mathrm{cq}}$ (Equation 2.5 on page 70) and the result is converted to dB. For $0 \leq k < 12$ we define:

$$x^{\mathrm{cqp}}[k] = \max(20 \log_{10}(\sum_{n \in \mathcal{N}(k)} |\mathbf{x}^{\mathrm{cq}}[36n + 3k + 1]|) + dbT, 0), \tag{3.3}$$

where $\mathcal{N}(k) = \{n < \frac{N}{36} : \mathbf{x}^{\mathrm{cq}}[36n + 3k + 0] < \mathbf{x}^{\mathrm{cq}}[36n + 3k + 1] > \mathbf{x}^{\mathrm{cq}}[36n + 3k + 2]\}$, N being the number of bins of the CQ-transform. dbT is a threshold value that should be chosen in a way so that contributions from background noise stay in the negative range. If not otherwise stated we use $dbT = 170$. The effect of this summation process is similar to working with Shepard tones in Leman [1995]'s system.

A triple CQ-bin per scale tone provides a possibility of detecting inharmonic frequency contributions like onset noise or crackling. A spectral contribution that is reasonable in our sense has a frequency near the center frequency of some middle CQ-bin (middle with respect to the triple). So in this case the magnitude of the middle bin towers above the side magnitudes. On the other hand frequencies stemming from a note onset blast are spread over several CQ-bins with the strongest intensity at the note's frequency. This simple criterion for filtering is realized by taking the sum in (3.3) only over those CQ-bins whose magnitude is a local maximum. Figure 3.21 shows a selection of a constant Q transform of a *forte* minor third c – e♭ (on a dB-scale) played on the piano.

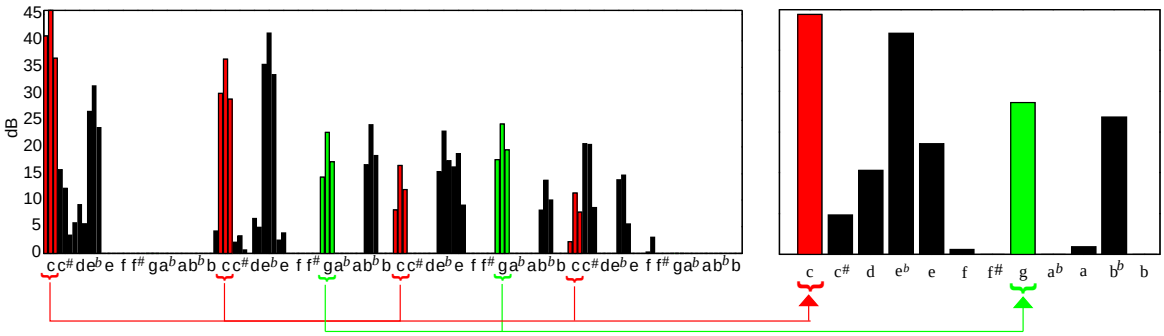


FIGURE 3.21: The constant Q transform is calculated from a minor third c – e♭ (played on piano) with three bins per half-tone (left figure). We yield the constant Q profile (right figure) by summing up middle bins for each tone over all octaves.

In most applications of analysis filters the trade-off between frequency and time resolution is delicate. Since in the present application subsequent filter outputs are summed up anyway, we are free to choose such a fine frequency resolution. The resulting time resolution would be too coarse, for instance, for note identification.

A *long-term CQ-profile* summarizes successive short-term CQ-frames. Given a sampled signal, short-time CQ-profiles are calculated every 50 ms if not stated otherwise.

Depending on the intention of the analysis, these frames might just be summed up (giving a kind of overall tonal orientation) or the sum might be weighted by a window function. To simulate memory one can use, for example, an exponential window. We use unweighted sums in all our simulations. For better comparability we scale all profiles to have mean 0 and standard deviation 1. Equalizing the deviation is problematic for the comparison of profiles stemming from very different musical styles. For example, the average standard deviation of the preludes of Hindemith's Ludus Tonalis is more than four times as much as the deviation of any other piece of traditional tonal music discussed in this thesis.

Assuming approximately stable equal temperament in the sound sample, a simple heuristics is used to fine tune the minimum frequency f_0 of the constant Q transform. Within 128.75 Hz $\pm$ a quarter tone, f_0 is varied. For each f_0 the long-term CQ-profile is calculated. If the bin triples are in the right position the middle bin captures most of the partials. Then the sum of the CQ-profiles is maximal, indicating the appropriate tuning of f_0.

The CQ-Reference Set

A *CQ-reference set* is a set of twenty-four CQ-profiles, one for each key. Every profile should reflect the tonal hierarchy that is characteristic for its key. A more precise description depends on the specific purpose for which the reference set is calculated. The reference set can be used to display structural information, for example by a geometrical arrangement of the profiles. We also employ the reference set for key-finding. It may be used, for instance, to determine the key of a piece or a tone center of a section of a piece. For the former task a CQ-reference set calculated from sampled cadences is appropriate while for the latter moderately modulating pieces would be a good basis. In Section 3.3.5 on page 126 and in the Appendix B on page 189 we present CQ-reference sets from various sets of pieces, among them Bach's WTC. The CQ-profile of each prelude reflects many characteristics of its key. But due to the individuality of each prelude, there is some variance in the profiles. For comparison the NCQ-profiles are used. Supposing that our tonal reference set should be invariant under transposition we calculate a major and a minor mean prototype profile as the average over all major respectively minor NCQ-profiles. An invariant reference set is established by transposing the mean prototypes to all keys. This can only be done since the CQ-profiles are transposable. Alternatively one could take different pieces of music for each key and calculate the average of their CQ-profiles. But then many more suitable pieces from music literature have to be found and more material has to be processed.

3.3.2 Fuzzy Distance and Key-Finding

How can a musical signal, an entire piece or a passage thereof, be classified according to a CQ-reference set? Generally we have the problem of matching an extracted CQ-profile with a profile of the CQ-reference set. Typical matching criteria

are strongest correlation or minimal Euclidean distance. Classification is also possible with a SOM that has been trained with the profiles of the CQ-reference set (Section 3.8.3). But there is some additional structural information that is neglected by these methods, namely the uncertainty of each rating value. When comparing the NCQ-profiles of various pieces of the same composer the variance will vary from component to component. For example the poor rating for the minor second will be quite fixed throughout most music samples. But for the major and minor sixth and seventh there will be a notable variance in different minor samples due to the variable appearances of the minor scale. This uncertainty in the rating can be quantized as standard deviation and should be attached as additional information to the CQ-reference set. To make use of this information, we present a different classification method. We define a distance measure that takes the uncertainty into account. Therefore we call it *fuzzy distance* although no allusion to fuzzy logic is intended. The classification of a CQ-profile is made by choosing the key of that profile of the CQ-reference set with minimum fuzzy distance to the given profile (cf. Appendix C.1).

Let y be a value subject to an uncertainty quantized by a value σ. Typically y is the mean and σ the standard deviation of some statistical data. The fuzzy distance of some value x to y regarding σ is defined by

$$d_\sigma(x,y) := |x-y| \cdot \left(1 - \frac{\sigma}{|x-y|+\sigma} e^{-\frac{|x-y|^2}{2\sigma^2}}\right). \tag{3.4}$$

The generalization to higher dimensions is accomplished by using the Pythagorean equation as in the Euclidean metric (Equation C.6 on page 196).

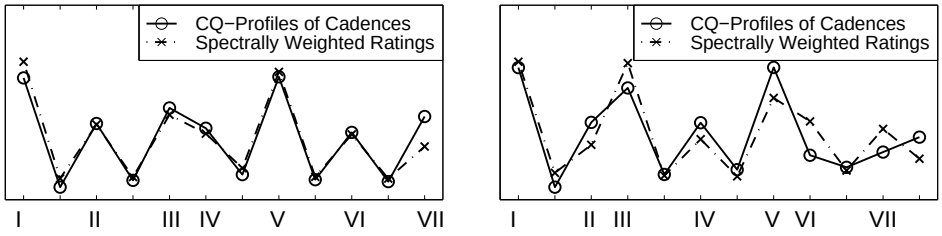
For $\sigma = 0$ the fuzzy distance equals the Euclidean metric. The greater the uncertainty the more relaxed is the metric, or rather the more relative becomes the meaning of "small".

To test the fuzzy classification method we process the preludes of Bach's WTC I recorded by Gould. With the proposed method, to every piece the right key is assigned, whereas there is one misclassification, when one of the other methods is employed.

On the 96 pieces of midi-synthesized audio (from Baroque to romanticism), 94 % of the keys are correctly assigned in a leave-one-out procedure. We take the first 15 s of each piece, weighted by a cosine window (from 0 to $\frac{\pi}{2}$). However, in this experiment, we use correlation for measuring the similarity to the reference vectors. Our approach ranks second in the MIREX 2005 audio key finding competition.

3.3.3 CQ-Profiles and Probe Tone Ratings

Krumhansl [1990] observes a remarkable correlation between the probe tone ratings and the profiles of frequency of occurrence of the twelve chromatic scale tones in musical compositions (cf. Section 1.2.9 on page 61). Krumhansl's observation gives rise to the conjecture that probe tone ratings and profiles of a CQ-reference set are



(a) Major ratings (b) Minor ratings

FIGURE 3.22: The CQ-profiles of sampled piano cadences compared with spectrally weighted ratings display high consistency. Cf. Figure 1.19 on page 62.

closely related also. But in order to establish a direct correspondence one fact has to be taken into consideration. In the CQ-profiles not only the played tones are captured but all harmonics. For piano tones the strongest frequency contribution falls (modulo octaves) on the keynote and its dominant in an approximate average ratio 3:1. Hence CQ-profiles should not be compared with the probe tone ratings, but with adapted ratings in which the harmonic spectrum of the analyzed tones is accounted for. Such a spectrally weighted rating is calculated by adding to the rating value for each tone one third of the rating value for the tone seven chromatic steps above (modulo octave). Figure 3.22 shows the highly consistent correlation of the average NCQ-profiles of sampled piano cadences (I-IV-V^7-I and I-VI-V^7-I) with the spectrally weighted ratings. In this experiment the NCQ-profiles are averaged across all major or minor keys respectively. For all cadences the octave positions of the samples are determined by random.

This close relationship indicates why CQ-profiles can be employed for automatic key-finding.

3.3.4 Musical Corpus

As musical material we choose Johann Sebastian Bach's WELL-TEMPERED CLAVIER (WTC) and various sets of preludes and fugues that encompass a piece in every key. The 226 constant Q profiles under investigation are:

1. The preludes of WTC, Book I and II, and the fugues of Book II by Bach (1685-1750) are interpreted by Glenn Gould. Another interpretation of the preludes of Book II is given by Samuil Feinberg. There are altogether 96 Bach profiles.

2. The PRÉLUDES op. 28 (1836-39) by Frédéric Chopin (1810-1849) are played by Alfred Cortot and Ivo Pogorelich, altogether 48 profiles.

3. The 25 PRÉLUDES op.31 (1847) by Charles-Valentin Alkan (1813-1888), including two pieces in C–major, are recorded from Olli Mustonen.

4. We use 21 profiles from the Preludes op. 11 (1888-1896, all pieces except the ones in A–major, b–minor, and f–minor) by Alexander Scriabin (1872-1915), played by Vladimir Sofronitsky, Heinrich Neuhaus, Vladimir Horowitz, and the composer, reproduced from a Welte-Mignon piano roll.

5. Mustonen played the Twenty-four Preludes op. 34 (1932-33) by Dmitri Shostakovich (1906-1975).

6. Ludus Tonalis (1942) by Paul Hindemith (1895-1963) contains one fugue and one interludium for each pitch class as a keynote, but neither major nor minor. The 12 fugues are part of the corpus, presented by Mustonen.

3.3.5 Composers and Pieces Portrayed in CQ-Profiles

It is an interesting question, what other references to Bach the latter pieces contain, apart from the concept as a set of preludes and fugues in all keys.

Investigations of tonal music give indications that the profile of frequency of occurrence of notes subsumed under pitch classes, eventually considering durations also, is very characteristic for its key. This should be reflected by the CQ-profiles. In Chopin and Bach their peaks in Figure 3.23 are related to the diatonic scale and

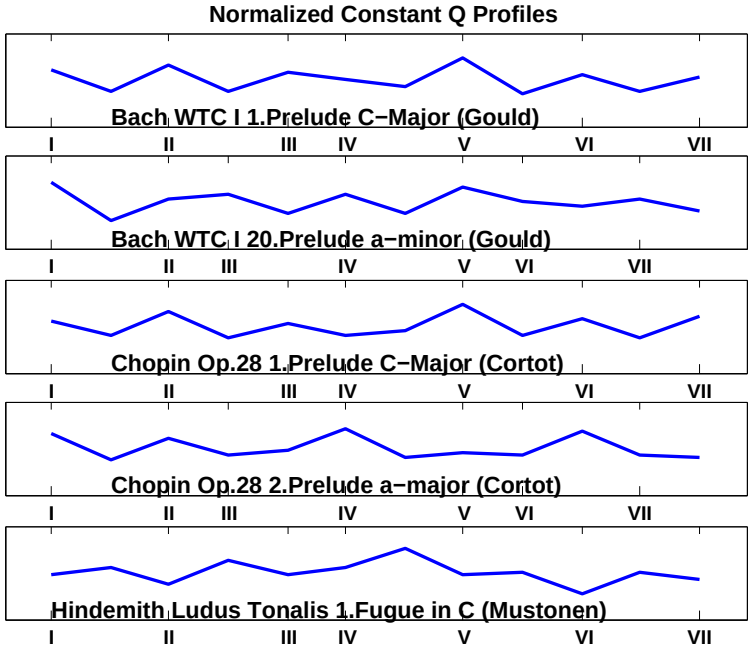


Figure 3.23: Normalized constant Q profiles for selected pieces from Bach, Chopin and Hindemith. Scale degrees are shown on the horizontal axis. In Chopin and Bach the peaks are related to the diatonic scale and to psychological probe tone ratings (Cf. Section 3.3.1). Hindemith deemphasizes the diatonic notes.

to psychological probe tone ratings (Cf. Section 3.3.1 on page 121). However, Hindemith deemphasizes the diatonic notes. We find strong conformity among CQ-profiles within WTC. Figure 3.24 and 3.25 show the CQ-profiles of all preludes of Book I, interpreted by Glenn Gould.

The high rating for the tonic triad tones and the poor rating for the minor second is consistent in most music samples. Minor sixths and sevenths exhibit a notable variability in different minor samples, due to the variable appearances of the minor scale as natural, harmonic or melodic.

An important question that arises regarding CQ-reference sets that are calculated from audio recordings is in what respect the results are affected by

(1) musical interpretation,
(2) the recorded instrument,
(3) the selected pieces of music,
(4) the data format.

For the examination of (1) we compare the radically different interpretations of the Chopin preludes by Alfred Cortot on one hand and by Ivo Pogorelich on the other hand. The mean profiles show a correlation of [10] 0.993/0.981. A comparison of the recordings of the preludes of WTC I, by Glenn Gould and Samuil Feinberg reveals

[10] When writing correlation values in the form x/y we use the convention that x refers to major profiles and y to minor profiles.

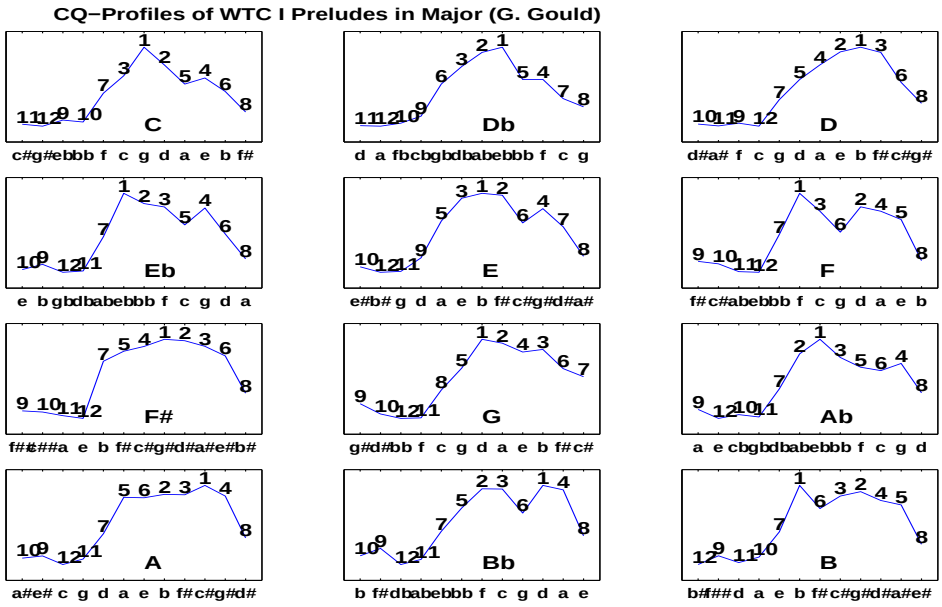


FIGURE 3.24: The CQ-profiles of the preludes of Bach's WTC I in major, played by Glenn Gould. It can be observed that the profiles can be approximately generated from the combination of two Gaussians or triangles, a big Gaussian/triangle centered at scale degree I or one or two fifths up in the circle of fifths and a small Gaussian/triangle centered at the major third.

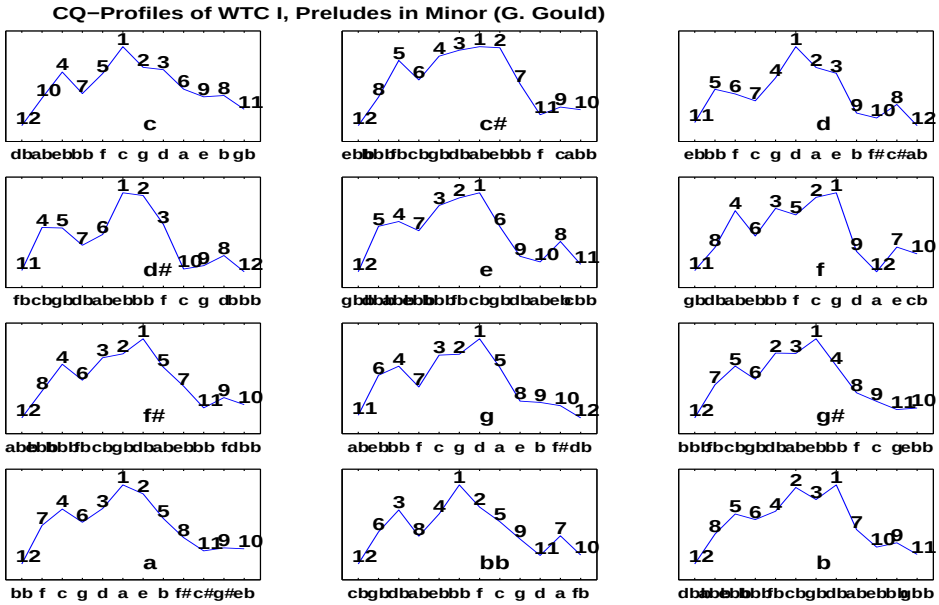


FIGURE 3.25: For the minor preludes of WTC I again two Gaussians or triangles appear. But this time the small one is on the left side centered at the the minor third.

even a correlation of 0.997 between the corresponding profiles (for both major and minor).

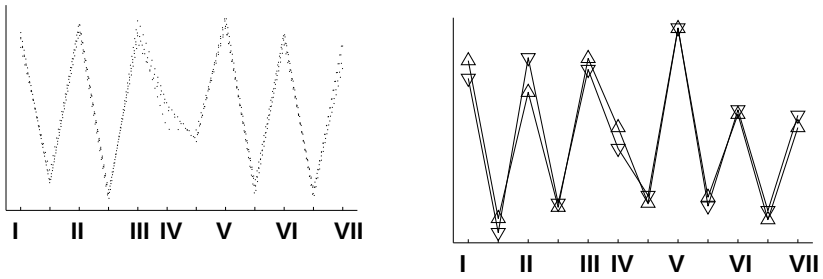
For the investigation of (2) we compare the recordings of the preludes of WTC I performed on modern pianos and on a Pleyel harpsichord by legendary Wanda Landowska. We find that the mean NCQ-profiles of the harpsichord recording correlate very highly with those of the piano recordings. The correlation with Gould's recording is 0.989/0.982.

To study the impact of the selection of music (3) on the corresponding reference sets, we perform some inter and some across epoch comparisons. Group 1 consists of four reference sets calculated from the preludes/fugues cycles (separately) of both books of WTC (Glenn Gould's recording). Group 2 consists of two reference sets derived from Alfred Cortot's recording of Chopin's PRÉLUDES op. 28 and from Olli Mustonen's recording of Alkan's PRÉLUDES op. 31. Group 3 consists of a reference set based on Scriabin's PRELUDES op. 11. The inter-group correlations are 0.992/0.983 for the Bach reference sets (mean value) and 0.987/0.980 between Chopin's and Alkan's preludes. The mean across group correlations are 0.924/0.945 between groups 1 and 2, 0.935/0.949 between groups 1 and 3 and 0.984/0.952 between groups 2 and 3.

How do the constant Q profiles depend on the data format the music is represented in? In Table 3.1, first the mean is calculated, then the correlation is performed. The following formats are compared: score (WTC fugues) encoded in humdrum format [Huron, 1999], expressively performed MIDI, spectrally weighted probe tone ratings, introduced in Section 3.3.3, Figure 3.22 on page 125 [Krumhansl and Kessler, 1982], and CQ-profiles of a digitized recording. On average, for the scores, the corre-

	Profile Correlation			
	WTC II Score	WTC I Exp	Spec Probe	Rec CQ
WTC I Score	0.994/0.994	0.993/0.996	0.905/0.914	0.970/0.967

TABLE 3.1: Correlations between profiles in various data formats: overall annotated durations ("Score"), expressively performed MIDI ("Exp"), spectrally weighted probe tone ratings ("Spec Probe"), and accumulated CQ-profiles from a digitized recording ("Rec CQ").



(a) WTC I+II (b) Romantizism

FIGURE 3.26: Comparison of mean NCQ-profiles in major from (*a*) the preludes and fugues of Bach's WELL-TEMPERED CLAVIER, Book I and II, and (*b*) the romantic preludes by Alkan op. 31 ("△") and Chopin op. 28 ("▽").

lation between cycles is in the range of the correlation between score and expressive MIDI, higher than the correlation between score and CQ-profiles, and a lot higher than the one between spectrally weighted ratings and score.

Figure 3.26 (a) compares four NCQ-profiles: the mean NCQ-profile of all preludes of Bach's WTC I, the mean profile of all fugues of WTC I, and the same for WTC II. As in Bach, the CQ-reference sets of preludes by Chopin and Alkan are mostly stable. Figure 3.26 (b) shows the averaged NCQ-profiles in Alkan and Chopin. The upright triangle stands for the 25 preludes by Alkan op. 31. The reverse triangle stands for Chopin's PRÉLUDES op. 28. Again, the mean NCQ-profiles are quite similar. Slight differences arise from the fact that Chopin more frequently uses the tritone in minor and the leading tone in major, whereas Alkan uses more third and fourth scale degrees in major and the leading tone in minor. (Minor is not shown in the figures.)

Comparing the CQ-profiles of Figures 3.26 (a) and (b) we find only small differences. There seems to be the general tendency in Bach to equally emphasize all diatonic scale degrees in major.

Even though the particular handling of pitch might be very special and expressively free, the overall distribution of pitch seems to be quite determined in the discussed composers. It might serve as the reference even though the momentary pitch use might differ.

3.3.6 Discussion

The constant Q technique is by no means limited to piano and tuning in equal temperament. The respective filter matrix can be adjusted to various instruments and tunings. For tunings significantly different from equal temperament the initial frequency estimation heuristics (p. 123) has to be modified. In addition, the efficient calculation of CQ-profiles requires the stability of the tuning throughout the piece, since the filter matrix (Equation 2.8 on page 70) has to be adjusted to the tuning and calculated beforehand. Frequency glides and vibrato may require special treatment. When analyzing other instruments we have to take their overtone structure into account, especially when comparing to the probe tone ratings as in Figure 3.22. Capability of generalization to other instruments is indicated by the experiments on page 124.

3.4 Mode, Style, and Composer Information in Profiles from Audio

Employing the key-finding method (Section 3.3.2 on page 123), we can transpose a CQ-profile so that the first component is the intensity of the keynote, yielding a NCQ-profile. For further investigation of the NCQ-profiles we apply various machine learning techniques. Classifiers are used to determine the composer. Clustering reveals the salient attribute in NCQ-profiles, the mode. Visualization by Isomap indicates interrelations between pieces and composers.

3.4.1 Composer Classification

With an appropriate classifier, composer classification works astonishingly well based on NCQ-profiles, especially considering that only very reduced and noisy data are provided. One composer is classified against all the rest. We apply a Matlab implementation of the RDA by Benjamin Blankertz with modifications by Guido Dornhege. For model selection by $n \times k$-fold cross-validation the optimization library CPLEX with a Matlab-C interface by Gunnar Rätsch is used. [11] Table 3.2 on the facing page shows the area under the curve of the receiver operating characteristics. Instead of applying LDA, classification performance improves when using RDA (cf. Section 2.3.1 on page 85), a compromise between LDA and QDA. An explanation could be (1) that the variance varies a lot between classes and (2) that some class sample sizes are quite small. E.g. for Hindemith there are only twelve samples with high variance. RDA and SVM (Sections 2.3.1 and C.2) with radial basis functions are even for Bach and Shostakovich. RDA beats SVM slightly for Scriabin and Hindemith. SVM beats RDA for Chopin and Alkan. RDA and SVM beat k-nearest neighbor (k-NN, Section 2.3.1).

[11] Rätsch et al. [2001]. Cf. also Section 2.3 of this thesis.

	LDA	RLDA	RDA	SVMrbf	k-NN
Bach	0.79	0.79	**0.95**	**0.95**	0.91
Chopin	0.52	0.52	0.64	**0.73**	0.63
Alkan	0.43	0.43	0.72	**0.76**	0.57
Scriabin	0.65	0.65	**0.72**	0.69	0.61
Shostakovich	0.81	0.85	**0.86**	**0.86**	0.71
Hindemith	0.93	0.93	**0.97**	0.95	0.72

TABLE 3.2: For the classification of one composer versus the rest, the performance of various classifiers is evaluated by a measure called the area under the curve describing the receiver operating characteristics (cf. Section 2.3). The best classifiers are emphasized in boldface. Regularized discriminant analysis and support vector machines with radial basis functions as kernels perform equally well. (Cf. Section 2.3.1 on page 84.)

3.4.2 Which Scale Degree Does Signify Mode?

What does distinguish major and minor? According to the concept of position finding (Section 1.2.2 on page 48), single pitch classes may have the potential to signify the key. To identify the significance of different scale degrees for mode separation we apply two methods: (1) two sided t-test [Witting, 1966] with error level 1%, (2) linear

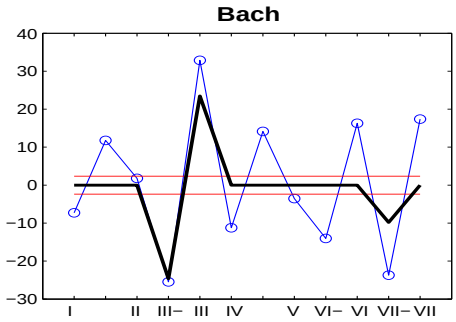
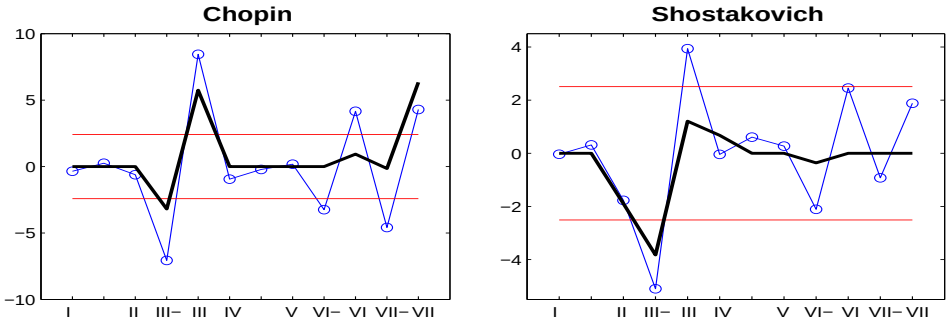


FIGURE 3.27: Significance of scale degrees for mode discrimination. The horizontal axis denotes the scale degrees, "-" indicating the minor intervals. The vertical axis denotes significance. The horizontal red lines indicate the significance level for the t-test of error level 1%. If the blue line ("o") is above the upper or below the lower red line, the scale degree contributes significantly to major/minor discrimination. The thick black line indicates to what extent which scale degree is emphasized for discrimination in a sparse linear programming machine (LPM). In Bach both t-test and LPM emphasize the thirds and the minor seventh. In addition t-test identifies every scale degree to be very significant, except the major second and the fifth.



(a) (b)

FIGURE 3.28: In Chopin (a) the t-test identifies only thirds, sixths, and sevenths to be significant for mode discrimination. LPM emphasizes thirds and the major seventh. In Shostakovich (b), according to the t-test, only the thirds allow for mode discrimination. (Cf. Figure 3.27 for an explanation of the curves.)

programming machine (LPM, Bennett and Mangasarian [1992]). The latter method is a classification algorithm that is similar to the SVM, but has a linear goal function. A special feature of the LPM is that it seeks to produce sparse projection vectors, i.e., it tries to find a good classification that uses as little feature components (in our case scale degrees) as possible. The components that are chosen by the LPM are thus most important for the discrimination task. We apply an LPM implementation coded by Benjamin Blankertz, adopting programs by Sebastian Mika and Gunnar Rätsch.

According to the t-test, in Bach (Figure 3.27) we see the outstanding significance of both thirds and the minor seventh. All other scale degrees are equally significant, except the (almost) insignificant major second and the fifth. In Chopin (Figure 3.28 a) the significance of these scale degrees is much less. In the t-test only thirds, sixths, and sevenths are significant. The minor sixth is close to insignificance. LPM emphasizes the major seventh. In Shostakovich (Figure 3.28 (b)) only the thirds are significant according to the t-test. In Alkan thirds are significant. The minor seventh is slightly significant. For Scriabin thirds have high significance. But also sixths and sevenths have some significance.

As expected the thirds and to some degree the sixths and sevenths are significant for major/minor discrimination. During music history from Bach to Shostakovich the significance of all pitch classes gets eroded due to increasing chromaticism and the continuing extension of tonality.

3.4.3 Mode as Salient Feature in Clustering

To evaluate the outcome of the clustering procedure we introduce the following class membership function $m_i(\mathbf{p})$, with data point $\mathbf{p}$. We scale down k-means clustering

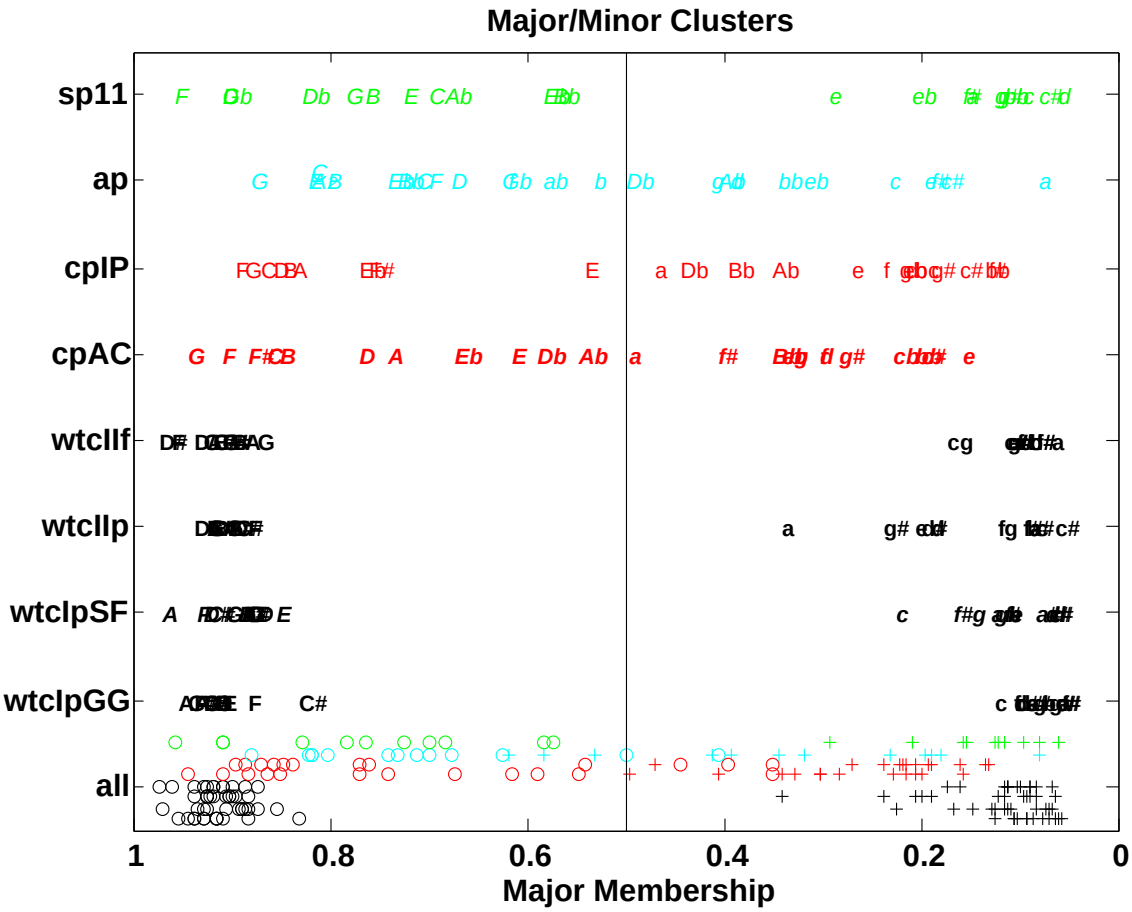


FIGURE 3.29: "Majorness" in k-means clustering ($k = 2$) of normalized constant Q profiles. The vertical line in the middle separates the different groups of pieces. From bottom to top the rows mean: a summary of all pieces, then Bach's WTC (preludes – "p" and fugues – "f", black) performed by G. Gould ("GG"), and S. Feinberg ("SF", only WTC I preludes), Chopin's preludes ("cp", red), performed by A. Cortot ("AC") and I. Pogorelich ("IP"), Alkan's preludes ("ap", cyan), and Scriabin's preludes ("sp11", green). At the bottom of the graph, the pieces are shown again without labels, "o" indicating major, "+" indicating minor, with the typical colors assigned to each composer. The horizontal axis indicates the mode membership for major (1 meaning typical major, 0 minor, 0.5 ambiguous). The vertical line in the middle at membership 0.5 splits the profiles into a left major and a right minor cluster. Bach and Scriabin clearly separate. In Chopin and Alkan more major/minor ambiguous pieces can be found, e.g. Chopin's a–minor prelude (cf. text).

with the $k = 2$ class centers $\mathbf{c}_1, \mathbf{c}_2$ of the data points:

$$m_1(\mathbf{p}) = \frac{\|\mathbf{p} - \mathbf{c}_2\|}{\|\mathbf{p} - \mathbf{c}_1\| + \|\mathbf{p} - \mathbf{c}_2\|}. \tag{3.5}$$

A value near 1 is assigned to data points $\mathbf{p}$ close to the center $\mathbf{c}_1$ of cluster 1 by function m_1. Points that ambiguously lie between the clusters get values near 0.5 in both functions, m_1 and m_2.

K-means clustering is performed on the corpus excluding Hindemith and Shostakovich, since they are the least tonal. Using Equation 3.5 as a major membership function yields clustering into a major (left) and minor (right) cluster (Figure 3.29). The results indicate a degree of "majorness" of the keys. Pieces which lie on the borderline between major and minor may not be very typical representatives of that key. Bach's pieces concentrate in a smaller region. Bach clearly separates with a wide margin. This is due to the fact that in general Bach modulates only to tone centers that lie closely around the tonic on the circle of fifths. Usually he modulates in the range of a fifth lower and three fifths upper. Only the chromatic a–minor prelude of WTC II is a little off the Bach cluster. This will be discussed in Section 3.4.4 on the facing page. Chopin's a–minor prelude is (for Pogorelich and Cortot) almost on the borderline. This is related to the wired harmonic content of the piece that makes it ambiguous in mode.

Analyzing pieces that are found in the "wrong" cluster reveals musical particularities of these pieces. There are three minor NCQ-profiles (f–minor, b–minor, and a♭–minor by Alkan) on the side of the major cluster and six major NCQ-profiles (Chopin's B♭–major played by Cortot and Pogorelich, D♭–major, A♭–major, and Alkan's D♭–major and A♭–major) within the minor cluster. For some of these pieces there is an intuitive musical explanation for their position in the other cluster. In Chopin's B♭–major prelude the middle part in G♭–major emphasizes the minor third and deemphasizes the leading note *a* in B♭–major. In addition, the left hand figuration introduces chromaticism that later is taken over by the right hand also. Chopin's A♭–major prelude is once found in the "wrong" cluster (Pogorelich). In Cortot's interpretation it is the major piece second closest to the cluster border. The closeness to the minor cluster is especially due to modulations to E–major and therefore the emphasis of the minor third of A♭–major. Also in the A♭–major parts the minor seventh chord on *a♭* is frequently used. Alkan's A♭–major prelude is chromatic. Within the length of one bar, all four flats, *b♭*, *e♭*, *a♭*, *d♭*, alternate with notes a half tone above. Therefore[12] the major and minor third are heard all the time. Alkan's D♭–major prelude has a middle part in d♭–minor. Alkan's f–minor prelude has an F–major middle part. Alkan's a♭–minor has a middle part in A♭–major.

Major/minor mode is the most salient property in the NCQ-profiles. K-means clustering with the major membership function reveals the degree of "majorness". For most pieces around the borderline between clusters chromaticism or extended tonality holds. Mode separability of different composers is reflected in the clustering.

[12] Lacking a proper score, the musical interpretation of the clustering is done with MIDI files at hand that do not contain enharmonic spelling. Therefore we can only refer to enharmonic equivalence classes, instead of enharmonically spelled out notes.

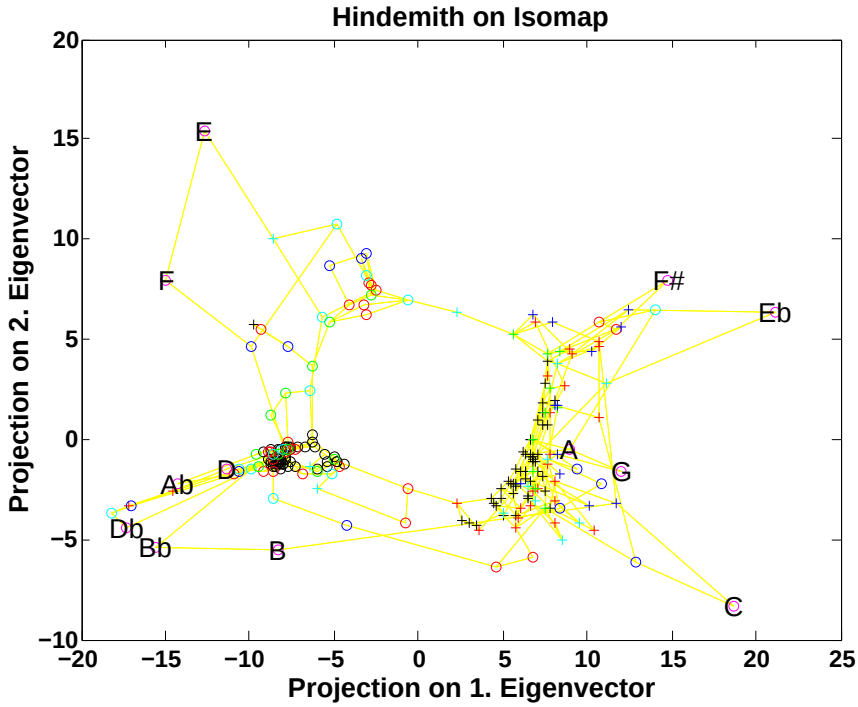


FIGURE 3.30: "Landscape" of all pieces. A two-dimensional projection of the Isomap with k-nearest neighbor topology ($k = 2$), trained with the entire corpus: Bach (black), Chopin (red), Alkan (cyan), Scriabin (green), Shostakovich (blue), and Hindemith (magenta). As in Figure 3.29, "o"/"+" denote major/minor respectively. The two-dimensional projection accounts for explaining 86.2 % of the variance in the data. In this graph the projected pieces by Hindemith are labeled with their keynote. Hindemith occupies the outlier positions in the Isomap projection. Figures 3.31, 3.32, 3.33, 3.34, and 3.35 are all the same curve, but with labels that belong to another composer.

3.4.4 Visualizing Inter-Relations of Modes, Composers, and Pieces

How can we measure similarity among pieces? For each pair of pieces let us calculate the Euclidean distances between their NCQ-profiles. A dissimilarity measure is then given by the shortest path length via pieces that have smallest Euclidean distance to their predecessor. We would like to analyze the net structure that is build up by connecting the NCQ-profile of each piece to its two closest neighbors. We use the Isomap (Section 2.3.4 on page 91), implemented by Josh Tenenbaum, with k=2 nearest neighbors for visualization (Figures 3.30, 3.31, 3.34, 3.35, and 3.33). Under the assumption of the above given dissimilarity, Isomap reveals that the pieces form a circular structure with two "poles" (major and minor) and some extensions to outliers. This two-dimensional visualization accounts for 86.2 % of the variance in the data.

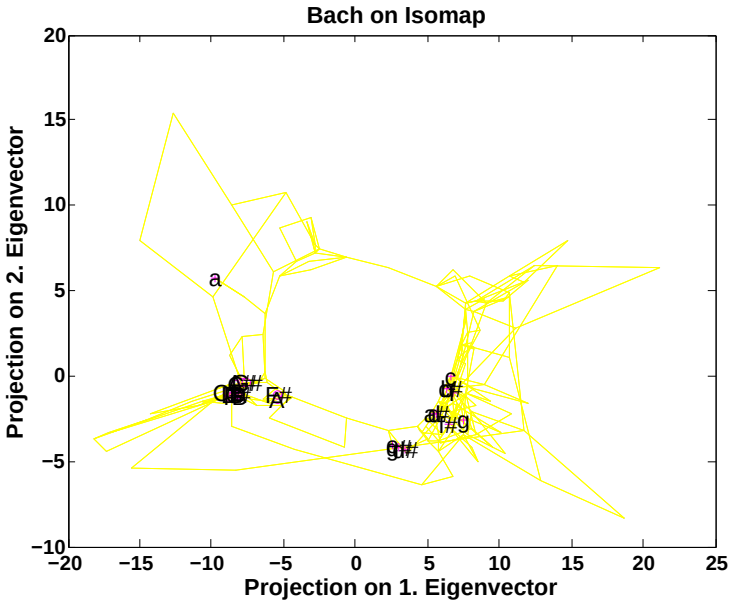


FIGURE 3.31: Bach's preludes of WTC II are labeled. Two densely concentrated well separated clusters are formed: a highly compressed major cluster and a minor cluster in the form of a strip. The very chromatic a–minor prelude of WTC II is an outlier relative to these clusters. (Cf. Figure 3.30 for all Bach pieces labeled as black symbols.)

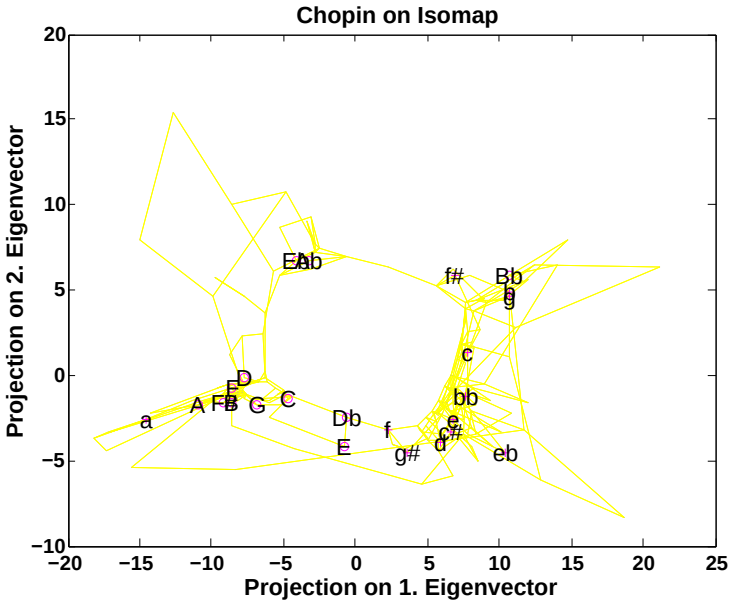


FIGURE 3.32: Chopin's PRÉLUDES op. 28 in Cortot's interpretation are highlighted. They distribute around the circle. Major and minor are not clearly separated. See all Chopin preludes marked as red symbols in Figure 3.30.

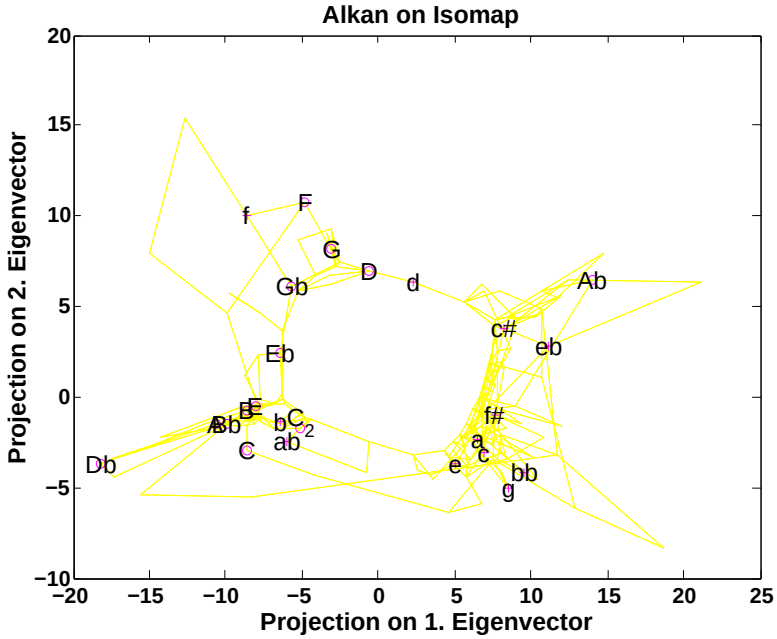


FIGURE 3.33: Alkan concentrates somewhat on two wide clusters, but major and minor intermingle a little. D♭–major and f–minor are outliers due to their modulation to the parallel of the tonic in the middle part, A♭–major is a chromatic prelude. (Cf. Figure 3.30 for details.)

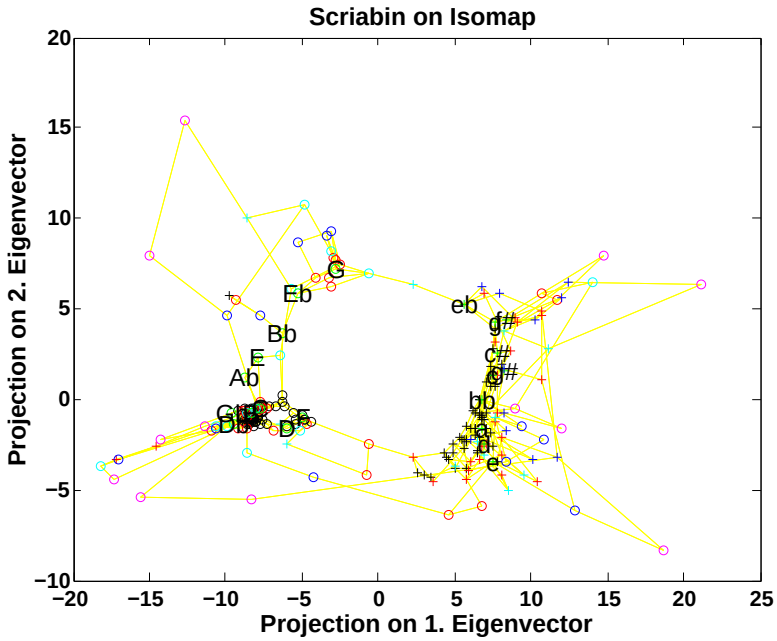


FIGURE 3.34: Scriabin clearly separates into a major and a minor cluster. (Cf. Figure 3.30 for details.)

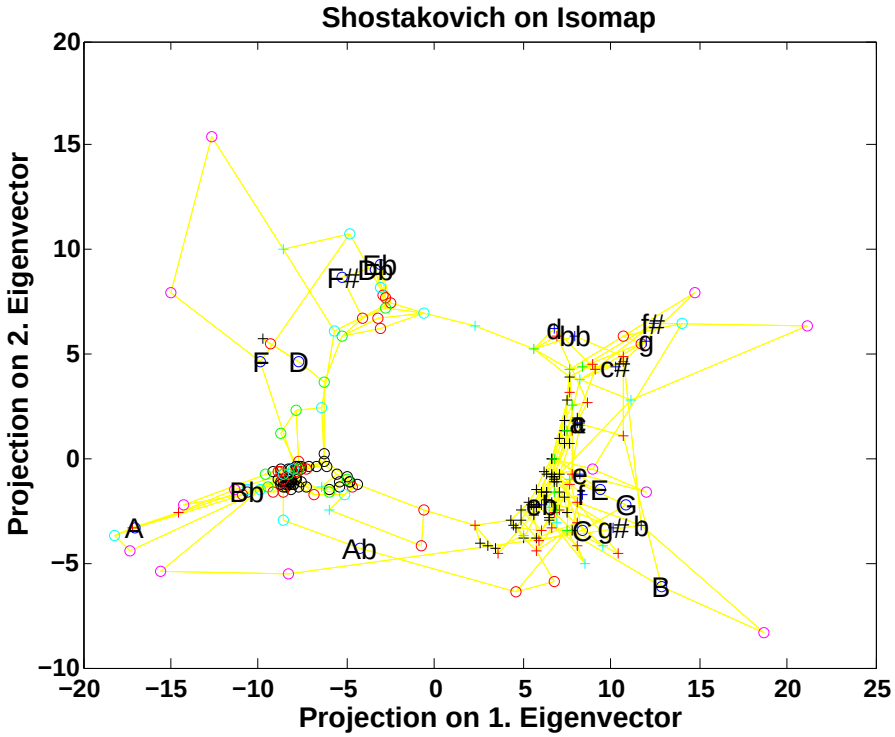


FIGURE 3.35: Shostakovich is widely distributed with A–major as an outlier. (Cf. Figure 3.30 for details.)

Compare the positions of the pieces in the visualization of Figure 3.30 to the clustering in Figure 3.29. On the Isomap Hindemith (Figure 3.30) occupies outlier positions very far away from the center. Only A(–major) lies more closely to the center. Bach's major preludes cluster densely on a small spot, while his minor preludes concentrate on a prolonged strip (Figure 3.31 on page 136). As in *k*-means clustering in Figure 3.29, the margin between the two clusters is wide. It is not arbitrary that the a–minor prelude of WTC II is the only prelude outside the two main Bach clusters. This prelude also stands aside from the Bach minor cluster in Figure 3.29. The special position of this prelude is due to the chromaticism that dominates it. In Chopin (Figure 3.32 on page 136) major and minor intermingle. In Alkan (Figure 3.33 on the preceding page) a major and a minor cluster is found, but they are not clearly separated. Due to its chromaticism A♭–major is very far in the minor region. On the other hand, f–minor and a♭–minor are in the major region, because they modulate to the parallel key in their middle parts. D♭–major is far off for the same reason. Accidentally, it is next to Hindemith's fugue in D♭. Scriabin (Figure 3.34 on the previous page) makes up two major/minor clusters clearly separated by a wide margin. Shostakovich (Figure 3.35) does not separate well between major and minor. The A–major prelude is off.

The Isomap reveals stylistic characteristics of composers and pieces. Again, distinguished positions on the "landscape" often coincide with high chromaticism and a

character unusual for that composer. Also stylistic coherence in a composer's work is visualized.

3.4.5 Discussion

Audio in the compressed format of the NCQ-profile has been processed by a wide range of machine learning methods. We successfully applied classification, clustering, and visualization to NCQ-profiles for composer, mode, and style analysis.

The classification results show that normalized constant Q profiles can substantially contribute to composer identification based on audio, provided the use of appropriate classifiers: RDA and SVMrbf performed best among a group of classifiers.

K-means clustering of the NCQ-profiles reveals that the "majorness" is the most salient property in the data. Applying an appropriate mode membership function yields two clearly separated clusters: major and minor. The mode membership can be interpreted musically. For some of the pieces with "ambiguous" membership to both the major and minor cluster center, musical analysis reveals an ambiguous major/minor content, chromaticism, or modulations to the parallel major/minor key or to tone centers far away from the tonic on the circle of fifths. The compactness of clusters can be musically interpreted in the way that the composer consistently uses the same modulation range in all pieces, e.g. stays in a small compact neighborhood of tone centers around the tonic.

The Isomap visualization allows a stylistic comparison of pieces and groups of pieces by different composers. Some composers reside in the outlier positions (Hindemith) whereas other densely concentrate (Bach).

Considering that the NCQ-profiles reflect only a very special aspect in music, reducing high musical complexity to a small cue, they reveal a remarkable variety of information. It is promising to use this feature in combination with rhythmic and other attributes.

Based on an extended corpus of musical data, other questions of interest include the investigation to what extent certain features do manifest in NCQ-profiles, like key character (cf. Section 1.2.8) and performer. We can test the hypothesis, whether NCQ-profiles signify the key character as perceived by a particular composer (Section 1.2.8 on page 59). It is also interesting whether the performer can be determined with a recording at hand.

The suggested machine learning methods are by no means restricted to the analysis of NCQ-profiles for composer, mode, and style analysis. They can be promisingly used for instrument identification, beat weight analysis, and harmonic analysis as well, just to name a few other possible applications.

3.5 Circle of Fifths

Often it is more reasonable to display pitch class profiles in the order of the circle of fifths than, of the pitch class circle. Let us arrange the probe tone ratings in

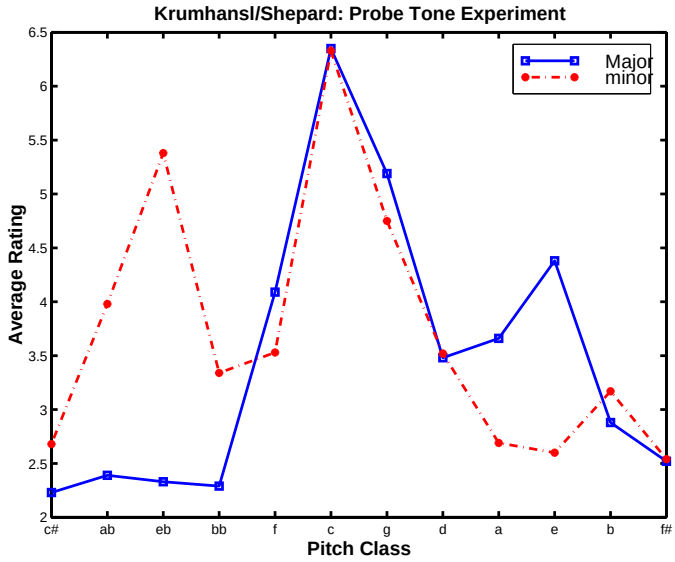


FIGURE 3.36: Probe tone ratings mapped on the pitch class axis in circle of fifths order. Compared to Figure 1.19 on page 62, the hierarchies describe a smoother graph that resembles an addition of two hills centered at the tonic and the major/minor third.

Figure 1.19 according to the circle of fifths (Figure 3.36). Then the zigzag pattern is replaced by a composition of two hills, one centered at the tonic, and a smaller second one centered at the third scale degree, the major or minor third respectively.

Heinichen [1728]'s circle of fifths can be deduced from Euler [1739]'s tonnetz and Weber [1817]'s chart of tone centers (cf. Section 1.2.6 on page 54). By enharmonic identification the straight line of fifths bends and turns into a closed circle of fifths (cf. Section 1.2.7 on page 56).

First we will see how the circle of fifths is supported by equal temperament. Then we will observe the emergence of the circle of fifths in a couple of experiments, in correspondence analysis (WTC I fugues score; WTC II fugues audio) and in Isomap (entire WTC audio).

3.5.1 Equal Temperament for Infinite Tonal Progression

In Pythagorean temperament (p. 46) a modulation along the line of fifths would be perceived like the sawtooth pattern in Figure 3.1 on page 93. The sequence ascends smoothly with a disruption at the wolf. How can this break be eliminated? Although already Adrian Willaert (c. 1490–1562), in the chromatic duo QUID NON EBRIETAS, takes the singer through the entire circle of fifths, the closure of the latter becomes a manifest in equal temperament, enabling instruments to modulate to tone centers infinitely far away. In the tonnetz (Figure 1.8) the tones of just intonation are generated by stepping forward in just fifths and just octaves. We have seen in Section 1.2.1 that twelve fifths and seven octaves differ by the Pythagorean

comma. In just intonation, from tone *c*, we can progress infinitely far in pure fifths and octaves towards either flat or sharp keys without ever reaching initial *c* again. The finite number of keys of a keyboard instrument, customarily twelve per octave, constricts either tuning or smooth progression along the circle of fifths. On instruments with stable tuning, well-tempered tuning, or, to a even higher degree, equal temperament, continuous progression is enabled across the entire circle of fifths, by equal distribution of the Pythagorean comma to all fifths. Thereby there is no discontinuity in the chain of fifths. Modulations may start from c-minor, following the circle of fifths, until reaching b♯–minor, which – useful enough – physically is identical to c–minor on the keyboard. Some pieces ground in that idea, e.g. the CANON CIRCULARIS PER TONOS of Bach's MUSIKALISCHES OPFER. The first entry of the canon starts in c-minor and modulates towards d-minor. For the second entry voices are transposed up a major second to d-minor. The second entry of the canon ends in e–minor, and so forth. After six entries the canon reaches the initial ton center c–minor again. Hofstadter [1979] arranges the canon with Shepard tones to emphasize the circularity of the modulation. ZWEI PR LUDIEN DURCH ALLE TONARTEN op. 39 (first performed 1789) by Ludwig van Beethoven also circumnavigate the circle of fifths. Franz Schubert's SONATE in B♭-major op. posth., Movement 3, SCHERZO, bars 1–67, modulates in the following way: B♭-major → E♭-major → f-minor → A♭-major → b♭-minor → D♭-major → f♯/g♭-minor → A-major → d-minor → B♭-major. According to Werts [1983], p. 17, Figure 42, the trajectory of these tone center transitions can be displayed on the circle of fifths rather than in Weber [1817]'s chart of tone centers.

3.5.2 Circle of Fifths Emergence in Correspondence Analysis

We will now investigate the set of preludes & fugues in Bach's WTC. For each part of WTC there is a one-to-one mapping between all 24 pairs of preludes & fugues and all 24 major and minor keys. Table 2.2 on page 86 shows how each key – that implies each prelude & fugue pair also – can be represented by a frequency profile of pitch classes. The pitch class frequency profiles can either contain the overall annotated durations from the score or the accumulated CQ-profiles from a performance of that piece. Correspondence analysis visualizes inter-key relations on keyscapes based on pitch class profiles. The projection of pitch classes homogeneously displays the circle of fifths for both score and performance.

Circle of Fifths from Score

Humdrum [CCARH, 2003] is a digital format that has been developed in order to fully and precisely represent the essential content of a composition as it is indicated in notation in sheet music. All fugues of Bach's WTC are encoded in Humdrum **kern format. Instead of analyzing the co-occurrence table of frequencies of keys and pitch classes (Table 2.2) we look at the *overall annotated duration* of pieces in a particular key and the overall annotated duration of pitch classes across all fugues in

WTC. Annotated duration is measured in multiples and fractions of quarter notes, rather than in seconds. Measuring duration in seconds would imply that the results would vary a lot depending on the choice of tempo. But the issue of tempo in Bach is heavily debated.

The correspondence analysis of WTC (Figure 3.37) reveals a two-dimensional structure which allows for an adequate representation in a plot based on the first two factors corresponding to the two largest singular values (Figure 3.39). The 24 pitch class frequency profiles are optimally projected onto a two-dimensional plane, such

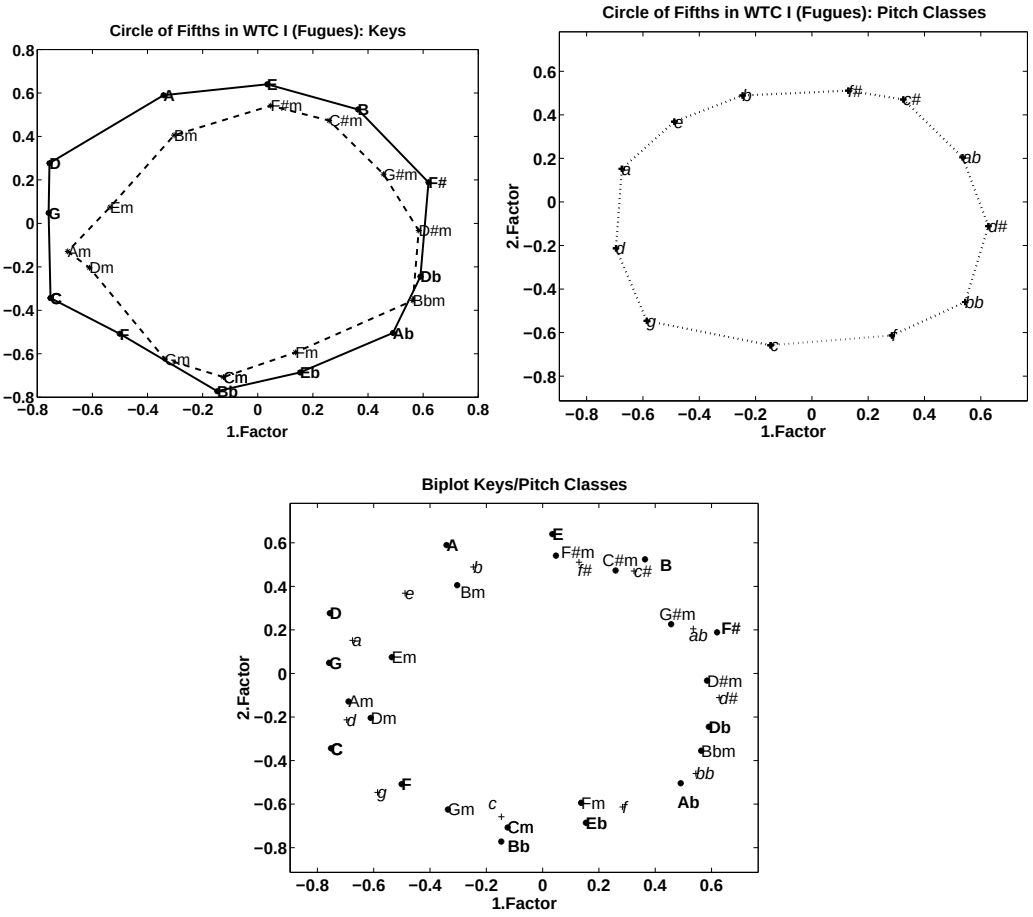


FIGURE 3.37: Annotated durations of keys and pitch classes are derived from the score of the fugues of Bach's WTC I and then projected onto the factors of correspondence analysis. *Upper left*: The emerged circle of fifths is lined out for major (solid) and minor (dashed) keys ("m" denoting minor). *Upper right*: Likewise, pitch classes appear in the order of the circle of fifths. *Lower:* The biplot of keys and pitch classes derived from putting both transparent upper graphs on top of each other. We observe that the pitch classes are close to the corresponding minor keys.

that the χ^2-distance between profiles is minimally distorted and the χ^2-distance of the two-dimensional projections matches the original profile distance as well as possible. In the fugues of WTC, the circle of fifths emerges clearly and homogeneously (upper left Figure 3.37). However, in the upper left Figure 3.37 some inter-key distances are smaller (D♭–A♭) than others (D–A), due to different χ^2-distances between pitch class profiles of overall annotated durations in these pieces. In addition the minor keys form a circle of fifths inside the circle of fifths of the major keys. This shows that the pitch prominence patterns for major keys are more distinct than for minor keys according to the metric employed.

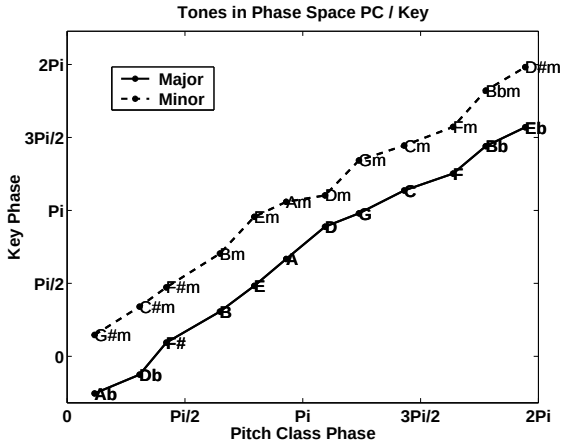


FIGURE 3.38: Phase of the circles of fifths described by major and minor keys (upper left Figure 3.37) relative to the phase of the circle of fifths in pitch classes (upper right Figure 3.37). The graph displays almost straight parallel lines. The angular co-ordinates of pitch classes and minor keys are "in phase". The angular co-ordinates of pitch classes and major keys are offset by a constant value slightly less than $\pi/2$.

In co-occurrence Table 2.2, pitch classes are represented as columns of overall annotated durations in the different keys, that means overall annotated durations in each fugue of WTC I, since in WTC I a one-to-one correspondence of fugues and the 24 keys is given. The same factors with maximal singular values as in the upper Figure 3.37 are used to optimally project the 24-dimensional pitch class vectors upon a plane. We observe the pitch classes forming a circle of fifths as well (upper right Figure 3.37). We can now consider the biplot (lower Figure 3.37) by putting the transparent plot of pitch classes (upper right Figure 3.37) on top of the plot of keys (upper left Figure 3.37). We have three circles of fifths, one each for the major and minor keys and one for the pitch classes. We change the co-ordinates of the factor plane to polar co-ordinates in terms of a polar angle (on the circle of fifths) and the distance to the origin. Consider the angles of both the major and minor keys relative to the angles of the fundamental pitch class (Figure 3.38).

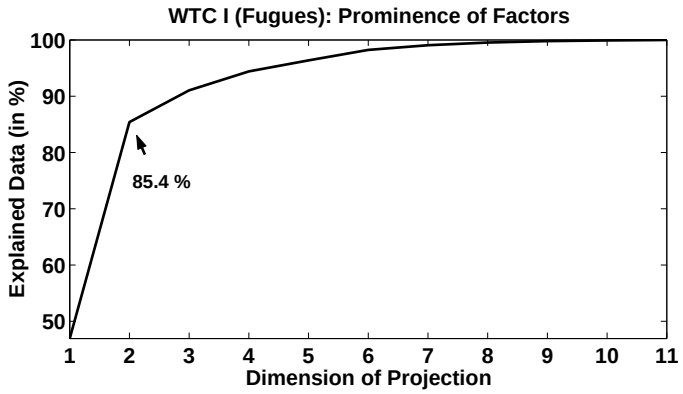


FIGURE 3.39: A low dimensional representation is sufficient for displaying the high dimensional data: For WTC I (fugues) in score representation the two-dimensional projection of the overall annotated durations of keys and pitch classes accounts for 85.4 % of the variance of the high dimensional vectors. Each of the two most prominent factors have approximately the same singular value.

The plot shows two almost straight parallel lines, reflecting that pitch classes and keys proceed through the circle of fifths with almost constant offset. The graph for the minor keys is almost the identity, indicating that pitch classes and keys are "in phase": The pitch classes can be identified with the fundamentals of the minor keys. The explanation lies in the relatively high overall annotated duration of the fundamental pitch class in the annotated duration profile in minor compared to major. Also the overall annotated duration of the fugues in minor is longer than the one in major (Figure 2.8 on page 88). We conclude that the minor keys induce the circle of fifths in the pitch classes. In Figure 3.39 the singular values to the factors are shown. They indicate how much of the variance in the data is captured if correspondence analysis projects the data onto the factors with highest singular values. It is interesting that the explanatory values, e.g., the singular values, of the two most prominent factors in WTC I (fugues) are almost equal, in contrast to the other singular values, whose explanatory value is much smaller.

Circle of Fifths from Performance

We choose a recording of Glenn Gould playing the preludes and fugues of Bach's WTC, Book II. We calculate long-term CQ-profiles (p. 122), one for each of the 24 fugues of WTC II in all major and minor keys (cf. Figure 2.7 on page 87). Instead of containing frequencies of co-occurrences (Table 2.2) or annotated durations (p. 141 and Figure 3.37), the co-occurrence table now consists of the intensities of each pitch class accumulated for each of the 24 fugues in all 24 major and minor keys (Figure 2.7). Pitch classes are represented by 24-dimensional key intensity vectors. In

the same way as on p. 141, in correspondence analysis a singular value decomposition is performed yielding the factors as a new co-ordinate system. As in upper right Figure 3.37 on page 142, the pitch class vectors are projected onto a two-dimensional plane, spanned by the two most prominent factors. The circle of fifths evolves in pitch classes in performance data as well. The two factors of performed WTC II (lower Figure 3.40) capture an even higher percentage (88.54 %) of the variance of the data, than those for the score data of WTC I (cf. Figure 3.39). Both factors are

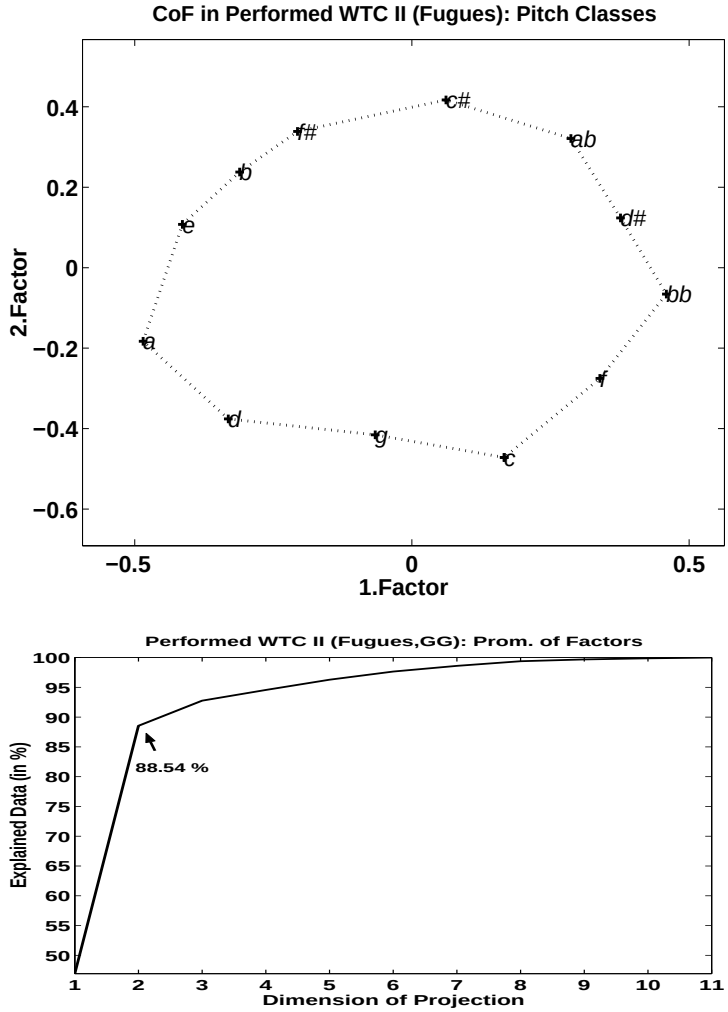


FIGURE 3.40: The circle of fifths (lined out) appears also in performed WTC (*upper*). The analyzed data are the overall intensities of pitch classes in the fugues of Bach's WTC II in the recording of Glenn Gould shown in Figure 2.7. The same procedure as in Figure 3.37 (upper right) is applied to project the 24-dimensional pitch class vectors onto a two-dimensional plane, spanned by the two most prominent factors. These two factors of performed WTC II capture an even higher percentage (88.54 %, *lower*) of the variance of the data than those for the score data of WTC I (cf. Figure 3.39).

high and almost equal. Therefore the two–dimensional projection appears to be a very appropriate representation of pitch classes.

Comparisons have been made with other cycles of musical pieces like Chopin's Préludes op. 28 and Hindemith's Ludus Tonalis: In these cycles one singular value alone is by far most prominent. That means that key frequency space as well as pitch class space can be reduced to one dimension still being able to explain the majority of the data.

3.5.3 Circle of Fifths Evolving in Isomap

Using the CQ-profiles of all pieces from WTC in our corpus, the circle of fifths clearly emerges in k-nearest neighbors Isomap ($k = 4$, Figure 3.41). The projection to two dimensions accounts for 53% of the variance in the data. As an extension to similar experiments with correspondence analysis in Section 3.5.2, here more performed pieces are taken into account.

3.5.4 Discussion

In WTC score and audio, the circle of fifths emerges in correspondence analysis as well as in Isomap. Besides fifths relations we can consider parallel and relative major/minor relations also. Then we will arrive at toroidal models of inter-key relations (TOMIR, Section 3.8 on page 156). For further discussion of the simulations in this section please cf. the discussion for TOMIR in Section 3.8.5 on page 166.

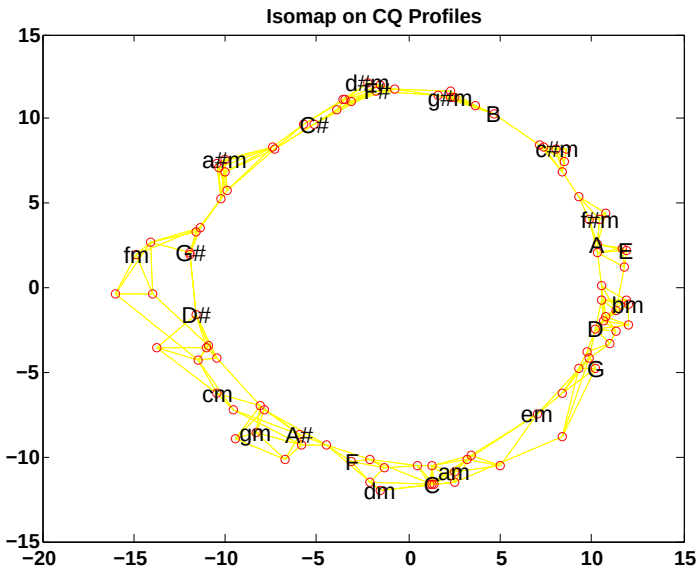


Figure 3.41: Keys arrange in a circle of fifths in an Isomap trained by the CQ-profiles of all Bach sets in the corpus.

3.6 Composers Characterized by Key Preference

In the following experiment the interplay of key preference and composer, rather than the interplay of key duration and pitch class duration is considered. For seven composers we provide the frequencies of pieces each composer wrote in each of the 24 major and minor keys. We will discuss key preference in the light of key character and display the relations between different composers and between composers and keys.

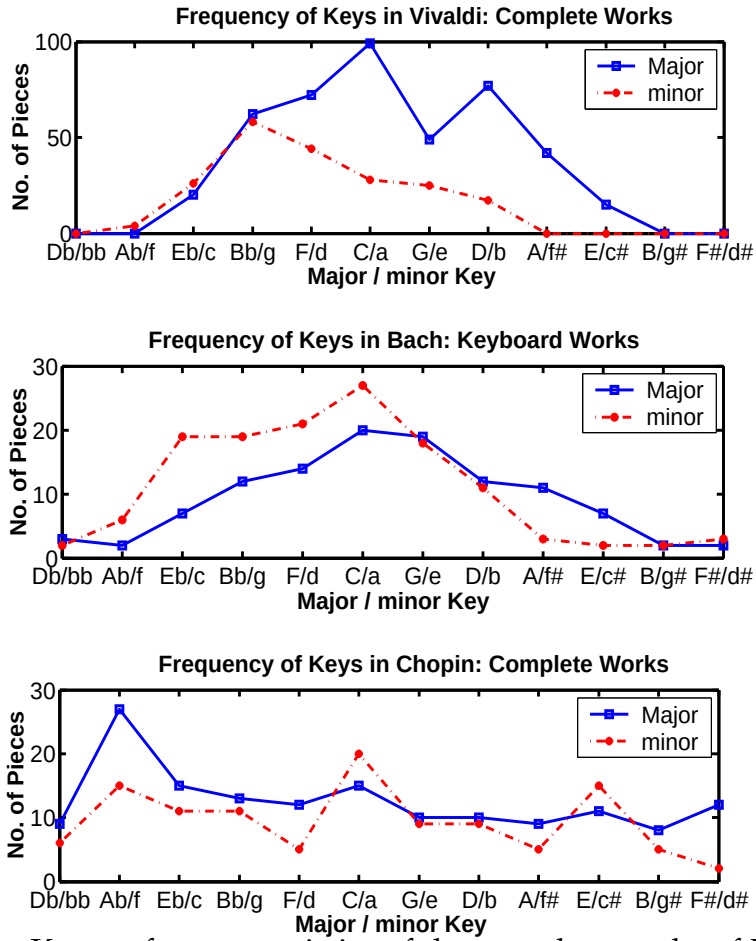


FIGURE 3.42: Key preference statistics of the complete works of Vivaldi and Chopin and the keyboard works of Bach. In this and the two subsequent figures, small letters indicate minor keys. Vivaldi prefers C–major and D–major, Bach prefers a–minor, c–minor, and C–major. Some rare keys are only used in the WTC: F♯–major, B–major, A♭–major, c♯–minor, g♯–minor, and b♭–minor. 45.5% of the pieces are written in a major key. The most frequent keys in Chopin are A♭–major, a–minor, C–major, E♭–major, and c♯–minor. There are only two pieces in d♯–minor. 57% of the pieces are in major.

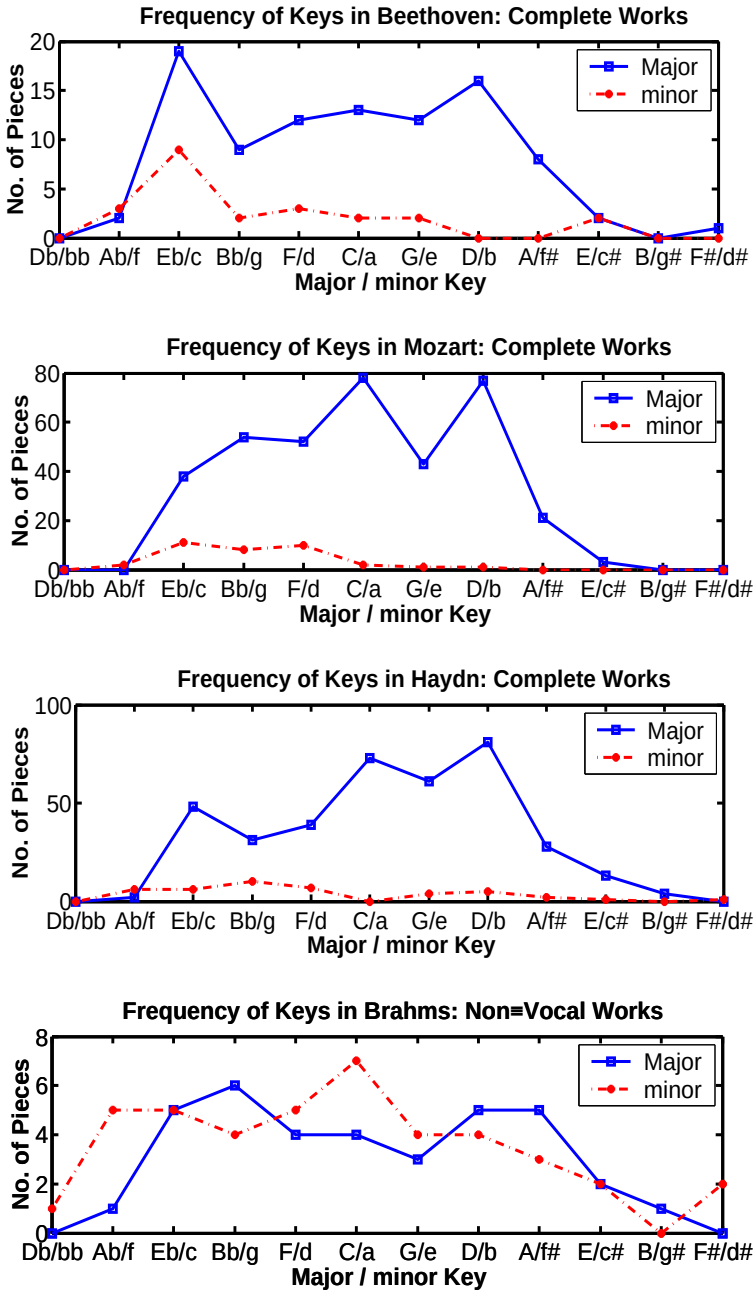


FIGURE 3.43: Beethoven prefers the "heroic" E♭–major and D–major. As in Vivaldi (cf. Figure 3.42) the key preference of Mozart and Haydn takes the "Cologne Dome" shape, favoring C–major and D–major. 90% of the keys are in major.

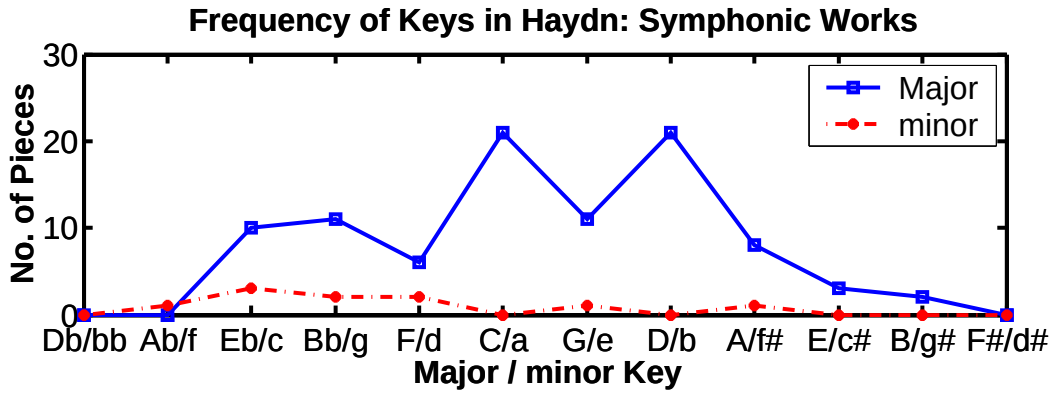


FIGURE 3.44: Haydn Symphonies show the "Cologne dome" distribution in contrast to the lesser preference of D–major in his keyboard works (not displayed in this thesis).

3.6.1 Key Preference Statistics

For each key and for each composer the co-occurrence table 2.2 now contains the numbers of pieces written in that particular key by that particular composer. We identify enharmonically equivalent keys (p. 46). Key preference statistics are collected in the following composers: Antonio Vivaldi (AV, 1678-1741), Johann Sebastian Bach (JSB, 1685-1750, only works for keyboard), Joseph Haydn (JH, 1732-1809), Wolfgang Amadeus Mozart (WAM, 1756-1791), Ludwig van Beethoven (LvB, 1770-1827), Frédéric Chopin (FC, 1810-1849), and Johannes Brahms (JB, 1833-1897, non-vocal works). If not stated otherwise all works of the composer are considered provided they contain the key name in the title of either the entire work or of single pieces in case the work consists of a cycle of several pieces. For instance, a sonata in C–major is accounted for once, but WTC, Book I is accounted for 24 times. These key preference statistics (Figures 3.42 and 3.43) are generated from complete work lists of these composers found on the Internet.[13] Auhagen [1983] counts the keys of every single movement in a set of works by Chopin. He counts enharmonically equivalent keys separately. The most frequently used key is A♭–major in our statistics as well as in Auhagen [1983]. Also the six most frequent keys are the same in our statistics and in Auhagen [1983].

[13] http://classical.efront.com/music/composer/works/vivaldi/
http://www.jsbach.net/catalog/
http://home.wxs.nl/ cmr/haydn/catalog/piano.htm
http://classical.efront.com/music/composer/index.html
http://www.classical.net/music/composer/works/chopin/#opus
http://w3.rz-berlin.mpg.de/cmp/brahms_works.html

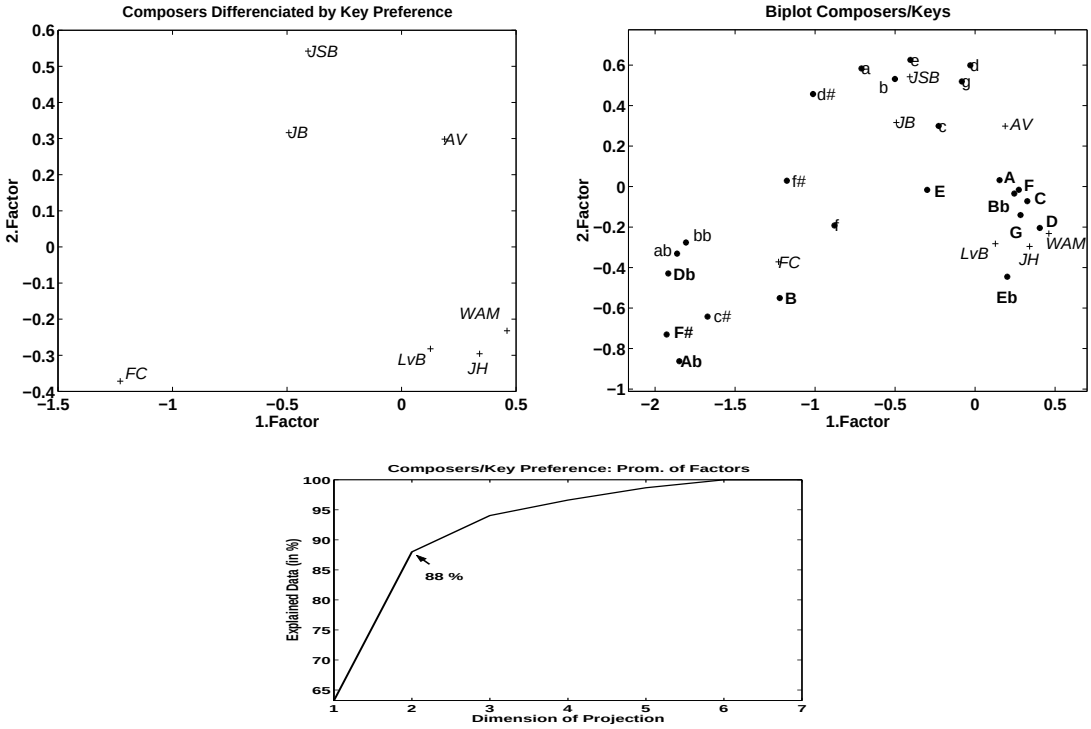


FIGURE 3.45: Based on the key preference profiles (cf. Figures 3.42 on page 147 and 3.43) for different composers the stylescape (*upper left*) and the biplot of composers/keys (*upper right*) with associated singular values (*lower*) are shown. Capital letters indicate major keys, small ones minor keys, italics indicate composers. *Upper left:* Haydn (JH) and Mozart (WAM) are very close. Beethoven (LvB) is nearby. Brahms (JB) and Bach (JSB) can be considered a group. Chopin (FC) is an outlier. *Upper right:* In the biplot composers/keys we observe that the Viennese classics (JH, WAM, LvB) gather in the region of D–major and G–major. Chopin (FC) maintains his outlier position due to the distant position of his favored A♭–major key.

3.6.2 Key Character and Key Statistics

Key character (Section 1.2.8 on page 59) is determined by several factors. One is the complex aspect of keyboard tuning (Section 1.2.1 on page 45). The composers considered here strongly differ in their preference for A♭–major. To give an example we will inspect the key character of this key and consider its impact on key preference.

In mean tone tuning w.r.t. C, the wolf fifth $g\sharp$ – $e\flat$ sounds very rough, since this fifth is 35.7 Cent higher than the just fifth (Meister [1991], p. 81 cited in Grönewald [2003]). Bach used Neidhardt's tuning [Lindley, 2001]. Even though in Neidhardt's tuning the wolf is eliminated, Bach uses A♭–major and g♯–minor only in WTC. Bach's avoidance of these keys could be due to the reminiscence that these keys would have sounded odd in a tuning with the wolf on $a\flat/g\sharp$.

On the other hand, A♭–major is the favored key of Chopin. What does make Chopin prefer this key? Does A♭–major have a special key character for Chopin? Lindley [2003] points out "the tradition of tender affects" for this key which may have appealed to Chopin's character. Chopin used equal temperament. For all major keys, all intervals relative to the tonic were the same. For Chopin key character may be influenced by 18th century keyboard tuning that preceded equal temperament. In tunings common at time when "Clementi was flourishing and Beethoven was not yet deaf", A♭–major had characteristic differences to tunings of keys with none or only few accidentals, e.g. C–major. First, the major third *a♭–c* in A♭–major is stretched compared to *c-e* in C–major. Second, the leading note – tonic step, *g–a♭*, is squeezed compared to *b–c*. Therefore this tuning endows A♭–major with a more nervous character and C–major with a more solid one. [14]

3.6.3 Analysis

Correspondence analysis is performed on the key preferences of the seven composers. It is noteworthy that correspondence analysis contributes to the varying amount of data available for the different composers. E.g. the few available pieces by Brahms weight less than the several pieces by Haydn. The projection of the composers onto the two-dimensional plane spanned by the two most prominent factors provides a stylescape: stylistically related composers are close to each other on this plane (cf. Figure 3.45 on the preceding page). In the biplot of composers/keys in Figure 3.45 composers are related to keys: Due to their shared "Cologne Dome" preference for C–major and D–major, Haydn and Mozart are very close and Vivaldi is not so far off. Beethoven is close to his favored E♭–major and proximate to Haydn and Mozart. Reflecting their preference for the minor mode, Bach and Brahms are positioned within a cluster of minor keys. Chopin maintains his outlier position due to the outlier position of his favored *A♭*–major key. The explanatory value for the first (63.23 %) and the first two factors (88 %) in correspondence analysis is high. The most important factor is very dominant.

3.6.4 Discussion

The data can be counted differently. It would be more appropriate, but more laborious, to take, in addition, pieces into account which do not use the key name in the title. It is a matter of preference, whether to count entire pieces or each movement in that key. To isolate the contribution of key character from factors like playability on the particular instrument, it would be interesting to investigate a heterogeneous set of keyboard pieces.

[14] The arguments in this subparagraph are from Lindley [2001, 2003] and from personal communication with Mark Lindley.

3.7 Visualization of Inter-Key Relations

We will first investigate how inter-key relations can be displayed by colors. Then we will introduce a new algorithm that globally arranges items according to a previously stipulated local neighborhood, such as closely related keys.

3.7.1 Key-Color Maps

> ...auch hatte ich gerade ein Kleid an, das ich einst im höchsten Unmut über ein mißlungenes Trio gekauft, und dessen Farbe in cis-Moll geht, weshalb ich zu einiger Beruhigung der Beschauer einen Kragen aus E-Dur-Farbe daraufsetzen lassen ... [15] (E. T. A. Hoffmann [1814]: Kreisleriana)

Synesthesia is the close connection between different perceptual modes. It can refer to various musical attributes including tones and keys. It is believed that Nikolai Rimsky-Korsakov had *synesthetic* tone-color perception. This kind of perception is quite rare, but shared e.g. by Oliver Messiaen. On the other hand, Scriabin supposedly had *pseudo-synesthetic* tone-color association. In contrast to genuine synesthesia, pseudo-synesthesia is a rather deliberate mapping between the attributes of different modes. [16] The tradition of tone-color mappings originates in the Pythagorean view of the world, identifying notes with intervals, intervals with relations of celestial bodies and celestial bodies with (apparent) colors. Jewanski [1999] [17] reconstructs the tone-color mapping of Aristotle and accounts for his followers in detail (Jewanski [1999], p. 67-266). The other, "prismatic", tradition can be traced back to Newton [1704] and builds on the analogy of electromagnetic and acoustic waves that is further emphasized by the observation that white sunlight is composed of seven spectral colors, with the widths of these colors in the spectrum being proportional to the intervals of the diatonic scale (Jewanski [1999], p. 7). Castel took up Newton's idea in constructing the "Clavecin oculaire" (Jewanski [1999], p. 267-449).

The color encoding schemes discussed in this section refer only to the keynote, not considering the mode (major or minor). They are all based on the prismatic color order. So the major difference in the schemes lies in the onset (which color to start with?, which color to assign to *c*?), the direction how to go through the color spectrum (from violet to red or vice versa), and how many keys are mapped to a particular region of the color spectrum.

We distinguish two types of key-color mappings. In the *prismatic chromatic fifth key mapping (PCFKM)* the color prism is mapped onto the entire circle of fifths consisting

[15] ...in addition I just wore a garment that I had once bought in the highest displeasure about a failing Trio, and whose color tends towards c♯–minor, wherefore, for calming down the spectator, I let a collar in E–major color sew onto it ...

[16] Kandinsky [1926]'s pseudo-synesthesia associates basic shapes with colors: circle – blue, square – red, triangle – yellow.

[17] Jewanski [1999], p. 86, citing Gaiser [1965], p. 190.

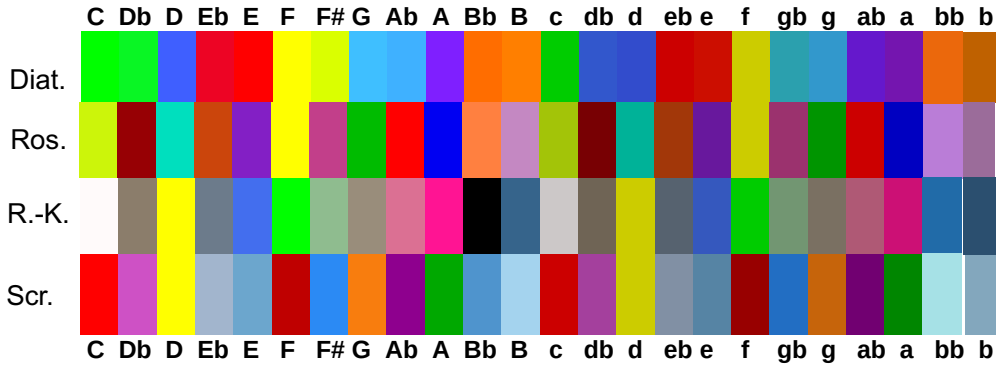


FIGURE 3.46: Key-color map according to the prismatic diatonic fifth key mapping ("Diat."), the Rosicrucians ("Ros."), Scriabin ("Scr."), and Rimsky-Korsakov ("R.-K.").

of twelve keys. In the *prismatic diatonic fifth key mapping (PDFKM)* the color prism is mapped onto the notes of the diatonic scale only. In the latter, alterations of the scale notes are recognizable as a derivate of the color that has been assigned to the non-altered tone. This color derivation can be performed by lightening or darkening or mixing the color more with brown or gray (e.g. Sapp [2001]). This scheme respects the enharmonic difference between $c\sharp$ and $d\flat$.

We will describe the following key-color mappings (cf. Figure 3.46).

- *Rainbow* This is the straight forward PCFKM. The *Rosicrucian* order is another straight forward PCFKM with onset c = "yellowish green". Variants of this scheme are presented by Charles Fourier [Lucy, 2000]. He maps the keys onto the flipped prism. Inspired by Fourier, Lucy [2000] assigns c to "dark green". For some reason f is mapped to the invisible light spectrum. Therefore no color is assigned to it.

- *Rimsky-Korsakov* His color scheme is basically a PDFKM. The onset is d = "yellow". The keys on the circle of fifths are mapped along the color sequence "yellow – rose – blue – dark blue – green". The alterations of the scale tones are colored by a grayish or brownish derivation of the original color of that scale note. In contradiction to the PDFKM, c is mapped to "white" and no color at all is assigned to $b\flat$.

- *Sapp* A strict PDFKM is given by Sapp [2001].

- *PDFKM for pitch classes* We want to apply the PDFKM to an audio based analysis. In that scenario we do not know the enharmonic spelling a priori (Section 1.2.1 on page 46). In our approach we first estimate the key of the piece. If we know the key we choose the most likely enharmonic spelling according to the scheme depicted in Table 1.1 on page 47.

We assign the basic colors to the diatonic pitch classes of the scale. The non-diatonic pitch classes are assigned with the (modulated) color of the pitch classes they are derived from. But in contrast to Sapp [2001] we darken the color when the pitch class is altered downwards.

- *Scriabin* This is a mixed PCFKM/PDFKM scheme with onset c = "red". His specific key-color mapping was rediscovered late. The so-called luce-string is a notation for the light-organ accompanying Scriabin's orchestra piece PROMETEI, POEM OF FIRE. The decryption of the luce-string is the basis of the insight in Scriabin's key-color mapping. Scriabin's specific pseudo-synesthetic approach is described in Shaw-Miller [2000]:

  > Scriabin did not ... experience color-tone relationships individually, but rather through chordal complexes, and, according to some sources, deduced the full cycle from his spontaneous recognition of C = red, D = yellow and F-sharp = blue.

  This key-color mapping traces back to Castel [Vanechkina, 1994]. The Scriabin key-color map is a PCFKM, with two exceptions. The grayish blue color codes of the keys (E♭–major, B♭–major) are a greyisch derivation from the color code of (E–major, B–major).

How can we derive a color for minor keys from given colors of major keys? We can use the same or a darker color as for parallel or relative major.

3.7.2 Topographic Ordering Map

We present a learning algorithm that uses Kohonen's idea of establishing a topology preserving map in a new and unusual manner. In the usual SOM (Kohonen [1982], cf. Section 2.3.4 on page 91) the objects under investigation are characterized by feature vectors in a high dimensional space which are assumed to lie approximately on a low dimensional sub-manifold outlined by the given metric on the formal neurons (after successful training). The correspondence between neurons and objects is established by the self-organizing learning process.

In the modified SOM, each neuron represents exactly one object in a fixed a priori correspondence. In the learning process, a suitable placement of the neuron vectors is to be found in a stipulated metrical space that realizes a given neighborhood relation on the neurons.

Let $\mathcal{K} = \{0, \ldots, n\}$ be a finite set of indices and $\mathcal{T}$ be a topological space with metric $d_{\mathcal{T}}$. [18] As given information we stipulate for every pair of indices $i, j \in \mathcal{K}$ a (symmetrical) degree of neighborhood $d_{\mathcal{K}}(i, j) \in \{0, 1\}$, a "0" meaning immediate neighborhood. [19] The learning algorithm is designed to place vectors $\mathbf{w}_0, \ldots, \mathbf{w}_n$

[18] Cf. also Appendices C.1 on page 195 and C.3.

[19] Our algorithm generalizes to the case of arbitrary neighborhood degrees $d_{\mathcal{K}}(i, j) \in \mathbb{R}^{\geq 0}$ in a straight forward manner.

(formal neurons) in the topological space $\mathcal{T}$ in a way that the relation of the immediate neighborhood of the $\mathbf{w}_i$'s in $\mathcal{T}$ coincides with the previously stipulated one, i.e. $\mathbf{w}_i$ and $\mathbf{w}_j$ should be immediate neighbors in $\mathcal{T}$ just in case $d_{\mathcal{K}}(i,j) = 0$. (The aim should be clear, in spite of the absence of an exact meaning of "immediate neighbor" in the metrical space $\mathcal{T}$.) Formally we use the SOM algorithm with certain specifications: Our input samples $\mathbf{w}_i$ are equally distributed across $\mathcal{T}$. Our special assumptions on what is close and remote are encoded in the function $d_{\mathcal{K}}(i,j)$, which is then used as the neighborhood function. Because $d_{\mathcal{K}}(i,j)$ is a function that only maps to $\{0,1\}$, in the update rule for the weights, we have to treat the winner neuron separately in order to approximate the shape of a triangle.

- Choose initial values for vectors $\mathbf{w}_i \in \mathcal{T} \quad (i = 0, \ldots, n)$ by random.
- Iterate the following steps while letting the learning rates η_1 and η_2 decrease towards 0.
  - Choose $\mathbf{x} \in \mathcal{T}$ by random (equally distributed over $\mathcal{T}$).
  - Determine the neuron of maximal excitation, i.e. choose

$$i_{\mathbf{x}} = \arg\min_i (d_{\mathcal{T}}(\mathbf{x}, \mathbf{w}_i)). \tag{3.6}$$

  - Update the neuron vectors according to the following rules:

$$\mathbf{w}_i := \begin{cases} \mathbf{w}_i + \eta_1 \cdot (\mathbf{x} - \mathbf{w}_i) & \text{if } d_{\mathcal{K}}(i, i_{\mathbf{x}}) = 0, i \neq i_{\mathbf{x}} \\ \mathbf{w}_i + \eta_2 \cdot (\mathbf{x} - \mathbf{w}_i) & \text{else if } i = i_{\mathbf{x}} \\ \mathbf{w}_i & \text{otherwise} \end{cases} \tag{3.7}$$

To measure the error (or stress) of a configuration of the neurons we define the neighborhood dissimilarity d_N on $\mathcal{T}$ by

$$d_N(\mathbf{w}_i, \mathbf{w}_j) := \begin{cases} 0 & \text{if } \mathrm{rk}_{d_{\mathcal{T}}}(\mathbf{w}_i, \mathbf{w}_j) \leq \sum_{i \neq k} (1 - d_{\mathcal{K}}(i,k)) \\ 1 & \text{otherwise} \end{cases} \tag{3.8}$$

where $\mathrm{rk}_{d_{\mathcal{T}}}(\mathbf{w}_i, \mathbf{w}_j)$ for $i, j \in \mathcal{K}$, $i \neq j$ is the rank order (Equation C.8) of the distances $d_{\mathcal{T}}$. According to Equation 3.8 we can calculate rank order distortion:

$$\sigma := \frac{1}{(n+1)n} \sum_{i \neq j} |d_{\mathcal{K}}(i,j) - d_N(\mathbf{w}_i, \mathbf{w}_j)|, \tag{3.9}$$

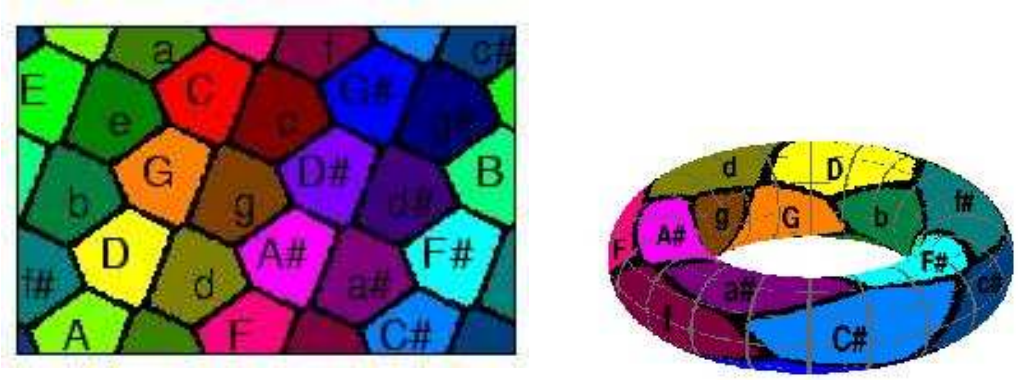
where $n+1$ is the number of weights $\mathbf{w}_i$. This error function gives the possibility to start the algorithm described above several times with different initializations, and to choose the result with minimum stress value.

3.7.3 Discussion

We have introduced means for visualizing inter-key relations, by key-color mappings and by an algorithm for deriving a global arrangement of keys from given key proximities. For a comprehensive coverage of color/tone interrelations cf. Jewanski [1999]. Applications of the topographic ordering map include logistic problems and integrated chip construction. If certain units depend on being adjacent to other ones an optimal global arrangement can be deduced by the topographic ordering map, for the localizations of logistic centers as well as the spacial positioning of the building blocks in an integrated chip.

3.8 Toroidal Models of Inter-Key Relations (TOMIR)

Fed with audio, the circle of fifths evolved in various cognitive models (Section 3.5). We arrive at a torus, referred to as a TOMIR, if we consider relative and parallel major/minor relations also. We will first discuss the topology of a torus. Stipulating two sets of key kinships, we will derive key configurations on a torus, using the topographic ordering map. Then simulations with various cognitive models of key acquisition are preformed. In all simulations a TOMIR emerges. A neuromimetic model consists of an auditory model and a SOM. The model processes piano cadences. Processing the Chopin PRÉLUDES, another cognitive model calculates average CQ-profiles. A SOM is then trained by these profiles. In yet another approach correspondence analysis is employed. The correspondence analysis is fed by profiles of frequency of occurrence from the scores of WTC.



(a) 2-dim. Representation (b) 3-dim. Representation

FIGURE 3.47: The torus as a geometric object can be seen in the 3-dimensional representation in Figure (b). We can reconstruct Figure (b) from Figure (a) by gluing the upper and lower sides and the left and right sides. (Cf. Section 3.8.1 for a note on the visualization.)

Topology of a Toroidal Surface A toroidal surface can be parameterized in different ways. The most prominent are the following:

• A 4-dimensional representation: In this form the toroidal key arrangement is first established in Krumhansl and Kessler [1982] (cf. Figure 1.20 on page 63). Since a four-dimensional space is hard to imagine, Krumhansl and Kessler use their finding about the structure to further scale down the data to a two-dimensional representation (see below).

• A 3-dimensional representation (cf. Figure 3.47 (b)): This is the geometrical object that one would usually think of, when talking about a torus. For the sake of a homogeneous configuration the 2- or the 4-dimensional representation should be used, see the remark below and Krumhansl and Kessler [1982], p. 345.

• A two-dimensional representation (cf. Figure 3.47 a): Each point of the toroidal surface is uniquely determined by two angles in Figure 1.20 and Figure 3.47 (b). So another parameterization is given by the set $[0, 2\pi r_1[\times [0, 2\pi r_2[$ endowed with the toroidal metric C.7 on page 197.

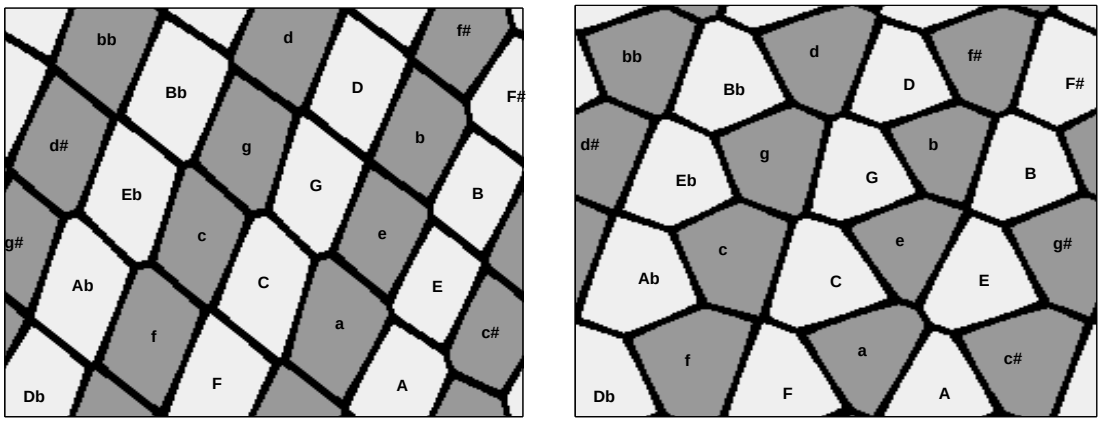
Mathematical Remark: The 2- and the 3-dimensional representation are isomorphic to each other, but they are not isomorphic to the 4-dimensional version. But the induced topological spaces are homeomorphic.

3.8.1 TOMIR Evolving in Topographic Ordering Map

In Section 1.2.7 on page 56 the curling of the strip of [Weber, 1817]'s chart of tone centers leads to a torus. Let us simulate this process using a computer experiment. The topographic ordering map (Section 3.7.2 on page 154) tries to find a toroidal configuration of the keys according to a given set of close relationships, with the understanding that closely related keys should become immediate neighbors in the spatial formation. This set is the only external information about the keys and their relations that is passed to the algorithm. For technical details on the used algorithm see Section 3.7.2.

A Note on our Visualization of SOMs

A SOM (Section 2.3.4 on page 91) as well as the topographic ordering map give rise to a mapping from the set of formal neurons to the set of input vectors by selecting an input vector with minimum Euclidean distance to the given neuron vector. A usual display of a SOM is a direct visualization of this mapping. To get a smoother display we extend this mapping to the whole manifold that is outlined by the neurons. The extension is done by usual interpolation. By this we associate with each input vector a possibly unconnected region on the manifold. In the simulations of Sections 3.8.2 and 3.8.3 that manifold is a toroidal surface and there are always 24 input vectors, each representing a major or minor key. We use 21×12 formal neurons supplied with the toroidal metric, Equation C.7 on page 197. Neurons are arranged in a mesh, gluing together opposite borders. A point is colored black if the distance (of its projection) to the nearest input vector is not sufficiently smaller



(a) Relationship $\mathcal{V}_1$ resembling the local structure of Weber [1817]'s chart of tone centers (Figure 1.14)

(b) Relationship $\mathcal{V}_2$ yields a Krumhansl like configuration (Figure 1.21)

FIGURE 3.48: Arrangement of keys evolving from sets of close relationships by the topographic ordering map. (Cf. p. 157 for the visualization technique used.)

than the distance to the second nearest. So each region gets a black border whose thickness corresponds to uncertainty, in a relative measure. The placement of the key name is determined by the position of the neuron with minimum distance to the input vector that corresponds to that key.

Simulations

We simulate the toroid configuration derived in Section 1.2.7 on page 56 from tone centers considered to be close to each other according to Weber [1817]. Those are the dominant, subdominant, relative, and parallel kinships (see Figure 1.14 on page 56). For the topographic ordering map, the appropriate set of close relations is

$$\mathcal{V}_1 := \{\text{C–G, C–F, C–a, C–c, c–g, c–f, c–E}\flat, \ \dots\} \tag{3.10}$$

where the dots stand for analogous relations between keys with different tonic keynotes. Even in this setting, where the set $\mathcal{V}_1$ is the only information given about the structure of inter-key relations, the algorithm ends up with the arrangement depicted in Figure 3.48 (a). We want to stress that in this simulation a global arrangement of tone centers, like in Weber [1817]'s chart, Figure 1.14, evolves, but is not presupposed.

In comparison, let us consider relations between tone centers resulting from maximization of common tones between adjacent scales as well as maximizing common tones between their respective tonic triads. The tonic's kin, the dominant, subdominant, relative, and parallel, is now additionally joined by the mediant. Under this

stipulation, e–minor is as well an immediate neighbor of C–major, as F–major, G–major, and a–minor are. To integrate the strengthening of the mediant relation we expand the set of close relationship to

$$\mathcal{V}_2 := \mathcal{V}_1 \cup \{\text{C–e, c–G}\sharp, \ldots\}.$$

This is consistent with Krumhansl's probe tone ratings showing a strong correlation of keys between the tonic and its mediant (Figure 1.21, Krumhansl [1990], p.39 & 46). The resulting configuration shown in Figure 3.48 (b) is quite close to Krumhansl's. Of course, the small amount of information that is used in this simulation is not sufficient to produce finer distinctions. In Krumhansl's arrangement a–minor, for example, is closer to C–major than e–minor.

3.8.2 TOMIR Evolving in a Neuromimetic Environment

The simple cadential chord progression, $\mathrm{I} - \mathrm{IV} - \mathrm{V}^7 - \mathrm{I}$, in all major and minor keys is used. The chords are sampled using Shepard tones (cf. Section 1.2.3 on page 48). A signal is preprocessed by digital filters modeling the outer and middle ear, a filter bank modeling the basilar membrane [Patterson et al., 1988], and the hair cell model (Meddis and Hewitt [1991], Section 2.2.1). The perceptual principle of the missing

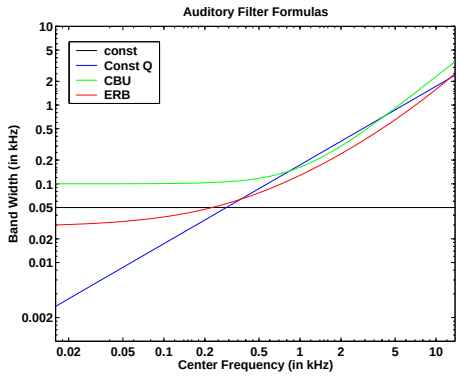


FIGURE 3.49: Modeling of the basilar membrane as a band pass filter bank: The discrete Fourier transform (e.g. FFT) gives a filter bank with a constant band width (e.g. 50 Hz). The constant Q transform (cf. Section 2.1) gives a filter bank with constant band width (minor third) in the *logarithmic* frequency domain. Two equations approximate the results in psychophysics experiments: 1) the critical bandwidth (Bark: Zwicker and Fastl [1990], Equation 1.3 on page 34), 2) the equivalent rectangular bandwidth equation (ERB, Moore and Glasberg [1983], Equation 1.2 on page 34). The equally spaced discrete Fourier transform is an approximation for the auditory filter bank for center frequencies below 500 Hz. The constant Q filter bank is an approximation above 500 Hz. It is a reasonable compromise between psychoacoustic modeling and an efficient implementation.

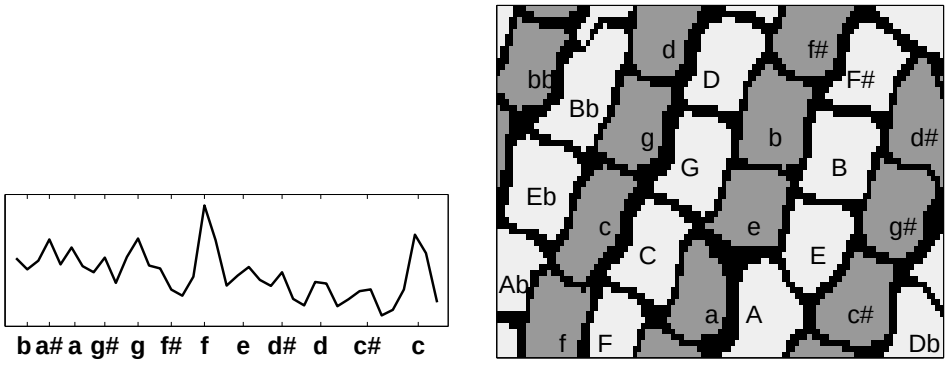
Par.	*Meaning in Model*	*Value*
A	Permeability	100.00
B	Permeability	6000.00
g	Release rate in units per s	2000.00
y	Replenishment rate in units per s	5.05
l	Loss rate in units per s	2500.00
x	Reprocessing rate in units per s	66.31
r	Recovery rate in units per s	6580.00
m	Max no. of transmitter packets in free pool	1

TABLE 3.3: Parameters used in the hair cell model (Meddis [1988], Section 2.2.1 on page 71).

fundamental (Section 1.1.6) is coarsely modeled by the application of autocorrelation (Section 2.21 on page 79). For temporal processing, memory is simulated by an exponentially weighted integration along the time axis ("leaky integrator").

We employ the Development System for Auditory Modeling (DSAM) by O'Mard et al. [1997]. For convenience we use the implementation as a Matlab function, by Benjamin Blankertz [Purwins et al., 2000a]. Fourth order gamma tone filters [Moore and Glasberg, 1983] for the filter bank, and the hair cell model [Meddis, 1988] are employed.

The choice of the center frequencies as well as the post-auditory processing are adopted from Leman [1995], with the exception of raising the downsampling rate from 2500 Hz to 5000 Hz and canceling the parabolic attenuation ($\alpha = 0$) in the



(a) Compact profile of C–major

(b) Toroidal arrangement of compact profiles

FIGURE 3.50: Cadential chord progressions made of Shepard tones are filtered by an auditory model and then autocorrelated (a). A toroidal SOM arranges the compact profiles of the different keys similar to Figure 1.17 (b).

windowed autocorrelation function. This is not necessary in view of the results but it yields nicer tone center profiles. Since convergence is no problem in our simulations we do not pay much attention to the control of the learning rate in the training of the SOM (Figure 3.50 (b)). The learning parameter in the batch learning is linearly decreased from 0.15 to 0.001 in 10000 iterations.

To obtain a relation to Krumhansl and Shepard's probe tone ratings (cf. p. 61 f.), we restrict the correlograms to components corresponding to the frequencies in the octave range 64 Hz – 128 Hz, as seen in Figure 3.50 (a). This octave is so low that the autocorrelation of all notes in the usual range have non-zero values within that range.

3.8.3 TOMIR Evolving in a Cognitive Model from Averaged Performance

For the sake of building a cognitive model for tone center perception, we use the constant Q profile technique in combination with the SOM. Since cycles with pieces in all keys are a suitable musical form for establishing tone centers, we choose Chopin's Préludes op. 28, for the calculation of a tonal reference set. To show stability of CQ-profiles with respect to sound quality, we take the historical recording of Alfred Cortot from 1933/34 (EMI References), in audio data format. We calculate the CQ-transform (Section 3.3.1 on page 121), within short-time frames, with 36 bins per octave. This corresponds to the basilar membrane filter bank. Then the constant Q transform is compressed into a CQ-profile according to Equation 3.3 on page 122. As a simple model of memory, the CQ-profile of a whole prelude is calculated by

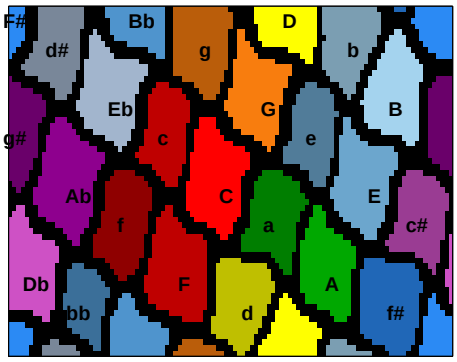
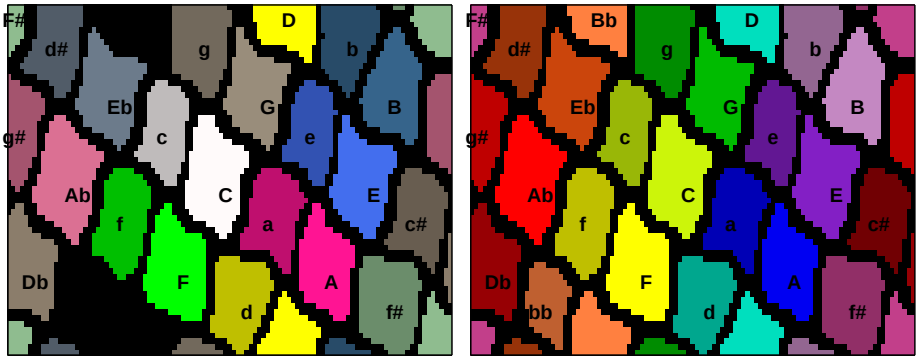


Figure 3.51: Emergence of inter-key relations. Inter-key relations are derived from a setup including constant Q profile calculation and a toroidal SOM trained by Chopin's Préludes op. 28 recorded by A. Cortot in 1932/33, in audio data format. The image shows a torus surface. Upper and lower, left and right sides are to be glued together. Keys with dominant, major/minor parallel, or relative relations are proximate. For color display Scriabin's key-color map is used. Cf. Figure 3.52 for other key-color maps.



(a) Rimsky-Korsakov (b) The Rosicrucians

FIGURE 3.52: The key configuration from Figure 3.51 displayed in different synesthetic key to color mappings (cf. Section 3.7.1 on page 152).

summing up CQ-profiles of short-time frames. The topographic principle in the SOM corresponds to tonotopy [Schreiner and Langner, 1988] in the auditory domain. A toroidal SOM is trained with the average vectors from all single preludes in Chopin op. 28. Convergence of the SOM does not critically depend on the learning parameter. They are chosen as in Section 3.8.2.

The configuration resembles Krumhansl [1990]'s, and even more closely the one derived from Weber [1817]'s chart of tone centers (Figure 1.14). The circle of fifths wraps around the torus three times. Closely related keys according to Schönberg [1969] are neighbors.

Tone Center Dissimilarities: Theory – Model Figures 3.54 and 3.53 show the dissimilarities implied by different music theory writings [Schönberg, 1969; Chew, 2001] compared to the dissimilarities implied by the SOM (Figure 3.51). Dissimilarities are given for major (Figures 3.53) as well as for minor (Figures 3.54). For classes (two top Graphs) and regions (Graph 3 and 4 from top) in Schönberg [1969] and in the model of Chew [2000] enharmonic spelling is ignored. The enharmonic equivalence class is represented by the equivalent tone center closest to the tonic. Schönberg [1969] makes a difference between a region of tone centers of a major (Figure 1.15) and minor key. He considers the region of tone centers of a minor key to be smaller. The region is so small that certain keys lie outside that region. In Figure 3.54 the outsider position is indicated by an assigned rank of 12 in Graph 3 or rank 5 in Graph 4. Graph 5 gives the Euclidean tone center distances in Chew [2001]'s model (Figure 1.22 on page 65) with the parameter set suggested by her. For a given key, tone center dissimilarities are qualitatively similar in Schönberg's classes (Graph 1) and regions (Graph 3). This holds in particular for minor tone centers. In Graph 6 the tone center distances are given on the surface of a 2-dimensional toroidal SOM (Figure 3.51). Ranks and dissimilarities are highly similar for the SOM (Graph 6) and Chew [2001] (Graph 5), especially for the minor tone centers. In Figure 3.54,

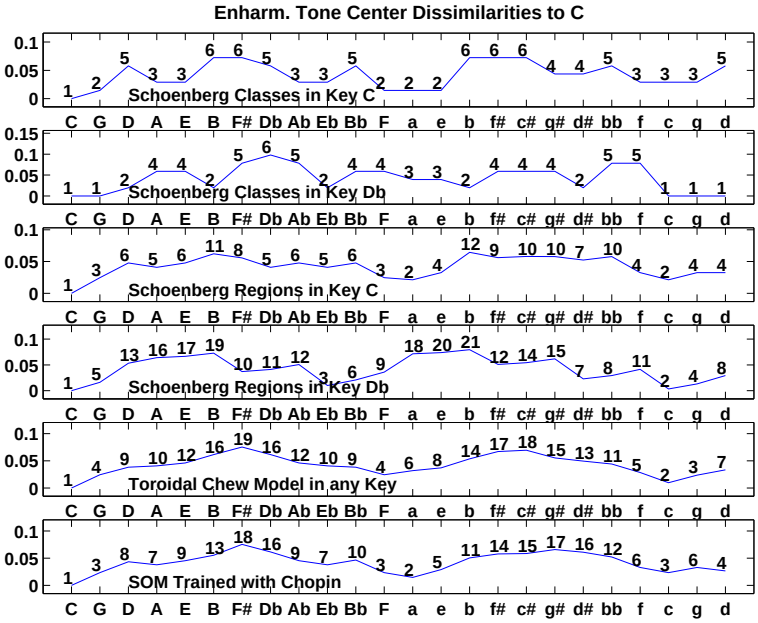


FIGURE 3.53: Dissimilarities between C–major and all other keys on a SOM compared to music theory. The five upper graphs show dissimilarities between tone centers implied by different music theory concepts. The lowest graph shows the dissimilarities on the surface of a torus, resulting from training a toroidal SOM with the Chopin preludes (cf. Figure 3.51 and Section 3.8.3).

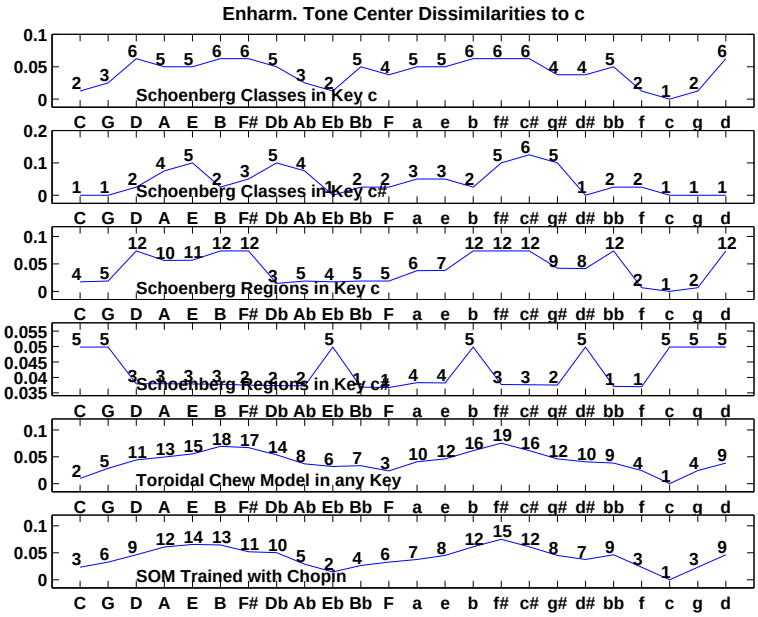


FIGURE 3.54: Dissimilarities between c–minor and all other keys on a SOM compared to music theory (cf. Figure 3.53 and Section 3.8.3).

Graphs 1, 3, 5, and 6, there are local peaks for $f\sharp$–minor and $b\flat$–minor. In Weber [1817]'s chart (Figure 1.14 on page 56) tone centers are spaced homogeneously. But in Schönberg [1969]'s chart (Figure 1.15) tone centers are more squeezed together farther away from the key. Dissimilarities of tone centers are relative to the key, e.g. compare the first and the third line (c–minor and C–major) to the second and fourth line ($c\sharp$–minor and $D\flat$–major) in Figures 3.53 and 3.54. Consider a piece modulating into a tone center far away from its key. Could the global key still be extracted from the average CQ-profile of that passage? That is a question for future research.

3.8.4 TOMIR Emergence in Correspondence Analysis from Score

Figure 3.37 on page 142 displays the projection of keys and pitch classes onto the plane spanned by the two most prominent factors. How can we visualize the projection onto the first four factors? We represent points on each of the planes spanned by the first & second and third & fourth factor, respectively, in polar co-ordinates, i.e., by their polar angle and by their distance to the origin. We then plot their angle in the 1-2-plane against their angle in the 3-4-plane (upper left Figure 3.55 on the facing page). Topologically, we can think of the resulting body as the surface of a torus, which can be parameterized by two angles. Upper left Figure 3.55 can be viewed as a torus if we consider vertical and horizontal periodicity, i.e., we glue together the upper and lower side as well as the right and left side. The three circles of fifths then meander around the torus three times as indicated by the solid, dashed, and dotted lines. In addition to the relationship regarding fifths in upper left Figure 3.55 we see that the projection on the 3. and 4. factor contains information about the interrelation between major keys and their parallel and relative minor keys.

Consistency with Chew's Geometric Model. It is fruitful to compare the toroidal interpretation (upper left Figure 3.55 on the next page) of the biplot of keys and pitch classes (lower Figure 3.37) with Chew [2000] (upper right Figure 3.55). In Chew [2000] heterogeneous musical quantities, in particular, tones, chords, and keys are embedded in a three-dimensional space, thereby visualizing their inter-relations (cf. Section 1.2.10 on page 64 and Appendix C.3.3 for technical details). We reduce this model to pitch classes and keys, assuming that the intermediate level of chords is implicitly given in the music.

Chew [2000] gives a set of parameters derived by optimization techniques from musically meaningful constraints. We choose a different set of parameters to fit the model to the result of our correspondence analysis as displayed in upper left Figure 3.55. (Cf. Appendix C.3.3 for parameters.)

In order to facilitate comparison we choose a two-dimensional visualization of the three-dimensional model in Chew [2000]. The projection of the model onto the X-Y-plane is circular. Therefore we can parameterize it as angle and length. We plot the vertical dimension (the elevation of the helix) versus the phase angle of the X-Y-plane (upper right Figure 3.55). We interpret the phase angle of the X-Y-plane as the first

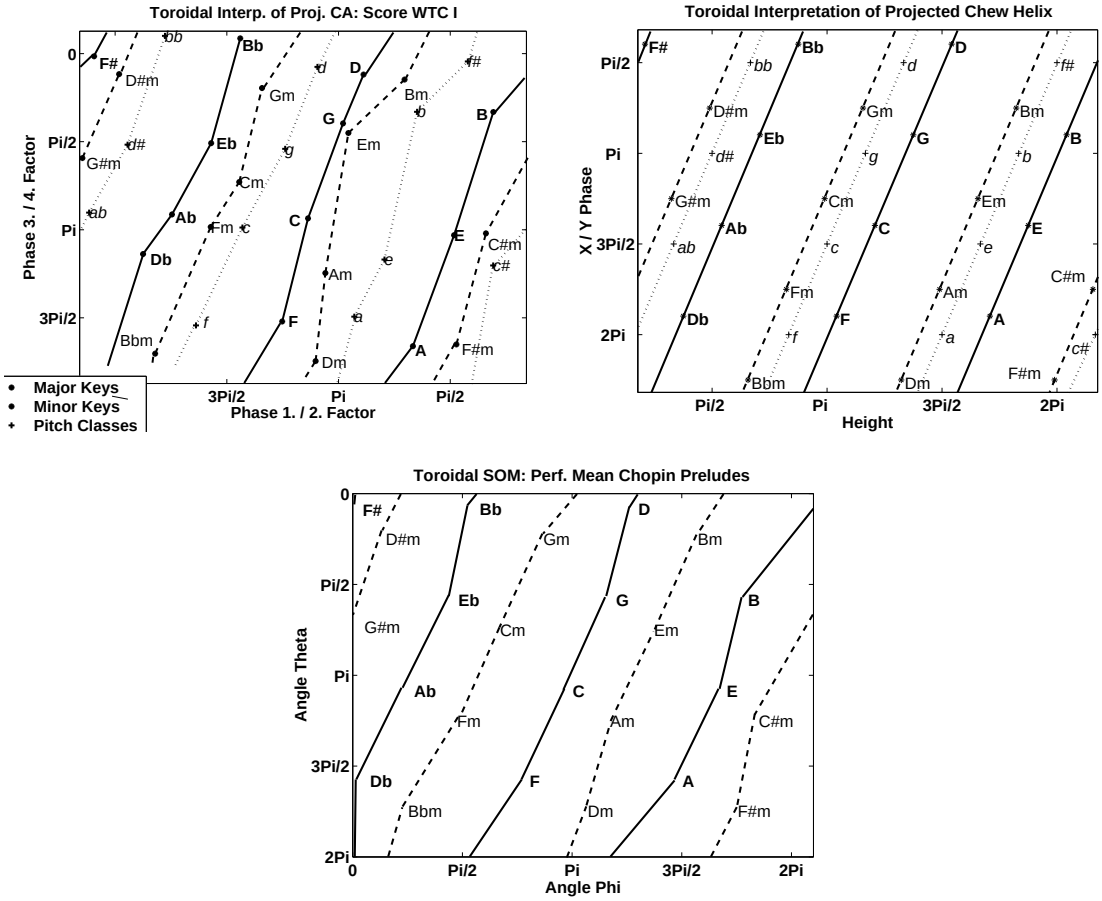


FIGURE 3.55: The toroidal interpretation of the projection onto the first four factors (*upper left*) is consistent with Chew [2000] (*upper right*) and Section 3.8.3 (*lower*, same Figure as 3.51 and 3.52 in a different display). Capital letters denote keys ("m" for minor). Small italics indicate pitch classes. A torus can be described in terms of two angles, displayed on the horizontal and vertical axis. Glue the upper/lower and right/left side of the plots together to obtain a torus (cf. text).

angle of a torus, and the vertical height in the helix as the second angle of a torus. The helix is mapped on the surface of a torus by applying modulo $12h$ to the height. Here h is the distance on the vertical co-ordinate of two successive tones in the circle of fifths. We observe that the upper Figures 3.55 are very similar: The circles of fifths in major and minor keys and in pitch classes curl around the torus three times. The only difference is that in the toroidal model derived from correspondence analysis major keys and their relative minor keys are nearby, whereas in upper right Figure 3.55 major keys are closer to their parallel minor keys.

3.8.5 Discussion

We have seen how distances between tone centers are defined by the number of common notes in the corresponding scales and by the identity of scale degree meaning of shared pitch classes (Section 1.2.5). From these distances, Weber [1817]'s chart of tone centers is derived (Section 1.2.6). By synchromatic and enharmonic identification a strip of the chart curls resulting in the toroidal model of inter-key relations (TOMIR, Section 1.2.7). A TOMIR can be found in the MDS analysis of the psychological probe tone ratings (Section 1.2.9). A TOMIR appears also in correspondence analysis of scores (WTC). The TOMIR emerges in a neuromimetic model consisting of an auditory model and the SOM fed with audio (Chopin). Furthermore, the TOMIR also emerges in other models (CQ-models) consisting of CQ-preprocessing and SOM, correspondence analysis, or Isomap fed with audio. These experiments support the hypothesis of *psycho-physical parallelism* that sees the "mental world as an epiphenomenon of the physical world" (Leman [1995], p. 182). The TOMIR is part of the mental world. That is indicated by the music theory considerations, the curling of Weber's chart resulting in a torus. Also the MDS analysis of probetone ratings shows that the TOMIR is part of the mental world. The physical world is represented by the neuromimetic model or the CQ-models operating on audio. In the experiments, the TOMIR evolves in the model of the physical world.

The central result of this thesis is that the hypothesis of psycho-physical parallelism for the TOMIR is supported in a generalized form, maintaining a comparable level of biological relevance. First, the hypothesis is demonstrated based on a broader variety of samples, actual recordings of performed music by Bach and Chopin, instead of a limited number of cadential patterns, synthesized with artificially sounding Shepard tones. Second, fewer assumptions are needed to verify the hypothesis. The toroidal structure is not stipulated in the architecture.

The neuromimetic model and the CQ-model are of comparable neuromimetic relevance. The specific details of the auditory process, as implemented in an auditory model, are not particularly relevant for the cognitive processing of inter-key relations (Section 2.2.2). One of the main functionalities of Meddis [1988]'s hair cell model is the response to onset, sustain phase, and offset. These features are of minor relevance for the overall tonal content. Since for autocorrelation the biological plausibility is questioned, we use CQ-profiles, without loosing confirmed biological relevance. CQ-profiles are consistent with the Weber-Fechner rule (Equation 1.1) of logarithmic perception of frequency and they reveal high coherence with probe tone ratings in music psychology (Section 3.3.3).

The auditory system is not yet well understood. Most auditory models cover only roughly the first stages of auditory processing. The SOM [Kohonen, 1982], being considered as a model of the cortex, is an extreme simplification, simulating just topography. We think more knowledge about neural processing of sound is needed to make a well grounded hypothesis on the exact representation of tone centers in the cortex. Our auditory model/SOM setup in Section 3.8.2 is related to Leman [1995], p. 92. But Leman uses Shepard tones, in effect a non-audio representation of

a tone. In addition, Leman employs Terhardt's simplified virtual pitch model. We use a different auditory model [O'Mard et al., 1997].

In correspondence analysis (Section 3.8.4), we have shown how meaningful parameters in the complex structure of music can be visualized, thereby revealing the interrelations of music looked upon in the perspective of a certain parameter. To demonstrate the high potential of this approach we have given examples in the domain of inter-key relations based on the frequency of pitch class usage. The benefit of the method is its simplicity. But nonetheless, it does require almost no assumptions, only the most rudimentary musical presuppositions, tuning and octave equivalence and no special artificial data. In Leman [1995], artificially generated cadential chord progressions constructed from Shepard tones are used as training data. In the CQ-profile based model (Section 3.8.3, Purwins et al. [2000a]), we used profiles of digitized sound samples averaged across the entire cycle (Chopin's Préludes op. 28). In contrast, in Sections 3.5.3 (Isomap) and 3.8.4 (correspondence analysis) we used the profiles of digitized sound samples of each piece separately (WTC). In the neuromimetic model (Section 3.8.2 and Leman [1995]) and the CQ-profile based model (Section 3.8.3, Purwins et al. [2000a]), the toroidal key structure is implicitly stipulated by training a *toroidal* self-organizing feature map. In the simulation with correspondence analysis, in Section 3.8.4, the circularity emerges from the data alone, without an implicit assumption of periodicity in the model. Likewise, the circle of fifths emerges in the simulation with Isomap in Section 3.5.3. In this sense, our analysis can be viewed as discovering a model of circular structure rather than merely fitting such a model that has been stipulated before.

The available data has not been exhaustively analyzed. Projections to different sub-planes could be explored and interpreted. Correspondence analysis can be used to model an experienced listener exposed to a new piece of music. In correspondence analysis, this would correspond to embedding the new pieces in a co-ordinate system obtained from analyzing the pieces heard before. As an example, Bach's WTC can be used to generate a tonal co-ordinate system. In this co-ordinate system other works can be embedded, such as Chopin's Préludes, Alkan's Préludes, Scriabin's Preludes, Shostakovich's Preludes, and Hindemith's Ludus Tonalis. In this way the method can be used to model how a listener who is familiar with Bach's WTC would perceive keys and pitches in the more recent works. In addition, alternative concepts of inter-key relations underlying Hindemith and Scriabin may be discovered.

We would like to emphasize that the use of correspondence analysis is by no means limited to tonality analysis. The method is a universal and practical tool for discovering and analyzing correspondences between various musical parameters that are adequately represented by co-occurrences of certain musical events or objects. Examples include pitch class, key, instrumentation, rhythm, composer, and style. Three-dimensional co-occurrence arrays, for instance of pitch class, key, and metric position can be analyzed. In particular, it seems promising to extend our analysis to temporal transitions in the space of musical parameters.

In an fMRI study, using the representation of a tone center parameterized by two

toroidal angles on the TOMIR, Janata et al. [2002] [20] find evidence for localizing brain activity related to tone center modulation. Text-book like modulations, synthesized with artificially sounding FM-clarinets, are played to the subjects. Two toroidal angles on the TOMIR seem to be a representation of tone centers preserving sufficient information to identify voxels that are sensitive to tone center transitions. At this stage of research, non-invasive brain imaging techniques seem to be not yet capable to indicate whether voxels that are sensitive to *distinct* tone centers are spatially arranged in the brain in a certain manner, e.g. in a torus or in any other, for example more dynamic, configuration. An actual neural correlate of a dynamic topography may not give equal space to all of the tone centers, since e.g. tonic and dominant occur more often than remote keys. The modulation from one scale degree to another one is not equally likely as the reverse transition. E.g. I – vi occurs more often than vi – I.

3.9 Sketch – First Steps Toward Assignment of Tone Center Transitions

Based on CQ-profiles we have investigated styles, keys, and visualizations of their inter-relations. We have only considered average CQ-profiles across entire pieces, neglecting sequential aspects. We will now analyze sequences of profiles averaged over small sections of a piece. The segmentation is consistent with musical structure apparent in beats or bars. Tone center transitions and modulations are assigned on multiple time scales. They are displayed as trajectories on landscapes of tone centers. Tonal ambiguity and polytonality is measured. The outlined framework for assignment of tone center transitions is demonstrated for a couple of examples such as Chopin's Prélude op. 28 No. 20. The presented methods in this section should be considered *sketchy auxiliary results*.

3.9.1 Timing and Tone Centers

We will describe a rudimentary way how to calculate onsets and beat weights.

Onset Detection The onset detection algorithm introduced here first calculates the quotient of absolute intensities downsampled at different sampling rates (Equations 3.11 and 3.12). From the result recursively the discrete time derivative is thresholded and fed to an indicator function twice (Equations 3.13 and 3.14). Finally clusters of onset points are replaced by the last onset. An efficient method for onset detection is suggested by Rodet and Jaillet [2001].

For a sound signal $\mathbf{s}$ with the original sampling rate at time t_0 the relative intensity $\mathbf{i}_{\text{rel}}$ is calculated as the quotient between the mean absolute intensities $\mathbf{i}_{w_1}$ and $\mathbf{i}_{w_2}$ of

[20] Petr Janata mentioned the employment of this particular representation in personal communication.

Parameter	*Symbol*	*Value*
Sampling Rate	f_s	5512.5
Cut-off Threshold	T_Δ	0.1
Cut-off Threshold	T_{Δ^2}	0.3
JN Time Lag	w_1	0.05 s
Mean Window	w_2	1 s
Unification Distance	d	3
Phase Correction	p	0.15 s

TABLE 3.4: Parameters of onset analysis of Chopin's op. 28 No. 20.

a short (w_1) and a long time frame (w_2)

$$i_{\text{rel}}[t] = \frac{i_{w_1}[t]}{i_{w_2}[t]} \tag{3.11}$$

with

$$i_w[t_0] = \frac{1}{w} \sum_{t=t_0-w+1}^{t_0} |s[t]|. \tag{3.12}$$

With the discrete time derivative $\frac{\Delta \mathbf{i}_{\text{rel}}}{\Delta \mathbf{t}}$ we extract what is above threshold T_Δ

$$\mathbf{i}' = \mathbf{1}_{\{\frac{\Delta \mathbf{i}_{\text{rel}}}{\Delta \mathbf{t}} - T_\Delta > 0\}}. \tag{3.13}$$

We calculate the onset **o** relative to threshold T_{Δ^2} with an indicator function

$$\mathbf{o} = \mathbf{1}_{\{\frac{\Delta \mathbf{i}'}{\Delta \mathbf{t}} - T_{\Delta^2} > 0\}}. \tag{3.14}$$

For time t, $o[t]$ indicates whether there is an onset ($o[t] = 1$) or not ($o[t] = 0$). By Equation 3.14, from multiple consecutive onsets the first one is singled out. In a last step accumulations of beats are unified: If $o[t_1]$ has a preceding onset $o[t_0]$ less than d resolution steps before (on the time scale downsampled by w_1), the preceding onset $o[t_0]$ will be removed

If $o[t_1] = 1$ **and** $\exists t_0 : t_1 - d + 1 \leq t_0 \leq t_1 - 1 : o[t_0] = 1$
then $o[t_0] = 0$.

In Figure 3.56, the beginning of the first prelude in C–major of WTC I, recorded by Glenn Gould, is analyzed. A JN time lag w_1 of 0.032 seconds and a mean window w_2 of 1 second are used in Equation 3.11. For Equations 3.12 – 3.14 parameters in Table 3.9.1 are used. All onsets are captured correctly, except in two instances (red arrows) the detected onset is distorted. The distortion is caused by piano pedaling that erodes the sharp onsets.

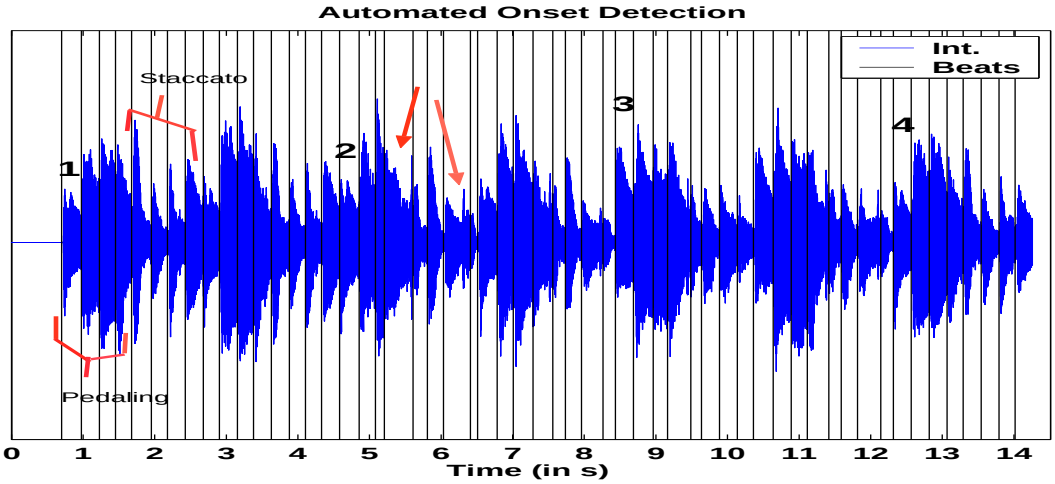


FIGURE 3.56: Automated onset detection in the first three and a half bars of the prelude C–major of WTC I recorded by Glenn Gould. The vertical bars indicate detected onsets. Numbers in the figure count the bar beginnings. All onsets are captured correctly, except twice (arrows) the onset is detected on a distorted position, since pedaling on the piano erodes the clear cut onsets. As a comparison, the sixteenth notes in the second half of the first half bar have sharp onsets in the time domain. The loudest notes do not occur on the first beat of a bar.

Beat Strength Extraction Based on successfully extracted onsets, by periodicity analysis it is possible to obtain further temporal information, for instance on beat strength, meter, and tempo. A diagram of beat strengths indicates a hierarchy of cycle durations. Fleischer [2002] proposes a model for the calculation of beat weights, based on calculating local meters from annotated note onsets. Therefore her approach cannot extract beat strengths from expressively varied note onsets in real performances. Also her model could not compensate for incorrectly extracted onsets. It is beyond the scope of this thesis to evaluate various beat weight calculation methods such as Scheirer [1998] based on an oscillator model, Seppänen [2001]; Tzanetakis et al. [2002]; Gouyon and Herrera [2003]; Eisenberg et al. [2004], Uhle and Dittmar [2004] using independent subspace analysis, and Lang and de Freitas [2004] employing graphical models.

To calculate beat strength, the binary onsets $o[t]$ are subject to autocorrelation (Equation 2.21 on page 79) yielding $\mathbf{x}^{\text{acf}}$. A linearly increasing window $\mathbf{w}$ is then multiplied component wise with $\mathbf{x}^{\text{acf}}$ to compensate for the summation of a different number of frames in the autocorrelation function resulting in the modified correlogram

$$p[t] = x^{\text{acf}}[t] \cdot w[t]. \tag{3.15}$$

To achieve an order of beat weight prominence, we sort the normed $p[t]$ decreasingly.

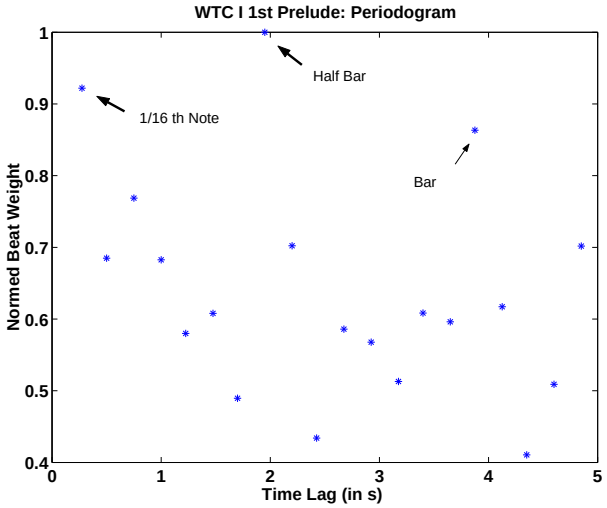


FIGURE 3.57: Periodogram of WTC I (Glenn Gould), first five seconds: The most prominent period is the half bar length, then sixteenth's length, then bar length. The half bar length corresponds to the melodic pattern repeating every half bar.

Again the beginning of the first prelude C–major in WTC I serves as an example. In Figure 3.57 from previously analyzed onsets, beat weights are calculated in a correlogram using Equations 2.21 and 3.15. Prominent cycle durations on different time levels are revealed. They correspond to the minimal duration of notes ($\frac{1}{16}$th notes), beats, half bars, and bars. The half bars are most prominent, since the same melodic pattern repeats every half bar. For longer sections of a piece also phrases and formal parts may be indicated in the correlogram.

We have developed only a rudimentary tool to aid the analysis of tone center sequences. It is beyond the scope of this thesis to optimize these tools in various ways, such as: 1) Onset detection can be performed on the outcome of a filter bank. 2) The artifacts of taking a discrete derivative of a previously thresholded function are to be discussed. 3) The binary onsets fed to the autocorrelation may be replaced by Gaussians centered at the onset in order to improve the capturing of tempo changes. [21] 4) The problem has to be explored that autocorrelation captures half of the speed of a given pulse but not double speed. This increases the difficulty of detecting fast tempi.

3.9.2 Chopin's c–Minor Prélude

Containing several modulations with clear cadences, Chopin's c–minor PRÉLUDE op. 28 No. 20 is well suited as a running example for the assignment of tone center transitions. In the score (Figure 3.58 on the following page) tone centers are

[21] Personal communication with Klaus Frieler.

marked. They were determined by a musicologist. Tone centers in parentheses indicate tonicizations on a very short time scale. The prelude is written in the key of c–minor. Each of the first four bars contains a cadences. A descending line follows in bars 5–9, echoed in bars 9–12 at *pp*, before the final chord in bar 13. The piece gives the impression of a withdrawing funeral march [Koczalski, 1909; Leichtentritt, 1921]. The same rhythmic pattern is maintained throughout the entire piece. A detailed scale degree analysis reveals: bar 1 (c): $\mathrm{I} - \mathrm{IV} - \mathrm{V}^{87}_{65} - \mathrm{I}$, bar 2 (A♭): $\mathrm{I} - \mathrm{IV} - \mathrm{V}^{87}_{65} - \mathrm{I}$, bar 3 (c): $\mathrm{V}^7 - \mathrm{I}^{\flat 7} - \mathrm{iv}^{87} - \mathrm{i(I)}$, bar 4 (G): $\mathrm{V}^7 - \mathrm{I} - \mathrm{V}^{7}_{65} - \mathrm{I}$, bar 5 (c): $\mathrm{i} - \mathrm{VI} - \mathrm{VII}^{\flat 7}_{\sharp 5 \flat 7} - \mathrm{v}$, bar 6 (G): $\mathrm{ii}^{7}_{\flat 7} - \mathrm{V}^{7}_{\flat 5 \flat 5} - \mathrm{I}^{54}_{32} - \mathrm{I}^{7}_{7}$, bar 7 (c): $\mathrm{i}_3 - \mathrm{iv} - \mathrm{V}^{87}_{3} - \mathrm{i}$, bar 8: $\mathrm{VI} - \mathrm{II}\flat - \mathrm{V}^{7}_{65} - \mathrm{I}$. In bar 1 there is a stereotype cadence in c–minor. In bar 2 as well there is a clear cadence, in A♭–major. A plagal cadence in c–minor occurs in bar 3. But the tone center here is less clear. Due to the interdominant C^7 (2nd beat) there is a short tonicization for f–minor. It is debated whether the last chord in bar 3 should be a C–major (e') or a c–minor chord ($e\flat'$).[22] However, in the recording that we will analyze Alfred Cortot plays the C–major chord. Again a clear cadence, in G–major, occurs in bar 4. From bar 5 beat 2 until bar 6 beat 3 the bass line descends, doubled in octaves, in chromatic steps every quarter note, while the leading voice

[22] In bar 3 on the last beat there is an e'. Zimmermann [1969] considers this a typo in Chopin's manuscript. He argues for replacing e' by $e\flat'$, because major and minor keys should alternate within the first four bars. In Chopin [1957] it is mentioned that Chopin himself manually corrected this in Jane Stirling's copy.

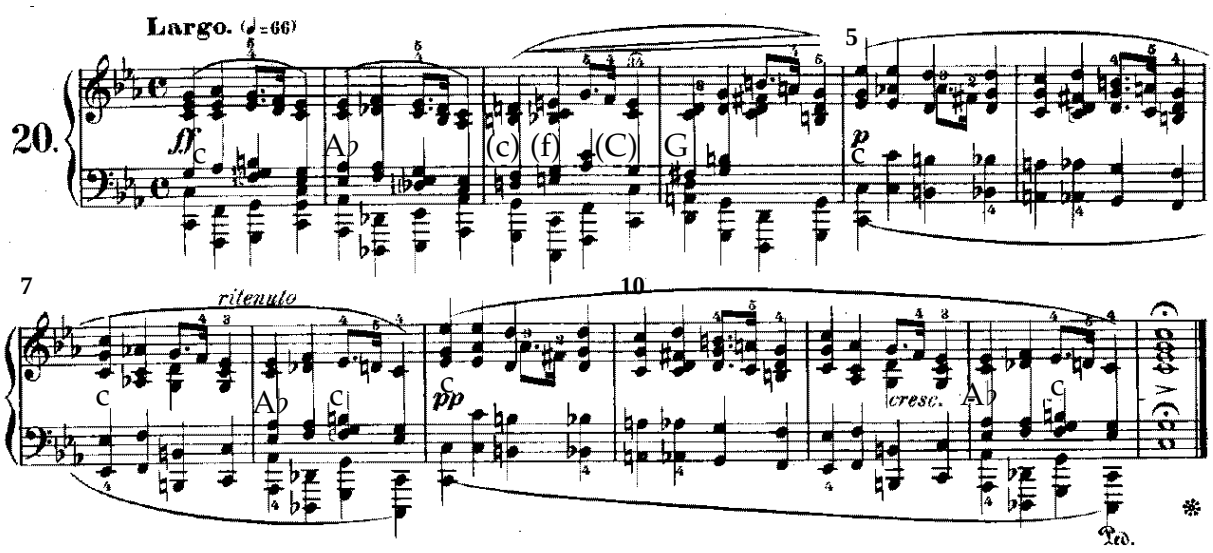


FIGURE 3.58: Chopin's c–minor PRÉLUDE op. 28 No. 20. Tone center analysis is given by a musicologist. Tone centers in parentheses indicate tonicizations on a very short time scale.

follows an ornamented line descending in seconds on the 1st and 3rd beat. In bar 7, there is a clear cadence in c–minor. The A♭–major and D♭–major chords in bar 8 (1st and 2nd beat) give a flavor of A♭–major. But the perception is ambiguous, since the D♭–major chord can be associated with the Neapolitan chord in c–minor, although the fundamental is in the bass.[23] Therefore, the perceived tone center oscillates from c–minor (bar 7) to ambiguous A♭/c–minor (1st and 2nd beat of bar 8) and back to c–minor (bar 8, 3rd and 4th beat).

3.9.3 Moving Average Analysis

In the simplest scenario for the assignment of tone center transitions, the moving average of a sequence of CQ-profiles is matched against a set of key prototypes, the reference set, employing an appropriate distance measure. From a recording of Chopin's Prélude op. 28 No. 20 in audio data format, performed by Alfred Cortot, a sequence of 20 CQ-profiles per s is calculated. CQ-profiles are averaged across a time frame of 4 s length using a cosine window. A reference set (Section 3.3.1 on page 123) is yielded as follows. For each mode an average NCQ-profile of all Chopin Préludes is generated. In turn, these average CQ-profiles are transposed to all keys of the same mode. Then each of the short-term average CQ-profiles is assigned to the key whose reference vector is closest according to maximal correlation.

[23] Cf. the footnote on p. 56.

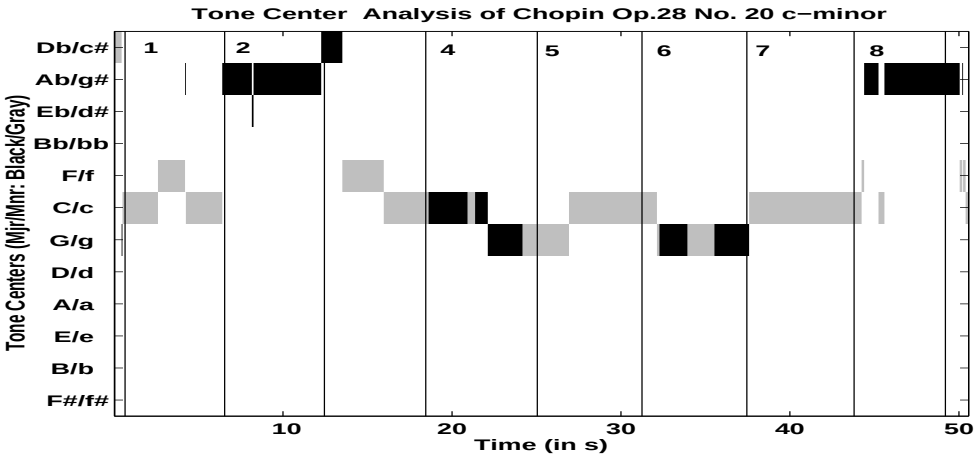


Figure 3.59: Moving average analysis of Chopin's c–minor Prélude op. 28 No. 20. Gray indicates minor, black indicates major. If there is neither black nor gray at a certain time, the significance of a particular key is below a given threshold. Vertical lines indicate bars with the number on top of the figure. There is no distinction between enharmonically equivalent keys. The analysis is highly consistent with a musicologist's analysis in Figure 3.58 on the facing page. (Cf. text for details.)

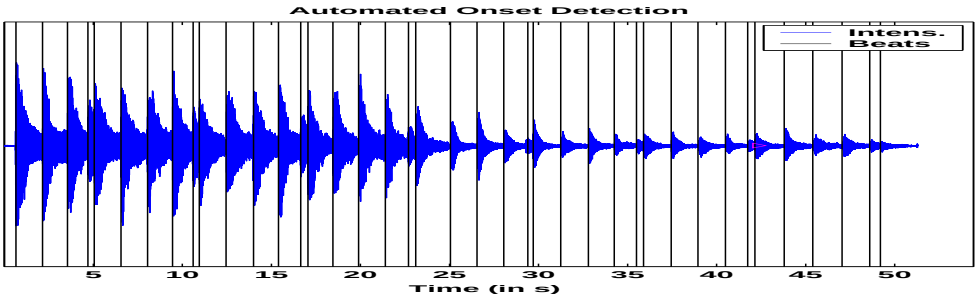
In general, the automated analysis of tone center transitions is highly consistent with the musicologist's assignment in Figure 3.58 on page 172. Some spontaneous misclassifications occur, i.e. before bar 1, A♭–major in bar 1, E♭–major in bar 2, f–minor and c–minor in bar 8. In bar 1 c–minor is captured, interrupted by a short f–minor misclassification. A♭–major is recognized correctly in bar 2. In bar 3 a D♭–major misclassification occurs in the beginning. Successfully the tonicization of f–minor and the tone center c–minor are recognized. In bar 4 in the beginning, C–major and c–minor are falsely identified, but later G–major is correctly assigned. In the beginning of bar 5 g–minor is assigned incorrectly. But later the right tone center c–minor is found. The chromatic descend in bar 6 misleads the algorithm to falsely identify G–major and g–minor. In bar 7 c–minor is correctly captured. In bar 8 the A♭–major tonicization is recognized, but not the tone center c–minor.

This is a good classification result, taking into account that we provide almost no explicit musical knowledge to the system except the concept of pitch classes, as a representation of the piece. The system receives musical knowledge only implicitly from the pieces of music providing the reference vectors. The choice of the analysis window does not depend on onset, rhythmical, and metrical structure.

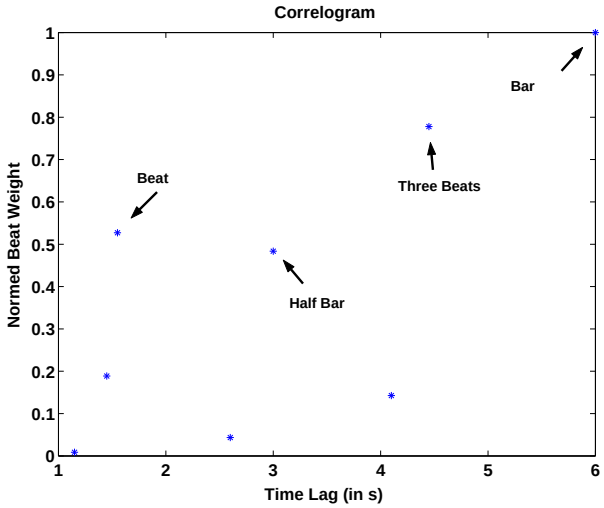
3.9.4 Multiple Time Scale Analysis

Instead of determining a tone center within a rigid moving average window it is desirable to calculate harmony and tone center of musically meaningful segments, such as chords, bars, or formal sections of a piece. For this purpose, we first calculate onsets (Section 3.9.1 on page 168). Then through beat weights (Section 3.9.1 on page 170) we generate a musically motivated segmentation on multiple time scales. The CQ-profiles averaged across these segments are matched with the key prototypes of the reference set in order to assign a tone center to them. Finally the tone center progressions on various time scales are displayed. For this purpose, we first determine onsets and beat weights according to Equations 3.11 – 3.14 using the parameter settings in Table 3.9.1 on page 169. The results are displayed in Figure 3.60 (a). Then we use autocorrelation in Equations 2.21 on page 79 and 3.15. The autocorrelation is calculated in a maximal time window of 6.5 seconds, for window $\mathbf{w}$ a linear function is employed, and cycle durations up to the order of 20 are considered. Time window step sizes from fine-grained to coarse temporal resolution are calculated in Figure 3.60 (b). The strongest beat weights are on time lags associated with bar, $\frac{3}{4}$ bar, half bar, and beat.

High beat strength in the correlogram in Figure 3.60 on the facing page gives rise to a temporal segmentation based on the referring time lag. Starting from onsets, recursively, temporal segmentation layers from fine to coarse resolution are constructed. Let us select a small time lag with high associated beat strength in the correlogram. The first onset is chosen to be the first segmentation mark. The next segmentation mark is given by the onset that lies in approximately a time lag distance to the first one. In the same way the third segmentation mark is constructed



(a)



(b)

FIGURE 3.60: Onset extraction (a) and correlogram (b) of Chopin's op. 28 No. 20. The onset extraction (a) works entirely correct. Time lags of high beat weights (b) can be related to musically meaningful cycle durations, such as beat, bar, and fractions thereof. Cf. Section 3.9.4 on the facing page for details.

from the second one, and so forth. From this segmentation a segmentation with coarser resolution is constructed. We yield segmentations on three levels of temporal resolution, for onsets, beats, and bars. Each segment defines a time window. Weighted by a cosine, CQ-profiles within this window are averaged. The average CQ-profiles of these segments are matched with the key prototypes from the reference set in the same way as in Section 3.9.3 on page 173, thereby assigning tone centers. For Chopin's PRÉLUDES op. 28 No. 20, automatically detected tone center transitions on various time scales are shown in Figure 3.61, using Scriabin's key-color map (cf. Section 3.7.1 on page 152). The top row indicates the tone center analysis on the bar level. It coincides with the analysis of a musicologist, except in bar 6 (cf. Section 3.9.2, Figure 3.58 on page 172). In this bar, C–major is falsely indicated due to the chromaticism in the descent. Segmentations in the lowest (onsets) and the

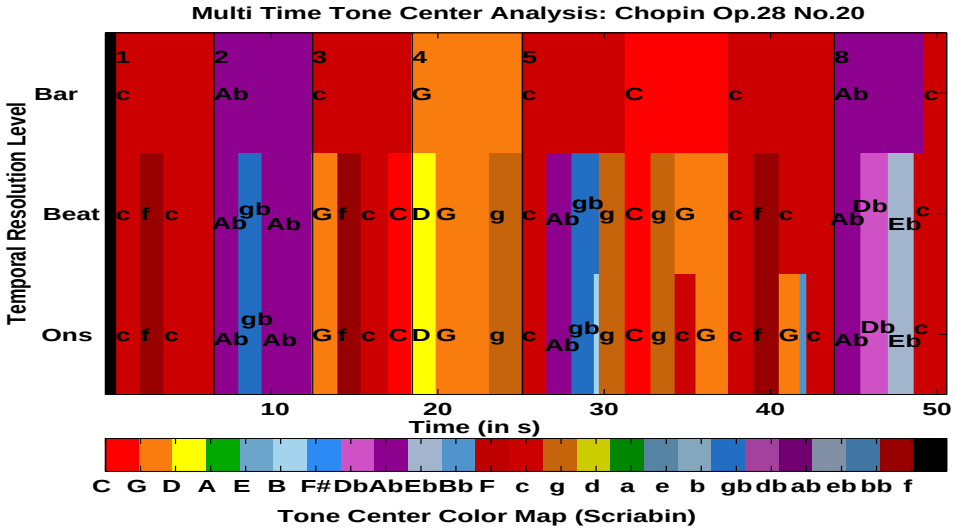


FIGURE 3.61: Analysis of tone center progression on multiple time scales, on the onset, beat, and bar level (from bottom to top). Vertical lines and numbers on top indicate bars. On the bar level, the analysis is always correct, except in bar 6. On the beat and onset level, often the correct chords are identified. (Cf. Section 3.9.4 and the score in Figure 3.58 on page 172.)

second lowest row (beats) in Figure 3.61 are almost the same, since the rhythm is the main beat, except on the 3rd beat a quarter note is replaced by a dotted eighth and a sixteenth note. In further refinement of this approach, both on onset and beat time level it would be more appropriate to use chord prototypes instead of tone center prototypes. However, the reference set generated from the averaged PRÉLUDES gives a reasonable approximation. Most chords, on the beat and onset level in bars 1, 2, 4, and 7, are assigned to the tone center suggested by the musicologist. The harmony of the chord is identified correctly in other instances, in bar 5 beat 2 & 4,bar 7 beat 2, and bar 8 beat 2. Sometimes suspension (bar 8 beat 3) misleads the harmonic analysis. It also occurs that a harmony closely related to the chord is indicated, such as the relative minor in bar 1 beat 2, the tonic of the 7th chord interpreted as a dominant in bar 3 beat 2 and bar 6 beat 2, and the parallel minor in bar 4 beat 4.

3.9.5 Ambiguity and Poly Tone Center Analysis

Instead of just considering the best matching tone center, we can gain additional information from the second, third, fourth, and so forth ... most prominent tone center.[24] In late romanticism and in polytonality several tonal references exist simultaneously. How clear or rather ambiguous is the best matching tone center? We will

[24] Leman [1995] in his tone center attractor dynamics, based on the concept of elastic snails, speaks of "attractors."

suggest a measure for tonal ambiguity.

For both poly tone center and ambiguity analysis we consider CQ-profiles averaged across bar segments such as in the top row in Figure 3.61 on the preceding page. *Poly tone center analysis* displays the n tone centers with the lowest dissimilarity values with respect to the sample at hand. Figure 3.62 displays the four most prominent tone centers in the first eight bars of Chopin's op. 28 No. 20.

Ambiguity a is calculated by dividing the dissimilarity value (d_1) of the closest tone center by the dissimilarity (d_2) of the second closest tone center relative to the analyzed sample

$$a = \frac{d_1}{d_2}. \tag{3.16}$$

In Figure 3.63 on the next page, high ambiguity is found in bars 3, 5, and 6. In bar 3 the chords g^7, c^7 induce tonicizations of c–minor and f–minor. From there the tone center is switched to C–major. In bars 5 and 6, chromaticism makes tonality unclear.

Sapp [2001] introduces the idea of ambiguity and poly tone center analysis, but without explicitly stating Equation 3.16. His analysis is based on MIDI rather than audio, not using content-driven segmentation.

3.9.6 Tone Center Path on the Torus

How can we spatially visualize the tonal content of a piece, including tone center progressions, ambiguity, and tonicization? We need to line out the topography of tone centers. The strength of a tone center in the piece appears as the height of a mountain elevating from the tone center's location in the tonal topography. A tonal modulation is viewed as a path across the mountains. A toroidal model of inter-key

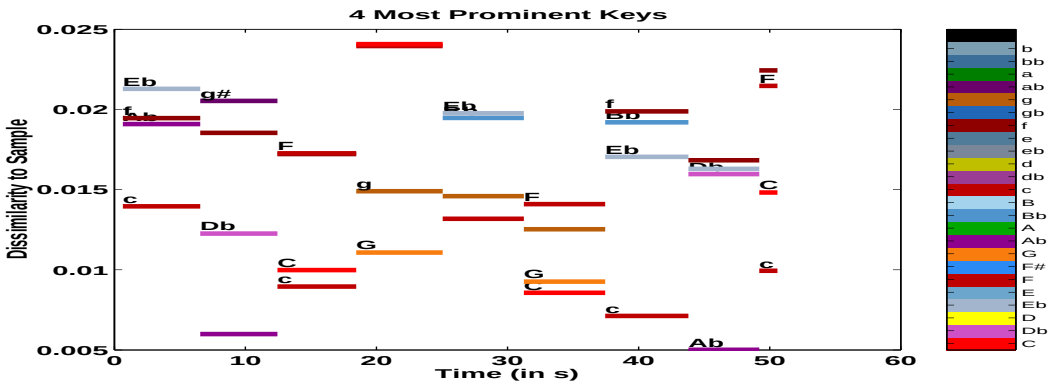


FIGURE 3.62: Poly tone center analysis for average CQ-profiles for each bar of Chopin's op. 28 No. 20. The lower the dissimilarity, the better the bar matches with the particular tone center. The lowest tone centers are identical with the ones in the top row of Figure 3.61 on the facing page.

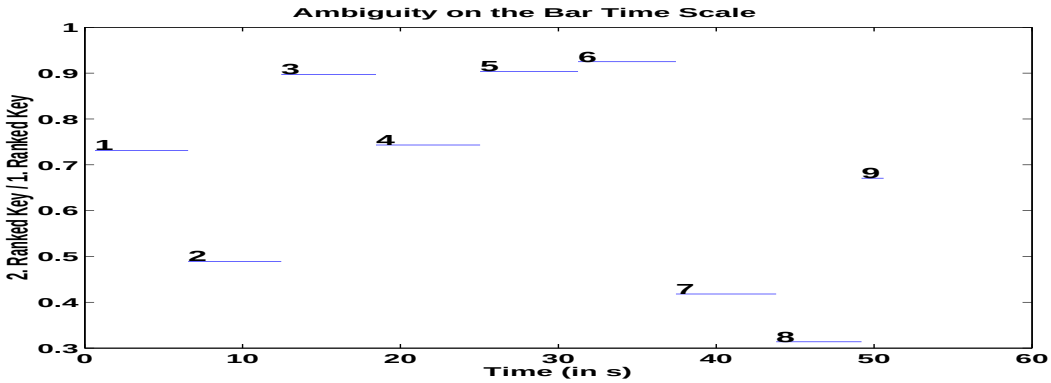


FIGURE 3.63: Tonal Ambiguity. Each level corresponds to a bar in Chopin's op. 28 No. 20. The bars are indicated by numbers. High ambiguity prevails in bar 3, and in the beginning of the chromatic descending passage in bar 5 and 6.

relations provides a tonal topography, e.g. the SOM trained with Chopin reference vectors (Figure 3.51 on page 161). We consider the four most prominent tone centers (Figure 3.62 on the preceding page) in each bar in order to incorporate tonal ambiguities. To build the mountains representing the tonal content, we accumulate the prominences, the inverse dissimilarity in Figure 3.62, of each tone center. The overall prominence of a tone center is encoded by a relief, contour lines, and a color scale from grey (low) to white (high). [25]

How can we read the musical landscape of PRÉLUDE No. 20 in Figure 3.64 on the facing page? The top of the mountain is at c–minor, the key of the piece. The crest at A♭-major reflects the two modulations towards this tone center. There are minor elevations for f–minor and C–major, reflecting short tonicizations in bar 3 and the C–major misclassification in bar 6. Low elevations appear at E♭–major, g–minor, G–major, F–major, D♭–major, and B♭–major.

3.9.7 Error Function

How can we measure the quality of the assignment of tone center transitions? We define a dissimilarity between the suggestion by a musicologist and the best matching tone center. These dissimilarities are then integrated over time.

For time t, be $k_{\text{theo}}[t]$ the tone center assigned by a musicologist. $k_{\text{emp}}[t]$ be the best matching tone center provided by automated analysis. Accordance of the musicologist's suggestion and the best matching tone center is indicated by

$$\mathbf{1}_{k_{\text{theo}}[t]}(k_{\text{emp}}[t]) = \begin{cases} 1 & \text{if } k_{\text{theo}}[t] = k_{\text{emp}}[t] \\ 0 & \text{otherwise} \end{cases}. \tag{3.17}$$

[25] Note that in Figure 3.51 on page 161 colors redundantly encode the position of a tone center on the TOMIR. But in Figure 3.64 the colors encode the *strength* of the tone center at that position.

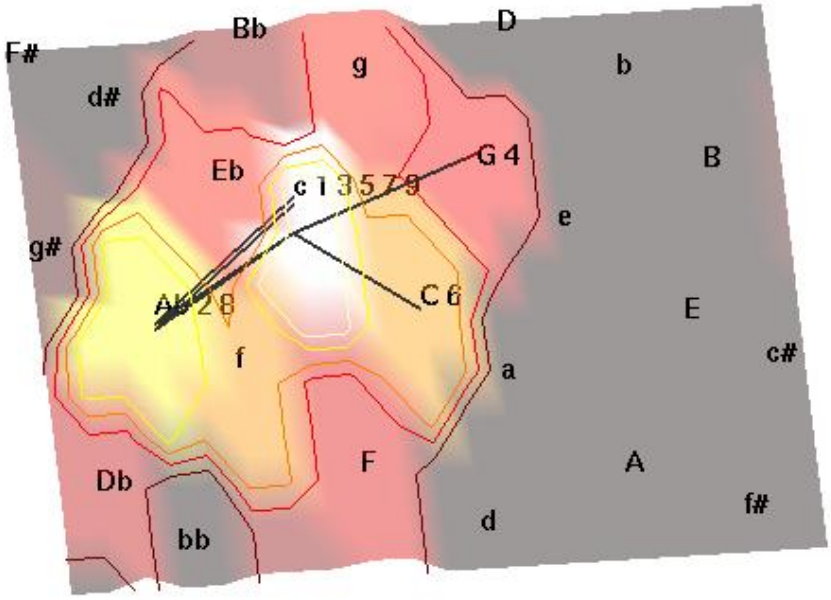


FIGURE 3.64: The path of tone center progression on the tonality "mountains" of Chopin's PRÉLUDE No. 20. The topography is provided by the SOM trained by average Chopin profiles (Figure 3.51 on page 161).The color scale brown-red-orange-white encodes low to high strength of tone centers. The numbers indicate bars. For each bar the strength of the four most prominent tone centers is accumulated. (Cf. text.)

A simple error function at time t_0 integrating over a back-looking time window Δt is given by

$$E = \sum_{t=t_0-\Delta t+1}^{t_0} (1 - \mathbf{1}_{k_{\text{theo}}[t]}(k_{\text{emp}}[t])). \tag{3.18}$$

A more refined misclassification analysis can be derived from integrating the distance (Equation C.7) on the torus, e.g. on Figure 3.51, between the musicologist's assignment and the best matching tone center. However, the slight refinement of error analysis by the latter concept can be compensated for by the simplicity of Equation 3.17.

3.9.8 Moving Average Analysis of Simple Modulations

The potential of moving average tone center analysis (Section 3.9.3 on page 173) is further explored. Textbook examples [Schneider, 1987] of a simple diatonic and a

chromatic modulation are played on a piano by Young-Woo Lee and recorded by Joo-Kyong Kim in the recording studio of Berlin University of the Arts.

The score in Figure 3.65 (Schneider [1987], p. 168) displays a diatonic modulation from C–major to G–major. Within the first two bars, a stereotype cadence $I - IV - V - I$ in C–major occurs. In bar 2, the second chord (C–major) is employed as a pivot chord. This chord acts as the tonic in preceding C–major and as the subdominant in subsequent G–major. From the third chord until the end we observe a cadence $IV - V^{4-3} - I$. Moving average tone center analysis (Figure 3.66 on the facing page) yields the following. In the beginning, tone center C–major is correctly captured. In the end, the final tone center G–major is recognized correctly after a small delay. For short moments, misclassifications occur in the beginning (F–major) and towards the end (D–major). For the second chord tonicization for F–major occurs.

A musicologist's detailed analysis of the chromatic modulation from C–major to A–major in Figure 3.68 on page 182 (Schneider [1987], p. 172) reveals the following. From the first chord (C–major) the second chord is derived by altering the c to leading note $c\sharp$ and adding the fundamental a. Thereby, this chord is turned into a dominant seventh chord on a, leading to the subdominant D–major (first chord in bar 2) of A–major. A $IV - V^7 - I$ cadence confirms the tone center A–major in bar 2 and 3. Moving average analysis in Figure 3.68 on page 182 grasps the D–major tonicization on the first chord in bar 2. Little errors occur in the beginning with short false F–major/f–minor and in the end with false D–major classification.

In general, the modulations are captured with some short spontaneous misclassifications and a delay. This could be compensated for by correcting size and weighting of the analysis window.

3.9.9 Discussion

In the tracking of tone center transitions, temporal aspects of changing CQ-profiles come into play. For simple temporal processing, we employ moving average. We have experimented also with the temporal Kohonen map [Chappell and Taylor, 1993]. But the algorithm is not practically useful for reasonably quantized input

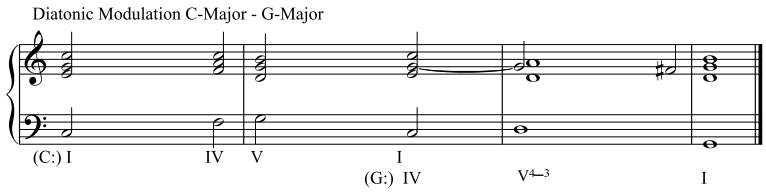


FIGURE 3.65: Textbook example of a diatonic modulation from C–major to G–major (from Schneider [1987], p. 168 bottom). In the second bar, the second (pivot) chord can be either heard as the tonic in preceding C–major or as the subdominant in subsequent G–major.

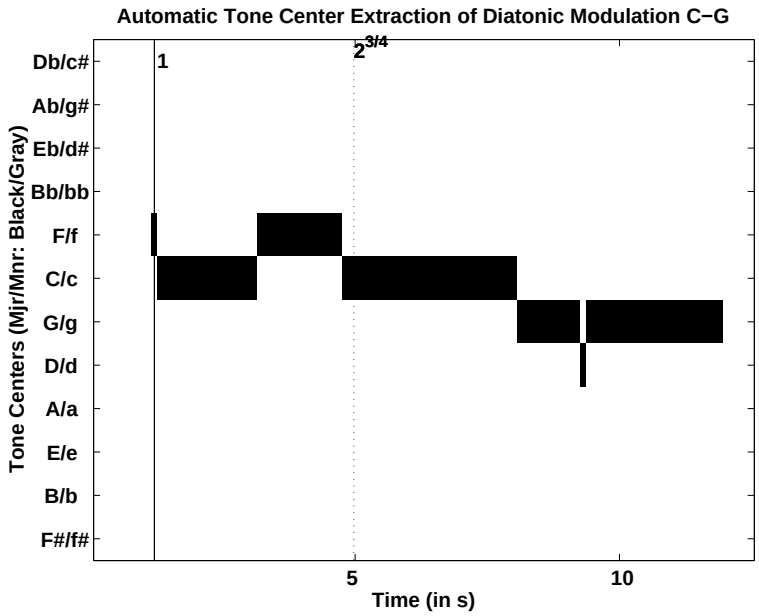


FIGURE 3.66: Moving average tone center analysis of the diatonic modulation in Figure 3.66 from C–major to G–major. All found tone centers are major. The dotted vertical line indicates the onset of the pivot C–major chord. Overall, the tone centers are captured correctly, but delayed and interrupted by short spontaneous misclassifications and a tonicization of the second chord.

data. Harmony is partly organized syntactically (Section 1.2.4 on page 53). If we can further reduce successive CQ-profiles to a short sequence of a limited number of states, hidden Markov models can be useful. [Sheh and Ellis, 2003] Multiple cues, e.g. tonal rhetoric (p. 61), characteristic dissonances, and other idiomatic patterns, can be extracted to further improve tone center transition analysis. Tonal rhetoric can be taken into account by weighting the CQ-profiles of the entire piece with an inverted cosine window emphasizing the beginning and the end of the piece and

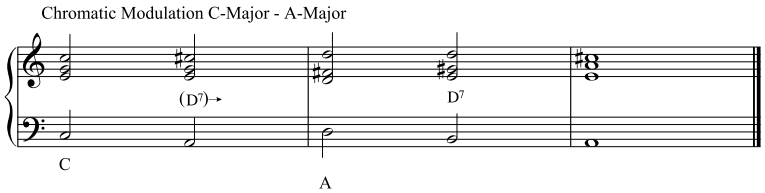


FIGURE 3.67: Chromatic modulation from C–major to A–major (from Schneider [1987], p. 172 top). By high alteration $c(\sharp)$ of the keynote and addition of the fundamental a, the second chord functions as a dominant seventh chord to the subsequent D–major chord. The new tone center A–major is confirmed by a IV $-$ V^7 $-$ I cadence.

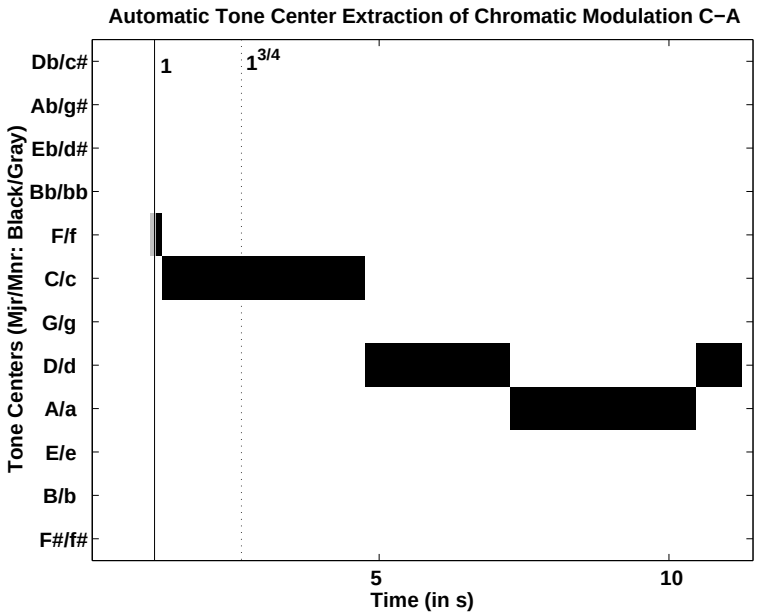


FIGURE 3.68: Moving average tone center analysis of the chromatic modulation in Figure 3.67. The dotted vertical line indicates the onset of the pivot chord, the dominant seventh chord on *a*. The analysis works well, except a tonicization on the third (D–major) chord and short misclassifications in the beginning and in the end.

deemphasizing the middle section.

Another problem is the availability of reliable tone center labels for supervised learning, for several reasons: (1) Historically, western harmony becomes more ambiguous towards the end of the 19th century. (2) Sometimes there is no "correct" label. The tonal interpretation of a musical excerpt is debated among musicologists applying different concepts of harmony. (3) Labeled music in digital form is not publicly available in high quantities. Labeling of musical data, e.g. with tone centers, is expensive. A workaround is to generate labels automatically. E.g. with a functioning MIDI-to-audio alignment at hand, one could automatically extract label information from music in MIDI format, align MIDI to audio, and then use the labels for audio also.

Summary

We have investigated various aspects of circular perception of relative pitch, CQ-profiles, and inter-key relations. The work in this thesis contributes to the convergence of distinct disciplines towards a unifying cognitive musicology. The results can be summarized under two areas as follows:

1. Circular perception of relative pitch
    a) generalization to harmonic complex tones (more natural in a musical context)
    b) verification in a psychoacoustic experiment
    c) theory in the framework of gestalt principles and perceptual quantization
    d) formalization and implementation as a computational psychoacoustic model
    e) design of an acoustical stimulus that is circularly perceived as ascending and descending in pitch at the same time
    f) analysis of music with respect to circularly perceived pitch, tempo, and loudness
2. CQ-profiles and visualizations
    a) fast, robust real time method of calculating CQ-profiles
    b) significance of pitch class profile or particular pitch class intensities for the composer and mode discrimination in a testing and classification framework
    c) clustering and visualization of inter-mode and inter-composer relations
    d) sketch of decomposition of pitch class profiles into components referring to composer, cycle, and piece
    e) identification of the circle of fifths in various cognitive modeling scenarios
    f) evolving toroidal models of inter-key relations (TOMIR) in various models of interaction between early auditory processing, schema, and music samples
    g) gathering of key preference statistics and the visualization of composer inter-relations based upon them

h) algorithm inducing a global topology from a stipulated local topology

i) first steps toward sound based assignment of tone center transitions

j) the latter also in conjunction with varying temporal resolution and beat weights.

We have provided a computational model simulating the perception of pitch differences and a framework for tone center perception established by previous exposure to tonal music.

Music analysis tools are built for style investigation and for tone center tracking. In traditional musicology, harmonic modulation tracking as well as stylistic differentiation require a thorough understanding of music theory. The methods presented here attack the problems in an elegant, nevertheless surprisingly effective, bottom-up approach.

The following links are shown between distinct aspects of cognitive musicology. A surprising correspondence between a particular digital filter bank and results in experimental work in music psychology is discovered: The CQ-profiles of cadential chord progressions played on the piano match very well with the spectrally weighted probe tone ratings.

Normalized constant Q profiles based on the audio signal are a significant indicator, revealing a lot about the stylistic characteristics of the music. The stylistic analysis by means of average CQ-profiles can be extended to other music, e.g. late romanticism, neo-baroque, and in a more limited fashion, to atonal, polytonal, and micro-tonal music. The method opens the door to the analysis of non-written music traditions, like Jazz, and perhaps also non-Western music (particularly of the South and Southeast Asian varieties).

An operational model for the acquisition of inter-key relations is exposed to pieces of tonal music. The circle of fifths and a homogeneous toroidal configuration of keys evolve. The model's basic assumptions are reduced to (1) Logarithmic frequency resolution grounded in the Weber-Fechner law, reflected in various aspects of the auditory periphery, (2) Contextual formation of semantics of tone centers, by application of the proximity gestalt principle in algorithms like Isomap and correspondence analysis, serving as a metaphor for a schema, (3) Identification of octave components based on the projection of pitch onto the pitch class circle in the helix model, and (4) Consideration of semi-tones. The explanatory power of Leman [1995]'s TOMIR is increased, first, by substantially extending the corpus of data it operates on, and, second, by omission of strong presumptions: (1) The sensory data that forms the internal representation of inter-key relations consists of a broad corpus of actual performances of pieces of music, e.g. Chopin's Préludes, instead of a very limited number of textbook-like examples in the artificially sounding timbre of Shepard tones. Due to the timbral simplicity of Shepard tones, the latter present a quasi symbolic score-like representation. (2) The toroidal topology is not presumed by using a toroidal SOM as an implementation of the schema. Instead, the circle of fifths and a homogeneous toroidal configuration evolve in correspondence analysis

and Isomap, neither of them assuming a toroidal topology a priori. In contrast to Leman [1995], the model used here does not rely on features which are of small relevance to inter-key perception and whose relation to biology is not clear, e.g. autocorrelation and the global comparison necessary in the SOM algorithm.

CQ-profiles in conjunction with SOMs, CA, or fuzzy distance are suited to automatically track tone center modulations. Such an approach yields a very cheap, elegant, and powerful method. It can easily be applied to different tunings. As a built-in analysis module it forms the backbone of automatic accompaniment of recorded audio sources in real time.

First and foremost, we have concentrated on the analysis of circular pitch and pitch class profiles. We emphasize, however, that for organization of large data bases the given methods of classification, clustering, and visualization are by no means restricted to pitch classes. The employed machine learning techniques for significance testing (t-test, sparse linear programming machine), classification (support vector machine, regularized discriminant analysis), clustering, and visualization (self-organizing feature map, SOM; topographic ordering map, correspondence analysis, CA; Isomap) are not at all limited to CQ-profile analysis. The methods are universal and provide a practical tool for discovering and analyzing correspondences among various musical parameters that are adequately represented by co-occurrences of certain musical events or objects: e.g. pitch classes, keys, instrumentation, rhythm, composers, and styles. In particular, it seems promising to extend our analysis to temporal transitions in the space of musical parameters.

An increasing problem in music technology is the management of huge databases of music, including elaborated searching methods. The labeling of music with various features facilitates navigation in the database. From recorded music, numerous distinct features are automatically extracted yielding information about pitch (class), tonality, and rhythm. As a result, engines searching for the appropriate matching sample for a music arrangement, a sound track, or other particular requirements are constituted or enhanced. These methods are embedded in the MPEG-7 framework. The request for such technology in music (industry) and research is apparent, as a workbench for musicology, a psychologically based composition environment, interactive music making (like DJ-tools, music minus one, and karaoke systems), compression methods, content retrieval, music recommendation systems, and hearing aids.

In opposition to a purely Platonic approach in music theory, our findings are supported by physics, psychoacoustics, statistics, and auditory neuroscience. The strongest evidence for our results is given by psychological experiments. Biologically inspired algorithms are employed to such an extent as they are available at the current (early) state of auditory neuroscience and relevant to the modeled problem, and in particular the modeling of some features of the auditory periphery and the design of schemata. The models are operational and act on the raw sound signal, instead of a high-level symbolic representation. The main algorithms are capable of working in real time and can react to the input signal. The core results are based on the analysis of a statistically significant number of music samples. The great number

of pieces that can be processed in a data base and the help of statistics and machine learning enables musicology hypothesis testing with defined levels of statistical significance instead of musical discussion based on isolated musical case studies.

Although here and there landscapes become apparent, wide areas on the map of the mysterious topography of music remain white, attracting adventurous explorers:

> Überhaupt bleibt hier dem sinnigen, von höherem Geiste beseelten Physiker und Mechaniker noch ein weites Feld offen, und ich glaube, daß bei dem Schwunge, den die Naturwissenschaft erhalten, auch tieferes Forschen in das heilige Geheimnis der Natur eindringen, und manches, was nur noch geahnet, in das rege Leben sichtlich und vernehmbar bringen wird. [26] (E. T. A. Hoffmann [1821]: Die Automate)

[26] Generally a wide domain of research remains open to the witted physicist and mechanic who is animated by a higher spirit. Due to the drive that science receives, in addition, I think that more profound research will penetrate into the holy mystery of nature and some that merely has been divined will visibly and audibly be brought into brisk being.

A Theory and Solution Ideas of the Binding Problem

The binding problem is the general formulation of the task to group partials to a holistic tone perception (cf. Section 1.1.7 and its relevance for Shepard tones, Section 1.2.3). Two hypotheses regarding the neural implementation of the binding problem are given. Independent component analysis (ICA) and possible applications to the cocktail party problem are briefly introduced.

The first hypothetical solution to the binding problem, *hierarchical organization*, works via integration by anatomic convergence. This model assumes that at an early stage, basic object features such as frequency components are detected. Through progressive convergence of the connections, cells emerge with more specific response properties on a higher processing level. For example, they respond to tones, chords, harmonies, and keys. This corresponds to hierarchical artificial intelligence approaches (cf. context-free grammars and Section 1.2.4 on page 51). Even though hierarchical structures in the brain are joined by lateral connections, in practice a strictly hierarchical concept is successful, e.g. within a framework of a knowledge database and a Bayesian network [Kashino et al., 1998].

Another way of trying to explain feature binding is through *neural synchronization*. The temporal binding model assumes that assemblies of synchronously firing neurons represent objects in the cortex. For example, such an assembly would represent a particular speaker. These assemblies comprise neurons, which detect specific frequencies or amplitude modulation frequencies. The relationship between the partials can then be encoded by the temporal correlation among these neurons. The model assumes that neurons, which are part of the same assembly, fire in synchrony whereas no consistent temporal relation is found between cells belonging to representations of different speakers. Evidence for feature binding by neural synchronization in the visual cortex is given by Engel et al. [1997].

Terman and Wang [1995]; Wang [1996]; Brown and Cooke [1998] supply an implementation based on time log-frequency representation of music. Their model consists of a set of oscillators in the time frequency domain or the correlogram (Section 2.2 on page 71). Oscillators which are close to each other are coupled strongly (Hebbian rule and principle of proximity). An additional global inhibitor stops oscillators belonging to different streams being active at the same time. This approach can be used for vowel segregation and also for segregation of different voices according to the proximity principle (Cooke and Brown [1999], Figure 1.6). A more promising approach is based on the "integrate and fire" model [Maass, 1997]. An ensemble of such models displays synchronous spike patterns.

A.1 Implementations of Acoustical Source Separation

A mixture of several sound sources (speech, music, other sounds) is recorded by one or several microphones. The aim is the decomposition into the original sources. Since an important application is the development of hearing aids, the goal is demixing with at most two microphones.

There are some successful applications of source separation for artificial mixtures (sources are recorded separately and are then digitally mixed by weighted addition). On the other hand, mixtures in real environments are more difficult to demix. The different approaches can be roughly divided into two categories:

1. Mimicking the auditory system and
2. Employment of techniques of digital signal processing without reference to biology.

Okuno et al. [1999] aims to combine (1) and (2) synergetically. A possible treatment similar to (1) is as follows: An auditory model is used for preprocessing. From the output of the auditory model (cochleagrams and correlograms) harmonic substructures are extracted. By the use of gestalt principles spectral units are built from this. From these separated units sound can be resynthesized [Nakatani et al., 1995]. In another approach the spectral units are determined as a sequence of correlograms, and the auditory transformation is inversely calculated [Slaney, 1994, 1998].

The larger portion of methods according to (2) deals with the realization of the *independent component analysis* (ICA, Comon [1994]; Cardoso [1998]; Müller et al. [1999]). Sanger [1989] indicates a reference to biological systems. But his approach is largely a purely statistical model. So we need at least as many microphones as sound sources. It is a problem that the model does not account for real reverberation. So decomposition of mixtures in a real acoustic environment works only under very special conditions. First approaches are Lee et al. [1998]; Casey and Westner [2000]; Parra and Spence [2000]; Murata et al. [2000]. In addition, it is still a tough problem to separate one speaker from a cocktail party environment with several sound sources using only two microphones, as all applicable algorithms up to now require as many microphones as sound sources.

B More, Explicit CQ-Profiles

First we will present some more CQ-profiles, of Chopin and Alkan, in addition to Bach's WTC (Figures 3.24 on page 127 and 3.25 on page 128). Then we will sketch out very coarsely an idea how to decompose a profile into portions that can be attributed to mode, the composer's style manifested in the cycle, and the individual character of the particular piece.

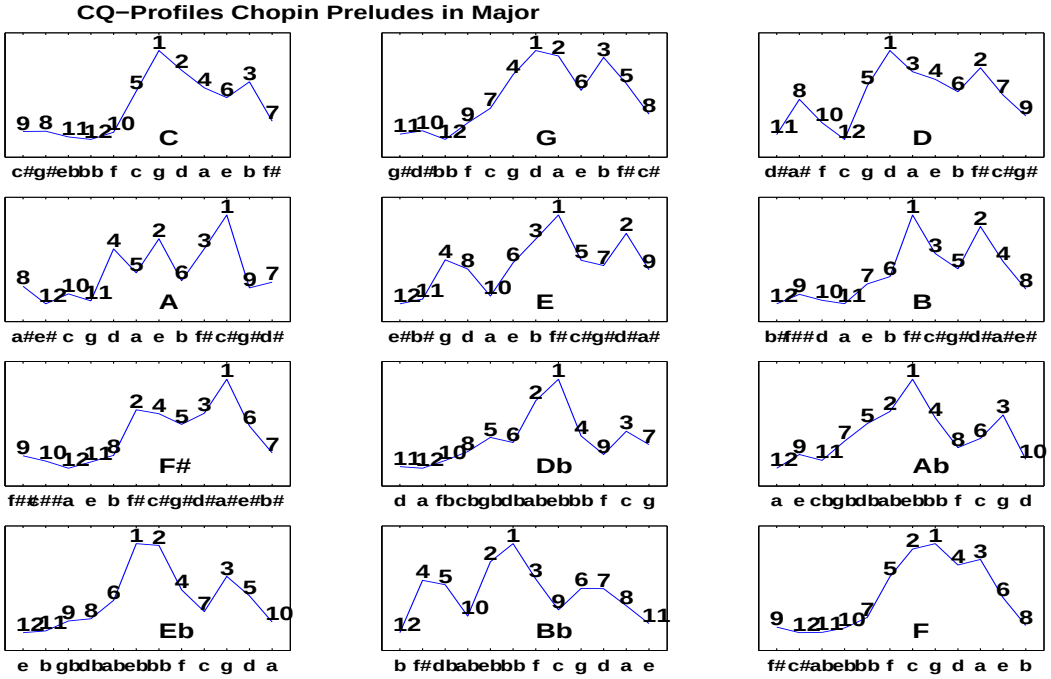


FIGURE B.1: CQ-profiles of Chopin's PRÉLUDES op. 28 (major) played by Alfred Cortot.

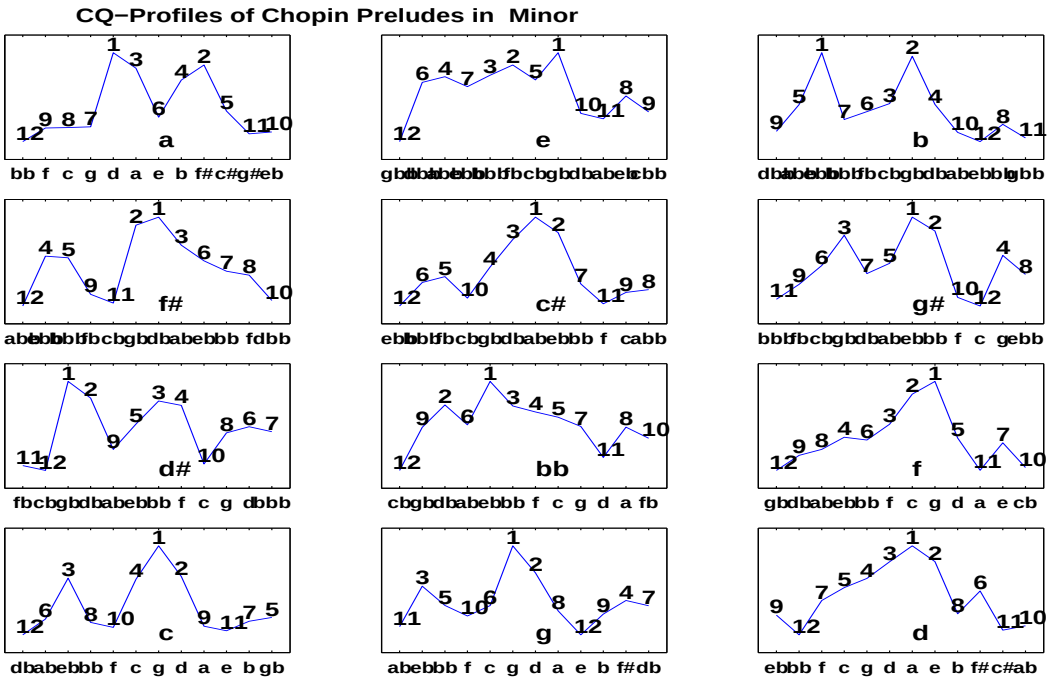


FIGURE B.2: CQ-profiles of Chopin's PRÉLUDES op. 28 (minor) played by Alfred Cortot.

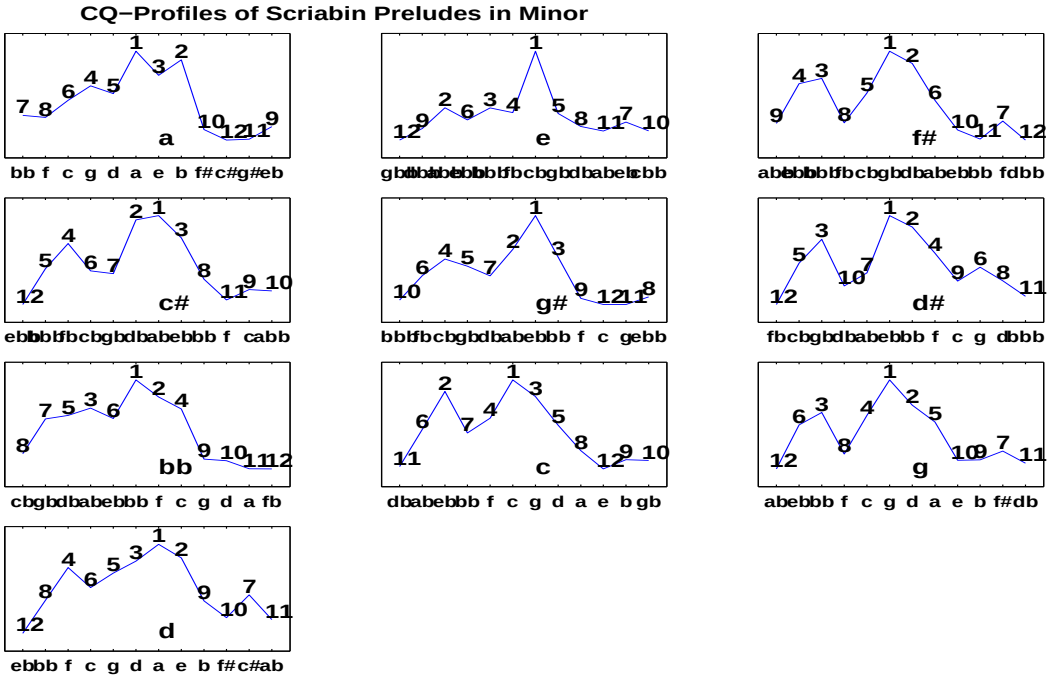


FIGURE B.3: Scriabin's PRELUDES op. 11 (minor), played by various pianists (cf. Section 3.3.4 on page 125).

B.1 Coarse Sketch of Profile Decomposition

We have studied what a CQ-profile can reveal about key and composer (Sections 3.3.1 - 3.5, and 3.8). However, a CQ-profile contains more information, e.g. about the stylistic characteristics of a particular piece. In order to single out the information that can be attributed to mode, the composer's style (as manifested in a musical cyle), and the individual piece we decompose an average NCQ-profile of a piece into following components:

profile = mode prototype + cycle residuum + individual residuum.

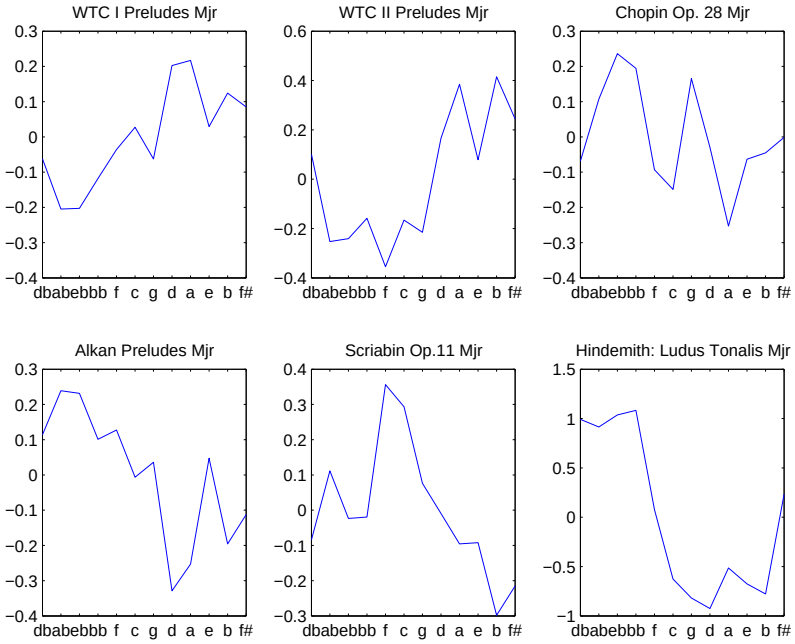


FIGURE B.4: The cycle residua in major, calculated from normalized CQ-profiles, transposed to tone center *c*. The various cycles show specific shapes.

The major mode prototype (not shown here) is calculated from Bach, Chopin, Alkan, and Scriabin. The mode prototype displays two peaks located at the dominant in the middle and the major third on the left. The shape of the major mode prototype can be approximately thought of as a sum of two Gaussians or triangles centered at the dominant ("g", *dominant blob*) and the major third ("e", *(major) third blob*). Refering to blobs instead of peaks is motivated by the observation that the average CQ-profiles are relatively smooth due to the fifth relation, effective both musically and acoustically, because of the strength of the third partial. In the cycle residua for major, Figure B.4, one can observe the following. For Bach (cf. also CQ-profiles, Figure 3.24 on page 127) the left side is decreased. The valley between

Residua Histogram: Statistical Features in Major

	WTC I Preludes	WTC II Preludes	Chopin Op. 28	Alkan Preludes	Scriabin Op. 11	Hindemith: Ludus Tonalis
Std.	0.19	0.39	0.18	0.23	0.2	1.5
Skew.	0.086	-0.016	0.86	0.46	0.2	1.1
Kurt.	3.2	2.8	3	2.1	1.7	4.6
Range	0.42	0.77	0.49	0.57	0.65	2

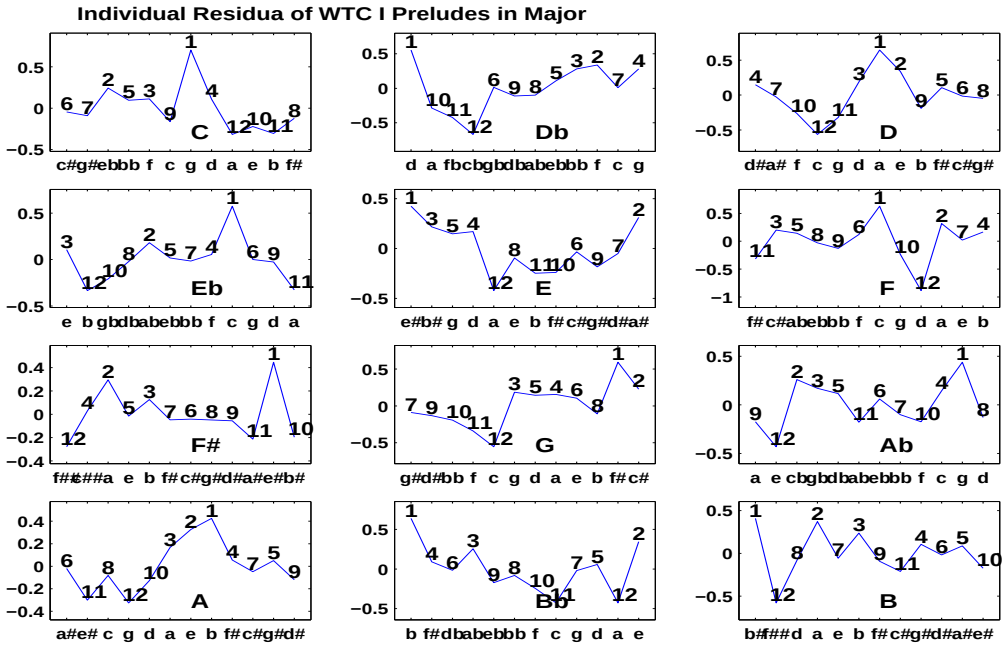
TABLE B.1: Skewness ("Skew.") and kurtosis ("Kurt.") are normed by the standard deviation ("Std."). Hindemith has by far the broadest standard deviation and the widest range. Apart from Hindemith, Chopin stands out w.r.t. skewness. WTC has very low skewness.

dominant blob and major third blob is almost eroded. For Chopin the opposite is the case (cf. also CQ-profiles, Figure B.1 on page 189). In major, the valley between the dominant and the right major third blob is deepened. In addition on the left side, the minor third blob (at "eb") occurs. This indicates an "androgynous" major-minor mode flavor. Therefore, major and minor do not separate so clearly in clustering (Figure 3.29 on page 133) and visualization (Figure 3.32 on page 136). Similarly, for Alkan the dominant and the major third blob are separated by a wide gap. Also an additional left minor third blob occurs. Scriabin has a high value at scale degree IV ("f", Figure B.3). Hindemith equalizes the distribution of pitch class intensities. But his fugues are not considered in the mode prototype calculation, since stylistically Hindemith is on the edge of major-minor tonality.

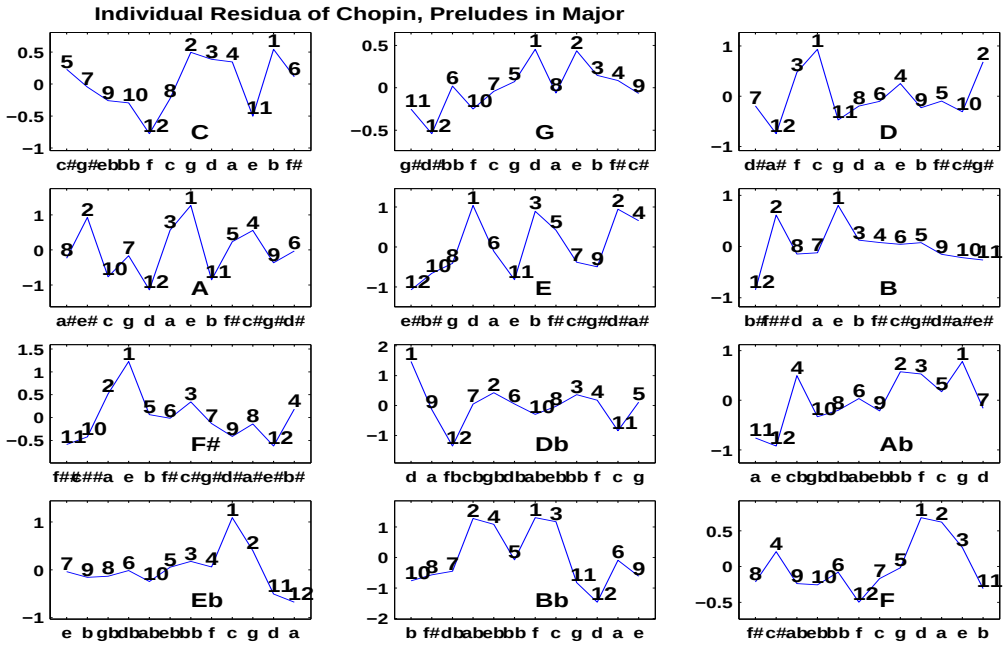
A look at the residua of the individual pieces (Figures B.5-B.6) indicates why in Section 3.4.1 classification according to the composer sometimes fails. At the same time, these individual residua reveal characteristics of the particular pieces. E.g. for the major preludes of Bach's WTC I, Figure B.5 (a), we observe three different groups of pieces, emphasizing the falling leading note, the rising leading note, or the dominant.

Can we describe the residua of the individual pieces statistically? Table B.1 indicates that range, standard derivation, and skewness are significant for distinguishing between various cycles.

Decomposition of CQ-profiles into mode prototype, cycle residuum, and individual residuum helps to understand classification, clustering, and visualization results in Section 3.4. We have indicated how to use CQ-profile decomposition for refined quantitative music analysis, discovering, for instance, the characteristics of a piece. However, the method and its musical interpretation are by no means thoroughly investigated yet.

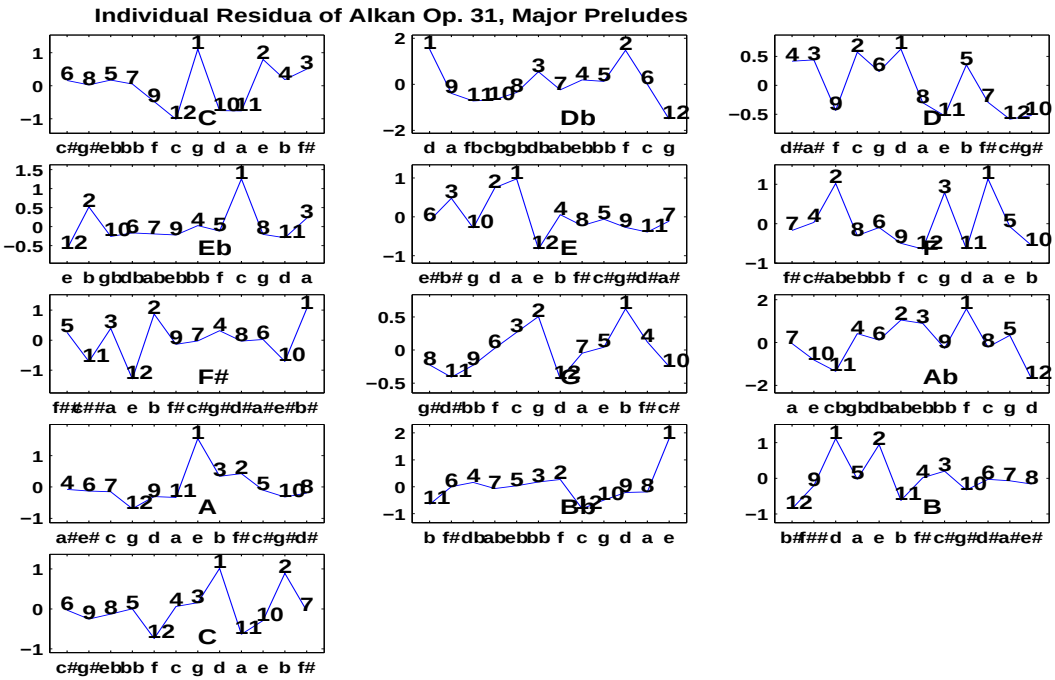


(a)

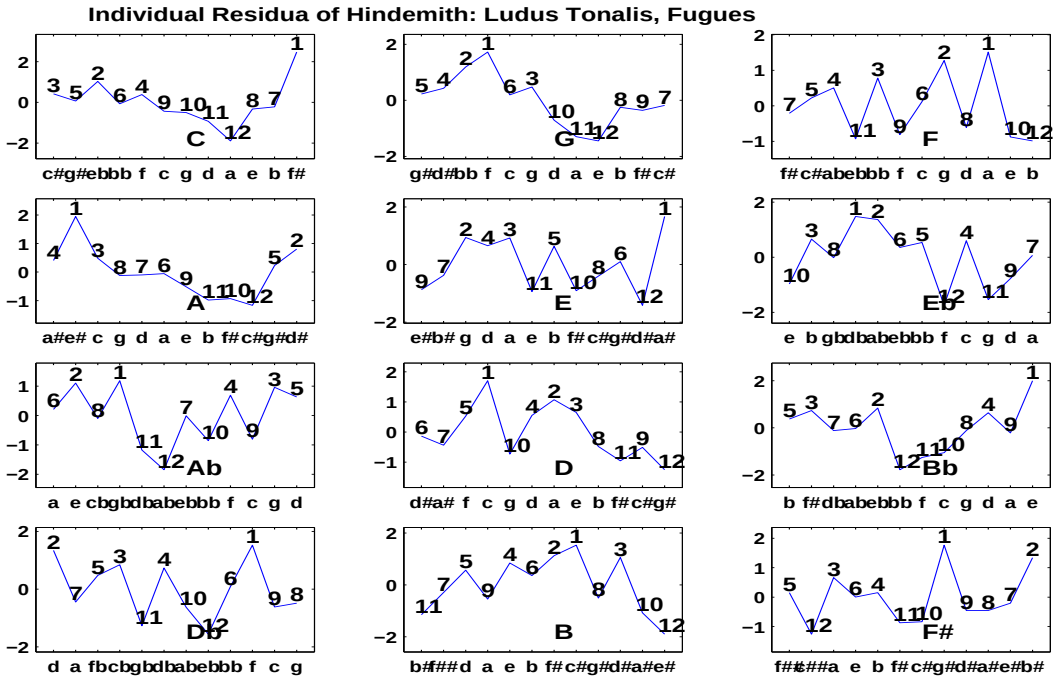


(b)

FIGURE B.5: Individual residua revealing characteristics of the particular pieces. (a) In the major preludes of WTC I (Gould), three patterns can be observed: the emphasis of the falling leading note(D♭–major, E–major, B♭–major, B–major), the rising leading note (F♯–major, G–major, A♭–major), or the dominant (C–major, D–major, F–major). Cf. Figure 3.24 on page 127 for the original CQ-profiles. (b) Preludes of Chopin's op. 28 performed by Alfred Cortot. Cf. Figure B.1 on page 189 for the original CQ-profiles.



(a)



(b)

FIGURE B.6: Individual residua for the major preludes of Alkan (a) and the Hindemith fugues (b).The range for Hindemith is much wider than for the other composers.

C Technical Details of Dissimilarities, Classification, and Visualization

After explaining the difference between categorical, ordinal, and interval data, we will introduce the notions of dissimilarity and metric, and give examples of the latter. Then we will provide technical details of the support vector machine, correspondence analysis, Isomap, and Chew's geometric model of tones, triads, and keys.

C.1 Data and Metrices

Data may be accessed as categorical, ordinal, or interval variables. Let a *categorical* variable "musician" have the value "violin player", "singer", or "piano player". But there is no straightforward way to order the values from highest to lowest, in contrast to *ordinal* variables. An example for the latter is "musical education" with a value "no musical education", "some musical training", or "professional musical training". Such variables can be compared by the "$\geq$" relation. If the order is equally spaced the variable is an *interval* variable, e.g. "income" measured in dollars. The order of the variables can be quantified, e.g. by specifying "10 000 $ more". It is sometimes useful to "convert" between different variable types, e.g. to consider only the order of interval data given in a dissimilarity matrix, or to induce interval relations on categorical data by comparing associated feature vectors. For metrical data we require additional conditions, such as the triangle inequality (Equation C.3).

Keys are in itself categorical data. How can we relate two keys to each other? To two keys, denoted by integers $i, j \in \mathcal{K}$, we assign a number by a *dissimilarity* $d_{\mathcal{K}}(i, .) \colon \mathcal{K} \to \mathbb{R}$. Thereby we endow the keys with interval relations. How can we derive a dissimilarity? One way is to explicitly stipulate it, e.g. $d_{\mathcal{K}}(1, j) = 0$, if keys $1, j \in \mathcal{K}$ are "closely related" and $d_{\mathcal{K}}(1, j) = 1$ otherwise. We apply such a stipulated neighborhood relation in Section 3.7.2. Another way is to characterize keys $i, j \in \mathcal{K}$ by particular profiles $\mathbf{x}, \mathbf{y}$ in an n-dimensional feature space $\mathcal{R}^n$. The axes of the feature space may represent psychological variables as well as the output of a filter bank or a music theoretic description. We can then derive the dissimilarity by a function on $\mathcal{R}^n$, $d(\mathbf{x}, \mathbf{y}) \colon \mathcal{R}^n \times \mathcal{R}^n \to \mathbb{R}$.

A *metric* d on $\mathcal{R}^n = \mathbb{R}^n$ is a dissimilarity so that for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

$$\mathbf{x} = \mathbf{y} \Leftrightarrow d(\mathbf{x}, \mathbf{y}) = 0, \tag{C.1}$$

$$d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x}) \quad \text{(symmetry)}, \tag{C.2}$$

$$d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z}) \quad \text{(triangle inequality)}. \tag{C.3}$$

Typically we define a metric by a positive-semidefinite matrix $\mathbf{A} \in \text{Mat}(n \times n, \mathbb{R})$ by

$$d(\mathbf{x}, \mathbf{y}) = \langle \mathbf{a}, \mathbf{b} \rangle_{\mathbf{A}} := \mathbf{a}^t \mathbf{A}^{-1} \mathbf{b} \tag{C.4}$$

where $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ are derived from columns $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{y} = (y_1, \ldots, y_n)$ of a data matrix $\mathbf{X} \in \text{Mat}(n \times l, \mathbb{R})$ (i.e. l n-dimensional feature vectors), for instance, in one of the following ways:

1. $\mathbf{a} = \mathbf{b} = (\sqrt{|x_i - y_i|})_{1 \leq i \leq n}$, $\mathbf{A}$: unity matrix, (absolute distance)
   for d restricted on $\mathcal{R}^n = \{0, 1\}^n$ being the Hamming distance,
2. $\mathbf{a} = \mathbf{b} = \mathbf{x} - \mathbf{y}$, $\mathbf{A}$ unity matrix, (squared Euclidean distance)
3. for $\mathbf{F} \in \text{Mat}(n \times n, \mathbb{R})$ being the diagonal matrix of the marginal distribution of the n features in $\mathbf{X}$ across all l samples (cf. Section 2.3.2):

$$\mathbf{a} = \mathbf{b} = \frac{\mathbf{x}}{\|\mathbf{x}\|} - \frac{\mathbf{y}}{\|\mathbf{y}\|}, \ \mathbf{A} = \mathbf{F}^{-1}, \qquad (\chi^2\text{−distance}) \tag{C.5}$$

4. $\mathbf{a} = \mathbf{b} = \mathbf{x} - \mathbf{y}$, $\mathbf{A} = \text{COV}(\mathbf{X})$ covariance matrix of $\mathbf{X}$; $\mathbf{x}, \mathbf{y}$ with 0 mean,
   (Mahalanobis distance)
5. $\mathbf{a} = \mathbf{b} = (\sqrt{|x_i - y_i|})_{1 \leq i \leq n}$, $\mathbf{A} = \text{diag}(\mathbf{d})$ diagonal matrix with diagonal components

$$d_{i,i} = \frac{1}{n}\left(1 - \frac{\sigma_i}{|x_i - y_i| + \sigma_i} e^{-\frac{|x_i - y_i|^2}{2\sigma_i^2}}\right) \qquad \text{(fuzzy distance)} \tag{C.6}$$

   with σ being the variance vector in the n features of $\mathbf{X}$ across samples (cf. Section 3.3.2).

The χ^2-distance has the advantage that it does not sensitively depend on whether samples are represented by many features or whether the features are gathered to a smaller number of feature groups, since the unification of several dimensions in the feature space is compensated by addition of the respective components in the marginal distribution on the diagonal of $\mathbf{F}$. Considering components with high variance less significant, Mahalanobis as well as fuzzy distance weight them less. Avoiding the calculation of the full covariance matrix, fuzzy distance leads to good results in tracking of tone centers transitions (cf. Section 3.9).

We can define a distance measure operating on a manifold, e.g. on a torus $\mathcal{T} = [0, 2\pi r_1[\times [0, 2\pi r_2[$. For $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}$ the *toroidal distance* yields

$$d_{\mathcal{T}}(\mathbf{x}, \mathbf{y}) := \sqrt{\sum_{i=1}^{2} \min(|x_i - y_i|, r_i - |x_i - y_i|)^2}. \tag{C.7}$$

The *rank order* for columns $\mathbf{x}, \mathbf{x}'$ of $\mathbf{X}$ with respect to a dissimilarity d is defined as

$$\mathrm{rk}_d(\mathbf{x}, \mathbf{x}') = \#\{\mathbf{y} \text{ column of } \mathbf{X} : \mathbf{y} \neq \mathbf{x}, d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{x}')\} \tag{C.8}$$

$\mathrm{rk}_d(\mathbf{x}, \mathbf{x}') = 1$ means that $\mathbf{x}'$ is the vector closest to $\mathbf{x}$ among the columns of $\mathbf{X}$. $\mathrm{rk}_d(\mathbf{x}, \mathbf{x}') = 2$ indicates that $\mathbf{x}'$ is the second closest, and so forth. In general, rk_d is not symmetric. A rank order $\mathrm{rk}_d(x, .)$ with fixed x reduces an interval variable to an ordinal one.

C.2 Support Vector Machine

After having outlined the basic idea of the *support vector machine* (SVM, Vapnik [1998]) in Section 2.3.1 we will now give a more technical description. The SVM provides a means for learning labeled data, e.g. training vectors $\mathbf{x}_i \in \mathbb{R}^n, i = 1, \ldots, l$ with a label vector $\mathbf{y} \in \{1, -1\}^l$. The algorithm learns the classification by minimizing Equation C.9 under the constraints C.10 and C.11:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2}\mathbf{w}'\mathbf{w} + c \sum_{i=1}^{l} \xi_i, \tag{C.9}$$

$$y_i(\mathbf{w}'\phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \tag{C.10}$$

$$\xi_i \geq 0 \quad (i = 1, \ldots, l) \tag{C.11}$$

with vector $\mathbf{e}$ containing only ones, penalty constant c, some unknown function ϕ, and parameters $\mathbf{w}, b, \xi_1, \ldots, \xi_l$. We cannot solve this problem directly, but we can minimize the *dual formulation*, Equation C.12 under constraints C.13 and C.14, if there exists a kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)'\phi(\mathbf{x}_j)$ and vector α:

$$\min_{\alpha} \frac{1}{2}\alpha'\mathbf{Q}\alpha - \mathbf{e}'\alpha, \tag{C.12}$$

$$0 \leq \alpha_i \leq c \quad (i = 1, \ldots, l), \tag{C.13}$$

$$\mathbf{y}'\alpha = 0 \tag{C.14}$$

with positive semidefinite $\mathbf{Q} = (y_i y_j K(\mathbf{x}_i, \mathbf{x}_j))_{\substack{1 \leq i \leq l \\ 1 \leq j \leq l}}$. For more details cf. Müller et al. [2001].

C.3 Visualization

In some high dimensional feature space $\mathcal{R}^n$, it may be desirable to visualize the columns $\mathbf{x}, \mathbf{x}'$ of a data matrix $\mathbf{X} \in \text{Mat}(n \times l, \mathcal{R})$ in one or two dimensions, preserving the rank order rk_d (Equation C.8). Then at a glance, one could grasp the rank orders of items represented by the feature vectors x, x'. The question of scaling down to a lower dimension and of visualization can be formulated as follows. We look for a mapping $\pi\colon \mathcal{R}^n \to \mathcal{M}$, with given dissimilarities d_n on $\mathcal{R}^n$ and $d_{\mathcal{M}}$ on the low dimensional manifold $\mathcal{M}$, so that the rank order $\text{rk}_{d_n}(\mathbf{x}, \mathbf{x}')$ is maintained by $\text{rk}_{d_{\mathcal{M}}}(\pi(\mathbf{x}), \pi(\mathbf{x}'))$ in $\mathcal{M}$. A good overview on embedding and clustering algorithms is given by Ripley [1996].

C.3.1 Correspondence Analysis

We now provide some more technical details of singular value decomposition and correspondence analysis (Greenacre [1984]; Kockelkorn [2000], cf. Sections 2.3.2 and 2.3.4). The following theorem is crucial for analyzing the co-occurrence matrix in correspondence analysis:

Theorem 1 (Generalized Singular Value Decomposition) *Let* $\mathbf{A}$ *be a positive definite symmetric* $m \times m$ *matrix and* $\mathbf{B}$ *a positive definite symmetric* $n \times n$ *matrix. For any real-valued* $m \times n$ *matrix* $\mathbf{F}$ *of rank* d *there exist an* $m \times d$ *matrix* $\mathbf{U} = (\mathbf{u_1}, \ldots, \mathbf{u}_d)$*, a* $n \times d$ *matrix* $\mathbf{V} = (\mathbf{v_1}, \ldots, \mathbf{v}_d)$ *with* $\mathbf{U}'\mathbf{A}\mathbf{U} = \mathbf{V}'\mathbf{B}\mathbf{V} = \mathbf{I}_d$*, and a diagonal* $d \times d$ *matrix* $\mathbf{\Delta} = (\delta_{ij})$ *so that*

$$\mathbf{F} = \mathbf{U}\mathbf{\Delta}\mathbf{V}' = \sum_{k=1}^{d} \delta_{kk}\mathbf{u}_k\mathbf{v}_k'. \tag{C.15}$$

Cf. Greenacre [1984] for a proof. For $\mathbf{A} = \mathbf{I}_m, \mathbf{B} = \mathbf{I}_n$, Theorem 1 yields the ordinary singular value decomposition. If furthermore $\mathbf{F}$ is symmetric, we get the familiar eigenvalue decomposition.

The columns $\mathbf{u}_k$ of $\mathbf{U}$ can be viewed as the column factors with singular values δ_{kk}. Vice versa the rows $\mathbf{v}_k$ of $\mathbf{V}$ are the row factors with the same singular values δ_{kk}. The magnitude of $\mathbf{F}$ in each of the d dimensions in the co-ordinate system spanned by the factors $\mathbf{u}_k$ is then given by δ_{kk}.

For the matrix of relative frequencies $\mathbf{F}^{\mathcal{P},\mathcal{K}} = (f_{ij}^{\mathcal{P},\mathcal{K}})$ and positive definite diagonal matrices $(\mathbf{F}^{\mathcal{P},\mathcal{P}})^{-1}$ and $(\mathbf{F}^{\mathcal{K},\mathcal{K}})^{-1}$ with the inverted relative frequencies of row and column features, respectively, on their diagonal Theorem 1 yields:

$$\mathbf{F}^{\mathcal{P},\mathcal{K}} = \mathbf{U}\mathbf{\Delta}\mathbf{V}', \tag{C.16}$$

with

$$\mathbf{U}'(\mathbf{F}^{\mathcal{P},\mathcal{P}})^{-1}\mathbf{U} = \mathbf{V}'(\mathbf{F}^{\mathcal{K},\mathcal{K}})^{-1}\mathbf{V} = \mathbf{I}_d. \tag{C.17}$$

Defining

$$\mathbf{S} = (s_{ij}) := \mathbf{\Delta V}' \left(\mathbf{F}^{\mathcal{K},\mathcal{K}}\right)^{-1} \tag{C.18}$$

we get

$$\mathbf{F}^{\mathcal{P}|\mathcal{K}} = \mathbf{F}^{\mathcal{P},\mathcal{K}} \left(\mathbf{F}^{\mathcal{K},\mathcal{K}}\right)^{-1} = \mathbf{U\Delta V}' \left(\mathbf{F}^{\mathcal{K},\mathcal{K}}\right)^{-1} = \mathbf{US}. \tag{C.19}$$

Taking the i-th column on both sides of Equation C.19 we get

$$\mathbf{f}^{\mathcal{P}|\mathcal{K}=i} = \sum_{k=1}^{d} \mathbf{u}_k s_{ki}. \tag{C.20}$$

The profile $\mathbf{f}^{\mathcal{P}|\mathcal{K}=i}$ is described in terms of co-ordinates s_{ki} on the axes $\mathbf{u}_k$. s_{ki} is the projection – in the χ^2-metric – of profile $\mathbf{f}^{\mathcal{P}|\mathcal{K}=i}$ onto the axis $\mathbf{u}_k$.
Vice versa we have

$$\mathbf{F}^{\mathcal{K}|\mathcal{P}} = \underbrace{\mathbf{V\Delta U}' \left(\mathbf{F}^{\mathcal{P},\mathcal{P}}\right)^{-1}}_{=:\mathbf{Z}=(z_{ij})} = \mathbf{VZ}. \tag{C.21}$$

The profile $\mathbf{f}^{\mathcal{K}|\mathcal{P}=j}$ is described in terms of co-ordinates z_{kj} on the axes $\mathbf{v}_k$.

Each key $i \in \mathcal{K}$ is given by its pitch class profile $\mathbf{f}^{\mathcal{P}|\mathcal{K}=i}$. In Figure 3.37 on page 142 key i is represented by its first two coordinates (s_{1i}, s_{2i}).

Each pitch class $j \in \mathcal{P}$ is given by its key profile $\mathbf{f}^{\mathcal{K}|\mathcal{P}=j}$. In Figures 3.37 on page 142 and 3.40 on page 145, pitch class j is represented by its first two coordinates (z_{1j}, z_{2j}).

C.3.2 Isomap

For an intuitive description of the Isomap [Tenenbaum et al., 2000] cf. Section 2.3.4 on page 91. Be $\mathbf{X}$ a $n \times l$-matrix, with l vectors of dimension n. For $\mathbf{D} = (d_{ij})_{1 \le i,j \le l}$ be $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$ the Euclidean distance between columns $\mathbf{x}_i, \mathbf{x}_j$ of $\mathbf{X}$. Generate a dissimilarity matrix $\mathbf{T}$ in the following way: Initialize $\mathbf{T} = (t_{ij})_{1 \le i,j \le l}$ by

$$t_{ij} := \begin{cases} d_{ij} & \text{for } \mathbf{x}_i \in \mathcal{G}_{\mathbf{x}_j}(k) \\ \infty & \text{otherwise} \end{cases} \tag{C.22}$$

with $\mathcal{G}_{\mathbf{x}_j}(k)$ being the set of k nearest neighbors to $\mathbf{x}_j$ (Equation 2.23 on page 85). Enforcing the triangle inequality, iterate

$$t_{ij} = \min_{k \in \{1,\ldots,l\} \setminus \{i,j\}} (t_{ij}, t_{ik} + t_{kj}). \tag{C.23}$$

Finally perform multidimensional scaling $\tau(\mathbf{T})$ on $\mathbf{T}$ and project $\mathbf{X}$ onto the most prominent eigenvectors of $\tau(\mathbf{T})$.

C.3.3 Chew's Model with Choice of Parameters

We use a simplified instance of Chew's more general model (Section 1.2.10, Chew [2000]). It proposes a spatial arrangement such that tones, triads and keys are represented as vectors in three-dimensional space. Tones $j \in \mathbb{N}$ proceed in steps of fifths, that is $j \sim c$, $j+1 \sim g$, and so forth. The tones' spatial location is denoted by $\mathbf{t}(j) \in \mathbb{R}^3$, the location of major and minor triads by $\mathbf{c}_{\mathrm{M}}(j), \mathbf{c}_{\mathrm{m}}(j) \in \mathbb{R}^3$, and of major and minor keys by $\mathbf{k}_{\mathrm{M}}(j), \mathbf{k}_{\mathrm{m}}(j) \in \mathbb{R}^3$. The tones are arranged in a helix turning by $\pi/2$ and rising by a factor of h per fifth:

$$\mathbf{t}(j) = \left(\sin\left(\frac{j\pi}{2}\right), \cos\left(\frac{j\pi}{2}\right), jh\right). \tag{C.24}$$

Spiraling up (Figure 1.22 on page 65), every four fifths $\mathbf{t}(j)$ returns to the same X-Y co-ordinates. Both major and minor triads are represented as the weighted mean of their constituent tones:

$$\mathbf{c}_{\mathrm{M}}(j) = m_1\mathbf{t}(j) + m_2\mathbf{t}(j+1) + m_3\mathbf{t}(j+4), \tag{C.25}$$
$$\mathbf{c}_{\mathrm{m}}(j) = m_1\mathbf{t}(j) + m_2\mathbf{t}(j+1) + m_3\mathbf{t}(j-3). \tag{C.26}$$

The keys are represented as weighted combinations of tonic, dominant, and subdominant triads, with the minor keys additionally incorporating some of the parallel major triads:

$$\mathbf{k}_{\mathrm{M}}(j) = m_1\mathbf{c}_{\mathrm{M}}(j) + m_2\mathbf{c}_{\mathrm{M}}(j+1) + m_3\mathbf{c}_{\mathrm{M}}(j-1) \tag{C.27}$$
$$\mathbf{k}_{\mathrm{m}}(j) = m_1\mathbf{c}_{\mathrm{m}}(j) + m_2\left(\frac{3}{4}\mathbf{c}_{\mathrm{M}}(j+1) + \frac{1}{4}\mathbf{c}_{\mathrm{m}}(j+1)\right) \tag{C.28}$$
$$+ m_3\left(\frac{3}{4}\mathbf{c}_{\mathrm{m}}(j-1) + \frac{1}{4}\mathbf{c}_{\mathrm{M}}(j-1)\right). \tag{C.29}$$

We choose the parameters so as to obtain a good fit with the data from Bach's WTC I (fugues), resulting in the following parameter settings:

$$\mathbf{m} = (m_1, m_2, m_3) = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4}) \quad \text{and} \quad h = \frac{\pi}{6}. \tag{C.30}$$

The general model of Chew [2000] allows the weights to be independent from each other for $\mathbf{c}_{\mathrm{M}}, \mathbf{c}_{\mathrm{m}}, \mathbf{k}_{\mathrm{M}}, \mathbf{k}_{\mathrm{m}}$. In this application, weights $\mathbf{m}$ are set equal, across major and minor chords and keys (cf. Chew [2001]). In our instance of the model, we have only $\mathbf{m}$ and h as free parameters. We use values different from Chew [2001].

Bibliography

al Farabi, A. N. (c. 950). *Kitab al-Musiqa al-Kabir*. Reprint Cairo, 1967.

Aristoxenus (BC c. -330). *Harmonika stoicheia*. Athens?

Arom, S. (2003). A cognitive approach to the study of musical scales in polyphonies of Central Africa. Lecture Notes. *European Soc. for the Cognitive Sciences of Music (ESCOM)*. Hannover.

Auhagen, W. (1983). *Studien zur Tonartencharakteristik in theoretischen Schriften und Kompositionen vom späten 17. bis zum Beginn des 20. Jahrhunderts*, volume 36(6) of *Europäische Hochschulschriften*. Peter Lang, Frankfurt/M.

Auhagen, W. (1994). *Experimentelle Untersuchungen zur auditiven Tonalitätsbestimmung in Melodien*, volume 180 of *Kölner Beiträge zur Musikforschung*. Gustav Bosse, Kassel.

Aures, W. (1985). Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale. *Acustica*, 59:130–141.

Austin, J. L. (1962). *How to Do Things with Words*. Harvard University Press, Cambridge, MA.

Barlow, K. (1995). Seminar. Lecture Notes. Darmstädter Arbeitstagung Institut für Neue Musik und Musikerziehung.

Behrendt, J. E. (1983). *Nada Brahma. Die Welt ist Klang*. Rowohlt, Reinbek.

Bennett, K. P. and Mangasarian, O. L. (1992). Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software*, 1:23–34.

Benzécri, J.-P. (1977). Histoire et préhistoire de l'analyse des données. *Cahiers de l'Analyse des Données*, 2:9–40.

Beran, J. (2003). *Statistics in Musicology*. Chapman & Hall/CRC.

Bharucha, J. and Krumhansl, C. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, 13:63–102.

Bharucha, J. J. and Todd, P. M. (1989). Modeling the perception of tonal structure with neural nets. *Computer Music J.*, 13(4).

Blankertz, B., Purwins, H., and Obermayer, K. (1999a). Constant Q profiles and toroidal models of inter-key relations. CD-ROM. CCRMA Affiliates meeting, Stanford.

Blankertz, B., Purwins, H., and Obermayer, K. (1999b). Toroidal models of inter-key relations in tonal music. In *6th Int. Conf. on Systematic and Comparative Musicology*, Oslo. Unpublished Manuscript.

Bod, R. (2002). A unified model of structural organization in language and music. *J. of Artificial Intelligence Research*, 17:289–308.

Bregman, A. S. (1990). *Auditory Scene Analysis*. MIT Press, Cambridge, MA.

Bregman, A. S. and Dannenbring, G. (1973). The effect of continuity on auditory stream segregation. *Perception & Psychophysics*, 13:308–312.

Britannica (2003a). Deconstruction. In *Encyclopaedia Britannica*.

Britannica (2003b). Ockham's razor. In *Encyclopaedia Britannica*.

Britannica (Accessed 2004). *Britannica Concise Encyclopedia*. http://concise.britannica.com.

Brown, G. J. and Cooke, M. (1998). Temporal synchronization in a neural oscillator model of primitive auditory stream segregation. In Rosenthal, D. F. and Okuno, H. G., editors, *Computational Auditory Scene Analysis*, pages 87–103. Lawrence Erlbaum Associates, Mahwah, NJ.

Brown, J. (1991). Calculation of a constant Q spectral transform. *J. of the Acoustical Soc. of America*, 89(1):425–434.

Brown, J. C. and Puckette, M. S. (1992). An efficient algorithm for the calculation of a constant Q transform. *J. of the Acoustical Soc. of America*, 92(5):2698–2701.

Brown, J. C. and Puckette, M. S. (1993). A high resolution fundamental frequency determination based on phase changes of the Fourier transform. *J. of the Acoustical Soc. of America*.

Browne, R. (1981). Tonal implications in the diatonic set. *In Theory Only*, 5:3–21.

Buteau, C. (1997). *Motivic Topologies and Their Meaning in the Motivic Analysis of Music*. PhD thesis, Université Laval Québéc.

Butler, D. (1989). Describing the perception of tonality in music: A critique of the tonal hierarchy theory and a proposal for a theory of intervallic rivalry. *Music Perception*, 6:219–242.

Camurri, A., Coletta, P., Ricchetti, M., and Volpe, G. (2000). Synthesis of expressive movement. In *Proc. of the Int. Computer Music Conf.*, pages 270–273. ICMA.

Cardoso, J.-F. (1998). Blind signal separation: statistical principles. In *Proc. of the IEEE, special issue on blind identification and estimation*.

Casey, M. A. and Westner, A. (2000). Separation of mixed audio sources by independent subspace analysis. In *Proc. of the Int. Computer Music Conf.*, pages 154–161. ICMA.

Castellano, M. A., Bharucha, J. J., and Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *J. of Experimental Psychology*, 113(3):394–412.

CCARH (2003). Muse data. Center for Computer Assisted Research in the Humanities. http://www.musedata.org.

Chappell, G. J. and Taylor, J. G. (1993). The temporal Kohonen map. *Neural Networks*, pages 441–445.

Chew, E. (2000). *Towards a Mathematical Model for Tonality*. PhD thesis, MIT Sloan School of Management.

Chew, E. (2001). Modeling tonality: Application to music cognition. In *Proc. of the 23rd Annual Meeting of the Cognitive Science Society*.

Chomsky, N. (1988). Language and problems of knowledge. In Martinich, A. P., editor, *The Philosophy of Language*. Oxford University Press.

Chopin (1957). 24 préludes op. 28. volume 5049. Editions Salabert, Paris, New York.

Chowning, J. M. (1980). Computer synthesis of the singing voice. In *Sound Generation in Winds, Strings, Computers*, volume 29. Royal Swedish Academy of Music, Stockholm.

Cohen, M. A., Grossberg, S., and Wyse, L. L. (1995). A spectral network model of pitch perception. *J. of the Acoustical Soc. of America*, 98(2):862–879.

Cohn, R. (1997). Neo-Riemannian operations, parsimonious trichords, and their tonnetz representations. *J. of Music Theory*, 41(1):1–66.

Comon, P. (1994). Independent component analysis, a new concept ? *Signal Processing*, 36:287– 314.

Cooke, M. P. and Brown, G. J. (1999). Interactive explorations in speech and hearing. *J. of the Acoustical Soc. of Japan*.

Dahlhaus, C. (1967). *Untersuchungen über die Entstehung der harmonischen Tonalität*, volume 2 of *Saarbrücker Studien zur Musikwissenschaft*. Bärenreiter-Verlag, Kassel.

Dai, H. (2000). On the relative influence of individual harmonics on pitch judgment. *J. of the Acoustical Soc. of America*, 2:953–959.

Dallos, P. (1992). The active cochlea. *J. Neuroscience*, 12:4575–4585.

de la Motte, D. (1980). *Harmonielehre*. Bärenreiter, Basel, 3rd edition.

Deutsch, D. (1986). A musical paradox. *Music Perception*, 3:275–280.

Diels, H. (1903). In Kranz, W., editor, *Die Fragmente der Vorsokratiker*, volume 1. Weidmann, Dublin, 12th (1966) edition.

Douthett, J. and Steinbach, P. (1998). Parsimonious graphs: A study in parsimony, contextual transformations, and modes of limited transposition. *J. of Music Theory*, 42(2):241–263.

Drabkin, W. (Accessed 21 Oct. 2004). Urlinie. In Macy, L., editor, *Grove Music Online*. http://www.grovemusic.com.

Drobisch, M. W. (1855). Über musikalische Tonbestimmung und Temperatur. *Abhandlungen der Mathematisch-Physischen Klasse der Sächsischen Akademie der Wissenschaften*, 2.

Duifhuis, H. (1982). Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception. *J. of the Acoustical Soc. of America*, 89:2837–2842.

Ehrenfels, C. v. (1890). Über Gestaltqualitäten. *Vierteljahrsschrift für wissenschaftliche Philosophie*, 14:249–292.

Eisenberg, G., Batke, J.-M., and Sikora, T. (2004). Beatbank – an MPEG-7 compliant query by tapping system. In *Audio Engineering Soc. 116th Convention*, Berlin.

Elliott, L. (2003). Environmentalism. In *Encyclopaedia Britannica*.

Engel, A., Roelfsema, P. R., Fries, P., Brecht, M., and Singer, W. (1997). Role of the temporal domain for response selection and perceptual binding. *Cerebral Cortex*, 7:571–582.

Epperson, G. (2003). Music. In *Encyclopaedia Britannica*.

Euler, L. (1739). Tentamen novae theoriae musicae. In Rudio, F., Krazer, A., Speiser, A., and du Pasquier, L. G., editors, *Opera Omnia*, volume 3(1). Teubner, Stuttgart. 1926.

Euler, L. (1774). De harmoniae veris principiis per speculum musicum repraesentatis. In Rudio, F., Krazer, A., Speiser, A., and du Pasquier, L. G., editors, *Opera Omnia*, volume 3(1), pages 568–586. Teubner, Stuttgart.

Farey, P. (Accessed 2004). Peter Farey's Marlowe page. http://www2.prestel.co.uk/rey/.

Felder, D. (1977). An interview with Karlheinz Stockhausen. *Perspectives of New Music*, 16(1):85–101.

Fielder, L. D., Bosi, M., Davidson, G., Davis, M., Todd, C., and Vernon, S. (1995). AC-2 and AC-3: Low-complexity transform-based audio coding. In *Collected Papers on Digital Audio Bit-Rate Reduction*.

Fisher, R. A. (1940). The precision of discriminant functions. *Annals of Eugenics*, 10:422–429.

Fleischer, A. (2002). *Die analytische Interpretation: Schritte zur Erschließung eines Forschungsfeldes am Beispiel der Metrik*. PhD thesis, Humboldt-Universität, Berlin. dissertation.de.

Fletcher, H. and Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *J. of the Acoustical Soc. of America*, 5:82–108.

Floros, C. (1996). *György Ligeti – Jenseits von Avantgarde und Postmoderne*. Lafite, Wien.

Forte, A. (1973). *The Structure of Atonal Music*. Yale University Press, New Haven.

Friedman, J. H. (1989). Regularized discriminant analysis. *J. of the American Statistical Association*, 84(405).

Fucks, W. and Lauter, J. (1965). *Exaktwissenschaftliche Musikanalyse*. Westdeutscher Verlag, Köln.

Fujishima, T. (1999). Realtime chord recognition of musical sound: a system using Common Lisp Music. In *Proc. of the Int. Computer Music Conf.*, pages 464–467. ICMA.

Gaiser, K. (1965). Platons Farbenlehre. In Flashar, H. and Gaiser, K., editors, *Synusia. Festausgabe für Wolfgang Schadewaldt zum 15. März 1965*, pages 173–222. Pfullingen.

Gang, D. and Berger, J. (1999). A unified neurosymbolic model of the mutual influence of memory, context and prediction of time ordered sequential events during the audition of tonal music. In *Hybrid Systems and AI: Modeling, Analysis and Control of Discrete + Continuous Systems*. American Association for Artificial Intelligence (AAAI) Technical Report SS-99-05.

Gann, K. (1995). *The Music of Conlon Nancarrow*. Cambridge University Press.

Gauss, C. F. (1866). Nachlass: Theoria interpolationis methodo nova tractata. In *Werke*, volume 3, pages 265–327. Königliche Gesellschaft der Wissenschaften, Göttingen.

Glasberg, B. R. and Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47:103–138.

Goldstein, J. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *J. of the Acoustical Soc. of America*, 54:1496–1516.

Gouyon, F. and Herrera, P. (2003). Determination of the meter of musical audio signals: Seeking recurrences in beat segment descriptors. In *Audio Engineering Soc. 116th Convention*.

Graepel, T., Burger, M., and Obermayer, K. (1997). Phase transitions in stochastic self-organizing maps. *Physical Review E*, 56(4):3876–3890.

Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis*. Academic Press, London.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbre. *J. of the Acoustical Soc. of America*.

Griffith, N. (1994). Development of tonal centers and abstract pitch as categorizations of pitch use. In *Connection Science*, pages 155–176. MIT Press, Cambridge.

Grönewald, J. (2003). 128 musikalische Temperaturen im mikrotonalen Vergleich. http://www.groenewald-berlin.de/.

Guttman, L. (1941). The quantification of a class of attributes: A theory and method of scale construction. In Horst, P., editor, *The Prediction of Personal Adjustment*. Social Science Research Council, New York.

Harris, F. J. (1978). On the use of windows for harmonic analysis with discrete Fourier transform. In *Proc. IEEE*, volume 66, pages 51–83.

Harrison, D. (1994). *Harmonic Function in Chromatic Music*. The University of Chicago Press.

Hartmann, W. M. (1997). Signals, sound, and sensation.

Haykin, S. (1999). *Neural Networks*. Prentice-Hall, Upper Saddle River, NJ, 2nd edition.

Heideman, M. T., Johnson, D. H., and Burrus, C. S. (1984). Gauss and the history of the fast Fourier transform. *IEEE ASSP Magazine*, 1(4):14–21.

Heinichen, J. D. (1728). *Der General-Baß in der Composition*. Dresden. Reprint 1969, Georg Olms Verlag, Hildesheim.

Helmholtz, H. v. (1863). *Die Lehre von den Tonempfindungen als Physiologische Grundlage für die Theorie der Musik*. Vieweg, Braunschweig.

Herbart, J. F. (1850). *Psychologie als Wissenschaft, neu gegründet auf Erfahrung, Metaphysik und Mathematik*, volume 2. E. J. Bonset, Amsterdam. Reprint 1968.

Hirschfeld, H. O. (1935). A connection between correlation and contingency. *Cambridge Philosophical Soc. Proc. (Math. Proc.)*, 31:520–524.

Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.*, 117:500–544. London.

Hoffmann, E. T. A. (1814). *Fantasiestücke in Callot's Manier*, volume 1, chapter Kreisleriana.

Hoffmann, E. T. A. (1821). *Die Serapions-Brüder*, chapter Die Automate. Reimer, Berlin.

Hofstadter, D. R. (1979). *Gödel, Escher, Bach: an Eternal Golden Band*. Basic Books, New York.

Holtzmann, S. R. (1977). A program for key determination. *Interface*, 6:29–56.

Horst, P. (1935). Measuring complex attitudes. *J. Social Psychol.*, 6:369–374.

Houtsma, A. J. M. and Goldstein, J. L. (1972). Perception of musical intervals: Evidence for the central origin of the pitch of complex tones. *J. of the Acoustical Soc. of America*, 51:520–529.

Huffman, C. (Accessed February 2005). Archytas. In *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/archytas.

Huovinen, E. (2002). *Pitch Class Constellations : Studies in the Perception of Tonal Centricity*. PhD thesis, Turku University.

Huron, D. (1999). *Music Research Using Humdrum: A User's Guid*. Center for Computer Assisted Research in the Humanities, Stanford, California.

Huron, D. and Parncutt, R. (1993). An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology*, 12(2):154–171.

Hyer, B. (1995). Reimag(in)ing Riemann. *J. of Music Theory*, 39(1):101–138.

Immerseel, L. V. and Martens, J.-P. (1992). Pitch and voiced/unvoiced determination with an auditory model. *J. of the Acoustical Soc. of America*, 91(6):3511–3526.

Izmirli, O. and Bilgen, S. (1996). A model for tonal context time course calculation from acoustical input. *J. of New Music Research*, 25(3):276–288.

Jain, A. K. and Dubes, R. C. (1988). *Algorithms for Clustering Data*. Prentice Hall.

Janata, P., Birk, J. L., Horn, J. D. V., Leman, M., Tillmann, B., and Bharucha, J. J. (2002). The cortical topography of tonal structures underlying Western music. *Science*, pages 2167–2170.

Jewanski, J. (1999). *Ist C = Rot? Eine Kultur- und Wissenschaftsgeschichte zum Problem der wechselseitigen Beziehung zwischen Ton und Farbe – Von Aristoteles bis Goethe*. Sinzig.

Kandel, E., Schwartz, J., and Jessell, T. (1991). *Principles of Neural Science*. Prentice-Hall, 3rd edition.

Kandinsky, W. (1926). *Punkt und Linie zu Fläche*. Benteli.

Kashino, K., Nakadai, K., Kinoshita, T., and Tanaka, H. (1998). Application of the Bayesian probability network to music scene analysis. In Rosenthal, D. F. and Okuno, H. G., editors, *Computational Auditory Scene Analysis*, pages 115–137. Lawrence Erlbaum Associates, Mahwah, NJ.

Keller, H. (1965). *Das wohltemperierte Klavier von Johann Sebastian Bach*.

Kellner, D. (1737). *Treulicher Unterricht im General-Bass*. 2nd edition.

Kirnberger, J. (1776). *Die Kunst des reinen Satzes in der Musik*. Berlin.

Kockelkorn, U. (2000). Multivariate Datenanalyse. Lecture Notes. Unpublished.

Koczalski, R. v. (1909). *Frederic Chopin*. P. Pabst, Leipzig.

Koelsch, S., Gunter, T., Friederici, A. D., and Schröger, E. (2000). Brain indices of music processing: "Nonmusicians" are musical. *J. of Cognitive Neuroscience*, 12(3):520–541.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biol. Cybernetics*, 43:59–69.

Krumhansl, C. (1990). *Cognitive Foundations of Musical Pitch*. Oxford University Press, Oxford.

Krumhansl, C. L. and Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89:334–368.

Krumhansl, C. L. and Shepard, R. N. (1979). Quantification of the hierarchy of tonal function with a diatonic context. *J. of Experimental Psychology: Human Perception and Performance*.

Kurth, E. (1913). *Die Voraussetzungen der theoretischen Harmonik und der tonalen Darstellungssysteme.* Musikverlag Emil Katzbichler. Reprint 1973.

Lakatos, S. (2000). A common perceptual space for harmonic and percussive timbres. *Perception and Psychophysics*.

Lang, D. and de Freitas, N. (2004). Beat tracking the graphical model way. In *Advances in Neural Information Processing Systems (NIPS)*, volume 18.

Langner, G. (1997). Neural processing and representation of periodicity pitch. In *Acta Otolaryngol*, pages 68–76. Stockholm. Suppl. 532.

Laske, O. (1987). Eine kurze Einführung in die Kognitive Musikwissenschaft: Folgen des Computers in der Musik. In *Computermusik: Theoretische Grundlagen - Kompositionsgeschichtliche Zusammenhänge – Musiklernprogramme*, pages 169–194. Laaber.

Lee, T.-W., Girolami, M., Bell, A. J., and Sejnowski, T. J. (1998). A unifying information-theoretic framework for independent component analysis. *Int. J. on Mathematical and Computer Modeling*.

Lehrdahl, F. (2001). *Tonal Pitch Space*. Oxford University Press.

Leibniz, G. W. (1734). Epistolae ad diversos. volume 1 of 2, page 240. Leipzig.

Leichtentritt, H. (1921). *Analyse von Chopins Klavierwerken*, volume 1.

Leman, M. (1990). Emergent properties of tonality functions by self-organization. *Interface*.

Leman, M. (1994). Schema-based tone center recognition of musical signals. *J. of New Music Research*, 23:169–204.

Leman, M. (1995). *Music and Schema Theory*, volume 31 of *Springer Series in Information Sciences*. Springer, Berlin, New York, Tokyo.

Leman, M. and Carreras, F. (1997). Schema and Gestalt: Testing the hypothesis of psychoneural isomorphism by computer simulation. In Leman, M., editor, *Music, Gestalt, and Computing*, number 1317 in Lecture Notes in Artificial Intelligence, pages 144–168. Springer, Berlin.

Lerdahl, F. and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. MIT Press.

Levine, M. (1996). *The Jazz Theory Book*. Sher Music.

Lévy, F. (2004). *Complexité grammatologique et complexité aperceptive en musique*. PhD thesis, EHESS, Paris.

Lewin, D. (1987). *Generalized Musical Intervals and Transformations*. Yale University Press.

Licklider, J. C. R. (1954). Periodicity 'pitch' and place 'pitch'. *J. of the Acoustical Soc. of America*, 26:945.

Lidov, D. (1999). *Elements of Semiotics*. Signs and Semaphores. St. Martin's Press.

Lindley, M. (1987). Stimmung und Temperatur. In *Hören, Messen und Rechnen in der frühen Neuzeit*, number 6 in Geschichte der Musiktheorie. Wissenschaftliche Buchgesellschaft Darmstadt.

Lindley, M. (2001). Temperaments. In Sadie, S., editor, *The New Grove Dictionary of Music and Musicians*, volume XVI, pages 205–206. Grove, Taunton, MA, 2nd edition.

Lindley, M. (2003). A rudimentary approach to the history of the major and minor keys. Unpublished Manuscript.

Lindley, M. (Accessed 1 Dec. 2004). Well-tempered clavier. In Macy, L., editor, *Grove Music Online*. http://www.grovemusic.com.

Linsker, R. (1989). How to generate ordered maps by maximizing the mutual information between input and output signals. *Neural Computation*, 1:402–411.

Lischka, C. (1987). Connectionist models of musical thinking. In *Proc. of the Int. Computer Music Conf.*, pages 190–197. Urbana, IL.

Longuet-Higgins, H. and Steedman, M. (1970). On interpreting Bach. *Machine Intelligence*, 6:221–239.

Lucy, C. E. H. (2000). Pitch, pi, and other musical paradoxes. http://www.harmonics.com/lucy/lsd/colors.html.

Lyon, R. and Shamma, S. (1996). Auditory representations of timbre and pitch. In *Auditory computation*. Springer.

Maass, W. (1997). Networks of spiking neurons: the third generation of neural network models. *Neural Networks*, 10:1659–1671.

Mach, E. (1886). Beiträge zur Analyse der Empfindungen. Jena.

Marin, O. S. M. and Perry, D. W. (1999). *The Psychology of Music*, chapter Neurological aspects of music perception and performance, pages 653–724. Academic Press, San Diego, 2nd edition.

Mattheson, J. (1735). *Kleine General-Bass-Schule*. J. Chr. Kissner.

Mayer, L. (1947). *Die Tonartencharakteristik im geistlichen Vokalwerk Johann Sebastian Bachs*. PhD thesis, Wien.

Mazzola, G. (1990). *Geometrie der Töne*. Birkhäuser Verlag, Basel.

Mazzola, G. (2002). *The Topos of Music*. Birkhäuser, Basel.

McAdams, S., Winsberg, S., Donnadieu, S., Soete, G. D., and Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychological Research*, 58:177–192.

Meddis, R. (1988). Simulation of auditory-neural transduction: Further studies. *J. of the Acoustical Soc. of America*, 83(3):1056–1063.

Meddis, R. and Hewitt, M. J. (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *J. of the Acoustical Soc. of America*, 89(6):2866–2882.

Meister, W. T. (1991). *Die Orgelstimmung in Süddeutschland vom 14. bis zum Ende des 18. Jahrhunderts*. Orgelbau-Fachverlag Rensch, Lauffen am Neckar.

Mendenhall, T. C. (1901). A mechanical solution of a literary problem. *The Popular Science Monthly*, 60(7):97–105.

Moore, B. (1973). Some experiments relation to the perception of complex tones. *The Quarterly J. of Experimental Psychology*, 15:451–475.

Moore, B. and Glasberg, B. (1983). Suggested formulae for calculating auditory filter bandwidths and excitation patterns. *J. of the Acoustical Soc. of America*, 74:750–753.

Muellensiefen, D. and Frieler, K. (2004). Cognitive adequacy in the measurement of melodic similarity: Algorithmic vs. human judgments. In Hewlett, W. B. and Selfridge-Field, E., editors, *Computing in Musicology*, volume 13. Center for Computer Assisted Research in the Humanities and MIT Press, Menlo Park.

Müller, K.-R., Mika, S., Rätsch, G., Tsuda, K., and Schölkopf, B. (2001). An introduction to kernel-based learning algorithms. *IEEE Neural Networks*, 12(2):181–201.

Müller, K.-R., Philips, P., and Ziehe, A. (1999). JADE-TD: Combining higher-order statistics and temporal information for blind source separation (with noise). In *Proc. of the 1st Int. Workshop on Independent Component Analysis and Signal Separation (ICA-99)*, pages 87–92, Aussios, France.

Murata, N., Ikeda, S., and Ziehe, A. (2000). An approach to blind source separation based on temporal structure of speech signals. *Neurocomputation*.

Nakatani, T., Okuno, H. G., and Kawabata, T. (1995). Residue-driven architecture for computational auditory scene analysis. In *Proc. of the 14th Int. Joint Conf. on Artificial Intelligence (IJCAI'95)*, volume 1, pages 165–172.

Nestke, A. (2004). Paradigmatic motivic analysis. In Mazzola, G., Noll, T., and Luis-Puebla, E., editors, *Perspectives in Mathematical and Computational Music Theory*, Osnabrück Series on Music and Computation, pages 343–365. Electronic Publishing Osnabrück.

Newton, I. (1704). *Opticks: or, a Treatise of the Reflexions, Refractions, Inflexions and Colours of Light*. London. Reprint New York 1952.

Noll, T. (1995). Fractal depth structure of tonal harmony. In *Proc. of the Int. Computer Music Conf.*, Banff. ICMA.

Noll, T. (2002). Tone apperception, relativity and Weber-Fechner's law. In Belardinelli, M. and Olivietti, M., editors, *Proc. of the 3rd Int. Conf. Understanding and Creating Music*, Caserta.

Noll, T. (2003). A mathematical model of tone apperception. *American Mathematical Society (AMS) Sectional Meeting,* Baton Rouge.

Noll, T. (2005). Ton und Prozess. Habilitation Thesis, Berlin University of Technology. to appear.

Noll, T. and Brand, M. (2004). Harmonic path analysis. In Mazzola, G., Noll, T., and Luis-Puebla, E., editors, *Perspectives in Mathematical and Computational Music Theory*, Osnabrück Series on Music and Computation, pages 399–431. Electronic Publishing Osnabrück.

Normann, I. (2000). Tonhöhenwahrnehmung: Simulation und Paradoxie. Diploma Thesis, University of Tübingen.

Normann, I., Purwins, H., and Obermayer, K. (2001a). Interdependence of pitch and timbre perception for octave ambiguous tones. Deutsche Jahrestagung für Akustik (DAGA-01).

Normann, I., Purwins, H., and Obermayer, K. (2001b). Spectrum of pitch differences models the perception of octave ambiguous tones. In Schloss, A., Dannenberg, R., and Driessen, P., editors, *Proc. of the Int. Computer Music Conf.*, Havana.

Obermayer, K., Blasdel, G. G., and Schulten, K. (1991). A neural network model for the formation of the spatial structure of retinotopic maps, orientation- and ocular dominance columns. In Kohonen, T., Mäkisara, K., Simula, O., and Kangas, J., editors, *Artificial Neural Networks I*, pages 505–511. North Holland.

Obermayer, K., Ritter, H., and Schulten, K. (1990). A principle for the formation of the spatial structure of cortical feature maps. *Proc. of the National Academy of Science USA*, 87:8345–8349.

Oettingen, A. v. (1866). *Harmoniesystem in dualer Entwickelung*. Dorpat.

Okuno, H. G., Ikeda, S., and Nakatani, T. (1999). Combining independent component analysis and sound stream segregation. In *Proc. of the IJCAI-99 Workshop on Computational Auditory Scene Analysis (CASA'99)*, pages 92–98, Stockholm, Sweden.

O'Mard, L. P., Hewitt, M. J., and Meddis, R. (1997). *LUTEar 2.0.9 Manual*. http://www.essex.ac.uk/psychology/hearinglab/lutear/manual/Manual.html.

Oppenheim, A. V. and Schafer, R. W. (1989). *Discrete-Time Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ.

Palisca, C. V. and Moore, B. C. J. (Accessed 1 Dec. 2004). Consonance. In Macy, L., editor, *Grove Music Online*. http://www.grovemusic.com.

Pardo, B. and Birmingham, W. P. (2002). Algorithms for chordal analysis. *Computer Music J.*, 26(2):27–49.

Parra, L. and Spence, C. (2000). Convolutive blind separation of non-stationary sources. In *IEEE Transactions Speech and Audio Processing*, pages 320–327.

Patterson, R., Holdsworth, J., Nimmo-Smith, I., and Rice, P. (1988). An efficient auditory filterbank based on the gammatone function. Technical Report 2341, APU. Annex B of the SVos Final Report: The auditory filter bank.

Peeters, G., McAdams, S., and Herrera, P. (2000). Instrument sound description in the context of MPEG-7. In *Proc. of the Int. Computer Music Conf.* ICMA.

Petroni, N. C. and Tricarico, M. (1997). Self-organizing neural nets and the perceptual origin of the circle of fifths. In Leman, M., editor, *Music, Gestalt, and Computing*, number 1317 in Lecture Notes in Artificial Intelligence, pages 169–180. Springer, Berlin.

Petsche, H. (1994). In Bruhn, H., Oerter, R., and Rösing, H., editors, *Musikpsychologie*, chapter Zerebrale Verarbeitung, pages 630–638. Rowohlts Enzyklopädie, Reinbek.

Plomp, R. (1965). Detectability threshold for combination tones. *J. of the Acoustical Soc. of America*, 37:1110–1123.

Plomp, R. (1967). Pitch of complex tones. *J. of the Acoustical Soc. of America*, 42:1526–1533.

Provost, F., Fawcett, T., and Kohavi, R. (1998). The case against accuracy estimation for comparing induction algorithms. In *Proc. 15th Int. Conf. on Machine Learning*, pages 445–453. Morgan Kaufmann, San Francisco, CA.

Purwins, H., Blankertz, B., Dornhege, G., and Obermayer, K. (2004a). Scale degree profiles from audio investigated with machine learning techniques. In *Audio Engineering Soc. 116th Convention*, Berlin.

Purwins, H., Blankertz, B., Graepel, T., and Obermayer, K. (2006). Pitch class profiles and inter-key relations. In Hewlett, W. B. and Selfridge-Field, E., editors, *Computing in Musicology*, volume 15. Center for Computer Assisted Research in the Humanities and MIT Press, Menlo Park. In print.

Purwins, H., Blankertz, B., and Obermayer, K. (2000a). Computing auditory perception. *Organised Sound*, 5(3):159–171.

Purwins, H., Blankertz, B., and Obermayer, K. (2000b). A new method for tracking modulations in tonal music in audio data format. In Amari, S.-I., Giles, C. L., Gori, M., and Piuri, V., editors, *Int. Joint Conf. on Neural Networks (IJCNN-00)*, volume 6, pages 270–275. IEEE Computer Society.

Purwins, H., Blankertz, B., and Obermayer, K. (2001a). Constant Q profiles for tracking modulations in audio data. In Schloss, A., Dannenberg, R., and Driessen, P., editors, *Proc. of the Int. Computer Music Conf.*, Havana. ICMA.

Purwins, H., Blankertz, B., and Obermayer, K. (2001b). Modelle der Musikwahrnehmung zwischen auditorischer Neurophysiologie und Psychoakustik. Technical report.

Purwins, H., Blankertz, B., and Obermayer, K. (2002). Automated harmonic music analysis. In *Int. Neuroscience Summit*. Berlin.

Purwins, H., Graepel, T., Blankertz, B., and Obermayer, K. (2004b). Correspondence analysis for visualizing interplay of pitch class, key, and composer. In Mazzola, G., Noll, T., and Luis-Puebla, E., editors, *Perspectives in Mathematical and Computational Music Theory*, Osnabrück Series on Music and Computation, pages 432–454. Electronic Publishing Osnabrück.

Purwins, H., Normann, I., and Obermayer, K. (2005). Unendlichkeit – Konstruktion musikalischer Paradoxien. In Stahnke, M., editor, *Mikrotöne und mehr – Auf György Ligetis Hamburger Pfaden*, pages 39–80. Bockel Verlag, Hamburg.

Rameau, J. P. (1722). *Traité de l'harmonie réduite à ses principes naturels*. Paris.

Rätsch, G., Onoda, T., and Müller, K.-R. (2001). Soft margins for AdaBoost. *Machine Learning*, 42(3):287–320.

Ratz, E. (1973). *Einführung in die musikalische Formenlehre*. Universal Edition, Wien.

Riemann, H. (1877). *Musikalische Syntaxis*. Leipzig.

Riemann, H. (1913). *Handbuch der Harmonie- und Modulationslehre*, volume 15 of *Max Hesses illustrierte Handbücher*. Max Hesses Verlag, Leipzig, 5th edition.

Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press.

Risset, J.-C. (1985). Computer music experiments 1964- ... *Computer Music J.*, 9(1):11–18.

Ritsma, R. J. (1962). Existence region of the tonal residue. *J. of the Acoustical Soc. of America*, 34:1224–1229.

Ritsma, R. J. (1970). Periodicity detection. In Plomp, R. and Smoorenberg, G. F., editors, *Frequency Analysis and Periodicity Detection in Hearing*. Sijthoff.

Ritter, H. and Schulten, K. (1988). Convergence properties of Kohonen's topology conserving maps: Fluctuations, stability, and dimension selection. *Biol. Cybernetics*, 60:59–71.

Rockstro, W. S., Dyson, G., Drabkin, W., Powers, H. S., and Rushton, J. (Accessed 2 Dec. 2004). Cadence. In Macy, L., editor, *Grove Music Online*. http://www.grovemusic.com.

Rodet, X. and Jaillet, F. (2001). Detection and modeling of fast attack transients. In *Proc. of the Int. Computer Music Conf.*, pages 30–33. ICMA.

Roederer, J. G. (1995). *The Physics and Psychophysics of Music*. Springer, 3rd edition.

Rosenthal, D. F. and Okuno, H. G., editors (1998). *Computational Auditory Scene Analysis*. L. Erlbaum Assoc.

Sabbe, H. (1987). *György Ligeti – Studien zur kompositorischen Phänomenologie*, volume 53 of *Musik-Konzepte*. edition text+kritik.

Sadie, S., editor (2001). *The New Grove Dictionary of Music and Musicians*. Grove, Taunton, MA, 2nd edition.

Salzer, F. (1962). *Structural Hearing: Tonal Coherence in Music*. Dover Publications.

Sanger, T. D. (1989). Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 2:459–473.

Sapp, C. S. (2001). Harmonic visualizations of tonal music. In Blanco, J., Blanco, E., Schloss, A., and Trevisani, M., editors, *Proc. of the Int. Computer Music Conf.* ICMA.

Saslaw, J. (Accessed 3 Dec. 2004). Modulation. In Macy, L., editor, *Grove Music Online*. http://www.grovemusic.com.

Scheffers, M. (1983). Simulation of auditory analysis of pitch: An elaboration on the DWS pitch meter. *J. of the Acoustical Soc. of America*, 74:1716–1725.

Scheirer, E. D. (1998). Tempo and beat analysis of acoustic musical signals. *J. of the Acoustical Soc. of America*, 103(1):588–601.

Schenker, H. (1935). *Der freie Satz*, volume 3 of *Neue musikalische Theorien und Phantasien*. Wien.

Schneider, A. (2005). Was haben Ligetis Études pour piano mit Shepard-Skalen zu tun? Über "auditorische Illusionen", Vertige und Columna infinită. In Stahnke, M., editor, *Mikrotöne und mehr – Auf György Ligetis Hamburger Pfaden*. Bockel Verlag, Hamburg.

Schneider, N. J. (1987). Durmollharmonik im 18./19. Jahrhundert. In Salmen, W. and Schneider, N. J., editors, *Der Musikalische Satz*, pages 143–185. Edition Helbling, Innsbruck.

Schönberg, A. (1966). *Harmonielehre*. Universal Edition, Wien, 2nd edition.

Schönberg, A. (1969). *Structural functions of harmony*. Norton, New York, 2nd edition.

Schopenhauer, A. (1859). *Die Welt als Wille und Vorstellung*. Frankfurt.

Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (1962). Pitch of the residue. *J. of the Acoustical Soc. of America*, 34:1418–1424.

Schreiner, C. E. and Langner, G. (1988). Coding of temporal patterns in the central auditory nervous system. In G. M. Edelman, W. G. and Cowan, W., editors, *Auditory Function: Neurobiological Bases of Hearing*. John Wiley and Sons, New York.

Seifert, U. (1993). *Systematische Musiktheorie und Kognitionswissenschaft*, volume 69 of *Orpheus-Schriftenreihe zu Grundfragen der Musik*. Verlag für systematische Musikwissenschaft, Bonn.

Seppänen, J. (2001). Computational models of musical meter recognition. Master's thesis, Tampere University of Technology.

Shaw-Miller, S. (2000). Skriabin and Obukhov: Mysterium & La livre de vie – The concept of artistic synthesis. http://www.aber.ac.uk/tfts/journal/december/skria.html.

Sheh, A. and Ellis, D. P. (2003). Chord segmentation and recognition using EM-trained hidden Markov models. In *Int. Conf. on Music Information Retrieval*, Baltimore.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*.

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *J. of the Acoustical Soc. of America*, 36:2346–2353.

Simon, H. A. (1968). Perception du pattern musical par auditeur. *Science de l'Art*, 5:28–34.

Slaney, M. (1994). Auditory model inversion for sound separation. In *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 563–569, Adelaide, Australia.

Slaney, M. (1998). Connecting correlograms to neurophysiology and psychoacoustics. In Palmer, A. R., Rees, A., Summerfield, A. Q., and Meddis, R., editors, *Psychophysical and Physiological Advances in Hearing*. Whurr Publishers.

Smolensky, P. (1991). Connectionism, constituency, and the language of thought. In Loewer, B. and Rey, G., editors, *Meaning in Mind, Fodor and his Critics*, pages 201–227. Blackwell.

Stevens, S. S. and Newman, E. B. (1936). On the nature of aural harmonics. *Proc. of the National Academy of Science*, 22:668–672.

Strang, G. and Nguyen, T. (1997). *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA.

Stuckenschmidt, H. H. (1969). *Twentieth Century Music*. Weidenfeld & Nicolson, New York.

Temperley, D. (2001). *The Cognition of Basic Musical Structures*. MIT Press.

Tenenbaum, J. B., de Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323.

Terhardt, E. (1974). Pitch, consonance and harmony. *J. of the Acoustical Soc. of America*, 55:1061–1069.

Terhardt, E. (1992). Zur Tonhöhenwahrnehmung von Klängen. II. Ein Funktionsschema. *Acustica*, 26:187–199.

Terhardt, E. (1998). *Akustische Kommunikation*. Springer.

Terhardt, E. and Seewann, M. (1984). Auditive und objektive Bestimmung der Schlagtonhöhe von historischen Kirchenglocken. *Acustica*, 54:129–144.

Terhardt, E., Stoll, G., and Seewann, M. (1982). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *J. of the Acoustical Soc. of America*, 71(3):679–688.

Terman, D. and Wang, D. L. (1995). Global competition and local cooperation in a network of neural oscillators. *Physica D*, 81:148–176.

Todd, N. (1999). Implications of a sensory-motor theory for the representation and segregation of speech. *J. of the Acoustical Soc. of America*, 105(2):1307.

Toop, R. (1999). *György Ligeti*. 20th Century Composers. Phaidon Press.

Tzanetakis, G., Essl, G., and Cook, P. (2002). Human perception and computer extraction of musical beat strength. In *Proc. of the 5th Int. Conf. on Digital Audio Effects (DAFx-02)*, Hamburg.

Uhle, C. and Dittmar, C. (2004). Generation of musical scores of percussive unpitched instruments from automatically detected events. In *Audio Engineering Soc. 116th Convention*.

Vanechkina, I. (1994). Castel and Scryabin: Evolution of light-music ideas. In Naranjo, M., editor, *From Castel to our Days*, pages 23–29.

Vapnik, V. (1998). *Statistical Learning Theory*. Jon Wiley and Sons, New York.

Walliser, K. (1968). *Zusammenwirken von Hüllkurvenperiode und Tonheit bei der Bildung der Periodentonhöhe*. PhD thesis, Technische Hochschule München.

Wang, D. L. (1996). Primitive auditory segregation based on oscillatory correlation. *Cognitive Science*, 20:409–456.

Weber, C. (2000). *Maximum a Posteriori Models for Cortical Modeling: Feature Detectors, Topography and Modularity*. PhD thesis, Technical University Berlin.

Weber, G. (1817). *Versuch einer geordneten Theorie der Tonsetzkunst*. B. Schott, Mainz.

Werckmeister, A. (1697). *Hypomnemata musica*. Quedlinburg.

Werts, D. (1983). *A Theory of Scale References*. PhD thesis, Princeton.

Whitfield, I. C. (1970). Central nervous processing in spatio-temporal discrimination. In Plomp, R. and Smoorenberg, G. F., editors, *Frequency Analysis and Periodicity Detection in Hearing*. Sijthoff, Leiden.

Winograd, T. (1968). Linguistics and the computer analysis of tonal harmony. *J. of Music Theory*, 12:2–49.

Witting, H. (1966). *Mathematische Statistik*. Teubner.

Yost, W. A. and Hill, R. (1979). Pitch and pitch strength of ripple noise. *J. of the Acoustical Soc. of America*, 66:400–410.

Yost, W. A. and Nielsen, D. W. (1989). Fundamentals of hearing.

Zimmermann, E. (1969). *Chopin Préludes op. 28: Kritischer Bericht*. Henle.

Zwicker, E. and Fastl, H. (1990). *Psychoacoustics – Facts and Models*. Springer.

Zwicker, E. and Fastl, H. (1999). *Psychoacoustics – Facts and Models*. Springer, 2nd edition.