Vetenskapsrådet

# *Application*

2015-03-31

| 2015-03770 | Friberg, Anders | NT-2 |
|---|---|---|

## Information about applicant

**Name:** Anders Friberg

**Doctorial degree:** 1995-05-26

**Birthdate:** 19580609

**Academic title:** Docent

**Gender:** Male

**Employer:** Kungliga Tekniska högskolan

**Administrating organisation:** Kungliga Tekniska högskolan

**Project site:** TMH, Tal, musik och hörsel

## Information about application

**Call name:** Forskningsbidrag Stora utlysningen 2015 (Naturvetenskap och teknikvetenskap)

**Type of grant:** Projektbidrag

**Focus:** Fri

**Subject area:**

**Project title (english):** Recognition of genre and emotion in music using perceptual features

**Project start:** 2016-01-01

**Project end:** 2019-12-31

**Review panel applied for:** NT-2, NT-14

**Classification code:** 10299. Annan data- och informationsvetenskap

**Keywords:** perceptual features, music perception, machine learning, computational models

## Funds applied for

| Year: | 2016 | 2017 | 2018 | 2019 |
|---|---|---|---|---|
| Amount: | 1,559,000 | 1,670,000 | 1,684,000 | 1,424,000 |

**Descriptive data**

## Project info

### Project title (Swedish)*

Igenkänning av genre och emotion i musik med perceptuella särdrag

### Project title (English)*

Recognition of genre and emotion in music using perceptual features

### Abstract (English)*

Today, computers and the Internet are commonly used for all aspects of music culture from production to listening. This implies that new computer tools are needed for characterizing, and indexing music audio. This is the focus of the new research field Music Information Retrieval (MIR).

In our previous projects, a new approach was successfully tested using perceptually defined features for describing music in a four-layered approach with audio, low-level features, mid-level perceptual features, and semantic descriptions. Recently we have modelled some of these proposed perceptual features with high accuracy (typical explained variance 90%).

In this project, we will further expand the modelling of perceptual features introducing a new layer for object recognition. A major focus will be on the development of computational models for various subtasks such as source separation, instrument recognition, and dissonance/tonal tension. We will use signal-processing techniques in combination with machine learning methods building on the battery of tools that we developed in the previous project. The different perceptual models will then be used for predicting the high-level semantic descriptions emotion and genre. Finally, we will develop small proof-of-concept applications in order to disseminate and test the result. A successful result will bridge the semantic gap between earlier studies in music psychology and contemporary data-mining projects leading to new ways of understanding and modelling music audio.

### Popular scientific description (Swedish)*

Musikbranschen har på kort tid ställt om till digital hantering av musik för alla led i produktionskedjan från musiker till lyssnare. Denna omsvängning har fött nya behov av verktyg för att hantera musiken. Stora musikdatabaser är svåra att orientera sig i utan adekvata sökmöjligheter. Framför allt finns det ett behov av att förstå det musikaliska innehållet i audiofiler, så att de kan bli sökbara på samma sätt som man använder sig av Google för att söka i text. Dessa aspekter behandlas i det nya forskningsområdet Music Information Retrieval (MIR) som är mycket aktivt inom EU och övriga världen.

I tidigare projekt utvecklade vi en ny modell för datorbaserad analys av musikfiler genom att införa perceptuellt baserade ljudparametrar. Den resulterande analysen består av fyra nivåer: audio, audioparametrar (ljudnivå, onsets…), perceptuella parametrar (hastighet, energi, rytmisk komplexitet), semantisk beskrivning (känslouttryck, musikstil…). Vi utvecklade modeller för flera av dessa perceptuella parametrar med hög precision (typisk förklarad varians 90%). Det visade sig att metoden har potential att signifikant förbättra både analysmetodernas effektivitet och den teoretiska förståelsen för hur vi upplever musik.

I det här projektet kommer vi att expandera modelleringen av de perceptuella parametrarna genom att introducera en ny analysnivå av grundläggande objektperception. Fokus kommer att ligga på att utveckla modeller för olika deluppgifter som källseparering, instrumentigenkänning och dissonans/tonal spänning. Där kommer vi att använda audiobaserad signalprocessning i kombination med maskininlärningsmetoder genom att bygga vidare på de verktyg som har utvecklats hittills. De olika perceptuella modellerna kommer sen att användas för att predicera de semantiska beskrivningarna känslouttryck och genre. Slutligen kommer vi utveckla små prototypsystem för att demonstrera hela konceptet.

Vi tror att resultaten kommer att bidra till att överbrygga gapet mellan tidigare studier inom musikperception och nyare forskning inom MIR (ibland kallad ”the semantic gap”) och potentiellt leda till ett nytt sätt att modellera och förstå hur människor upplever musik.

---

## Project period

### Number of project years*

4

### Calculated project time*

2016-01-01 - 2019-12-31

## Classifications

Select a minimum of one and a maximum of three SCB-codes in order of priority.

Select the SCB-code in three levels and then click the lower plus-button to save your selection.

**SCB-codes***

1. Naturvetenskap > 102. Data- och informationsvetenskap (Datateknik) > 10299. Annan data- och informationsvetenskap

Enter a minimum of three, and up to five, short keywords that describe your project.

### Keyword 1*

perceptual features

### Keyword 2*

music perception

### Keyword 3*

machine learning

### Keyword 4

computational models

### Keyword 5

**Research plan**

## Ethical considerations

Specify any ethical issues that the project (or equivalent) raises, and describe how they will be addressed in your research. Also indicate the specific considerations that might be relevant to your application.

### Reporting of ethical considerations*

We will conduct listening experiments with people listening to music examples at normal music listening levels.
These listening experiments will all be submitted to the regional ethical review board (regionala etikprövningsnämnden) for approval before doing any experiment.
The following ethical guidelines by law and advice will be followed carefully :
Codex(2013). Codexrules and guidelines for research. http://www.codex.vr.se/en/index.shtml.
SFS (2003:460) Lag om etikprövning av forskning som avser människor (including the additional regulations).
Vetenskapsrådet (2011). God forskningssed. Vetenskapsrådets rapportserie 1:2011. Stockholm: Vetenskapsrådet.

### The project includes handling of personal data

Yes

### The project includes animal experiments

No

### Account of experiments on humans

No

---

## Research plan

# Recognition of genre and emotion in music using perceptual features

## 1 Introduction

This is a follow-up of the on-going project *Computational Modelling of Perceptual Music Features* (ends 2015, see separate report). It is a multi-disciplinary project combining research in music cognition, analysis, and psychology with music signal processing, computer science tools, and machine learning techniques.

In a previous project, we developed a new conceptual model analysing music recordings, mainly working with audio representation. The new approach was to include *perceptually determined mid-level features* (e.g. perceptual speed, rhythmic clarity, modality, energy) into the analysis process. The results show that this method works well in that these features are possible to determine with high accuracy by perceptual evaluation (listening tests), and that they predict higher-level semantic descriptions such as emotional expression substantially more accurately than current state-of-the-art methods. This was very promising results given that this approach had never been tested previously. To our knowledge this approach is still unique within the research community.

In the on-going project we have continued to develop computational models for our perceptually defined features. Even if we anticipated that it would be a complex task, it was necessary to also approach several classical problems in music audio content analysis. This went better than expected and resulted in the improvement of the state-of-the-art in two classic problems; tempo estimation and polyphonic transcription. For the tempo detection one of the key factors was the inclusion of the perceptual feature 'speed' (see separate report).

In the new project we will continue in the same promising direction and complete the modelling of the perceptual features as well as extending it with a more thorough analysis of tonal tension and complexity. We will also extend the analysis to several basic perceptual principles within audio content analysis. In our analysis we have repeatedly discovered that basic source perception plays an important part to accurately model the perceptual features. Thus, for a more thorough understanding of music content we need to also approach aspects such as source separation, predominant melody estimation and instrument recognition. By adding this layer of abstraction we will utilize representations of music that are also perceived by human listeners. This suggests that we will be able to analyse the music in a way similar to human perception.

## 2 Purpose and aims

Our general goal is both to contribute to the basic understanding of music perception and cognition and to target future applications in the Music Information Retrieval (MIR) field. The resulting knowledge and models can then be used for developing intelligent musical applications both for MIR, music production and music pedagogy.

Subgoals:

- To further develop and evaluate new computational features predicting previous perceptual ratings (e.g. rhythmic clarity, modality, energy).
- To extend the basic analysis and work on source separation, instrument recognition, rhythm recognition, predominant instrument/voice recognition (finding the melody in a complex mix).
- To work on tonal complexity relating to perceptual concepts like dissonance, tonal tension, and roughness.
- If the above mentioned goals are met to an extent that we have a complete system, we will also develop a proof-of-concept application in form of a visualization of analysed features and/or in form of a music browser.

The long-term goal is to be able to analyse music and audio in the similar way as has been obtained in image and video processing research, where advanced models of facial recognition, eye movements and body gestures are used in a variety of different applications. Thus, this project is a continuation of a new research direction providing the gap between perceptual studies and computational models, and it will offer new insights for understanding how we perceive music.

# 3 Survey of the field

## 3.1 Music Information Retrieval

The Internet is becoming the primary source of information in our society. However, since there is no comprehensive subject structure (such as in a library) most of the information would not be accessible without the advanced data-mining technology used in search engines such as Google (Brin & Page, 1998). This technology is largely text-based, thus, content of multimedia is not directly accessible. Music distribution has made its way to the Internet and large music databases have been assembled, triggering the need for content-based music search capabilities. This has spawned the introduction of the new research field Music Information Retrieval (MIR) (e.g. Ellis, 2006; Leman, 2002). A number of research questions have emerged, such as musical beat analysis, genre classification, main melody extraction, music search, feature extraction, audio analysis, timbre classification, and user behaviour. However, despite the relatively large effort in this area, there is less research relating the feature extraction to semantic descriptions such as the emotional character. Search for similar musical pieces in databases, *music search*, are popular research areas within MIR. However, independent evaluations have demonstrated that the hitherto developed models suffer from lack of cognitive and musicological analysis (Müllensiefen, 2004). Recently, a number of commercial applications have appeared that offer new ways of interacting with music databases. Most of these systems rely on text metadata to describe the music, rather than direct analysis of the music audio. This demonstrates the need for further research improving on existing models. It also indicates the need for characterizing music in terms of the listeners' impression, rather than focussing on the actual notes or the harmonic progression.

Music Information Retrieval is part of what is now called Sound and Music Computing (SMC)[1]. A good overview of this research field is the Roadmap (Bernadini et al., 2007) funded by the EU and elaborated by the S2S2 consortium consisting of 11 research groups in Europe, including our group[2]. In the Roadmap, a number of future challenges are identified. The current project will address especially the challenge identified as the semantic gap: **"**The Semantic Gap in SMC - the discrepancy between what can be recognised in music signals by current state–of–the–art methods and what human listeners associate with music - is the main obstacle on the way towards truly intelligent and useful musical companions. […] The bridging of the semantic gap will require a radical re–orientation (1) towards the integration of top–down modelling of (incomplete) musical knowledge and expectations, and (2) towards a widening of the notion of musical understanding."

## 3.2 Semantic descriptions and musical features

The *semantic level* of music representation is thought to be a high-level verbal description of the musical character (Camurri et al., 2005). Semantic descriptions can be in terms of emotion terms, genre specifications, rhythmic terms or other musical descriptions. Certainly, as shown in several studies, the most important aspect of music is the communication of emotional expression (Bresin & Friberg, 2011; Eerola et al. 2013; Friberg, 2008; Juslin & Laukka, 2004). Overwhelming evidence from these studies show that for basic emotions such as happy, sad or angry, there is a rather simple relationship between the emotional description and the lower level representation in terms of musical performance parameters (tempo, sound level or articulation). One mapping model is the so-called lens model, relating the performer to the listener via a set of cues (musical features) using multiple regression (Juslin, 2000). The lens model was used for making a computer program that teaches musicians how to express emotions. While this was initially considered a rather provocative idea, it turned out that the program was more efficient than a real teacher (Juslin et al., 2006). Our group participated in this project and developed the feature extraction system (Friberg et al., 2007).

An alternative way of analysing expressive features using semantic descriptions was addressed in the European project MEGA. In that context we built an expressive mapper, which in real-time predicts emotional expression from either a set of audio features or a set of body motion features (Friberg, 2004). This mapper extends the previous linear models and uses a combination of fuzzy

[1] See http://smcnetwork.org/

[2] S2S2 project. http://www.s2s2.org/

sets for the prediction. It has been used in several applications including the Groove Machine – a dancer controls musical expressivity, presented at Kulturhuset 2002 (Lindström et al. 2005); the Expressiball – a visualization of music performance (Friberg et al. 2002); and the collaborative computer game Ghost in the Cave (Rinman et al. 2004)[3] - a game using only non-verbal input in terms of either body motion or singing.

*Musical genre* can be regarded as a sematic description as well. Important aspects for genre classification can be divided into three groups: texture, pitch and rhythm (Scaringella et al. 2006). An important part for determining genre is the recognition of different instruments or instrumentations, for example a string quartet or a classic rock guitar setting (McKay & Fujinaga, 2005). Therefore, instrument recognition will be included in this project.

A majority of previous studies have used an *intermediate feature level* for analysing high-level music expression, more or less corresponding to the audio features level discussed below. A common approach has been to assemble a large collection of features and to then use various data selection methods such as multidimensional scaling, Bayesian regression or Support Vector Machines (SVM) to model higher-level descriptions (e.g. Leman et al., 2005). Recently, new machine learning methods commonly referred to as deep learning has made it possible to automatically obtain several layers of intermediate data representations (Lee et al., 2009). This improves the state-of-the-art but do not necessarily increase the knowledge of the underlying perceptual principles. Our method is contrasting and complementary with respect to deep learning since it is based on *perceptually validated* music features. This gives the advantage that potentially a smaller number of low-level acoustic features can be used for each perceptual feature. This would yield fewer problems with over-fitting of data and the possibility to use smaller databases.

In the on-going project we have built models of the perceptual features, and laid a foundation for futures studies of *basic source perception*. Our model of speed successfully captured about 90 % of the variance of listener ratings (Elowsson and Friberg, 2013a; Elowsson et al., 2013). The model uses spectral fluctuations, onset densities and high-level rhythm features, which are combined in a linear regression framework. Furthermore, we have built a model of dynamics that captures about 85 % of the variance in listener ratings (Elowsson & Friberg, 2015b). The model is based on a combination of spectral flux-features, computed from source-separated audio. These results are well above human level performance, quantified as the ability for one listener to predict the mean rating of all the other listeners. Furthermore, we are seeing high results for rhythmic clarity and rhythmic complexity, with models expected to be finished during 2016. We have also been able to show the validity of the perceptual features to higher-level descriptors by using perceptual speed as a mid-level representation when computing tempo (Elowsson & Friberg, 2013b; Elowsson & Friberg, 2015a).

### 3.3 Source Perception

To understand an acoustical scene, we infer the identity of the sounding sources. This enables us to distinguish between e.g. friend or foe, essential from an evolutionary point-of-view (Friberg, 2012; Fhurmann, 2012). We can recognize a friend's voice at varying acoustical settings (e.g. busy street, cocktail party) by isolating the frequency components that are responsible for his or hers specific timbre (Bregman, 1994). Source perception is facilitated by multi-layered auditory processing, where frequency components are combined into higher-level perceptual cues. We perceive the complete harmonic series of a struck tone, with frequencies that can cover the whole frequency spectrum, as one single pitch. This is important to our perception of music at the note level. The relative amplitudes of the different partials across time are essential to characterize the tones (Handel, 1995), and these parameters can be conceived as the timbre of that tone.

Previous models for the *identification of different musical instruments* have been using a variety of different features. For example, Eronen (2001) used low-level audio features such as mel-frequency cepstral coefficients (MFCCs) and amplitude envelopes. It is also possible to use multiple fundamental frequency estimation as a form of source separation (all harmonics of most pitched instruments can be derived from the fundamental frequency), from which instrument recognition

[3] See also http://www.speech.kth.se/music/projects/Ghostgame/

can be facilitated (Heittola et al., 2009). Instrument recognition is important for correctly identify semantic descriptors (McKay and Fujinaga, 2005), especially to identify the correct genre of a music excerpt. Furthermore instrument recognition can also aid in a top-down source perception scheme; by remembering the timbre of different sources (such as a musical instrument) it becomes much easier to identify that source in a complex setting.

### 3.4 Tonality perception

Tonality is another important aspect of music perception. This includes a number of different aspects of pitch-related perceptual phenomena, such as vertical dissonance (from simultaneous notes), or melodic dissonance with respect to the harmony (Friberg & Battel, 2002; Krumhansl, 1990). A common principle is that a sense of tension is generated when the notes are perceived as distant. For the case of simultaneous notes (chords), one simple acoustical reason is that dissonant intervals have partials adjacent to each other in frequency, resulting in so-called 'beating'.

Krumhansl (1990) investigated the relation between harmony and melody in a series of seminal experiments. The resulting *key-profiles* have extensively been used in different types of tonal models, such as chord recognition in music audio. Similar concepts were used in our early models of harmonic and melodic tension within music performance modelling (Friberg et al. 2006).

Low-level perceptual phenomena relating to this concept, such as roughness and sensory pleasantness have been investigated extensively within psychoacoustics (Fastl & Zwicker, 2007). Within a MIR context, models for these phenomena have been used as features for predicting higher-level semantic descriptions. However, the perceptual relevance for these models within a real musical context is still unknown. Thus, there is a gap between the low-level aspects found in psychoacoustics and a real musical situation.

For our purposes, *musical dissonance* seems to be the most relevant concept (in contrast to roughness or sensory dissonance). Musical dissonance is partly culturally determined and relates more to higher-level processing rather than e.g. roughness that might relate more to peripheral processing in the ear (Fishman et al., 2001). However, there have been few attempts to estimate musical dissonance in a systematic way using listening experiments. Often the purpose has been to compare with other similar concepts such as roughness (Vassilakis, 2001) or evaluate the use of EEG for estimating dissonance (Fishman et al., 2001). What is needed is a systematic estimation of musical dissonance using a large number of different tone and timbre combinations that can serve as a ground-truth for further modelling. Therefore, in an ongoing pilot study we estimate musical dissonance by having 32 listeners rate 92 two-note and three-note combinations using recordings of a real piano. Preliminary results show consistent data and that dissonance was easy for the listeners to rate, which is promising for future more extensive experiments. This also indicates that dissonance is an interesting candidate to add to our set of perceptual features.

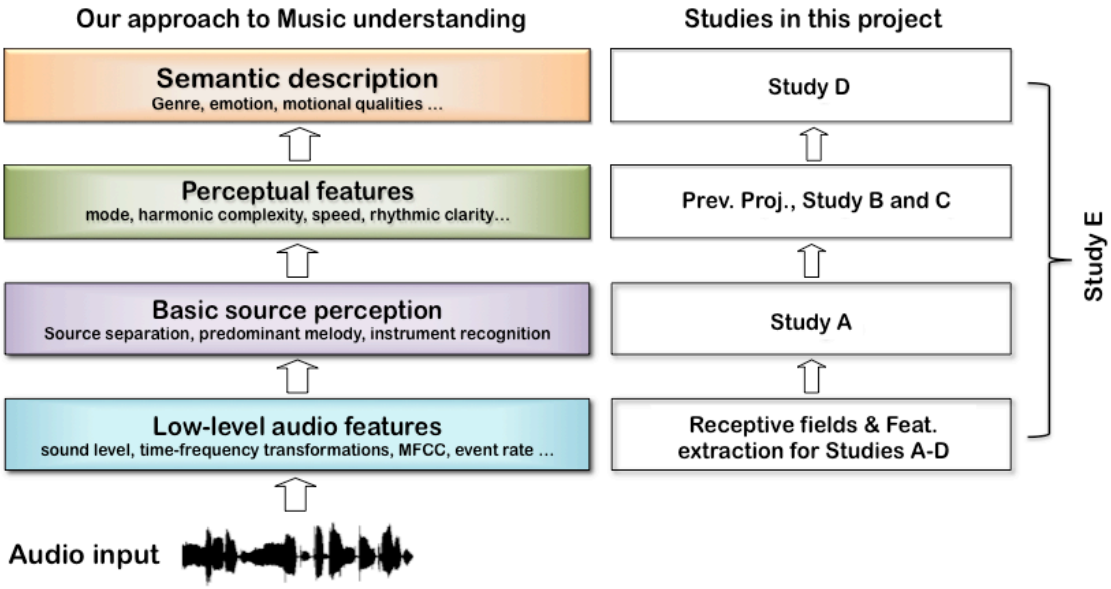
## 4 Project description

### 4.1 Theoretical framework

Previously, most studies for analysing for example emotion expression or genre have used a two-step procedure in which first computational features are extracted from audio and then the target high-level description is modelled using some selection of contemporary data mining methods such as SVM or similar. This corresponds to a three-layer framework for analysis that we used in previous studies (Friberg, 2004, Camurri et al. 2005). Starting with the previous project we introduced a middle layer of *perceptual features*. Thus, first low-level features are extracted from audio, second the perceptual features are modelled, and lastly, the semantic description is predicted. This was a new approach in that the perceptual features are modelled separately using a combination of low-level computational audio features optimized on perceptual data. This means that we move the main prediction effort from target prediction (semantic description) to a lower level.

In this project we go one step further and introduce the *basic source perception layer* (see figure below). This level emerged primarily from our previous computational models but is also motivated from a perceptual viewpoint. For example, in the models of speed and tempo it was necessary to separate the percussive content from the harmonic content (drums and rest), and analyse each part separately. This indicates that the perceptual process is following a similar two-

step process. This is also according to the ecological theoretical framework (see below). In this view, identification and characterization of the different sound sources are a primary perceptual goal, i.e. source identification.




The perceptual features were selected according to what has been identified as important for the musical character taken from three different fields: music theory, ecological perception, and emotion research. There is a large number of studies about the emotional expression which implies that the basic knowledge about which musical features that are important for the semantic description already exists and can be described in a qualitative way (for a summary see Gabrielsson and Lindström, 2001). In addition, we recently studied the quantitative relation of music performance parameters and emotional expression in several studies within the EU BrainTuning project (Bresin & Friberg, 2011; Eerola, Friberg & Bresin, 2013). Results in the previous project indicate that the prediction from perceptual features to semantic description is trivial given that the middle level is accurately modelled. For example, the explained variation on predicting emotional expression from perceptual features was about 90 % using standard multiple regression applied on the two data sets (Friberg et al., 2014).

A new finding in the previous project was the relevance of the ecological approach to music perception. Contrary to emotion research the ecological approach has been used in sound perception (Gaver, 1993), but not been applied to music perception (Friberg, 2012). Very briefly, the idea is that we try to decode the source properties of a sound (or the music) rather than just appreciate the sound itself. For example, in perceptual judgements ratings of the loudness of different voice sounds, it was the energy of the source (subglottal pressure and flow) and not the sound level in the room that explained the ratings the best (Ladfoged & McKinney, 1963). This approach presents a completely new exiting direction within the research field. Given that we can show this influence of the ecological approach in music listening some of the basic truths in music perception are actually challenged.

### 4.2 Methodology

The starting point will be our three databases (two collected previously and one being collected right now) in which perceptual data on both features and emotions are available. They will serve as the ground truth for modelling both audio and symbolically based perceptual features. In one of the databases we have access to both audio and symbolic representation making it suitable for comparing the different methods. In addition, we can use (and have been using) the available databases provided in the MIR community, such as the ones used in the MIREX competition for

evaluating and comparing MIR algorithms[4]. We are also providing the community with new data annotations of tempo in our most recent publication (Elowsson & Friberg, 2015a).

Our databases have been carefully selected in order to maximise the variance of different features across all examples. However, in this project we will also construct databases specifically for each study (A and B) in order to generate ground truth data for training the new layers of basic source perception. One possibility is to use a full-factorial design. In this way the relative importance of each feature can be estimated in a reliable way in listening experiments, and this was used in our study about emotional expression (Eerola, Friberg & Bresin, 2013).

We will use different methodological approaches following our previous developed methods for feature analysis of audio signals. We have now a large number of in-house developed features and analysis methods that will be serve as starting point for the development of new models.

One interesting approach that we will include in this project is the recently developed framework for *auditory receptive fields* (Lindeberg & Friberg, 2015a-b). It is modelling the different receptive field responses observed in neurons in the auditory and visual pathway. It is a general unified theory of perception that was derived from a set of invariance assumptions. Thus, various perceptual tasks can be modelled using a similar more coherent approach in comparison to current state-of the-art audio processing, which is of a more ad-hoc character. Currently, we are using the auditory receptive fields toolbox (ARF) for predicting articulatory parameters in unusual voice production within the EU-project Skat-VG[5]. In this project, we will use the ARF-toolbox as an alternative way of defining basic time-frequency analysis as well as to apply the receptive fields on the spectrogram in order to extract and enhance different aspects such as onsets, harmonic versus noise components, and spectral characterizations such as formants. We will do this work in collaboration with Prof. Tony Lindeberg, Computational Biology, KTH.

From the on-going project we have developed a general approach that we will continue to use. In the first stage, the time-frequency transformation, we will use a selection of transform methods including FFT, Constant-Q, and the receptive fields. Then for extracting different features a multi-layer approach will be used in which a higher-level abstractions (features) are developed in each layer. In same cases, machine learning will be used to optimize the specific features in that layer. In this respect it will be similar to the multi-layer approach used in so called deep learning (e.g. Lee et al., 2009). This approach was used in the tempo model (Elowsson & Friberg, 2015a). We will for the final step (from perceptual features to semantic description in the figure) test non-linear methods in accordance with the experimental results in e.g. Eerola et al. (2013).

It can be noted that we have reached well above human-level performance for the perceptual features that we have modelled. In fact we are in some cases close to the theoretical expected performance when modelling the mean of all human listeners. We have as a goal to reach the same performance for the rest of the perceptual features. To reach further performance gains, we need to extend the number of listeners (to reach an even more accurate mean rating) or extend the number of rated songs (to reach a larger sample to build the model on).

The project consists of the following studies A-E. The relation to the different layers is shown to the right in the figure above.

### 4.3 Study A. Modelling Source Perception

As outlined in Section 3.3, source perception is a fundamental aspect in music and a perceptual mechanism that possibly take place before further processing. Here we will work with different types of source recognition and separation. A simple but effective method is the separation of harmonic and percussive components, which we have been applied in all the previous models. By creating features from source separated audio it is possible to create more specialised representations that can capture a more specific perceptual phenomenon.

In addition we will work with instrument identification, e.g. identification of instrument families such as strings, woodwinds, electric guitar and percussion. The reason for modelling instrument recognition is that it adds a missing part for predicting genre in Study C.

[4] http://www.music-ir.org/mirex/2009/
[5] http://skatvg.iuav.it/

The singing voice is an essential element in most western music, so this will be a special focus both for source separation and recognition following our previous work on vocal identification in music (Elowsson et al., 2014).

We will collect ground-truth data for a large number of music excerpts specifically designed for each task. In this case we will e.g. make music examples in which different instruments are interchangeable so that different combinations can be analysed.

### 4.4 Study B. Estimation and modelling of tonality perception

As a perceptual features describing tonality we previously used 'harmonic complexity'. It received rather low consistency among the raters, which was not surprising given that this feature was of a more complex character. It actually assumed some music-theoretic knowledge contrarily to our overall approach of using ecologically valid features that are understood by lay people (i.e. non-experts). Our aim is to find new concepts that capture both (1) the perceptual tension of several simultaneous tones (chords), and (2) the tonal tension created with horizontal motion (harmonic and melodic progression) in a better more intuitive way.

In the first stage we will collect ground-truth data for a large number of music excerpts. These will be collected in listening experiments in the same manner as previously (regarding perceptual features). Thus, the listener will rate different aspects (e.g. dissonance) on unipolar continuous scales. From the results in the pilot experiment regarding dissonance, we conclude that it is relatively easy to get perceptual data for a large number of note combinations; it took only about 10 min to make ratings of 92 chord examples.

In the second stage we will develop different computational models predicting the perceptual judgements, starting from the audio representation of the excerpts. We will use similar methods as previously used for modelling perceptual features, thus, starting with different types of time-frequency representations and then extract various parameters.

In the evaluation, we will also compare with existing psychoacoustic models of e.g. roughness.

### 4.5 Study C. Refined modelling of perceptual features

Using the results in study A an B we will further refine and extend our battery of models of perceptual features. The modelling techniques will start from the methods and basic audio features that have been developed in previous projects. Thus, a set of low-level features will be combined using contemporary machine learning methods in a multi-layered approach (see Methodology section). Since the annotated properties are continuous we will use regression-type models rather than classification models. The evaluation is straightforward using the goodness of fit to the perceptual judgements with cross-validation within and between different data sets.

### 4.6 Study D. Semantic mapping to genre and emotion

In this part, the perceptual models developed in studies A-C (as well as from previous projects) will be used for estimating overall semantic descriptions of music. Given that the perceptual features are accurately modelled this will a rather simple task at least for modelling emotional expression. However, due to the extended available data sets we have also the opportunity to refine current state-of-the art models relating to the perceptual non-linearity of these judgements with respect to the selection of salient perceptual features. Drawing on the experience from the on-going project different mapping techniques will be considered, such as multiple regression, Bayesian modelling, or SVM. Optimisation and testing is performed using the databases with perceptual data. We will primarily use our own databases, which have annotations for emotion and genre. In addition we can use public databases with these annotations available within the MIR community.

### 4.7 Study E. Visualisation and prototype application (optional)

Finally, we will develop specific software primarily for visualizing and browsing music using perceptual features. These prototypes will be constructed primarily for checking the validity of the model and for the dissemination of the final result. One possibility is to make a music browser using the perceptual features in which the user can both select ranges of features as well as semantic descriptions such as emotional expression. Here it would be advantageous to use existing commercial databases such as Spotify or iTunes and make a plug-in structure within these environments. The realization of this step will depend on the results of the other sections, thus we

will do it only if we have a rather complete and accurate set of perceptual models. Also, it depends on the final budget.

### 4.8 Time plan

Study A, B will be the major part of the work that will be developed in parallel during the first three years. Starting in year two we work also in the Study C and D as soon as we have the necessary models from A and B ready. Study E will be in the fourth year. Each step will be disseminated in terms of research publications and collected data will be available on our website.

### 4.9 Dissemination of results

The results will be continuously disseminated in international publication and conferences. The project will also be presented on the web including demonstrations, and data download. There will be a continuous dissemination to technology and music students in the co-operation between KTH and the Royal College of Music, Stockholm (KMH). Such cooperation currently operates both in terms of research and by an exchange of teachers (e.g. Friberg teaches acoustics at KMH).

There are a number of local companies in the music area, and a number of commissioned master thesis projects for exchange of information and stimulation of business development are envisaged, for example a recommendation system study for Spotify (Bernhardsson, 2009).

In the previous project we collaborated with the company DoReMIR[6]. The company emerged from a previous research project supported by VR (Friberg, Dnr 2005–3696). We developed a new method for translating audio to symbolic representation of notes.

We have recently initiated a regular national conference in Sound and Music Computing, which had its first meeting in April 2011[7]. It gathered a large number of researchers and companies within the field and thus will provide a natural venue for presenting the new results.

## 5 Significance

A deeper understanding and modelling of music structure will provide a solid foundation for a number of applications in the area of Music Information Retrieval. It will bridge the gap between the previous focus on computer-oriented approaches and music theoretic/cognitive knowledge. In addition, the results should improve the scientific understanding of non-verbal communication in a general sense and contribute to the strong developing field of affective computing. The transfer of knowledge from musicians to computers is essential for preserving the previously largely intuitive understanding of expressiveness in the current society in which computers are used for music production. It is as appropriate as important that music culture be given the chance to benefit from todays and tomorrows technical development. Indeed, within the recent large structural changes in music production and music listening, there is a need for new computer tools in a variety of different applications. In particular the following applications can be mentioned.

- In consumer music databases a measure of music similarity can be used to suggest new music.
- A musical mood browser, for example as a tool in an Internet radio station, such as last.fm.
- Contribute to the development of tools for musical performance by developing representations of the perceived musical fundamental structure.

## 6 Preliminary results

The important findings related to the project are described in section 3 and in separate report.

## 7 Equipment

The principal site for the research will be the department of Speech, Music and Hearing, KTH, which will provide all basic resources including music studio facilities.

## 8 International collaboration

Within the ongoing project we have a collaboration with Prof. Stephen Downie and his group at the University of Illinois. For the work on tonal tension (Study B), we will continue a collaboration

[6] http://scorecleaner.com/
[7] http://smcsweden.se/

with Prof. Richard Parncutt and Erica Bisesi at the University of Graz. The research of our group at KTH has resulted in several invitations from leaders of European projects. We are currently collaborating within the EU-projects Skat-VG and Eunison. We have recently been participating in the European research projects SAME (Sound And Music For Everyone Everyday Everywhere Every Way), and the COST actions SID (Sonic Interaction Design), and AVFA (Advanced Voice Function Assessment). Previous European projects include MEGA (Multisensory Expressive Gesture Applications), SOb (Sounding object), AGNULA (A GNU Linux Audio distribution), IMUTUS (Interactive Music Tuition System), VEMUS (Virtual European Music School), BrainTuning (Tuning the Brain for Music), COST actions ConGAS (Gesture Controlled Audio Systems), and Coordination Action S2S^2 (Sound to Sense, Sense to Sound). The research group has been awarder a number of national projects, mainly funded by the Swedish Research Council (VR). Our research group participated in two EU projects concerning researcher mobility. We were a Marie Curie training site of excellence 2001-2005 and we participated in the project MOSART IHP Network (Music Orchestration Systems in Algorithmic Research and Technology). We have been continuously hosting about 5 visiting students or researchers annually over the last years. We also had a research exchange program with the University of York funded by STINT, 2002-2006. See also: http://www.speech.kth.se/smc/

## 9 References


Bernardini, N., Serra, X., Leman, M., Widmer, G., & De Poli. G., Eds. (2007). A Roadmap for Sound and Music Computing. Available at: http://smcnetwork.org/roadmap

Bernhardsson, E. (2009). *Implementing a scalable music recommender system*. Master thesis, KTH.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Bresin, R., & Friberg, A. (2011). Emotion rendering in music: Range and characteristic values of seven musical variables. *Cortex*, 47(9), 1068-1081.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer networks and ISDN systems*, *30*(1), 107-117.

Camurri A, De Poli G, Friberg A, Leman L & Volpe,G (2005). The MEGA project: analysis and synthesis of multisensory expressive gesture in performing art applications. *J. of New Music Research*, 34(1), 5-21.

Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music: contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology, 4*(487), 1-12.

Ellis, D.P.W. (2006). Extracting information from music audio. *Communications of the ACM*, 49(8), 32-37.

Elowsson, A., & Friberg, A. (2013a). Modelling Perception of Speed in Music Audio. *In Proceedings of the Sound and Music Computing Conference 2013*, SMC 2013, Stockholm, Sweden (pp. 735-741).

Elowsson, A., Friberg, A., Madison, G., & Paulin, J. (2013). Modelling the Speed of Music Using Features from Harmonic/Percussive Separated Audio. *In Proceedings of the International Symposium on Music Information Retrieval*. (pp. 481-486).

Elowsson, A,. & Friberg, A. (2013b). Tempo Estimation by Modelling Perceptual Speed. In MIREX Audio Tempo Estimation task 2013, 3 pages.

Elowsson, A., Schön, R., Höglund, M., Zea, E., & Friberg, A. (2014). Estimation of vocal duration in monaural mixtures. In *Proceedings of ICMC-SMC, 2014*, (pp. 1172-1177).

Elowsson, A. & Friberg, A. (2015a). Modeling the Perception of Tempo. *Submitted to Journal of Acoustic Society of America.*

Elowsson, A. & Friberg, A. (2015b). Modeling the Perception of Dynamics in Music. *Article in preparation.*

Eronen, A. (2001). Comparison of features for musical instrument recognition. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the* (pp. 19-22). IEEE.

Fastl, H., & Zwicker, E. (2007). *Psychoacoustics: Facts and models* (3rd ed.). Berlin: Springer.

Fuhrmann, F. (2012). *Automatic musical instrument recognition from polyphonic music audio signals*. Doctoral dissertation. Universitat Pompeu Fabra.

Fishman, Y. I., Volkov, I. O., Noh, M. D., Garell, P. C., Bakken, H., Arezzo, J. C., Howard, M.A., & Steinschneider, M. (2001). Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. *Journal of Neurophysiology*, *86*(6), 2761-2788.

Friberg A (2004) A fuzzy analyzer of emotional expression in music performance and body motion. In *Proceedings of Music and Music Science*, Stockholm 2004.

Friberg, A. (2008). Digital audio emotions - An overview of computer analysis and synthesis of emotions in music. *In Proc. of the 11th Int. Conference on Digital Audio Effects (DAFx-08), Espoo, Finland* (pp. 1-6). (Invited keynote).

Friberg, A. (2012). Music Listening from an Ecological Perspective. Poster presented at the *International Conference on Music Perception and Cognition (ICMPC) and ESCOM.*

Friberg, A., & Battel, G., U. (2002). Structural Communication. In (R. Parncutt & G. E. McPherson, Eds.) *The Science and Psychology of Music Performance: Creative Strategies for Teaching and Learning.* (pp. 199-218) New York: Oxford University Press.

Friberg, A., Bresin, R., & Sundberg, J. (2006). Overview of the KTH rule system for musical performance. *Advances in Cognitive Psychology, Special Issue on Music Performance*, 2(2-3), 145-161.

Friberg, A., Schoonderwaldt, E., & Juslin, P. N. (2007). CUEX: An algorithm for extracting expressive tone variables from audio recordings. *Acoustica united with Acta Acoustica*, 93, 411-420.

Friberg A, Schoonderwaldt E, Juslin, PN & Bresin R (2002) Automatic Real-Time Extraction of Musical Expression. In *Proceedings of the International Computer Music Conference 2002* (pp. 365-367).

Friberg, A., Schoonderwaldt, E., Hedblad, A., Fabiani, M., & Elowsson, A. (2014). Using listener-based perceptual features as intermediate representations in music information retrieval. *Journal of the Acoustical Society of America, 136*(4), 1951-1963.

Gabrielsson A, & Lindström E (2001) The influence of musical structure on emotional expression. In P. N. Juslin, & J. A. Sloboda (eds.), *Music and emotion: Theory and Research* (pp. 223-248). OUP press.

Gaver, W.W. (1993). How Do We Hear in the World?: Explorations in Ecological Acoustics. *Ecological Psychology*, 5(4), 285-313.

Handel, S. (1995). Timbre perception and auditory object identification. *Hearing*, 425-461.

Heittola, T., Klapuri, A., & Virtanen, T. (2009). Musical Instrument Recognition in Polyphonic Audio Using Source-Filter Model for Sound Separation. In *ISMIR* (pp. 327-332).

Juslin, P.N. & Laukka, J. (2004). Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening. *Journal of New Music Research*, 33(3), 217-38.

Juslin, P. N., Karlsson, J., Lindström, E., Friberg, A., & Schoonderwaldt, E. (2006). Play it again with a feeling: Feedback-learning of musical expressivity. *J. of Experimental Psychology: Applied*, 12(2), 79-95.

Juslin, P.N (2000) Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception ad Performance,* 26, 1797-813.

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch,* Oxford University Press.

Ladefoged, P., & McKinney, N. P. (1963). Loudness, Sound Pressure, and Subglottal Pressure in Speech. *Journal of the Acoustic Society of America* 35(4), 454-460.

Lee, H., Pham, P., Largman, Y., & Ng, A. Y. (2009). Unsupervised feature learning for audio classification using convolutional deep belief networks. In *Adv. in neural information proc. systems* (pp. 1096-1104).

Leman, M. (2002) Musical Audio Mining, in: J. Meij (Ed.), *Dealing with the Data Flood: Mining data, text and multimedia*, Rotterdam: STT Netherlands Study Centre for Technology Trends.

Leman, M., Vermeulen, V., De Voogdt, L. & Moelants, D. (2005). Prediction of Musical Affect Attribution Using a Combination of Structural Cues Extracted From Musical Audio. *Journal of New Music Research*, 34(1).

Lindeberg, T., & Friberg, A. (2015a). Idealized computational models for auditory receptive fields. *PLOS ONE* (in press).

Lindeberg, T., & Friberg, A. (2015b). Scale-space for auditory signals. Paper accepted to the conference *Scale Space and Variational Methods in Computer Vision*.

Lindström, E., Camurri, A., Friberg, A., Volpe G. and Rinman, M.-L. (2005). Affect, attitude and evaluation of multi-sensory performances. *Journal of New Music Research*. in press.

McKay, C., & Fujinaga, I. (2005). Automatic music classification and the importance of instrument identification. *Proceedings of the Conference on Interdisciplinary Musicology*.

Müllensiefen, D. (2004). Variabilität och Konstanz von Melodien in der Erinnerung. Ein Beitrag zur musikpsychologischen Gedächtnisforschung.

Rinman M-L, Friberg A, Bendiksen B, Cirotteau D, Dahl S, Kjellmo I, Mazzarino B, & Camurri A (2004) Ghost in the Cave - an interactive collaborative game using non-verbal communication. In *Gesture-based Communication in Human-Computer Interaction, LNAI 2915*, 549–556, Springer Verlag.

Scaringella, N., Zoia, G., & Mlynek, D. (2006). Automatic genre classification of music content: a survey. *Signal Processing Magazine, IEEE*, *23*(2), 133-141.

Vassilakis, P. (2001). Auditory roughness estimation of complex spectra—Roughness degrees and dissonance ratings of harmonic intervals revisited. *The Journal of the Acoustical Society of America*, *110*(5), 2755-2755.

## Interdisciplinarity

### My application is interdisciplinary

☐

An interdisciplinary research project is defined in this call for proposals as a project that can not be completed without knowledge, methods, terminology, data and researchers from more than one of the Swedish Research Councils subject areas; Medicine and health, Natural and engineering sciences, Humanities and social sciences and Educational sciences. If your research project is interdisciplinary according to this definition, you indicate and explain this here.

Click here for more information

---

## Scientific report

### Scientific report/Account for scientific activities of previous project

**Current status for the project Computational Modelling of Perceptual Music Features, Dnr 2012-4685, 2013-2015**

The project is in its last year and is in an intensive phase in which many new results has been found that is currently documented in papers. Due to a co-funding by KTH we were in this project able to hire a new PhD student Anders Elowsson that has been working in the project together with the project leader Anders Friberg. Elowsson has, in a very short time, been able to develop models including well-known tasks in the field of music information retrieval that improve the current state-of-the-art.

The current project has been developed with a few deviations according to the research plan but has also indicated new directions of research that was found to be necessary in order to reach the overall goal of the project. The main change is that we needed to go more in depth and widen the scope of the analysis to get a more thorough understanding of the different underlying perceptual concepts. For example, in order to model speed (see below) it was necessary to also go into the perception of tempo as well as to work with source separation of audio.

The focus in this project is the computational modelling of the previously suggested perceptual features. These features (currently nine, e.g. 'speed' and 'rhythmic complexity') are defined as a middle layer of music perception (or proximal cues) and have been estimated in two different databases of music examples using listening tests. This new conceptual model of music perception was developed within the previous VR project, 2009-4285.

The results of the first part of the overall project was analysed and summarized in a recent paper (Friberg et al. 2014) where we tested the validity of the new conceptual method and did preliminary audio and MIDI (symbolic music data) models of the rated perceptual features. The result was very positive in terms of model and approach, but less so in predicting the perceptual features from audio. This publication generated considerable media attention due it its new concept of perceptual features and was reported in several media such as Vice Television, Hungarian edition of the National Geographic, Swedish radio (P1 and P3) and French radio.

The results indicated the need to develop our own audio analysis to better get from audio to perceptual features. Our modelling started with the rhythmic aspects and the first perceptual feature was the perceived *speed* of the music (Elowsson & Friberg 2013; Elowsson et al. 2013). In short, the algorithm consists of source separation (percussive and harmonic part), clustering of percussive sounds, measures of spectral fluctuations, inter-onset durations and onset densities, and a final estimation of speed using machine-learning. It resulted in more than 90 % of explained variation, using cross-validation and tested on different datasets in collaboration with Guy Madison and Johan Paulin, Umeå University. Interestingly, this exceeded to a large extent the performance of individual subjects.

A model of tempo was developed using similar audio features and approaches as used in the speed model (Elowsson & Friberg 2015a). One of the key features was to use the model-estimated speed in order to reduce the previous well-know problem of 'octave error' that has been the major obstacle for obtaining good results in previous work. The resulting tempo model was submitted to MIREX competition within the ISMIR (International society of music information retrieval) conference. The tempo model obtained substantially better results than any other algorithm, also in the past (p-score = 0.86)[1]. The development of rhythmic features will be continued in a similar way in the last year including also the modelling of beat detection, rhythmic clarity and rhythmic complexity.

The basis for analysis of tonal features is the recognition of the played fundamental pitches (notes). Since each note is represented by a series of overtones across the spectrum and several notes are added together this is a non-trivial classic problem of audio analysis, which is very difficult also for trained musicians. A new model for such polyphonic pitch extraction was developed within the project. It was submitted to MIREX 2014 and got the highest overall result of all systems, also in the past[2]. The method is still not published and is currently considered for a patent application. Anders Elowsson, who developed the model, recently started a company to commercialize it.

In parallel (co-financed by a grant from KTH) we started to work on a new way of analysing audio using the previously developed theory for visual perception of receptive fields by Tony Lindeberg. The theory was adapted to audio and a computational toolbox was developed (Lindeberg & Friberg 2015a,b). This constitutes a completely new way of analysing audio from a theoretical and perceptual viewpoint. Our plan is to incorporate this in the new project (see application).

Currently the validity and the selection of perceptual features are further explored within a collaboration with Prof. Stephen Downie and coworkers at the University of Illinois. It is one of the leading groups in music information retrieval that initiated both the ISMIR conference and the MIREX competition. The purpose is to check the cross-cultural validity using *k-pop* music (South Korean pop). It will also extend the available databases for optimization and validation of the models. It is co-founded within a collaboration grant between KTH and Illinois. Ragnar Schön is working on this study in his master thesis and has been visiting University of Illinois in the spring.

Right now we are making a model of *dynamics* exploring different groups of features related to different perceptual concepts. Preliminary results indicate a fit better than 85% on rated data (Elowsson & Friberg, in prep.). More perceptual features will be modelled in the last year. We also tested alternative ways of collecting perceptual data within a distributed game context (Bellec et al. 2103).

To summarize,

- A substantial progress was made towards the modelling of perceptual features.

- A firm foundation of low level features and new analysis methods has been developed that will facilitate the final modelling in the last year.

- We were also able to improve the state-of-the-art in two classic MIR tasks, estimation of tempo and polyphonic pitch.

**Publications**

Bellec, G, Elowsson, A, Friberg, A, Wolff, D, & Weyde, T (2013). A social network integrated game experiment to relate tapping to speed perception and explore rhythm reproduction. In *Proc. of SMC* (pp. 19-26).

Elowsson, A, & Friberg, A (2015a, in review). Modeling the Perception of Tempo. *J. of the Acoustical Society of America.*

Elowsson, A, & Friberg, A (2013). Modelling Perception of Speed in Music Audio. In *Proc. of SMC* (pp. 735-741).

Elowsson, A, Friberg, A, Madison, G, & Paulin, J (2013). Modelling the Speed of Music Using Features from Harmonic/Percussive Separated Audio. In *Proc. of ISMIR.*

Elowsson, A, Schön, R, Höglund, M, Zea, E, & Friberg, A (2014). Estimation of vocal duration in monaural mixtures. In *Proc. of, ICMC and SMC 2014* (pp. 1172-1177).

Friberg, A, Schoonderwaldt, E, Hedblad, A, Fabiani, M, & Elowsson, A (2014). Using listener-based perceptual features as intermediate representations in music information retrieval. *J. of the Acoustical Society of America, 136*(4), 1951-1963.

Lindeberg, T, & Friberg, A (2015a). Idealized computational models for auditory receptive fields. *PLOS ONE* (in press).

Lindeberg, T, & Friberg, A (2015b). Scale-space for auditory signals. Paper accepted to the conf. *Scale Space and Variational Methods in Computer Vision*.

[1] http://nema.lis.illinois.edu/nema_out/mirex2013/results/ate/

[2] http://www.music-ir.org/mirex/wiki/2014:Multiple_Fundamental_Frequency_Estimation_%26_Tracking_Results

## Budget and research resources

### Project staff

Describe the staff that will be working in the project and the salary that is applied for in the project budget. Enter the full amount, not in thousands SEK.

Participating researchers that accept an invitation to participate in the application will be displayed automatically under Dedicated time for this project. Note that it will take a few minutes before the information is updated, and that it might be necessary for the project leader to close and reopen the form.

#### Dedicated time for this project

| | Role in the project | Name | Percent of full time |
|---|---|---|---|
| 1 | Applicant | Anders Friberg | 30 |

#### Salaries including social fees

| | Role in the project | Name | Percent of salary | 2016 | 2017 | 2018 | 2019 | Total |
|---|---|---|---|---|---|---|---|---|
| 1 | Applicant | Anders Friberg | 30 | 249,000 | 256,000 | 262,000 | 268,000 | 1,035,000 |
| 2 | Other personnel without doctoral degree | Anders Elowsson | 28 | 238,000 | 262,000 | 216,000 | | 716,000 |
| 3 | Other personnel without doctoral degree | Ragnar Schön | 80 | 437,000 | 480,000 | 529,000 | 581,000 | 2,027,000 |
| | Total | | | 924,000 | 998,000 | 1,007,000 | 849,000 | 3,778,000 |

### Other costs

Describe the other project costs for which you apply from the Swedish Research Council. Enter the full amount, not in thousands SEK.

### Premises

| | Type of premises | 2016 | 2017 | 2018 | 2019 | Total |
|---|---|---|---|---|---|---|
| 1 | Offices | 110,000 | 119,000 | 120,000 | 101,000 | 450,000 |
| | Total | 110,000 | 119,000 | 120,000 | 101,000 | 450,000 |

### Running Costs

| | Running Cost | Description | 2016 | 2017 | 2018 | 2019 | Total |
|---|---|---|---|---|---|---|---|
| 1 | Travels | Conferences | 30,000 | 30,000 | 30,000 | 30,000 | 120,000 |
| 2 | Equipment | | 20,000 | 10,000 | 10,000 | 10,000 | 50,000 |
| | Total | | 50,000 | 40,000 | 40,000 | 40,000 | 170,000 |

### Depreciation costs

| Depreciation cost | Description | 2016 | 2017 | 2018 | 2019 |
|---|---|---|---|---|---|

## Total project cost

Below you can see a summary of the costs in your budget, which are the costs that you apply for from the Swedish Research Council. Indirect costs are entered separately into the table.

Under Other costs you can enter which costs, aside from the ones you apply for from the Swedish Research Council, that the project includes. Add the full amounts, not in thousands of SEK.

The subtotal plus indirect costs are the total per year that you apply for.

### Total budget

| Specified costs | 2016 | 2017 | 2018 | 2019 | Total, applied | Other costs | Total cost |
|---|---|---|---|---|---|---|---|
| Salaries including social fees | 924,000 | 998,000 | 1,007,000 | 849,000 | 3,778,000 | | 3,778,000 |
| Running costs | 50,000 | 40,000 | 40,000 | 40,000 | 170,000 | | 170,000 |
| Depreciation costs | | | | | 0 | | 0 |
| Premises | 110,000 | 119,000 | 120,000 | 101,000 | 450,000 | | 450,000 |
| Subtotal | 1,084,000 | 1,157,000 | 1,167,000 | 990,000 | 4,398,000 | 0 | 4,398,000 |
| Indirect costs | 475,000 | 513,000 | 517,000 | 434,000 | 1,939,000 | | 1,939,000 |
| Total project cost | 1,559,000 | 1,670,000 | 1,684,000 | 1,424,000 | 6,337,000 | 0 | 6,337,000 |

## Explanation of the proposed budget

Briefly justify each proposed cost in the stated budget.

### Explanation of the proposed budget*

Anders Elowsson is a PhD student with about 50% left of his studies. He will work mainly with the source modelling in Study A and then in Study C-D. He will be partly financed by other sources at KTH.

Ragnar Schön is a potential PhD student and will work primarily with Study B and C,D.

The PhD students will be teaching up to 20% funded by KTH.

Anders Friberg will do the work in collaboration with the PhD students and also work on the receptive fields analysis within Study A and B.

Travel expenses are calculated based on one travel to a conference for each person and year.

Equipment are calculated for one computer and various hardware such as sound cards.

## Other funding

Describe your other project funding for the project period (applied for or granted) aside from that which you apply for from the Swedish Research Council. Write the whole sum, not thousands of SEK.

### Other funding for this project

| Funder | Applicant/project leader | Type of grant | Reg no or equiv. | 2016 | 2017 | 2018 | 2019 |
|---|---|---|---|---|---|---|---|

## CV

# Curriculum Vitae - Anders Friberg

**Personal data**
Date of birth: 1958-06-09
Place of birth: Ulricehamn, Sweden
Gender: Male
Nationality: Swedish
email: afriberg@kth.se

Anders Friberg is part of the music group at Speech, Music and Hearing, KTH. He has been working mainly with the synthesis and analysis of music performance, leading to a patented rule system translating the score to a performance. Recently, he has been focusing on automatic extraction of music parameters from audio and its relation to emotional/motional expression. Currently he is leading the nationally funded project Computational Modelling of Perceptual Music Features. He is author of more than 100 publications in peer-reviewed international journals and conferences as well as served as reviewer for a number of international journals, research proposals, and PhD dissertations. He is also an active pianist currently leading a jazz trio.

**1. Higher education qualifications**
1985 M.Sc. in Engineering Physics, , KTH Royal Institute of Technology
1991 Diploma in piano performance (Magna Cum Laude), Berklee College of Music, Boston.

**2. Doctoral degree**
1995 PhD in Music Acoustics, Dept. of Speech Music Hearing, KTH Royal Institute of Technology, ”A Quantitative Rule System for Musical Performance”. Supervisor: Prof. Johan Sundberg, Opponent: Prof. Giovanni De Poli, University of Padova.

**3. Postdoctoral positions:** NA

**4. Qualification required for appointments as a docent**
2004 KTH, Computer models for musical, expressive communication: The listener, the musician and new combinations.

**5. Current position: term of appointment and research portion**
Researcher (forskare) in music acoustics, Department of Speech, Music and Hearing, KTH, started 1985. Current research portion 75%.

**6. Previous positions and periods of appointment:** NA

**7. Interruption in research**
Music studies at Berklee College of Music 1989-1991 (two years totally).
Parental leave for each of my three children with about 18 months totally.

**8. Supervision of doctoral and postdoctoral students as primary supervisor**
*Anders Elowsson*, planned graduation 2016/17, Modelling the Perception of Music.
*PerMagnus Lindborg*, planned graduation 2015, Sound perception in multi-modal environments.
*Marco Fabiani*, main supervisor, graduated 2011, Interactive computer-aided expressive music performance. Analysis, control, modification, and synthesis.

**9. Teaching and presentations**
2009-2014 Director of studies at the department (Studierektor)
Teaching several courses at KTH, e.g. Music acoustics, Musical communication
Giving a course in acoustics at the Royal College of Music, Stockholm.
Guest lecturer in various courses at KTH, Uppsala University, University of Jyväskylä.
Supervisor for several master and candidate thesis projects, currently13 students spring 2015.
Invited keynote speech at DAFx 2008, Alto U., Helsinki. Invited presentations U. of Bologna, 2013, 2014, U. of Graz, 2012, U. of Curitiba, Brazil, 2013, Lund and Uppsala U.

**10. Reviews**
*Journal reviewer* for several scientific publications e.g. Journal of the Acoustic Society of America, IEEE Transactions On Audio, Speech And Language Processing, Music Perception, Journal of New

Music Research, Musica Scientiae, The Quarterly Journal of Experimental Psychology, Perception and Psychophysics, Psychology of Music, Logopedics Phoniatrics Vocology, PLOS ONE. Frontiers in Psychology.
*Conference proceedings reviewer* for ICMC, ISMIR, CIM, NIME, SMAC, SMC, Eurographics.
*Project reviewer* for The Bank of Sweden Tercentenary Foundation, The Research Council of Norway, The Leverhume Trust, London, The Technology Foundation STW, The Netherlands, Swiss national Science Foundation, Vlaanderen Research Foundation Flanders (FWO), Belgium, Humanities in the European Research Area (HERA). Application assessment, Government of Ireland Postdoctoral Fellowships in Science, Engineering and Technology.
*PhD assessment* for Steven Livingstone, Australia, 2008. Opponent for Peter Spissky 50% seminar, Malmö Faculty of Fine and Performing Arts, Lund University, 2014. Member of the grading board, Crispin Dickson, Engineering Sciences, KTH, 2014. Replacement member of the grading board, Christos Koniaris, CSC, KTH, 2012.

**11. Software**
Developed the main part of Director Musices, a computer program for modelling music performance, and pDM – a real time version of Director Musices. Both programs are regularly supported and are available from our website. Director Musices was awarded a price in the Bourges International competition for music software, 1999.

**12. Organizational activities (summary)**
2013- 2016 Project leader of the project Computational Modelling of Perceptual Music Features funded by the Swedish Research Council.
2013 Organised the Rencon international piano competition for computer models of music performance in Stockholm, Member of the organising committee of the SMAC-SMC conference.
2010- 2013 Project leader of SEMIR (Semantic understanding in Music Information Retrieval) funded by the Swedish Research Council, Dnr 2009-4285.
2006- 2008 Project leader of SYMIR (Symbolic Music Information Retrieval) funded by the Swedish Research Council, Dnr 2005–3696.
2002- 2005 Member of the steering committee for the music acoustic group at the department for Speech, music and hearing, KTH.
2003 Organizing committee of Stockholm Music Acoustic Conference 2003.
2001- 2003 Local coordinator (KTH) for the European project MEGA

**13. Scholarships**
1989 A one-year grant from the Sweden-America Foundation for studies at the Berklee College of Music, Boston

**14. Applied art-related projects**
2003- Developed gesture/sound interaction in the following applications:
*Ghost in the Cave* – an interactive collaborative game using gestures and sound for control. (a demo application within the EC-funded project MEGA)
*Fenix* - a similar interactive game (directed by Marie-Louise Rinman).
*Hoppsa universum*- an interactive dance installation modelled after children free movements (directed by Anna Källblad).
*CLOSE* – an interactive installation consisting of female voices from Palestina (directed by Anna Källblad and Anette Taranto).
*Flying Carpet* – a dance installation at Art's Birthday, Södra Teatern 2011.

**15. Publications and bibliometrics**
A complete list is found at http://www.speech.kth.se/staff/homepage/index.html?id=afriberg
In 2008 a bibliometric analysis was made of all research at KTH in a research assessment evaluation. In the group of human communication, inluding five departments, Friberg obtained together with his colleague Roberto Bresin the highest ranking in all important bibliometric measures. Currently Friberg has 10 papers with more than 100 citations each and an h-index = 28/29 in Google Scholar: http://scholar.google.com/citations?hl=en&user=Z3sShPkAAAAJ

## Publication list

# Anders Friberg: list of publications 2007-2014/15

The bibliometric analysis has been made in Google Scholar, March 30, 2015, see
http://scholar.google.com/citations?user=Z3sShPkAAAAJ&hl=en

**Citation indices**

| | All | Since 2010 |
|---|---|---|
| Citations | 3004 | 1172 |
| h-index | 29 | 18 |
| i10-index | 48 | 29 |

### *The five most cited publications*

Friberg, A., & Sundberg, J. (1995). Time discrimination in a monotonic, isochronous sequence, *Journal of the Acoustical Society of America*, 98(5), pp. 2524-2531. Number of citations: 191

Friberg, A. and Sundberg, J. (1999) Does music performance allude to locomotion? A model of final *ritardandi* derived from measurements of stopping runners, *Journal of the Acoustical Society of America*, 105, pp 1469-1484. Number of citations: 188

Bresin, R., & Friberg, A. (2000). Emotional Coloring of Computer-Controlled Music Performances. *Computer Music Journal,* 24(4), 44-63. Number of citations: 184

Friberg, A., Bresin, R., & Sundberg, J. (2006). Overview of the KTH rule system for musical performance. *Advances in Cognitive Psychology, Special Issue on Music Performance*, 2(2-3), 145-161. Number of citations: 173

Friberg, A. (1991) Generative Rules for Music Performance: A Formal Description of a Rule System, *Computer Music Journal,* 15(2), 56-71. Number of citations: 167

### *Submitted (but important for this study)*

*Elowsson, A., & Friberg, A. (submitted). Modeling the Perception of Tempo. *Journal of the Acoustical Society of America.*

### *1. Peer-reviewed articles*

*Lindeberg, T., & Friberg, A. (2015). Idealized computational models for auditory receptive fields. *PLOS ONE* (in press).

*Friberg, A., Schoonderwaldt, E., Hedblad, A., Fabiani, M., & Elowsson, A. (2014). Using listener-based perceptual features as intermediate representations in music information retrieval. *Journal of the Acoustical Society of America, 136*(4), 1951-1963. Number of citations: -

Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music: contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology, 4*(487), 1-12. Number of citations: 6

Istók, E., Friberg, A., Huotilainen, M., & Tervaniemi, M. (2013). Expressive timing facilitates the neural processing of phrase boundaries in music: evidence from event-related potentials. *PLOS ONE, 8*(1), e55150. Number of citations: 3

Kleber, B., Zeitouni, A., Friberg, A., & Zatorre, R. (2013). Experience-Dependent Modulation of Feedback Integration during Singing: Role of the Right Anterior Insula. *Journal of Neuroscience, 33*(14), 6070-6080. Number of citations: 5

Bresin, R., & Friberg, A. (2011). Emotion rendering in music: Range and characteristic values of seven musical variables. *Cortex*, 47(9), 1068-1081. Number of citations: 29

Fabiani, M., & Friberg, A. (2011). Influence of pitch, loudness, and timbre on the perception of instrument dynamics. *Journal of the Acoustical Society of America - Express Letters*, EL193-EL199. Number of citations: 9

Friberg, A., & Ahlbäck, S. (2009). Recognition of the main melody in a polyphonic symbolic score using perceptual knowledge. *Journal of New Music Research*, 38(2), 155-169. Number of citations: 3

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception*, 24(5), 433-454. Number of citations: 151

Friberg, A., Schoonderwaldt, E., & Juslin, P. N. (2007). CUEX: An algorithm for extracting expressive tone variables from audio recordings. *Acoustica united with Acta Acoustica*, 93, 411-420. Number of citations: 27

### *2. Peer-reviewed conference contributions*

Lindeberg, T., & Friberg, A. (2015). Scale-space theory for auditory signals. In *SSVM 2015: Scale-Space and Variational Methods in Computer Vision*. Number of citations: -

Elowsson, A., Schön, R., Höglund, M., Zea, E., & Friberg, A. (2014). Estimation of vocal duration in monaural mixtures. In *Proceedings of the 40th International Computer Music Conference, ICMC 2014 and 11th Sound and Music Computing Conference, SMC 2014* (pp. 1172-1177). Number of citations: -

Masko, J., Fischer Friberg, J., & Friberg, A. (2014). Software tools for automatic music performance. In *1st international workshop on computer and robotic Systems for Automatic Music Performance (SAMP14).* Venice, 2014. Number of citations: -

Bellec, G., Elowsson, A., Friberg, A., Wolff, D., & Weyde, T. (2013). A social network integrated game experiment to relate tapping to speed perception and explore rhythm reproduction. In *Proceedings of the Sound and Music Computing Conference (SMC) 2013, Stockholm, Sweden* (pp. 19-26). Number of citations: 1

Elowsson, A., & Friberg, A. (2013). Modelling Perception of Speed in Music Audio. In *Proceedings of the Sound and Music Computing Conference 2013, SMC 2013, Stockholm, Sweden* (pp. 735-741). Number of citations: 2

*Elowsson, A., Friberg, A., Madison, G., & Paulin, J. (2013). Modelling the Speed of Music Using Features from Harmonic/Percussive Separated Audio. In *Proceedings of the International Symposium on Music Information Retrieval.* Number of citations: 4

Parncutt, R., Bisesi, E., & Friberg, A. (2013). A Preliminary Computational Model of Immanent Accent Salience in Tonal Music. In *Proceedings of the Sound and Music Computing Conference 2013, SMC 2013, Stockholm, Sweden* (pp. 335-340). Number of citations: 1

Bresin, R., Askenfelt, A., Friberg, A., Hansen, K. F., & Ternström, S. (2012). Sound and Music Computing at KTH. TMH-QPSR special issue: *Proceedings of SMC Sweden 2012 Sound and Music Computing, Understanding and Practicing in Sweden*, 52(1), 33-35. Number of citations: -

Elowsson, A., & Friberg, A. (2012). Algorithmic Composition of Popular Music. *In Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music* (pp. 276-285). Number of citations: 1

Friberg, A. (2012). Music Listening from an Ecological Perspective. Poster presented at *the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*. Number of citations: -

Bisesi, E., Parncutt, R., & Friberg, A. (2011). An accent-based approach to performance rendering: Music theory meets music psychology. In *International Symposium on Performance Science* (ISPS 2011) (pp. 27-32). Number of citations: 6

Friberg, A., & Hedblad, A. (2011). A Comparison of Perceptual Ratings and Computed Audio Features. In *8th Sound and Music Computing Conference (SMC 2011)*, Padova, Italy. Number of citations: 11

Friberg, A., & Källblad, A. (2011). Experiences from video-controlled sound installations. In *Proceedings of New Interfaces for Musical Expression (NIME 2011)*. Number of citations: -

Friberg, A. (2008). Digital audio emotions — An overview of computer analysis and synthesis of emotions in music. In *Proc. of the 11th Int. Conference on Digital Audio Effects* (DAFx-08), Espoo, Finland (pp. 1-6). Number of citations: 20

Källblad, A., Friberg, A., Svensson, K., & Sjöstedt Edelholm, E. (2008). Hoppsa Universum – An interactive dance installation for children. In *Proceedings of New Interfaces for Musical Expression - NIME, Genova*, 2008. Number of citations: 2

Istok, E., Tervaniemi, M., Friberg, A., & Seifert, U. (2008). Effects of timing cues in music performances on auditory grouping and pleasantness judgments. In *The 10th International Conference on Music Perception and Cognition, Sapporo, Japan*. Number of citations: 2

Fabiani, M., & Friberg, A. (2007). A prototype system for rule-based expressive modifications of audio recordings. In *Proc. of the Int. Symp. on Performance Science 2007* (pp. 355-360). Porto, Portugal: AEC (European Conservatories Association). Number of citations: 3

Fabiani, M., & Friberg, A. (2007). Expressive modifications of musical audio recordings: preliminary results. In *Proc. of the 2007 Int. Computer Music Conf. (ICMC07)* (pp. 21-24). Copenhagen, Denmark: The International Computer Music Association and Re:New. Number of citations: 3

### *5. Peer-reviewed book chapters*

Friberg, A., & Bisesi, E. (2014). Using computational models of music performance to model stylistic variations. In Fabian, D., Timmers, R., & Schubert, E. (Eds.), *Expressiveness in Music Performance, A cross cultural and interdisciplinary approach* (pp. 240-259). Oxford: Oxford University Press. Number of citations: 2

Friberg, A., Bresin, R., & Sundberg, J. (2014). Analysis by synthesis. In Thompson, W. F. (Ed.), *Music in the Social and Behavioral Sciences*. SAGE. Number of citations: -

Friberg, A., Bresin, R., & Sundberg, J. (2014). Expressive timing. In Thompson, W. F. (Ed.), *Music in the Social and Behavioral Sciences*. SAGE. Number of citations: -

Bresin, R., & Friberg, A. (2013). Evaluation of computer systems for expressive music performance. In Kirke, A., & Miranda, E. R. (Eds.), *Guide to Computing for Expressive Music Performance* (pp. 181-203). London: Springer. Number of citations: 5

Fabiani, M., Friberg, A., & Bresin, R. (2013). Systems for interactive control of computer generated music performance. In Kirke, A., & Miranda, E. R. (Eds.), *Guide to Computing for Expressive Music Performance* (pp. 49-73). London: Springer. Number of citations: 3

*Friberg, A., Schoonderwaldt, E., & Hedblad, A. (2011). Perceptual ratings of musical parameters. In von Loesch, H., & Weinzierl, S. (Eds.), *Gemessene Interpretation - Computergestützte Aufführungsanalyse im Kreuzverhör der Disziplinen* (pp. 237-253). Mainz: Schott 2011, (Klang und Begriff 4). Number of citations: 10

Friberg, A., & Bresin, R. (2008). Real-time control of music performance. In Polotti, P., & Rocchesso, D. (Eds.), *Sound to Sense - Sense to Sound: A state of the art in Sound and Music Computing* (pp. 279-302). Berlin: Logos Verlag. Number of citations: 1

Fabiani, M., & Friberg, A. (2008). Rule-based expressive modifications of tempo in polyphonic audio recordings. In *Computer Music Modeling and Retrieval. Sense of Sounds* (pp. 288-302). Berlin: Springer Berlin. Number of citations: 3

Goebl, W., Dixon, S., De Poli, G., Friberg, A., Bresin, R., & Widmer, G. (2008). Sense in expressive music performance: Data acquisition, computational studies, and models. In Polotti, P., &

Rocchesso, D. (Eds.), *Sound to Sense - Sense to Sound: A state of the art in Sound and Music Computing* (pp. 195-242). Berlin: Logos Verlag. Number of citations: 28

### *7. Publicly available computer programs*

**Director Musices**
This is the main implementation of the KTH rule system for music performance. Friberg has been the primary developer and designer since its start in the 80' It is still supported and regularly updated with new features.

**pDM**
This is the real-time companion to Director Musices. Friberg designed and developed the main parts of it.

Both computer programs are available at our web site:
http://www.speech.kth.se/music/performance/download/

### *8. Popular-scientific articles/presentations*

Oct 10, 2014. Article by Ben Richmond in VICE/Motherboard.
http://motherboard.vice.com/read/the-search-for-the-words-to-describe-music

Oct 29, 2014. KTH news (in Swedish)
http://www.kth.se/aktuellt/nyheter/de-foradlar-framtidens-musiktjanster-1.514096

Oct 30, 2014, Swedish Radio P4. Interview with Anders Friberg.

Nov 11, 2014, Swedish Radio P3. Interview with Anders Elowsson.

Nov 18, 2014. KTH news (in English)
https://www.kth.se/en/aktuellt/nyheter/de-foradlar-framtidens-musiktjanster-1.514096

Radio France International. A program is under preparation based on an interview with Anders Friberg and Anders Elowsson.

National Geographic, Hungarian edition. Article in preparation.

Belluck, P. (2011). To Tug Hearts, Music First Must Tickle the Neurons. New York Times, April 18, 2011. (An article discussing music and motion including Friberg's work.)

Hamer, M. (2000) All that jazz . New Scientist 23/30 December 2000, 48-51. (An overview of Friberg's measurements on swing)

Friberg A & Sundström J (1999). Jazz drummers' swing ratio in relation to tempo. Publ online as *Acoust Soc Am ASA*/EAA/DAGA '99 Berlin Meeting Lay Language Papers.
http://www.acoustics.org/137th/friberg.html

In addition, Friberg has presented the research results numerous times, for example, he organised a symposium at the EuroScience Open Forum in Stockholm, 2004.
He also made regular lectures for school children at Tekniska Muséet, Stockholm.
The group arranged a "Känslor och musikdag" May 9, 2009, which was an open symposium at KTH. (http://www.speech.kth.se/music/kom2009).
Friberg organised the open symposium Music is motion at KTH, 2011.
In 2012 the group organised the national Sound and Music Conference at KTH. In 2013 Friberg organized the public Rencon international competition and were part of the organization for SMC2013/SMAC2013 in Stockholm.

**CV**

| | |
|---|---|
| **Name:**Anders Friberg | **Doctorial degree:** 1995-05-26 |
| **Birthdate:** 19580609 | **Academic title:** Docent |
| **Gender:** Male | **Employer:** Kungliga Tekniska högskolan |

## Research education

**Dissertation title (swe)**
Ett kvantitativ regelsystem för musikalisk interpretation

**Dissertation title (en)**
A quantitative rule system for musical expression

| Organisation | Unit | Supervisor |
|---|---|---|
| Kungliga Tekniska Högskolan, Sweden<br>Sweden - Higher education Institutes | TMH, Tal, musik och hörsel | Johan Sundberg |

| Subject doctors degree | ISSN/ISBN-number | Date doctoral exam |
|---|---|---|
| 10299. Annan data- och informationsvetenskap | ISSN 1104-5787 | 1995-05-26 |

## Publications

| | |
|---|---|
| **Name:**Anders Friberg | **Doctorial degree:** 1995-05-26 |
| **Birthdate:** 19580609 | **Academic title:** Docent |
| **Gender:** Male | **Employer:** Kungliga Tekniska högskolan |

Friberg, Anders has not added any publications to the application.

## Register

### Terms and conditions

The application must be signed by the applicant as well as the authorised representative of the administrating organisation. The representative is normally the department head of the institution where the research is to be conducted, but may in some instances be e.g. the vice-chancellor. This is specified in the call for proposals.

The signature*from the applicant*confirms that:

- the information in the application is correct and according to the instructions form the Swedish Research Council
- any additional professional activities or commercial ties have been reported to the administrating organisation, and that no conflicts have arisen that would conflict with good research practice
- that the necessary permits and approvals are in place at the start of the project e.g. regarding ethical review.

The signature *from the administrating organisation* confirms that:

- the research, employment and equipment indicated will be accommodated in the institution during the time, and to the extent, described in the application
- the institution approves the cost-estimate in the application
- the research is conducted according to Swedish legislation.

The above-mentioned points must have been discussed between the parties before the representative of the administrating organisation approves and signs the application.

*Project out lines are not signed by the administrating organisation. The administrating organisation only sign the application if the project outline is accepted for step two.*

*Applications with an organisation as applicant is automatically signed when the application is registered.*