

Technische Wetenschappen

**PracticeSpace:  
Exploration and training of music performance skills by means of  
adaptive monitoring and visual feedback.**

Research Proposal for the Open Technology Program.

## **1 Applicants**

Dr. ir. P. Desain  
Nijmegen Institute for Cognition and Information (NICI)  
University of Nijmegen  
P.O.Box. 9104  
6500 HE Nijmegen  
tel: 024-361 5885  
fax: 024-361 6066  
email: [desain@nici.kun.nl](mailto:desain@nici.kun.nl)

Dr. H. Honing  
Music Department / Institute of Logic, Language, and Computation (ILLC)  
University of Amsterdam  
Spuistraat 134  
1012 VB Amsterdam  
tel: 020-525 4698  
tel: 020-525 4429  
email: [honing@hum.uva.nl](mailto:honing@hum.uva.nl)

## **2 Embedding**

The project will be realized within four institutes: the Nijmegen Institute for Cognition and Information (NICI), which will develop the signal processing and conduct the experiments, the University of Amsterdam, which will be responsible for the musicological and didactical aspects, the Conservatory of Amsterdam and the Royal Conservatory of The Hague, whose teachers will contribute the example performances, and whose students will participate in the experiments and evaluation of the software.

## **3 Other funding**

No other sources for financing this project are currently considered.

## **4 Keywords**

Machine listening, music education software, interactive performance systems, computational auditory scene analysis, dynamic Bayesian networks, machine Learning, online learning and adaptation, multimedia information retrieval.

## **5 Project summary**

### **5.1 Research**

Learning to perform music is about much more than just ‘playing the right notes’. After acquiring this basic skill, in many subsequent years of intensive teaching and practice the complex skill of musical expression and style is developed. These important aspects of music performance are not prescribed in the score and have to be decided on by the musician. But even more important than choosing a musical interpretation is the amount of motor control needed for achieving the chosen expressive timing, dynamics, intonation, articulation, etc. These motor skills are hard to learn. For example, it is not unusual for the timing of notes (in the contexts of a fluctuating tempo and expressive delays) to be accurate in the order of around a hundredth of a second. This kind of precision is needed in order to perceive a fluent performance.

Correct interpretation of music of a certain genre is learned usually by imitation of master musicians and by reinforcement (i.e., following a teachers critics) without explicit instruction about the physical parameters of the sound. Today, with the advances in computer technology, it is feasible to provide detailed real-time feedback about these physical parameters such as duration, pitch level or timing deviations. However in a self practice or teaching situation, such detailed information has limited value, first of all because it is not directly a part of the conceptual system of a musician, and, more importantly, because there are so many complex interactions between the parameters that they become meaningless in isolation.

In this project a system will be developed that monitors a student during practice and provides feedback on the success of imitation in an integrated visual way. It allows teachers to build libraries of exercises in the form of recorded audio fragments, examples and counter examples of a specific style or aspect (e.g. ‘swing’, agogic accent, vibrato) instead of explicit instructions. From the sound fragments (examples and counter examples) the features are deduced that are relevant to explore and learn, and these are extracted in real time while the student is practicing. Feedback is provided as a mapping from these relevant features to parameters of a visual form (shape, texture, orientation, color, position etc.). This helps in exploring through the space of possible performances. How the method helps in learning to control the instrument will be evaluated in several experiments. The best design choices will be put in a prototype for music students that (depending on available technology) may be accessible globally via the Internet or on computer game consoles.

### **5.2 Utilization**

A system aimed at learning the high-level skills of controlling a musical performance more quick and efficient than can be done with feedback from a teacher alone will be constructed. Teachers contribute their lessons in the form of performed examples and the success of the student’s imitation of the various features is displayed online as dynamic visual shapes. After the prototype is fine-tuned in a real educational environment (with conservatory students) and the courseware is established (by conservatory teachers), the application can be commercially exploited by educational publishers, either as software or via the internet.

Using a similar approach, characterizing proper patterns of stress and intonation for correct pronunciation of a foreign language is also possible. Such a education aid has a broader market and is of interest of parties that are involved in language education.

### 5.3 Dutch summary

Titel: PracticeSpace (OefenRuimte): Exploratie en training van muziekuitvoeringsvaardigheden door middel van adaptieve monitoring en visuele feedback.

Het leren uitvoeren van muziek gaat om veel meer dan alleen “het spelen van de juiste noten”. Na het aanleren van deze basisvaardigheid volgen vele jaren van intensieve training om meer complexe aspecten zoals expressie en stijl onder de knie te krijgen. Deze belangrijke aspecten van muziekuitvoeringen staan niet in de partituur voorgeschreven, maar moeten door de musicus zelf worden bepaald.

Nog belangrijker dan het kiezen van een muzikale interpretatie is de beheersing over de motoriek die nodig is om de gekozen expressieve timing, dynamiek, intonatie, articulatie, etc. te kunnen uitvoeren. Deze vaardigheden zijn moeilijk te leren. Het is bijvoorbeeld niet ongebruikelijk dat de timing van noten (in de context van een fluctuerend tempo en expressieve verschuivingen) precies is tot op de honderste seconde nauwkeurig. Zulke precisie is nodig om een vloeiende uitvoering waar te kunnen nemen.

De juiste interpretatie van muziek van een bepaald genre wordt meestal geleerd door imitatie van een expert musicus of door aanmoediging (bijv. het opvolgen van kritiek van de docent) zonder dat daarbij expliciete instructies worden gegeven over de fysieke parameters van het geluid. Vandaag de dag, met de huidige computertechnologie, is het mogelijk om gedetailleerde real-time feedback te geven over de fysieke parameters van de klank zoals duur, toonhoogte of timing verschillen. In een zelfstudie- of lessituatie is zulke gedetailleerde informatie echter niet erg waardevol, omdat deze parameters geen onderdeel uitmaken van het conceptuele systeem van een musicus en omdat er zoveel complexe interacties tussen de parameters plaatsvinden dat ze betekenisloos worden als ze geïsoleerd worden gepresenteerd.

In dit project wordt een systeem ontwikkeld dat een leerling monitort tijdens het oefenen en feedback geeft over het succes van imitaties op een geïntegreerde visuele manier. Het systeem stelt leraren in staat om oefeningen samen te stellen in de vorm van opgenomen geluidsfragmenten van voorbeelden en tegenvoorbeelden van een bepaalde stijl of een bepaald aspect (bijv. swing, agogies accent, vibrato). Specifieke instructies zijn dan niet nodig. Uit de geluidsbestanden worden de kenmerken die relevant zijn voor het exploreren en imiteren herleid. Deze worden in real-time geëxtraheerd terwijl de student aan het oefenen is. Feedback wordt gegeven door deze relevante kenmerken te vertalen naar parameters van een drie-dimensionaal object (vorm, textuur, orientatie, kleur, positie, etc.). Dit maakt het mogelijk om de ruimte van mogelijke uitvoeringen te exploreren. Hoe succesvol de methode is om te helpen bij het leren beheersen van het instrument wordt in een aantal experimenten geëvalueerd. De beste ontwerpkeuzes worden in een prototype software-applicatie voor muzikleerlingen verwerkt dat (afhankelijk van de beschikbare technologie) ook als internetapplicatie of voor spelletjesconsoles ontwikkeld kan worden.

## 6 Research Group

### 6.1 Current research group

Dr. P. Desain and Dr. H. Honing combined their background in computer science, psychology and systematic musicology to develop a multi-disciplinary approach to the computational modeling of rhythm perception and production in the PIONIER project ‘Music Mind, Machine’ (MMM). This project (which involved (between 5 and 10 fte) in now in its final year. Projects and publications can be found at <http://www.nici.kun.nl/mmm>.

Dr. P. Desain is affiliated with the NICI and is an expert in auditory perception and computational modeling. In addition, he applies modern signal processing techniques to the study of time perception using brain imaging methods. He will contribute to the experimental and modeling aspects of the project. Dr. H. Honing is affiliated with the Music Department, University of Amsterdam. His research is part of the ‘Cognitive systems and information processing’ program of the Institute of Logic, Language and Computation (ILLC). In addition, he is site-coordinator for the European MOSART network on music technology. He will contribute to the music performance and music technology aspects of the project.

### 6.2 Candidates

A potential candidate for the Post-doctoral fellow position is A. Taylan Cemgil, who has been working during his Ph.D. in SNN Nijmegen on music transcription and tempo tracking using dynamic Bayesian networks. He made several important contributions to the field (Cemgil et al., 2000a, Cemgil et al., 2001, Cemgil and Kappen, 2002a, Cemgil and Kappen, 2003), including a best paper award at the international computer music conference (Cemgil et al., 2000b) and development of real time software (Cemgil and Kappen, 2001). He is an expert in probabilistic modeling and fast stochastic and deterministic approximate inference techniques. Before joining our group, he has also conducted research in audio signal processing and recognition (Cemgil, 1995, Cemgil and Caglar, 1995, Cemgil and Gürgen, 1997, Cemgil and Erkut, 1997).

A potential candidate for the Junior Researcher position is Makiko Sadakata. She got her masters degree in Music Psychology at Kyoto City university of Arts in 2002. She already made several contributions to the field (Sadakata, Desain & Honing, 2002, Sadakata, Ohgushi & Desain, 2002) and is now working as a Junior Researcher at the Music Mind Machine project, where she is studying the relation between the perception and production of simple rhythmical patterns (Sadakata, Ohgushi & Desain, submitted).

A potential candidate for the Programmer / Music technologist position is Paul Trilsbeek. He has been working as a music technologist for the last 5 years in the Music Mind Machine project. He is an expert in music technology such as MIDI software and hardware, signal processing software and hardware and sound recording techniques. He developed some online web demos for some of the research conducted within the group and has programming skills in Lisp and Matlab. During the last three years, he also conducted research on tempo and timing analysis of piano performances (Trilsbeek, Desain & Honing, 2001; Trilsbeek, Desain & Honing, submitted).

## 7 Description of the project

### 7.1 Introduction

In the last decade music education has taken advantage of a variety of music software that stimulate music students in mastering their musical skills, especially in topics related to music theory (e.g., solfège and ear-training software). However, while the training of aural skills seems well supported by modern technology, computer-aided training of performance skills is still in its infancy. As an example, consider a drummer taking classes in jazz performance, trying to learn how to control his/her timing while playing a rhythmic pattern in various styles (e.g., with ‘swing’, ‘laid-back’, or ‘groovy’). This important aspect of musicianship can at the moment only be learned through feedback by an expert musician in a tutorial situation, and by a large amount of practice. Keeping up motivation during practice is quite difficult, especially when feedback from the tutor is only available to a limited extent. For instance, a music student might note that something is wrong with the timing, but does not succeed in finding a way to perform it as desired.

In a naive view, it would be just a technological problem to provide the student with auditory or visual feedback during or after a performance. Combining this with a game-like quality will even make the practice more involving. However, there is a fundamental problem with this approach that is the cause why such systems have never been realized successfully. This problem is the lack of understanding in how a musical rhythm interacts with the sense of timing and tempo. A certain musical character (say ‘laid-back’) cannot be represented as a timing profile or tempo curve independent of the rhythmic material, as done in most commercial sequencer software. We simply do not know, as yet, how rhythm, tempo, and timing interact. This lack of theory makes the practical issues involved in designing good computer aided teaching systems (as described above) rather difficult. We have been pointing to this perceptual and representational problem in much of our earlier research (Desain & Honing, 1993; 1994; Honing, 2001).

A way to address these issues directly might be to fully accept that the complex interactions between the many musical parameters cannot be fully understood and modeled. Moreover, this might be the only sensible approach if we want to have teachers and students interact with the software in a natural way, as the knowledge of music and playing style is never explicitly stated nor communicated in terms of physical parameters. The most natural form is for the teacher to demonstrate by playing and for the student to learn by imitating. Thus, the most knowledgeable approach one can take to the design of software for performance training is to support an architecture devoid of explicit knowledge. A system in which the teacher can create lessons by simply performing short fragments of varying character on his own instrument. Both examples and counter examples are collected. The system stores these as sound files and extracts the features that account for the differences. These features are extracted from the students’ performance, again on his own instrument, in real time and evaluated in the light of these target performances. Thus, only relevant features are extracted and displayed, providing feedback on the success of the student in controlling his playing style, to navigate the performance space, and approach the desired goals. Animated visual objects will be used to present the huge number of parameters as one integrated whole (in size, shape, color, transparency, texture, orientation, position). Here as well the mapping of parameters to visual dimensions is not intended to be simple and easy to grasp analytically. It is to be used solely as a visualization of the different target performances. The aim of this proposal is realization of working prototype of such a learning tool.

## 7.2 Performance analysis

Performance analysis refers to feature extraction from performed musical material where the focus is on characterization of aspects such as expressive timing, tempo fluctuations, articulation, intonation, etc. In this project, we aim at developing methodology for robust musical performance analysis. More concretely, we will design and implement the software for specialized inference methods for online sound analysis and recognition.

As the application domain, we will restrict ourselves to the field of music education where the task is to compare a “master” performance with a “student” performance and provide feedback about the important differences. We chose this relatively less explored area because of broad applicability and importance. Moreover, performance analysis provides a structured auditory scene analysis problem while being sufficiently deep to bring insight about the generality of the developed techniques. When people listen to sound, they are able to recognize the acoustic signal as symbolic events in the external world (Bregman, 1990, Scheirer, 2000, Ellis, 1996). Artificial intelligent devices and intelligent environments have not reached that level of distinction yet. Under fairly restricted conditions, certain listening tasks can be automated reasonably well, such as speaker localization or speech recognition, but music applications require different listening skills. Even persons without musical training are able to orient themselves in music and can make rapid judgments such as determining the style, performer, beat, complexity, emotional impact etc. There is currently no general theory of music perception that can explain this complex behavior, and we are still far from constructing robust listening devices with similar capabilities. In our view, one reason for this is the lack of appropriate techniques. Machine vision, robotics and speech recognition have always been central problems of AI, and due to decades of extensive research, sophisticated computational tools have been developed and operational systems are constructed. On the other hand, sophisticated statistical models and associated computational machinery developed in the construction of vision (Frey, 2000) or robot control systems (Fox et al., 1999) are not being widely applied in machine listening applications. On the other hand, only few researchers (e.g., Vercoe and Puckette, 1985, Grubb, 1998, Raphael, 2001b, Raphael, 2001a, Dannenberg and Mukaino, 1988) have taken an analogous approach in building intelligent systems with listening capabilities.

The hierarchical structure of musical sound (and the ease of collecting large amounts of data) makes audio a unique test-bed for mathematical models and computational methodology. Moreover, there are many interesting applications that would directly benefit from advances in this area, such as interactive music performance systems (Rowe, 1993) music information retrieval and content description of audio material (e.g. in the context of the MPEG-7 standard (Casey, 2001)), and tracking and characterization of stress and intonation patterns in natural speech.

## 7.3 The effect of real time feedback on the process of skill acquisition

The ability to monitor one’s own musical performance seems to be an important aspect in the process of musical skill acquisition (Drake & Palmer, 1997). It is well known that it is hard to listen to oneself critically while performing, using another modality for feedback may help. Off-line visual feedback has a noticeable effect on musical performance (Tucker, Bates, Frikberg, Howarth, Kennedy, Lamb, and Vaughan, 1977; Basmajian and Newton, 1974; Fourein & Abberton, 1971; Sunberg, 1977). Some empirical work also shows the influence of real-time visual feedback (Bresin & Juslin, 2002). This kind of feedback may assist the musicians’ self-monitoring and therefore facilitate the acquisition of musical skills. In this project, we will develop a computer-aided training tool, which can give objective visual feedback on the performance in real time, based on consistent criteria. Certain features will be extracted from the musical performance and mapped to certain parameters of a three-dimensional visual object.

Further research will be necessary to show for example how many visual cues musicians are able to process at the same time and how well they can reflect these cues in changes in their performance. Real-time visual feedback may also help music students in keeping up their motivation and enthusiasm during practice. Practice usually involves numerous repetitions of the same part of music and therefore often lacks the spontaneous and enjoyable aspects of musical performance (Ericsson et al., 1990). For many – especially beginning – students it is hard to tell the difference between those repetitions and to see a direct effect of their efforts. This computer-aided training tool can give them more information about their performances and indicate whether they are moving in the right direction.

## **7.4 Proposed research and development**

### **7.4.1 Analysis/Modeling**

The central challenges in musical performance analysis is the characterization of musical phrases that emerge from combination of discrete sound events (e.g. notes, glissandos, etc.) that occur at irregular time intervals and yet still have a local structure at the signal level. Such a hierarchical structure introduces very long range correlations that are practically impossible or at best very difficult to capture with traditional signal models such as hidden Markov models (HMM's) or Kalman filters models (KFM's).

A standard ad hoc approach for modeling such complex processes relies on implicit simplifications such as computing a set of relevant features on a sliding window (e.g. spectral cues, comodulation, cepstrum coefficients) and defining a process in some feature space rather than directly on the observed signal. However, the computation of many nontrivial features intrinsically requires the choice of an appropriate time scale and specification of several parameters. Since there are simply too many options, a priori identification of optimal features is usually a tedious task.

Many of these problems can be addressed by carefully defining a model that explicitly specifies the relationship between desired quantities, unknown parameters and observed quantities. In fact, such relationships can be easily specified, since we often have a qualitative understanding about the problem domain. In this modeling framework, the sound features or descriptors can be viewed as hidden variables that are observed not directly. The relation of descriptors to the actual observed signal is defined by a probabilistic observation model. Once the master performance is observed, one can infer a posterior distribution over the descriptors. This inferred posterior distribution represents our knowledge (and uncertainties) about the master performance. Based on this characterization, we can “rate” the student performance in terms of the likelihood under the master performance model.

Dynamic Bayesian networks (DBN's) (Kanazawa et al., 1995, Murphy, 2002) provide a very flexible and powerful formalism to represent highly structured stochastic processes such as musical performances. DBN's generalize many standard time series models such as ARMA models, hidden Markov models (Rabiner, 1989), linear dynamical systems (Kalman filter models) (Bar-Shalom and Li, 1993), independent component analysis (Hyvarinen et al., 2001), and many more. Moreover, online adaptation and parameter learning can be handled in the same framework.

DBN's are especially well suited to state-space modeling of sequential data. In a general state-space model, we assume an underlying hidden state of a system that gives rise to the actual observations. The hidden state vector can be interpreted as the collection of all unknown parameters and unobserved signal features. In many realistic applications, some prior knowledge about the state variables and parameters is present. DBN's allow us to specify arbitrary dependence relationships among all hidden variables and observations and in this respect generalize simple state space models, HMM's or KFM's.

## Example

In this section, we consider the following scenario for real-time performance analysis using a DBN: a teacher is playing a drum pattern on the drum kit that is recorded. The student tries to imitate the teachers interpretation and our task is to provide feedback to the student about his success. We assume that there is no conventional musical score available; we infer the representation solely from the master performance. Our solution strategy is to focus on the generative process underlying the musical performance and the physical properties of the instrument. This includes modeling both the instrument as well as the instrumentalist using a DBN. The problem can be conveniently described in a Bayesian framework: given the audio samples, we wish to infer the onset times, durations, tempo as well as the instrumentation individual events (e.g., cymbal, hi-hat, snare etc.).

In this setting, the drummer can be viewed as a fluctuating timer that creates “interrupts” at quasi-regular intervals obeying some probabilistic rule. The drum kit can be modeled as a dynamical system that creates sounds with certain spectral characteristics. At onset times, the drummer ejects energy into the system. The main components of a DBN that defines the generative model are shown in Figure 1. During the analysis stage, we infer the posterior distribution over the latent variables given the master performance. Although this step seems like a transcription step, we actually never generate an explicit notation. Instead, we obtain a probabilistic characterization of the master performance including uncertainties. Then, we use this inferred posterior distribution to compute the likelihood of the student performance. If the model can predict the student well, we can conclude that the performance is a good imitation. Otherwise, we can inform the student about points where the prediction is bad.

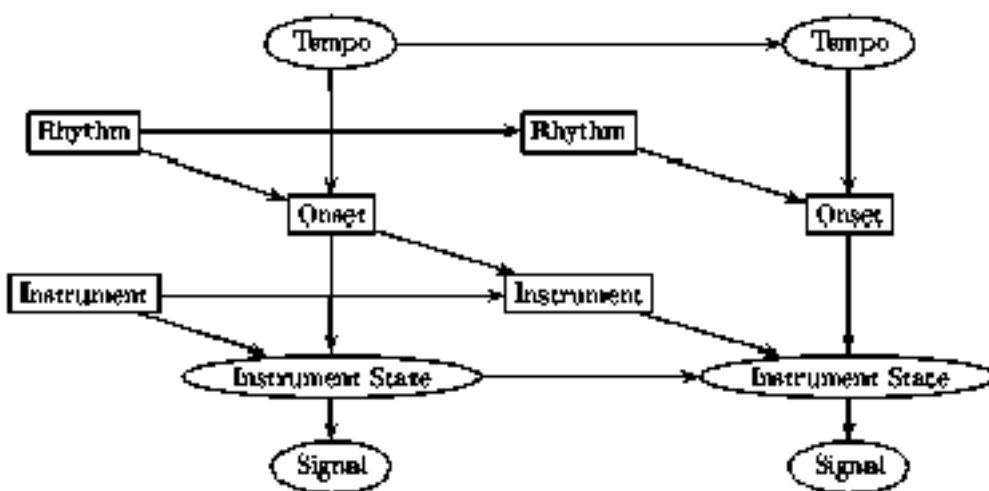


Figure 1: A (simplified) probabilistic generative model of an expressive performance on a drum kit with only one sound at a time. The timer model (drummer) determines the rhythm, instrumentation and the tempo. Given the tempo and the rhythm, we can determine at which instances an onset will occur. Each component (instrument) of the drum kit is modeled as dynamical system in state space form. The instrument indicator determines the transition model of the string state between consecutive time slices. When an onset is present, energy is inserted into the system, which changes the state vector abruptly. Otherwise, the instrument is in its normal mode of oscillation. The actual audio signal is a projection from the state vector. The task of performance analysis is, given the signal, to infer the hidden causes: the instrument label, rhythm and the tempo. For  $N$  note polyphony, we simply use  $N$  parallel chains of instruments and instrument states. During analysis, we infer a posterior distribution over the hidden states given the master performance. Then, we use this inferred posterior distribution to compute the likelihood of the student performance. If the model can predict the student well, we can conclude that the performance is a good imitation.

Although this model describes a specialized and hard real-time application for drum music, using the DBN formalism we can describe similar applications for other instruments. From a practical implementation point of view, the unifying framework of DBN's provides an environment to combine and test alternative models and inference algorithms and guides the development of reusable and efficient real-time program code.

Perhaps the most important aspect of our approach is that it conveniently blends a data driven approach with prior knowledge about musical structure. Since the exact characteristics of expressive timing, tempo deviations and instrument acoustics are not known, we consider these to be hidden variables that need to be learned from the prototypical target performances of a teacher. Given such a trained model, we can derive a distance-like measure that reveals to what extent a student approaches the prototypical target performances. Using such a measure, we can also provide a mapping to a space of suitably parametrized shapes to "visualize" the "abstract" performance characteristics.

## 7.4.2 Application

On the basis of the method described above, a software application will be developed which enables students to monitor their musical performances in real-time in a visual way. Relevant changes in the parameters extracted from the musical performances will be reflected in relevant changes in the properties of a three-dimensional object, such as position, size, shape, orientation, texture, color, etc. Target performances are entered by a teacher, or the user can input a performance by his or her favorite artist. These performances will be represented in the same visual way, and the presented three-dimensional objects converge once the input and target performances become more similar. A basic version of the application will be developed at the start of the project, this version will be used for the first experiments. During the course of the project, the results of the experiments will be used for further improvement and refinement of the application. Finally, the application may be implemented as an online web version such that it becomes accessible more easily to music students.

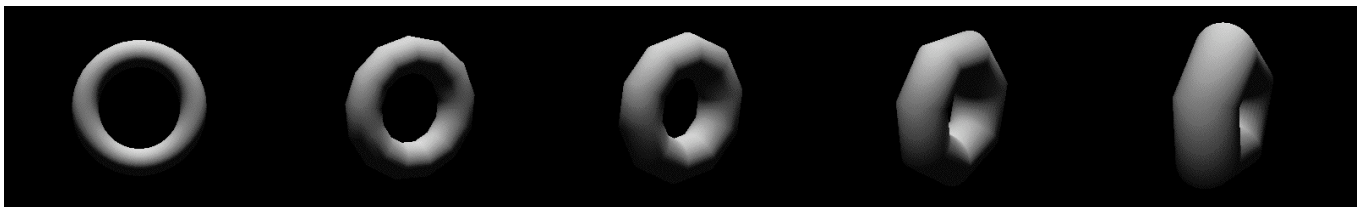


Figure 2: A possible visualization of several extracted performance features changing in time.

## 7.4.3 Experiments

The effect of real-time visual feedback on music performance will be investigated experimentally using musicians as participants. Through a series of experiments we aim at a) finding out the most effective visual presentation of the extracted features, b) investigating the processing of real-time visual feedback during performance and the difference between real-time and off-line feedback and c) investigating the cognitive capacities on processing real-time feedback of musicians of different skill levels.

In order to find an optimal visual presentation of the extracted performance features several steps need to be taken. First, a proper mapping needs to be found between the extracted performance parameters and the visual parameters, whereby the discrimination boundaries of both the performance and visual parameters need to be taken into account. Second, we will need to find out how visual parameters fit performance parameters best. One could imagine a combination of loudness and brightness of the sound be mapped to the distance of an object, tempo and timing fluctuations to the rotation speed, etc.

Besides a proper parameter mapping, analysis of the feedback data in combination with the performances will also give us some insight into human processing of real-time visual information during musical performance, for example if indeed many visual cues can be handled simultaneously in an integrated way and how fast musicians can reflect these cues into changes in their performance. The capacity of processing feedback during performance varies with the skill level of the performer (Palmer & Drake, 1997), therefore it is needed to conduct the experiments with musicians of various levels of skill. Towards the end of the project, experiments will be conducted to evaluate the usability of the software. Expertise of the teachers on evaluating the progress of the students will be needed here. Finally, interviews with users of the software will help to optimize the user interface and can learn us whether the use of the software has an effect on the musician's motivation during practice.

## **8 Personnel and equipment**

To be able to quickly design the complex signal analysis routines, a Postdoc, trained in machine learning and signal processing is needed. For the evaluation of the didactic qualities of aspects of the system a PhD project is foreseen. The realization and integration of the various components of the system can best be undertaken by a programmer/music technologist.

Supervisor Dr. ir. P. Desain, 4 years \* .1 fte (NICI)  
Supervisor Dr. H. Honing, 4 years \* .1 fte (UvA) (under consideration)  
Post-doctoral fellow, 4 years \* .8 fte (STW)  
Junior Researcher , 4 years \* .5 fte (STW)  
Programmer / Music technologist, 4 years \* .6 fte (STW)

Three high-end PC's with fast graphics and sound capabilities are needed. A laptop with fast graphics and sound capabilities is needed for conducting experiments on location (e.g. in conservatories, music schools, etc.). Some audio equipment for recording and playback will be purchased. Though the final input of the system will be audio, from any instrument, Midi equipment will make it possible to systematically generate approximate performances for testing during development of the analysis methods. Licenses for programming environments, audio processing and 3D visualization software and sound recording software are needed, insofar they are not already available at the participating institutes. For extensive experiments, student will have to be paid as subject. Teachers also need to be paid to participate in the experiments.

## **9 Time Schedules**

### **9.1 Time schedule Post-doc**

First year: Probabilistic models for audio analysis (particularly for percussion and singing voice),  
Second year: Development of efficient approximate inference algorithms  
Third year: Refinements to the model and extension to other instruments (and possibly speech)  
Fourth year: Implementation of a prototype

## 9.2 Time schedule Junior Researcher

First year: Protocol and collection of teaching example database  
Second year: Pilot experiment Real-time/Non Real-time comparison  
Third year: Evaluation experiments on mapping and success of method  
Fourth year: PhD thesis

## 9.3 Time schedule programmer / music technologist

First year: Visualization software, audio recordings, first prototype  
Second year: Real-time implementation of audio analysis  
Third year: Experiment software, data logging  
Fourth year: Documentation, final prototype, Web or game-console feasibility study

## 10 Available infrastructure

The research will be conducted mainly at the NICI and the UvA. At the NICI, a music recording studio and labs for conducting experiments are available. At the NICI, Dr. Desain will do the supervision, at the UvA this will be done by Dr. Honing.

## 11 Related research

Juslin et al. at Uppsala University in Sweden are working on a project called Feel-Me. The aim of this project is to develop an application for learning expressivity. They use a rule-based model for analyzing musical performances that are played with a certain emotional intention (sad, angry, happy, etc.) (Juslin et al., 2002). The output of this analysis (weights and parameter values of the rules) is compared to a listening panel model to determine the success of the communication of the particular emotion. If the communication is not successful, suggestions are given to the musician on how to improve. The performer also gets feedback about the performance via a 3D graphical representation of a selection of the parameters (Friberg et al., 2002).

Most MIDI sequencers provide a piano-roll visualization of piano performances. These show the pitch, onset time and the duration of performed notes. In separate axis the loudness of notes is displayed, or color-shading could indicate loudness of notes. Riley-Butler uses this visualization as a teaching method. She uses it as a feedback for piano students as well as a way to improve piano students' comprehension of dynamics and timing of artists' performances (Riley-Butler 2001; 2002). This visualization is complete and therefore redundant: it shows fixed elements of the score as well as expressive elements added by the performer.

A visualization of expression has also been developed by Goebel and colleagues (Dixon et al., 2002, Langner and Goebel, 2002). They represent the tempo and loudness of a performance and how it develops over time in a two-dimensional space. The axes represent tempo and loudness, while intensity of color represents time. The result is a worm-like representation of two expressive parameters: a line connects the points that follow each other in time, its color is light for past points and dark for recent points. Different performances show different trajectories of loudness and dynamics over time and these differences are reflected in differences in curvature of the worm. For two dimensions, the representation is intuitive and might be suitable for educative purposes as the authors mention in their discussion.

## **12 Utilization**

### **12.1 Practical use**

While our previous STW project extracted scores from the rich performance information, this one focuses on teaching to perform with this expressive richness not available in the score. Educational software companies and educational publishers will benefit from this new approach. We found a Dutch company (LCN) that can contribute its extensive experience in educational software via the user committee. The market for high-level training of musical skills may seem small at first sight, but a large community of youngsters spends much time in learning these skills, often trying to imitate their star performer. Computer game companies have recently been very successful with their musical toys and games for game halls in Japan. In these games, the performer is presented with clear targets (the notes to perform) that are shown in a graphical way. Points are given for each correct note played. Our proposal aims extends this approach to a higher level of musical skills like expression, style and mood, hopefully with the same sense of fun and motivation for the user.

A second area in which the methodology will be useful is the teaching of stress and intonation patters of a foreign language. These are notoriously hard to teach, maybe even because the perception of them has to be attained first. Online visual feedback may turnout to be an advantage, which would help a lot in the on-line teaching of foreign languages.

### **12.2 User committee**

#### **12.2.1 LCN (Drs. J. Ringelberg)**

Prins Bernhardstraat 7  
6521 AA Nijmegen  
Telephone : 024 3238130  
Fax : 024 3238074  
E-mail : Joop@lcn.nl

LCN Planning/Scheduling (1983-heden) is een software ontwikkelaar waar twintig mensen werken. LCN levert momenteel programma's aan organisaties in twee gebieden: onderwijs en zorg. We streven ernaar reguliere educatieve software te leveren, maar met een 'bite': extra inhoudelijke kwaliteit. Zo is ons programma 'Woordspel'(Malmberg) onderscheiden met een Maki prijs vanwege de combinatie van praktisch inzetbare software voor het spellingonderwijs met een verfijnde foutdetectie en gerichte feedback. LCN heeft programma's geleverd aan Malmberg, ThiemeMeulenhoff, Zwijsen en Cito. In de zorg leveren we internet applicaties die werkprocessen tussen geografisch verspreide professionals ondersteunen. Een bekend voorbeeld is ZorgDomein, waarmee ziekenhuizen hun aanbod afstemmen op de vraag van huisartsen. LCN leverde o.a. aan het AMC, Dijkzigt/ Sophia Kinderziekenhuis, en OLVG in Amsterdam.

LCN onderhoud goede contacten met de KUN om zich te verzekeren van instroom van personeel en om systematiek die zich in de academia heeft bewezen op te kunnen nemen in haar producten. Met NICI en het Expertisecentrum Nederlands werken we aan een doorlopend stageproject over spelling.

### **12.2.2 Royal Conservatory of The Hague (Dr. R. Timmers)**

Juliana van Stolberglaan 1  
2595 CA 's-Gravenhage  
Telephone: 070 3814251  
E-mail: renee74@xs4all.nl

The Royal Conservatory of The Hague will contribute their expertise in the teaching of expression in western classical music to steer the research of our project. Contact person is Dr. Timmers, who teaches music psychology at the institute of Sonology. She will help and organize the evaluation and fine-tuning of our method by students of the Conservatory.

### **12.2.3 Conservatory of Amsterdam**

Van Baerlestraat 27  
P.O. Box 78022  
1070 LP Amsterdam  
Telephone: 020 5277550  
Fax: 020 6761506  
E-mail: info@cva.ahk.nl

In the Conservatory of Amsterdam, the jazz department has been involved with the elaboration of extensive material for teaching the production of percussive expression. We will ask teachers to contribute example audio fragments. Furthermore, many students are eager to try to use new means of practicing the skills of timing in jazz. Their progress (with and without the help of PracticeSpace) can be easily monitored in the recurrent lessons.

## **12.3 Implementation**

A prototype software application will be developed using a combination of software packages. The analysis and modeling will be done using Matlab (<http://www.mathworks.com>). For the 3D visualization, initially a package called jMax ([www.ircam.fr/jmax](http://www.ircam.fr/jmax)) with DIPS extension (<http://www.dacreation.com/dips.html>) will be used since it allows for very fast development of real-time 3D graphics programs. In a later stage, this part can also be written in Matlab if this gives any performance advantages or when a Windows version is required. Using Matlab in combination with jMax, the prototype will be able to run on Linux and Mac OS X operating systems.

## **12.4 Past performance**

The applicants have collaborated in a previous STW-funded project on Automatic Music Transcription (NIFF 4494). This project will be finalized in 2003. The scientific impact of the project was considerable. A great number of publications were realized, one of the receiving a 'distinguished paper award' at the 2000 International Computer Music Conference (ICMC) in Berlin. The infrastructure and experience acquired in this project will be beneficial for the current application; both the music technologist and PhD-student involved in that project will contribute to the current proposed project.

## 13 Project budget

### 13.1 Personnel

Post-doctoral fellow, 4 years \* .8 fte

Junior Researcher , 4 years \* .5 fte

Programmer / Music technologist, 4 years \* .6 fte

...

## 14 Literature

### 14.1 Group publications

Cemgil, A., Desain, P., and Kappen, H. (2000a). "Rhythm quantization for transcription." *Computer Music Journal* 24:2:60-76.

Cemgil, A. and Erkut, C. (1997). "Calibration of physical models using artificial neural networks with application to plucked string instruments." *Proceedings of ISMA97, International Symposium on Musical Acoustics, Edinburgh UK.*

Cemgil, A. and Gürgen, F. (1997). "Classification of musical instrument sounds using artificial neural networks." *Proceedings of SIU97.*

Cemgil, A. and Kappen, H. J. (2001). "A dynamic belief network implementation for real-time music transcription." *Proceedings of the Belgian-Dutch Conference on Artificial Intelligence 2001, Amsterdam.*

Cemgil, A., Kappen, H. J., Desain, P., and Honing, H. (2000b). "On tempo tracking: Tempogram representation and kalman filtering." *Proceedings of the 2000 International Computer Music Conference, pages 352--355, Berlin.*

Cemgil, A. T. (1995). "Automated music transcription." Master's thesis, Bogazici University, Turkey.

Cemgil, A. T. and Caglar, H. Anarim, E. (1995). "Comparison of wavelet filters for pitch detection of monophonic music signals." *Proceedings of European Conference on Circuit Theory and Design, (ECCTD95).*

Cemgil, A. T. and Kappen, B. (2002a). "Tempo tracking and rhythm quantization by sequential monte carlo." In Dietterich, T. G., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Information Processing Systems 14*, Cambridge, MA. MIT Press.

Cemgil, A. T. and Kappen, H. J. (2003). "Monte Carlo methods for tempo tracking and rhythm quantization." *Journal of Artificial Intelligence Research* 18:45-81.

Cemgil, A. T., Kappen, H. J., Desain, P., and Honing, H. (2001). "On tempo tracking: Tempogram representation and Kalman filtering." *Journal of New Music Research*, 28:4:259--273.

Desain, P., & Honing, H. (1993). "Tempo curves considered harmful." In "Time in contemporary musical thought" J. D. Kramer (ed.), *Contemporary Music Review*. 7(2) 123-138.

Desain, P., & Honing, H. (1994). "Does expressive timing in music performance scale proportionally with tempo?" *Psychological Research*, 56, 285-292.

Desain, P., & Honing, H. (in press) "The formation of rhythmic categories and metric priming. Perception."

Honing, H. (2001) "From time to time: The representation of timing and tempo." *Computer Music Journal*, 35(3), 50-61.

Honing, H. (2002). "Structure and interpretation of rhythm and timing." *Tijdschrift voor Muziektheorie*, 7(3), 227-232.

Sadakata, M., Ohgushi, K. & Desain, P. (submitted). "A Cross-cultural Comparison Study of the Production of Simple Rhythmic patterns."

Sadakata, M., Desain, P. & Honing, H (2002). "The relation between rhythm perception and production: towards a Bayesian model." *Transaction of Technical Committee of Psychological and Physiological Acoustics, Acoustical Society of Japan*, Vol. 32, No.10, H-2002-92.

Sadakata, M., Ohgushi, K. & Desain, P. (2002). "A cross-cultural comparison study of the production of simple rhythmic patterns." *Proceedings of International Conference of Auditory Display 2002 Rencon Workshop*.

Trilsbeek, P., Desain, P. & Honing, H. (2001). "Spectral analysis of timing profiles of piano performances". *Proceedings of the 2001 International Computer Music Conference, Havana*, pages 286-289.

Trilsbeek, P., Desain, P. & Honing, H. (submitted). "How performance characteristics are revealed by spectral analysis of expressive timing."

## **14.2 Other references**

Bar-Shalom, Y. and Li, X.-R. (1993). "Estimation and Tracking: Principles, Techniques and Software." Artech House, Boston.

Basmajian, J. V., & Newton, W. J. (1974). "Feedback training of parts of buccinator muscle in man." *Psychophysiology*, 11, 92.

Bregman, A. (1990). "Auditory Scene Analysis." MIT Press.

Bresin, R., & Juslin, P. N. (2002). "Real-time visualization of musical expression." Abstract proposal accepted for 5th Triennial ESCOM Conference, Hanover, September 8-13, 2002.

Casey, M. A. (2001). "Mpeg-7 sound-recognition tools." In *IEEE Transactions on Circuits and Systems for Video Technology*, volume 11(6), pages 737--747.

- Clarke, E.F. (1999). "Rhythm and Timing in Music." In D. Deutsch (Ed.), *Psychology of Music*, 2<sup>nd</sup> Edition (pp. 473-500). New York: Academic Press.
- Dannenberg, R. B. and Mukaino, H. (1988). "New techniques for enhanced quality of computer accompaniment." *Proceedings of the International Computer Music Conference*, pages 243--248.
- Dixon, S. E., Goebel, W., & Widmer, G. (2002). "Real Time Tracking and Visualisation of Musical Expression" in: C. Anagnostopoulou & M. Ferrand & A. Smaill (Eds.), *Proceedings of the Second International Conference on Music and Artificial Intelligence (ICMAI'2002)*, Edinburgh. Berlin etc.: Springer, 58-68.
- Drake, C., & Palmer, C. (2000). "Skill acquisition in Music Performance: relation between planning and temporal control." *Cognition*, 74, 1-32.
- Ellis, D. P. W. (1996). "Prediction-Driven Computational Auditory Scene Analysis." PhD thesis, MIT, Dept. of Electrical Engineering and Computer Science, Cambridge MA.
- Ericsson, K. A., Tesch-Romer, C., & Krampe, R.T. (1990). "The role of practice and motivation in the acquisition of expert-level performance in real life." In M. J. A. Howe (Ed.), *Encouraging the development of exceptional skills and talents*. Leicester: British Psychological Society.
- Fouerein, A. J., & Abberton, E. (1971). "First applications of a new laryngograph." *Medical and Biological Illustration*, 21, 172.
- Fox, D., Burgard, W., and Thrun, S. (1999). "Markov localization for mobile robots in dynamic environments." *Journal of Artificial Intelligence Research (JAIR)*, 11.
- Frey, B. "Vision by inference and learning in graphical models." Tutorial at CVPR 2000, <http://www.psi.utoronto.ca/~frey/papers/cvpr00.ps.gz>.
- Friberg, A., Schoonderwaldt, E., Juslin, P., and Bresin, R. (2002). "Automatic Real-Time Extraction of Musical Expression." In *Proceedings of the 2002 International Computer Music Conference*, pages 365--367, Göteborg.
- Grubb, L. (1998). "A Probabilistic Method for Tracking a Vocalist." PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.
- Hyvarinen, A., Karhunen, J., and Oja, E. (2001). "Independent Component Analysis." John Wiley & Sons.
- Juslin, P. N., Friberg, A., & Bresin, R. (2002). "Toward a computational model of expression in music performance: The GERM model." *Musicae Scientiae, Special Issue 2001-2002*, 63-122.
- Kanazawa, K., Koller, D., and Russell, S. (1995). "Stochastic simulation algorithms for dynamic probabilistic networks." In *Proceedings of Uncertainty in AI*.
- Langner, J., & Goebel, W. (2002). "Representing expressive performance in tempo-loudness space." ESCOM 10th anniversary conference on Musical Creativity, April 5-8, 2002. Liège.

- Murphy, K. P. (2002). "Dynamic Bayesian Networks: Representation, Inference and Learning." PhD thesis, University of California, Berkeley.
- Palmer, C., & Drake, C. (1997). "Monitoring and Planning Capacities in the Acquisition of Music Performance Skills." *Canadian Journal of Experimental Psychology*, 51,
- Rabiner, L. R. (1989). "A tutorial in hidden Markov models and selected applications in speech recognition." *Proceedings of the IEEE*, 77(2):257--286.
- Raphael, C. (2001a). "A probabilistic expert system for automatic musical accompaniment." *Journal of Computational and Graphical Statistics*, 10(3):467--512.
- Raphael, C. (2001b). "Synthesizing musical accompaniments with Bayesian belief networks." *Journal of New Music Research*, 30(1):59--67.
- Riley-Butler, K. (2001). "Comparative performance analysis through feedback technology." *Proceedings of the 2001 Meeting of the Society for Music Perception & Cognition*, p. 27-28. Kingston Ontario, Canada.
- Riley, K. (2002). "Teaching expressivity: An aural/visual feedback/replication model." *Proceedings of the SRPMME conference Investigating Music Performance*, p. 37. London, England.
- Rowe, R. (1993). "Interactive Music Systems: Machine Listening and Composing." MIT Press.
- Scheirer, E. D. (2000). "Music-Listening Systems." PhD thesis, Massachusetts Institute of Technology.
- Sunberg, J. (1974). "Articulatory interpretation of the 'singing formant.'" *Journal of the Acoustical Society of America*, 55. 838-844.
- Tucker, W. H., Bates, R. H. T., Frykberg, S. D., Howarth, R. J., Kennedy, W. K., Lamb, M. R., & Vaughan, R.G. (1977). "An interactive aid for musicians." *International Journal of Man-Machine Studies*, 9, 653-651
- Vercoe, B. and Puckette, M. (1985). "The synthetic rehearsal: Training the synthetic performer." *Proceedings of the ICMC*, pages 275--278, San Francisco. International Computer Music Association.

## 15 List of Abbreviations

ARMA Autoregressive moving average  
 DBN Dynamic Bayesian network  
 HMM Hidden Markov model  
 KFM Kalman filter model  
 DSP Digital Signal Processor, Digital Signal Processing  
 MIDI Musical Instrument Digital Interface  
 MPEG Moving Pictures Experts Group  
 MPEG-7 an ISO/IEC standard developed by MPEG for Multimedia Content Description  
 PF Particle filtering

## SMC Sequential Monte Carlo