technische universität
dortmund

TAMPEREEN TEKNILLINEN YLIOPISTO

Workshop on Music Signal Analysis

November 3-4, 2008, Witten-Bommerholz

Organisation:

Claus Weihs (Dortmund, D), Anssi Klapuri (Tampere, FIN),

Uwe Ligges (Dortmund, D), Rainer Martin (Bochum, D)

Graduiertenkolleg
Statistische Modellbildung

technische universität dortmund

TAMPEREEN TEKNILLINEN YLIOPISTO

November 3, 2008

9:00 **Claus Weihs** (TU Dortmund, D)
Welcome and introduction

I) Music via hearing aids

Session I.1: Hearing aids (chair: A. Kohlrausch)

9:10 **Marcus Holmberg** (Otikon, DK)
Introduction to hearing aid processing and present challenges **(p. 8)**

9:50 **Rainer Martin** (Ruhr Universität Bochum, D)
Speech Enhancement in Hearing Aids - From Noise Suppression to Rendering of Auditory Scenes **(p. 13)**

10:30 Coffee break

Session I.2: Coding and music identification (chair: R. Martin)

11:00 **Gerald Schuller** (TU Ilmenau, D)
Perceptual Audio Coding with Low Delay **(p. 19)**

11:40 **Tuomas Virtanen** (Tampere University of Technology, FIN)
Automatic transcription, separation, and alignment of vocals in polyphonic music **(p. 23)**

12:20 Lunch break

Session I.3: Localization (chair: A. Klapuri)

13:40 **Armin Kohlrausch** (Philips Research Europe, Technische Universiteit Eindhoven, NL): The importance of monaural and binaural cues in human sound source separation **(p. 10)**

14:20 **Stephan Werner** (Fraunhofer IDMT, D)
Bio-inspired binaural sound source localization and separation for hearing aids **(p. 24)**

15:00 **Nilesh Madhu**, Rainer Martin (Ruhr Universität Bochum, D)
A Scalable Framework for Multiple Speaker Localization and Tracking **(p. 12)**

15:40 Coffee break

Graduiertenkolleg Statistische Modellbildung

II) Identification of Music Categories

Session II.1 (chair: G. Rudolph)

16:10 **Holger Blume** (Universität Hannover, D), **W. Theimer** (RIM, D), **M. Botteck**:
DSP-based Performance Analysis for Perceptual Feature-based Music Classification **(p. 6)**

16:50 **Wolfgang Theimer** (RIM, D):
Visualization of Personal Music Categories **(p. 21)**

17:30 **Jacob Abesser, Christian Dittmar, Holger Grossmann** (Fraunhofer IDMT, D)
High-Level Music Classification based on Automatic Transcription **(p. 5)**

18:10 Dinner

III) Poster

19:15 **Poster Session** (with bar open) **(p. 16)**

Group meeting

20:00 **Project planning**: Dortmund-Bochum Group (chair: H. Blume, T. Kemp)

Graduiertenkolleg
Statistische Modellbildung

November 4, 2008

IV) Automatic correcting feedback to singing/playing music

Session IV.1: Recognition I (chair: C. Weihs)

9:00 **Christopher Raphael, Kyung Ae Lim** (Indiana University, USA)
InTune: A Program for Visualizing Intonation **(p. 17)**

9:40 **Michael Oehler, Christoph Reuter, Jobst Fricke** (University of Cologne, D)
Recognizing and synthesizing wind instrument timbres: the role of formants and micro-modulations **(p. 15)**

10:20 Coffee break

Session IV.2: Recognition II (chair: U. Ligges)

10:50 **Gero Szepannek** (Dortmund University of Technology, D)
Phoneme Recognition in Popular Music **(p. 20)**

11:30 **Matthias Mauch** (University of London, UK)
Automatic chord identification **(p. 21)**

12:10 Lunch break

Session IV.3: Emotion and reverberation (chair: C. Raphael)

13:40 **Igor Vatolkin** (Dortmund University of Technology, D)
Emotion recognition **(p. 22)**

14:20 **Günther Rötter** (Dortmund University of Technology, D)
Emotion and vocal expression **(p. 18)**

15:00 **Berit Janssen** (Universität Hamburg, D)
Reverberation: Cognitive representation and musical applications **(p. 9)**

15:40 Coffee break

Session IV.4: Movement (chair: T. Virtanen)

16:10 **Frank Desmet** (IPEM, University of Gent, B)
Analysis of the gestural movement of a Guqin player **(p. 7)**

16:50 **Geoff Luck** (University of Jyväskylä, FIN)
Exploring relationships between the kinematics of a singer's body movement and the quality of their voice **(p. 11)**

17:30 **Claus Weihs** (TU Dortmund, D): Final words

17:45 Dinner

19:00 Final Discussion (at the bar)

Jacob Abesser, Christian Dittmar, Holger Grossmann (Fraunhofer IDMT, D)

Monday, Nov. 3, 17:30 h

High-Level Music Classification based on Automatic Transcription

We describe a set of high-level music features applicable for music-genre and artist-classification. The set consists of 148 single- and multidimensional features related to melodic, rhythmic, harmonic and structural properties observable via Automatic Transcription of four distinct instrumental domains. A simple but comprehensive instrumentation model is used to describe both the characteristics of melody and accompaniment as well as rhythmic and melodic interaction between them. To evaluate the features' discriminative power, an evaluation for both genre and artist classification has been performed. Two different classifier approaches have been utilized. First, a combination of Linear Discriminant Analysis with Support Vector Machines, and second, a Nearest Neighbour Classifier compare tempo-adaptive Rhythmical Structure Profiles. The results of both evaluations are presented and discussed accordingly. Concluding, suggestions for improvement and possible application scenarios are outlined.

Holger Blume (Universität Hannover, D), **W. Theimer** (RIM, D), **M. Botteck**:

Monday, Nov. 3, 16:10 h

DSP-based Performance Analysis for Perceptual Feature-based Music Classification

Today, more and more computational power is available not only in desktop computers but also in portable devices such as smart phones or PDAs. At the same time the availability of huge non-volatile storage capacities (flash memory etc.) suggests to maintain huge music databases even in mobile devices. Automated music classification promises to allow keeping a much better overview on huge data bases for the user. Such a classification enables the user to sort the available huge music archives according to different genres which can be either predefined or dynamically user defined. It is typically based on a set of perceptual features which are extracted from the music data. Feature extraction and subsequent music classification are very computational intensive tasks. Today, a variety of music features and possible classification algorithms optimized for various application scenarios and achieving different classification qualities are under discussion. In this paper results concerning the computational needs and the achievable classification rates on different processor architectures are presented. The inspected processors include a general purpose P IV dual core processor, heterogeneous digital signal processor architectures like a Nomadik STn8810 (featuring a smart audio accelerator, SAA) as well as an OMAP2420. In order to increase classification performance, different forms of feature selection strategies (heuristic selection, full search and Mann-Whitney-Test) are applied. Furthermore, the potential of a hardware-based acceleration for this class of application is inspected by performing a fine as well as a coarse grain instruction tree analysis. Instruction trees are identified, which could be attractively implemented as custom instructions speeding up this class of applications.

Frank Desmet (IPEM, University of Gent, B)

Tuesday, Nov. 4, 16:10 h

Analysis of the gestural movement of a Guqin player

The gestures of a Chinese Guqin player were captured using infrared cameras. The experiment is a part of a large scale experiment where subjects were asked to move along the Guqin music [1]. The movements of 11 different joints were registered for 3 different musical pieces. Investigation of the movement data revealed interesting information about the way the instrument is played with respect to the characteristics of the musical piece. Furthermore the movements of the different joints are correlated and depend on the difference between the head and shoulders, the left and the right arm. Principal Component Analysis enables to classify the different movements. Delay and anticipation patterns are present over the different joints. It was also found that the movement of the head reflects the intentionality of the player. Using phylogenetic analysis we were able to provide evidence that the head is the precursor of the consecutive movements of the other joints. The results are in accordance with previous research on the Physical Modeling of the Guqin [2].

[1] Sharing Musical expression through embodied listening – a case study using Chinese guqin music, Marc Leman, Frank Desmet, Frederik Styns and Leon van Noorden, JNMR, 2008 [paper accepted]

[2] A Gesture-based Typology of Sliding-tones in Guqin Music, Li Henbing, Marc Leman, Journal of New Music Research, Volume 36, Issue 2 June 2007 , pages 61 – 82.

Marcus Holmberg (Otikon, DK)

Monday, Nov. 3, 9:00 h

Introduction to hearing aid processing and present challenges

In this presentation I will briefly introduce the signal processing schemes that can be found in virtually every modern hearing aid: Which are they and what are the user benefits?
The second part of the talk will discuss present trends in hearing aid processing. Recently, wireless communication has been introduced in hearing aids. This opens up for connecting hearing aids directly to modern communication devices, but also for a whole new range of signal processing possibilities. The hearing aids are going from being bilateral to being binaural. Another trend is that “open fittings” (ear pieces with large vents) are getting more and more popular because of the wearing comfort. This is a different kind of challenge, partly because an open fitting typically reduces the effect of classical hearing aid processing.
In the last part I will use the example of a “music program” (special hearing aid settings for listening to music) to highlight some of the effects discussed earlier in the talk.

Berit Janssen (Universität Hamburg, D)

Tuesday, Nov. 4, 15:00 h

Reverberation: Cognitive representation and musical applications

Reverberation in rooms is simulated for various different purposes. It may be used as an improvement of immersion in a virtual environment, as for instance in computer games. In online conferences, an auditory virtual environment as invoked by reverberation may increase the sense of a natural conversation supported by room reflections. It is even conceivable that internet sites are connected as virtual rooms, the navigation through which can be more embodied if an auditory virtual environment accompanies the browser transition.
Spaces can even come close to being musical instruments. Traditionally, the performance rooms were often taken into consideration by composers, even before modern technology enabled musicians and composers to transfer their musical activity into any space. Nowadays, composition and sound design for film and radio, and acousmatic music make use of spatialization by reverberation.
Hence, versatile reverberation techniques are needed, which are dynamically and intuitively controllable. Strategies of artificial reverberation are manifold: they include convolution of a sound-object with an impulse response, the simulation of sound propagation by raytracing or image source algorithms, or the use of allpass and comb filter chains. None of these technologies enable the user to control realistic reverberation dynamically and in real time, however. A possible way to achieve a dynamic reverberation control may be a hybrid method, which for instance tests for psychoacoustically important room properties by raytracing algorithms, and emulates them using filters. In this sense, it is indispensable to learn which parameters enable human cognition to discern different reverberating spaces.
An experiment was carried out to test for these parameters. Comparable studies exist, but often focus on the acoustic quality of concert halls, and hence exclude most other rooms and spaces. In the present context it was more important to understand the dimensions underlying listeners' judgements, hence an approach of multidimensional scaling was chosen, processing the 49 participants' difference judgements of pairs of impulse response from nine inside and outside spaces.
The results imply that the temporal decrease of energy and the timbral development might be the parameters which are cognitively most prominent in the identification of spaces. These findings may be employed to reduce data in the emulation of auditory virtual environments, in order to move towards dynamic control of realistic reverberation.

Armin Kohlrausch (Philips Research Europe, Technische Universiteit Eindhoven, NL)

Monday, Nov. 3, 13:40 h

The importance of monaural and binaural cues in human sound source separation

The ability to segregate and identify sound sources in an auditory "scene" comes naturally to human listeners. Computationally, however, this has proven to be a difficult task. Basically, two approaches can be distinguished: An approach based on sound source properties (i.e. the statistical independence of the corresponding acoustic signals), often summarized under the term blind signal separation (BSS), and an approach trying to mimic human processing in such conditions, called Computational Auditory Scene Analysis (CASA). As for the human processing of auditory scenes, the cues derived from spatial properties of a sound are often considered to be dominant. This view does, however, neglect the strong role of monaural cues, e.g. temporal onset synchrony or co-modulations across frequency and the role of harmonic relations between spectral components belonging to the same source. In my presentation I want to give an overview about present-day thinking about the relative role of the various perceptual cues for human sound source separation (and support this overview by acoustic examples), in particular the relative contributions of monaural and binaural cues. Furthermore, I will give examples how this knowledge can be incorporated in algorithms that try to solve the sound source separation problem.

Geoff Luck (University of Jyväskylä, FIN)

Tuesday, Nov. 4, 16:50 h

Exploring relationships between the kinematics of a singer's body movement and the quality of their voice

Physical movement plays an important role in musical perception and production. It is generally agreed among professional singers and vocal teachers, for example, that there are relationships between the kinematics of a singer’s body and the quality of their voice. Thus, we might expect to find relationships between quantifiable indicators of a singer’s vocal performance and quantifiable features of their movements while they sing. High-resolution motion capture systems have been used in several studies to investigate connections between music and movement (e.g., Palmer & Dalla Bella, 2004; Wanderley, Vines, Middleton, McKay, & Hatch, 2005; Luck & Toiviainen, 2006). The overall orientation of different body parts and the amount of their movement can be estimated from the motion capture data, and these features can be subsequently modeled computationally. The aim of this study was to synthesize basic research on the singing voice, human movement, and quantification of audio and movement data in an exploration of relationships between bodily posture and singing quality. To this end, relationships between the spatial arrangement of the limbs and selected audio features of 15 singers’ performances of a song (“Tuulantei” by Oskar Merikanto) were examined statistically. Results indicated that, while there were individual differences in the relationships observed, features relating to timbre seemed to be frequently associated with the lateral angles of the head and neck. The frontal angles of the upper body, and the frontal angle and rotation of the head, were also important. The present study combines empirical methods of music psychology with mathematical, statistical, and signal processing methods to produce formalized knowledge on singing that has application areas in music education.

Nilesh Madhu, **Rainer Martin** (Ruhr Universität Bochum, D)

Monday, Nov. 3, 15:00 h

A Scalable Framework for Multiple Speaker Localization and Tracking

In this talk we present a novel, scalable approach to the localization and tracking of multiple speakers using microphone arrays. The approach is capable of localizing sources both in non-competing and in concurrent situations, and is based on the disjointness of speech in the short-time discrete Fourier domain (STFD). The algorithm operates on a narrowband localization cost function in the STFD and yields an estimate of the speaker activity per time frame k by applying a Mixture of Gaussians (MoG) fit to the narrowband localization estimates. The advantages of the proposed method are manifold: it allows us to use coarse search grids for the cost function evaluation, without compromising on the location accuracy; it allows for a soft-decision on the number and position of sources on the fly; the framework is scalable to multi-array systems; and it can also serve as a base framework for enhancement algorithms. In principle, this approach is not specific to speakers and will work for any source combination provided they exhibit some temporal and spectral disjointness.

Rainer Martin (Ruhr Universität Bochum, D)

Monday, Nov. 3, Tuesday, Nov. 3, 9:50 h

Speech Enhancement in Hearing Aids - From Noise Suppression to Rendering of Auditory Scenes

While early work in speech enhancement centred around the improvement of the target-to-interferer (or signal-to-noise) ratio recent insights into the function of the human auditory system suggest that the proper identification and formation of auditory objects and scenes is key to the successful rendering of the desired target sources. This is especially true when the listener suffers from an increased amount of energetic and informational masking due to a hearing impairment.
In this talk we like to explore auditory scene analysis and synthesis in the context of hearing instruments which, when equipped with a digital binaural audio interconnection, will enable the binaural rendering of auditory scenes. We will discuss contemporary speech enhancement techniques, such as single and dual channel noise reduction, beamforming microphone arrays, blind source separation, and source localization techniques in the light of their ability to provide consistent auditory access to the acoustic environment. This includes the spectro-temporal as well as the spatial properties of the rendered signals and emphasizes the perspective to improve the comprehension of the target source by rendering a consistent user-centered auditory scene with a reduced amount of energetic and informational masking.

Graduiertenkolleg
Statistische Modellbildung

Matthias Mauch (University of London, UK)

Tuesday, Nov. 4, 11:30 h

Automatic chord identification

Context information is vital in musical listening. We present a chord transcription algorithm using dynamic Bayesian networks (DBN) that achieves context-awareness by simultaneously estimating beat position, key, bass note and the chord itself. The given data are audio waves of pop songs. Several pre-processing steps including automatic tuning and automatic beat-segmentation enhance the quality of the features used as observations in the model. Qualitative evaluation of chord transcriptions of selected songs demonstrates the method's capability of reliably estimating more diverse chord types than our (and other authors') previous methods have while maintaining low fragmentation.

Graduiertenkolleg
Statistische Modellbildung

Michael Oehler, **Christoph Reuter, Jobst Fricke** (University of Cologne, D)

Tuesday, Nov. 4, 9:40 h

Recognizing and synthesizing wind instrument timbres: the role of formants and micro-modulations

An important cue for the discrimination of wind instrument sounds are formant areas that result from (instrument-) typical excitation impulses. In this context the ratio of pulse width and cycle duration is of particular interest: The pulse forming principle, a rediscovered model for the sound generating process of wind instruments rests upon the idea, that wind instrument sounds can basically be traced back to its excitation impulses, which always act upon the same principles, independent of the fundamental. Within several perception experiments it could be shown that besides characteristic formant areas, micro-modulations and vibrato can have a major effect on timbre perception.

Poster Session

Monday, Nov. 3, 19:15 h

Bernd Bischl (Dortmund University of Technology, D): A comparison of two methods for F0 estimation

Timo Gerkmann, Colin Breithaupt, Rainer Martin (Ruhr Universität Bochum, D): A Novel A Priori SNR Estimation Approach Based on Selective Cepstro-Temporal Smoothing

Tamás Harczos (Fraunhofer IDMT, D): Comparative Evaluation of Successive Cochlear Modelling Stages

Lydia Hennig (Universität Hamburg, D)**:** Neural Processing of Timbre Variations

Sebastian Krey (Dortmund University of Technology, D): SVM based instrument classification

Matti Ryynänen (Tampere University of Technology, FIN): Karaoke application with real-time tuning of singing voice to automatically transcribed melodies

Julia Schiffner (Dortmund University of Technology, D): Using Local Methods for Register Classification by Timbre

Katrin Sommer, Uwe Ligges, Claus Weihs (Dortmund University of Technology, D): Analysis of polyphonic music time series

Christopher Raphael, Kyung Ae Lim (Indiana University, USA)

Tuesday, Nov. 4, 9:00 h

InTune: A Program for Visualizing Intonation

We present our research in developing "InTune" --- a program to help instrumentalists and singers in visualizing and improving intonation. The program improves upon traditional pitch tracking techniques by using score matching to establish a correspondence between a musical audio excerpt and a known score. Using this information the program presents three synchronized views of the audio data: an annotated musical score, a pitch-sensitive piano roll representation, and a highlighted spectrogram. The user can move easily between these three views while maintaining focus on the current note or measure. We will present the results of a user study designed to evaluate the response to this program by students in the Jacobs School of Music at Indiana University. The presentation will also contain a live demonstration.

Günther Rötter (Dortmund University of Technology, D)

Tuesday, Nov. 4, 14:20 h

Emotion and vocal expression

One purpose of the human voice's emotional expression is the communication between the mother and her newborn child. Studies have shown that the emotional expression in the interaction between mother and child does not depend on a sociocultural context. The melodical contour of specific emotions (for example: if the mother wants to initialise a dialogue, she uses ascending melodic contoures) are the same in German, American and Chinese societies. It seems to be something like a 'human hardware', which is supported by studies of both monkeys and humans with brain lesions. These studies have yielded evidence of 'brain programmes' that function to

initiate and organize 'pre-wired' vocal expressions. There is further evidence that both encoding and decoding of emotion through voice prosody is handled by the right hemisphere of the brain. It is a highly speculative and controversial issue, that the neural organization underlying the recognition of vocal expression of emotions might also serve musical emotions.

Graduiertenkolleg Statistische Modellbildung

Gerald Schuller (TU Ilmenau, D)

Monday, Nov. 3, 11:00 h

Perceptual Audio Coding with Low Delay

A low delay audio coding scheme for communications applications is proposed. Its compression ratio is comparable to current state-of-the-art audio coding schemes, but with a much lower delay. The sources of delay in conventional audio coding are the filters for the subband coding, and the block switching of the filter bank. The block switching leads to high peaks in bit-rate which necessitates a large bit rate buffer to smooth the bit rate for a transmission channel. To avoid or reduce these delays, we replace the subband coding by predictive coding, and the hard switching of the filter bank by soft switching of the predictors. The overall delay becomes 6 ms at 32 kHz sampling rate. A subjective listening test with bit-rates around 64 kb/s for mono signals shows that the new scheme has a comparable quality to a conventional audio coder.

Graduiertenkolleg
Statistische Modellbildung

Gero Szepannek (Dortmund University of Technology, D)

Tuesday, Nov. 4, 10:50 h

Phoneme Recognition in Popular Music

Solving the task of phoneme recognition in music sound files may help for several practical applications: it will enable lyrics-transcription or an alignment of already known lyrics for karaoke applications. Furthermore, music file classification as it is done by several query-by-humming systems may be enriched by automatically extracted text features.

For the problem of automatic phoneme recognition in popular music several standard as well as non-standard auditory model-based feature sets are extracted, compared, combined and investigated as well as several competitive classification algorithms.

Wolfgang Theimer (RIM, D)

Monday, Nov. 3, 16:50 h

Visualization of Personal Music Categories

Managing large music collections on mobile devices poses challenges due to limited display and input capabilities. A new method to represent multiple personal music categories is introduced which assigns each music track to user-defined categories with respective similarity degrees: Examples from each category are selected in order to train a category-specific classifier using musical features as input. The classifier then ranks each song in the music collection according to its perceptional similarity to the category examples. Each category is graphically represented by a center of gravity on the display. An individual song appears as one point at a position defined by its similarity ratings for all music categories. Users can select a song by touching the device screen at the respective position. They can define a playlist by drawing a trajectory on the screen or by circling a region on the display.
The concept of multi-dimensional music similarity is extended to other input than audio signal-based features. The presentation concludes with an outlook.

Igor Vatolkin (Dortmund University of Technology, D)

Tuesday, Nov. 4, 13:40 h

Emotion recognition

On the one side, certain emotions can be activated by hearing music. On the other side, people prefer to listen to certain music pieces fitting to their current mood. Emotion recognition may provide the intuitive navigation through music collections and help the better understanding of music. However, emotion perception differs from user to user. Besides that there exist many different emotion models and no clear definitions. This talk gives an overview about the current research state in emotion recognition and discusses the experiences provided by a master thesis supervised at the Chair for Algorithm Engineering of TU Dortmund.

Tuomas Virtanen (Tampere University of Technology, FIN)

Monday, Nov. 3, 11:40 h

Automatic transcription, separation, and alignment of vocals in polyphonic music

This presentation introduces signal processing methods which can be used to analyze vocals in polyphonic music. Musical vocal signals are mainly pitched, and therefore a suitable fundamental frequency estimation is used as the first processing step in many analysis systems. We present a pitch estimation algorithm based on automatic melody transcription system which is genre-independent and produces accurate pitch estimates. Good pitch estimates facilitate many approaches which can be used to separate the pitched vocals from the mixture signal. We review the basic principles behind sinusoidal modelling and binary masking approaches. We also discuss advanced spectrogram factorization methods which can be used to learn a model for the interfering sounds and to improve the quality of the separated vocals. We present how the separated vocals can be further analyzed by temporally aligning them with textual lyrics. The presented approach uses hidden Markov model based phonetic speech recognition system which is adapted to singing. Audio demonstrations of the presented methods are provided.

Graduiertenkolleg
Statistische Modellbildung

Stephan Werner (Fraunhofer IDMT, D)

Monday, Nov. 3, 14:20 h

Bio-inspired binaural sound source localization and separation for hearing aids

The talk deals with the localization and separation of sound sources. The aim is to localize a preferred source and separate the sound signal from several disturbing others. To do this, the direction of the wanted source has to be extracted. An improvement in speech intelligibility and a clear presentation of the azimuthal direction is intended. This is reached by emphasizing interaural disparities from one preferred direction by a so-called “correlation curtain”. The localization of individual sound sources is realized by extraction and analysis of interaural time differences in a binaural model of hearing. The basis is a bio-inspired high-resolution monaural ear model, which transforms the amplitude-time-representation of the sound waves into space-time-representations (auditory images) of several processing stages of the cochlea. This model is used in a binaural signal processing chain. The chain contains a cross-channel-correlation module, a source tracking module based on the magnitude of the correlation results of former time steps, and a simple method to enhance localization results of a preferred direction in the current time step. The preferred direction of the current time step is defined by the results of the tracking module. As input signals for the localization chain the basilar-membrane oscillation and the neurotransmitter concentration in the synaptic cleft of the inner hair cells are used. These two kinds of signals are compared with each other in terms of achieved accuracy. The extended system is verified by simulations and tests in a real environment. Improvement in speech intelligibility is checked by an automatic speech recognition system employing hidden Markov models.

List of participants

NAME	PAGES
Jacob Abesser	3, 5
Bernd Bischl	16
Holger Blume	3, 6
Martin Botteck	3, 6
Frank Desmet	4, 7
Jobst Fricke	4, 15
Timo Gerkmann	16
Tamas Harczos	16
Karim Helwani	
Lydia Hennig	16
Marcus Holmberg	2, 8
Berit Janssen	4, 9
Jonas Kaiser	
Thomas Kemp	3
Anssi Klapuri	1, 2
Armin Kohlrausch	2, 10
Sebastian Krey	16
Uwe Ligges	1, 4, 16
Kyung Ae Lim	4, 17
Geoff Luck	4, 11
Nilesh Madhu	2, 12
Rainer Martin	1, 2, 12, 13, 16
Matthias Mauch	4, 14
Michael Oehler	4, 15
Christopher Raphael	4, 17
Günther Rötter	4, 18
Günter Rudolph	3
Matti Ryynänen	16
Gerald Schuller	2, 19
Gero Szepannek	4, 20
Wolfgang Theimer	3, 6, 21
Igor Vatolkin	4, 22
Tuomas Virtanen	2, 4, 23
Claus Weihs	1, 2, 4, 16
Stephan Werner	2, 24

Graduiertenkolleg
Statistische Modellbildung