In J. Sundberg & B. Brunson (Eds.) Proceedings of Music and Music Science, Stockholm, October 28-30, 2004

# A fuzzy analyzer of emotional expression in music performance and body motion

*Anders Friberg*
KTH – Royal Institute of Technology
Dept. of Speech, Music and Hearing
Lindstedtsvägen 24
100 44 Stockholm, Sweden
Email: andersf@speech.kth.se

### Abstract

A real time algorithm is presented for analyzing the emotional expression in music performance and body motion. It is primarily intended to be used as a real time controller in an artistic human-computer performance system. The analysis is done in three steps consisting of cue analysis, calibration, and fuzzy mapping. The cue analysis extracts tone parameters such as tempo, sound level and articulation from audio input or overall quantity of motion and vertical/horizontal motion from video input. The calibration step adjusts the system to a particular performer's expression in a semi-automatic procedure. The fuzzy mapper translates the cue values to three emotion outputs: happiness, sadness, and anger. The mapping to emotions was modeled using qualitative data from previous research, thus no training using control data was necessary. The fuzzy analyzer has been used in several applications including the Expressiball for visualizing music performance and the collaborative game Ghost in the Cave using gestures and voice as the main player input.

## Introduction

The development of and experimentation with new musical controllers (e.g. musical instruments) has been an integral part of the musical evolution. With the introduction of electronic sound manipulation these controllers were no longer bound to the acoustic properties of physical objects, thus opening up a wide field for any type of imagined sound processing. Traditional musical instruments have, however, provided natural restrictions in the sound production that we have learned to decode. For example, the physical energy used to excite an instrument can be at least partially decoded by the listener from the resulting timbre[1]. By removing these

[1] This is likely to work well for known instruments – the dynamic level of a piano recording corresponding to the energy used for pushing down the keys is easily decoded even if the listening volume is turned down.

restrictions the possible space of all sounds becomes much larger, but on the other hand, presumably a smaller subspace is musically/perceptually relevant.

In creating a new musical controller, a crucial issue is the mapping from the input device parameters to the sound parameters. It has been shown that it is important to make this mapping similar to how traditional musical instruments react in a general sense. More specifically, the mapping should preferably not be one-to-one but rather one-to-many or many-to-many (Hunt et al. 2004). One simple example in the piano is that increasing key speed increases sound level, high-frequency content, as well as changes the "bonk" sound obtained when the key is hitting the keybed.

The technology is not restricted to the invention of new musical instruments in the sense that each individual sound is triggered by the performer. A large number of different artistically oriented performance-computer interactions have been realized in the past (e.g. Wanderley and Battier 2000). One alternative is to have a system for performance control of the overall aspects of the computer-generated sound similar to a how a conductor controls an orchestra. In order to restrict the control space to perceptually meaningful variations, similar to the instrument control above, such system would benefit from having a more abstract higher-level expressive musical/gesture description defined as a part of the sound and input processing algorithms. Such high-level definitions was a goal of the recent European project MEGA[2], which resulted in a number of useful tools dealing with expressive features in music and dance aimed for artistic applications (Camurri et al. in press).

We will describe an algorithm that has been used for analyzing the emotional expression either in body motion or in music performance. This system is intended to be used as a real time controller in an artistic setting or in applications such as computer games. Before describing the algorithm, a short overview of emotional expression in music performance and gestures is given.

## Emotional expression in music performance and body motion

Most people agree that an important aspect of music is the communication of emotions. Recently, a number of scientific studies have explored this communication (Juslin and Sloboda 2001; Juslin and Laukka 2003). These studies indicate that for basic emotions such as happy, sad or angry, there is a rather simple relationship between the emotional description and the *cue* values (i.e. measured parameters such as tempo, sound level or articulation). Response data obtained from listeners rating the emotional expression of different music performances can be modeled by multiple regression analysis, thus, suggesting a linear relationship between cues and emotional expression (Juslin 2000). Since we are aiming at real-time playing applications we will focus here on performance cues such as tempo and dynamics. A complete set of cues should also include compositional features such as tonality, rhythm and instrumentation (Gabrielsson and Lindström 2001).

The emotional expression in body gestures has also been subject to research but to a lesser extent than in music. Camurri et al. (2003) analyzed and modeled the

[2] www.megaproject.org

emotional expression in dancing. Boone and Cunningham (2001) investigated children's movement patterns when they listened to music with different emotional expressions. Dahl and Friberg (2004) investigated movement patterns of a musician playing a piece with different emotional expressions. These studies all suggested particular movement cues related to the emotional expression, similar to how we decode the musical expression. We follow the suggestion that musical expression is intimately coupled to expression in body gestures and biological motion in general (Friberg and Sundberg 1999, Juslin et al. 2002). Therefore, we try to apply similar analysis approach to both domains.

Table 1 presents typical results from previous studies in terms of qualitative descriptions of cue values. As seen in the Table, there are several commonalities in terms of cue descriptions between motion and music performance. For example, anger is characterized by both fast gestures and fast tempo. The research regarding emotional expression yielding the qualitative descriptions as given in Table 1 was the starting point for the development of current algorithms.

| *Emotion* | *Motion cues* | *Music performance cues* |
|---|---|---|
| Anger | Large | Loud |
| | Fast | Fast |
| | Uneven | Staccato |
| | jerky | Sharp timbre |
| Sadness | Small | Soft |
| | Slow | Slow |
| | Even | Legato |
| | soft | |
| Happiness | Large | Loud |
| | rather fast | Fast |
| | | Staccato |
| | | Small tempo variability |

Table 1. A characterization of different emotional expressions in terms of cue values for body motion and music performance. Data taken from Dahl and Friberg (2004) and Juslin (2001).

## Fuzzy analyzer

The analysis of emotional expression in music performance/gestures is realized in three steps, see Figure 1. The first step is the extraction of basic cues. These cues are quite well defined for music input consisting of traditional tone parameters such as sound level, tempo, and articulation (staccato/legato). In the body motion analysis we have been interested in parameters describing general features of the movements rather than individual movements of each limb. A number of such general cues have been identified and algorithms have been developed for automatic extraction from video input (Camurri et al. 2004).

The second step is the semi-automatic calibration of cues. This is important for practical reasons in order to avoid otherwise lengthy optimization sessions trying to fine-tune internal parameters adapting to variations in musical content/body gestures or technical setup.

The third step is the mapping from cues to emotion description by the expression mapper in Figure 1. Instead of using the more common data-driven methods, such as Neural Networks or Hidden Markow Models, we suggest here fuzzy set functions, allowing the direct use of qualitative data obtained from previous studies. The three steps are described in the following.

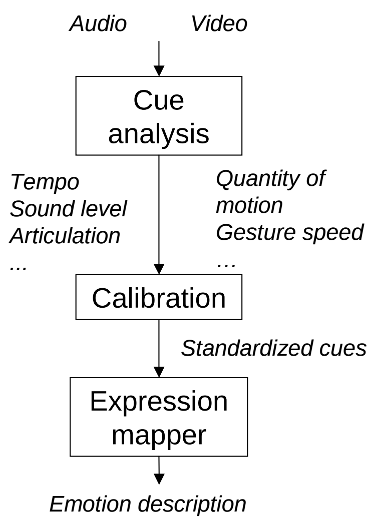


Figure 1. The three-step process of analyzing emotional expression in gestures or singing/playing.

**Music cue extraction**

The audio cue extraction was designed for monophonic playing or singing. A previous prototype was described in Friberg et al. (2002) and an improved version for non-real time use is found in Friberg et al. (2005). The first part of the cue extraction segments the audio stream into tones by computing two different sound level envelopes. The RMS sound level is computed on a hanning windowed buffer of 256 samples using a sampling rate of 22050 Hz. The resulting sound level envelope is then low-pass filtered using either a 40 Hz cut-off or a 1 Hz cut-off frequency. The first sound level envelope follows roughly the shape of each tone and the second sound level envelope follows the general shape of the phrase. The crossings of two envelopes define the tone onsets and offsets. For each segmented tone five different cues are computed: *sound level* (dB), *instant tempo* (tones/second), *articulation* (relative pause duration), *attack rate* (dB/ms), and *high-frequency content* (high/low energy).

**Body motion cue extraction**

Cues for describing body motion are not as clearly defined and are not as easily extracted as in music. Body motion is greatly multidimensional with several degrees of freedom for describing the motion of each body part. Another complication is that segmentation in terms of basic gestures is not as easily defined. However, there has been a considerable body of research for analyzing human motion from video cameras and a number of advanced algorithms have emerged (for an overview, see Moeslund and Granum 2001).

We use the analysis tools for gesture recognition developed at University of Genova included in the software platform EyesWeb (Camurri et al. 2000)[3]. The current version of the cue analysis uses only a few basic tools within EyesWeb. In order to remove the background, the first step is to compute the difference signal between consecutive video frames. This means that the algorithm just "see" something when there is a movement. This movement detection is improved by also computing the difference signal from the inverted picture. The use of difference signals makes the system quite robust and insensitive to e.g. light variations in the room. Three cues are computed from the difference signal. The total number of visible pixels constitutes the cue *Quantity of Motion* (QoM). The bounding rectangle defines the area in the picture that contains all non-zero pixels. The instant width and height of the bounding rectangle are computed and their peak-to-peak amplitude variations constitute the cues *width-pp* and *height-pp.* Figure 2 shows an example of the difference signal with the bounding rectangle.

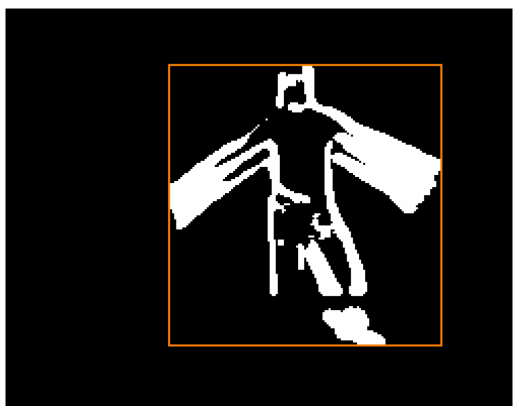

Figure 2. The difference video signal and the bounding rectangle used for computing the body motion cues. The number of white pixels is a measure of overall quantity of motion (QoM).

### Calibration

In the calibration module each cue is *standardized*, meaning that it is subtracted by its mean value and divided by its standard deviation: $CUE_{stand} = (CUE_{in} - m)/SD$. This results in cues with a mean of 0 and a standard deviation of 1. This is an important step implying that the following mapping does not need to be adjusted if the input conditions change, such as change of instrument, dancer, or artistic material. In order to obtain realistic values for the mean and standard deviation of each cue a calibration phase is needed before the recognition system is used.  In the calibration phase the user is asked to move or play in, for example, a happy, sad and angry way

[3] www.eyesweb.org

thus defining the space of all possible variations. All cue values are collected and the mean and standard deviation of all the values of each cue are computed and then used in the calibration module.

The emotional expression is not supposed to be changing from note to note or from gesture to gesture. Therefore the cues are averaged over time. A running average of 4-9 notes was used for the music analyzer. The amount of averaging can be adjusted according to the desired reaction speed versus data output stability.

**Expression mapper**

As mentioned above, previous mapping between cues and emotional expression in music has successfully been modeled using linear multiple regression (Juslin 2000). This is an example of a data-driven approach in which a measured data set is used to train the model. A number of other data-driven methods have been used for analyzing expression in music performance such as Neural Networks (Bresin 1998), Hidden Markow Models (Cirotteau et al. 2004) or Bayesian classification (Mion 2003). Due to lack of large databases the results of any of these methods are restricted and likely to be biased by e.g. coder variability. On the other hand, there is a large amount of qualitative data emerging from over 40 investigations regarding emotional expression in music (such as presented in Table 1). These data has recently been summarized and subjected to a meta-analysis (Juslin and Laukka 2003). Most often, cues have been characterized mainly in terms of being either high or low in relation to different emotional expressions. One interesting extension in the meta-analysis was to classify some cues in terms of three levels. It indicated that an intermediate cue level might be important at least for the sound level in happy expression. This is also in agreement with our informal experience in developing the mapping algorithms. However, it needs to be further examined in controlled experiments.

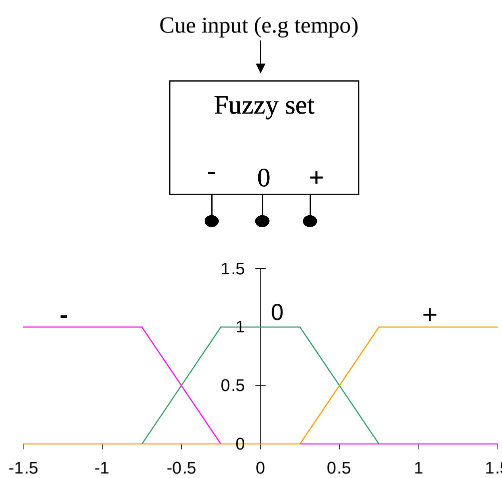


Figure 3. Qualitative classification of cues in terms of the three regions high (+), medium (0), and low (-), each with a separate output.

Following these ideas, we suggest here a method that uses the qualitative cue descriptions divided in three levels in order to predict the intended emotion. The same method is used both for musical and body motion cues. It uses *fuzzy set functions* to go from continuous cues to qualitative cues (e.g. Niskanen 2004, Zimmerman 1996). Fuzzy logic has been used in a wide variety of applications. Of related interest here can be mentioned the modeling of emotions in robots (Seif El-Nasr et al. 2000) and modeling music performance rules (Bresin et al. 1995). Each cue is divided into a three overlapping region functions. Within a region the corresponding output is one and outside the region it is zero with an overlapping linear area at the boundaries. The input is the standardized cues. An example dividing a cue in three regions is given in Figure 3. Thus, if the tempo is higher than 0.75SD from its mean, the "+" output is 1 and all the other outputs are 0.

The final emotion prediction output is computed by taking an average of a selected set of fuzzy set functions. This selection is done according to previous qualitative descriptions. This results in a continuous output for each emotion with a range 0-1. If the value is 1 the emotion is completely predicted if it is 0 it is not at all predicted. Using an average of a set of region functions makes the output smoothly changing between emotions depending on the number of "right" cues. This averaging method was selected with the final application in mind. For example, if this algorithm is used for controlling the musical expressivity in computer-played performance, the transitions between different emotions need to be smooth. Other applications might ask for a discrete classification of emotional expression. This could be easily modeled using fuzzy logic.

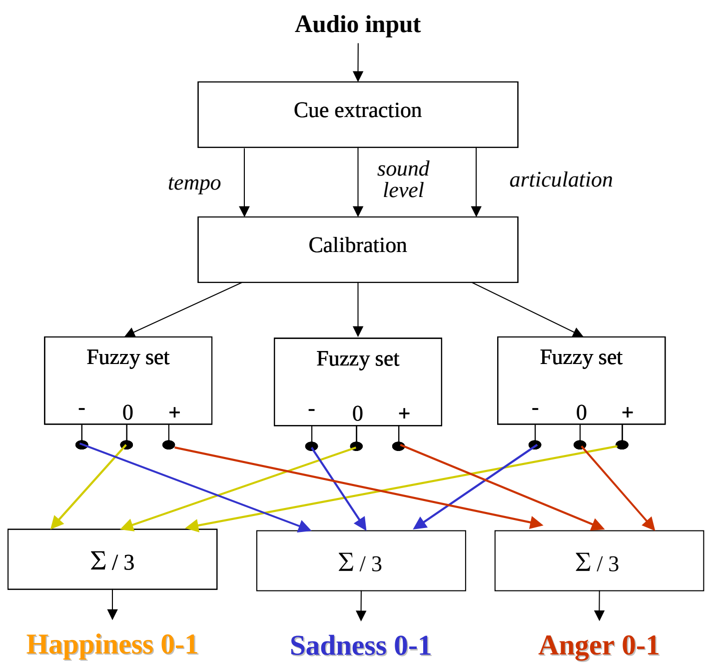


Figure 4. The complete system for estimating emotional expression in music performance using three cues. An audio input is analyzed in terms of tempo, sound level and articulation and the resulting prediction of emotional expression is output in terms of three functions ranging from 0 to 1.

The complete system is shown in Figure 4 using three cues extracted from audio input. The same setup is used for the body motion analysis using the body motion cues. The use of three cues with the mapping configuration indicated by the colored arrows in this example, has the advantage that emotion outputs are mutually exclusive, that is, if one output is 1, the other outputs are 0.

All parts of the fuzzy analyzer except the motion cue analysis have been implemented using the program `pd` (Puckette 1996). pd is a modular graphic environment for processing primarily audio and control information in real time. The fuzzy analyzer was implemented using preexisting blocks in the release pd-extended 0.37, complemented with the library SMLib made by Johannes Taelman. The video cue analysis was implemented as a patch in EyesWeb, a similar graphic programming environment primarily for video processing (Camurri et al. 2000). The audio analyzer makes modest claims on processing power and typically uses only a few percent of a Pentium 4 processor running Windows. Due to the cross-platform compatibility of pd, the audio cue analysis could easily be ported to MacOS or Linux. The video cue analysis is currently restricted to the Windows platform.

## Applications

The first prototype that included an early version of the fuzzy analyzer was a system that allowed a dancer to control the music by changing dancing style. It was called The Groove Machine and was presented in a performance at Kulturhuset, Stockholm 2002. Three motion cues were used, QoM, maximum velocity of gestures in the horizontal plane, and the time between gestures in the horizontal plane, thus slightly different from the description above. The emotions analyzed were (as in all applications here) anger, happiness, and sadness. The mixing of three corresponding audio loops was directly controlled by the fuzzy analyzer output. For a more detailed description, see Lindström et al. (in press).

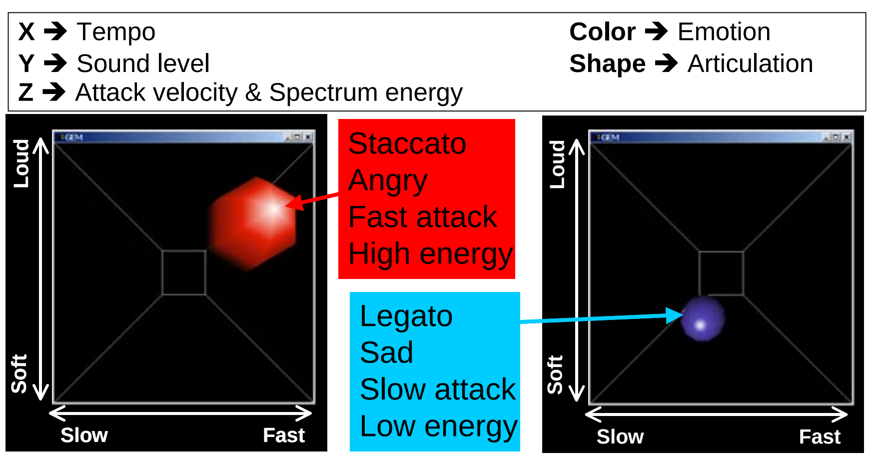


Figure 5. Two different examples of the Expressiball giving visual feedback of musical performance.

The ExpressiBall, developed by Roberto Bresin, is a way to visualize a music performance in terms of a ball on a computer screen (Friberg et al. 2002). A microphone is connected to the computer and the output of the fuzzy analyzer as well as the basic cue values are used for controlling the appearance of the ball. The position of the ball is controlled by tempo, sound level and a combination of attack velocity and spectral energy, the shape of the ball is controlled by the articulation (rounded-legato, polygon-staccato) and the color of the ball is controlled by the emotion analysis (red-angry, blue-sad, yellow-happy), see Figure 5. The choice of color mapping was motivated by recent studies relating color to musical expression (Bresin and Juslin 2005). The ExpressiBall can be used as a pedagogical tool for music students or the general public. It may give an enhanced feedback helping to understand the musical expression. A future display is planned at Tekniska museet, Stockholm.

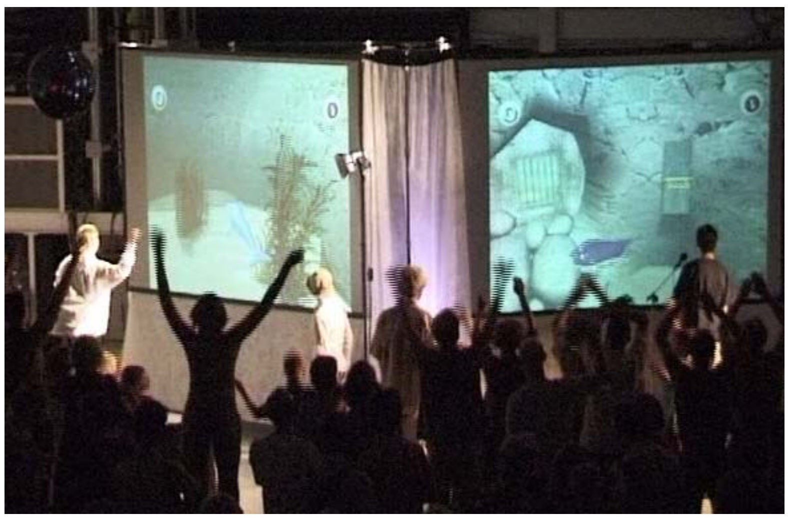

Figure 6. Picture from the first realization of the game Ghost in the Cave. Motion player to the left (in white) and voice player to the right (in front of the microphones).

The latest application using the fuzzy analyzer has been the collaborative game Ghost in the Cave (Rinman et al. 2004).[4] It uses as its main input control either body motion or voice. One of the tasks of the game is to express different emotions either with the body or the voice; thus, both modalities are analyzed using the fuzzy analyzer described above.

The game is played in two teams each with a main player, see Figure 6. The task for each team is to control a fish avatar in an underwater environment and to go to three different caves. In the caves there is a ghost appearing expressing different emotions. Now the main players have to express the same emotion, causing their fish to change accordingly. Points are given for the fastest navigation and the fastest

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

expression of emotions in each subtask. The whole team controls the speed of the fish as well as the music by their motion activity.

The body motion and the voice of the main players are measured with a video camera and a microphone, respectively, connected to two computers running two different fuzzy analyzers described above. The team motion is estimated by small video cameras (webcams) measuring the Quantity of Motion (QoM). QoM for the team motion was categorized in three levels (high, medium, low) using fuzzy set functions as shown in Figure 3. The music consisted of pre-composed audio sequences, all with the same tempo and key, corresponding to the three motion levels. The sequences were faded in and out directly by control of the fuzzy set functions. One team controlled the drums and one team controlled the accompaniment. The Game has been set up five times since the first realization summer at the Stockholm Music Acoustics Conference 2003, including the Stockholm Art and Science festival, Konserthuset, Stockholm, 2004, and Oslo University, 2004.

## Evaluation

The performance and validity of the fuzzy analyzer is not as easily tested as a data-driven system in which the outcome of different data sets with known classification can be compared. With the fuzzy mapping we know already the outcome for different cue values corresponding to the selected configuration in terms of connections to the different summation blocks in the bottom of Figure 4. Almost certainly a Neural Network-based mapper would perform better when using specific datasets for training and evaluation. Such datasets typically consists of performances in which performers are asked to express different emotions and corresponding listener data with emotion ratings of the same performances. These data are far from “ideal” in the sense that they contain several noise sources due to coder (performer) and rater (listener) variability. This will also influence a data-driven algorithm making the classification less distinct. This means that an evaluation using such datasets might give good test results but still results in poor usability in real situations.

One alternative to evaluate the algorithm would be to generate synthesized performances in which each cue is controlled in a systematic fashion. It is rather simple to synthesize music performances (e.g. Juslin et al. 2002) but more elaborate to synthesize body motion. The expressive content of the examples is then rated in listener experiments. Such a data set would allow a more systematic testing of, for example, the three cue levels used and will be considered in future work.

Another alternative is to make a usability test of the complete application asking the users to give feedback. This was done in terms of questionnaires both for *Groove Machine* and *Ghost in the Cave* (Lindström et al in press, Rinman forthcoming). These evaluations were in general very positive but were mostly concerned with overall questions such as market potential or cooperation issues. However, one specific question addressed the players’ controllability. In *Ghost in the Cave*, the team players believed that the main players were efficiently controlling the avatar giving a mean rating of 6.6 on a scale 0-10 with a confidence interval above 5. In the *Groove Machine* the audience strongly believed that the dancer was able to control the music giving a mean rating of 7.6 on the same scale (Lindström et al in press). This is an indication that the algorithm works but needs to be further investigated in future research.

## Summary and discussion

We describe a system that can be used for analyzing emotional expression both in music and body motion. The use of fuzzy mapping was a way of directly using previous results summarized in various investigations and turned out to be a robust mapping also in practical testing during the development of the applications. The advantages of the model can be summarized as

**Generic** - it is the same model for music performance and body motion.
**Robust** - The fuzzy set functions always stay between 0 and 1 implying that the emotion output is always between 0 and 1 as well. A collection of cues lowers the error influence from one cue proportionally.
**Handle nonlinearities –** This is not possible in e.g. a linear regression model.
**Smooth transitions between emotions** – This is achieved by the overlapping fuzzy set functions each with transition range.
**Flexibility –** It is easy to change the mapping using for example more cues since there is no need recalibrate the system.

The use of higher-level expression descriptions such as emotions has the advantage that it can provide a natural coherence between the controller’s expression (visual or auditive) and the expression of the control device (could be a synthesizer, visual effects etc.) in a public performance. Take an example with a dancer controlling the music performance with a one-to-one correspondence between high-level description in the body motion and music performance. When the movements of the dancer are aggressive - the music also sounds aggressive.

One potential disadvantage is that each cue has the same importance in predicting the emotions. From previous research it is obvious that this is not the case in human recognition of emotional expression. It can easily be modeled using different weights for each fuzzy input in the summation block in Figure 4. However, for simplicity and in order minimize the number of parameters this was not used.

The used cue extraction is rather simple and can easily be improved by, for example, using better tone onset detection for the audio analyser. However, due to the averaging effects (both over cues and over time) in the subsequent fuzzy mapper the resulting improvement will be small.

The KTH music performance rules has recently been complemented with a module called pDM that allows real time control of the rule system (Friberg 2005a). Thus, the expressivity of music performance can be dynamically controlled. pDM contains also mappers from emotional expression to the rule parameters. Since the current system analyzes expression in body motion the two could be easily connected yielding a future application that we call a *home conductor* system. Such a system would allow also laymen to be actively involved in the music process and “conduct” the music directly by gestures (Friberg 2005b).

## Acknowledgements

This work was supported by the EU project MEGA, IST-20410. I would like to acknowledge all people involved in the development of the Groove Machine and Ghost in the Cave, including Ambra Succi, Bounce dance company; Sofia Dahl, Marie-Louise Rinman, Kasper Marklund, KTH; Antonio Camurri, Gualtiero Volpe, Barbara Mazzarino, University of Genova; Ivar Kjellmo, Bendik Bendiksen, Octaga AS, Oslo; Damien Cirotteau, Hugh McCarthy, University of Padova; Erik Lindström, Uppsala University.

## References

Boone, R. T. and Cunningham, J. G. (2001). Children's expression of emotional meaning in music through expressive body movement. *Journal of Nonverbal Behavior*, 25(1).

Bresin, R. (1998) Artificial Neural Networks Based Models For Automatic Performance of Musical Scores. *Journal of New Music Research*, 27(3), 239-270.

Bresin, R. and Juslin, P. N. (2005). Rating Musical Expression With Colors. manuscript in preparation.

Bresin, R., De Poli, G. and Ghetta, R. (1995). A Fuzzy Formulation of KTH Performance Rule System. In J. Sundberg (ed.), *Proceedings of the 2nd International Conference on Acoustics and Musical Research* (pp. 15-36), Stockholm, Sweden: Speech, Music and Hearing, KTH.

Camurri A., Hashimoto S., Ricchetti M., Trocca R., Suzuki K., and Volpe G. (2000). EyesWeb – Toward Gesture and Affect Recognition in Interactive Dance and Music Systems. *Computer Music Journal*, 24(1), 57-69.

Camurri A., Lagerlöf I., and Volpe G. (2003). Recognizing Emotion from Dance Movement: Comparison of Spectator Recognition and Automated Techniques. *International Journal of Human-Computer Studies*, 59(1-2), 213-225.

Camurri A., Mazzarino B, and Volpe G. (2004). Analysis of Expressive Gesture: The EyesWeb Expressive Gesture Processing Library. In A. Camurri, G. Volpe (Eds.), *Gesture-based Communication in Human-Computer Interaction, LNAI 2915*, Springer Verlag.

Camurri, A., De Poli, G., Friberg, A., Leman, L., and Volpe, G. (in press). The MEGA project: analysis and synthesis of multisensory expressive gesture in performing art applications. *Journal of New Music Research*.

Cirotteau D., De Poli G., Mion L., Vidolin A., and Zanon P. (2004). Recognition of musical gestures in known pieces and in improvisations. In A. Camurri, G. Volpe (eds.) *Gesture Based Communication in Human-Computer Interaction*, LNAI 2915 (pp. 497-508), Berlin, Springer Verlag.

Dahl, S. and Friberg, A. (2004). Expressiveness of musician's body movements in performances on marimba. In A. Camurri, G. Volpe (eds.) *Gesture-based Communication in Human-Computer Interaction*, LNAI 2915 (pp.479-486), Berlin, Springer Verlag.

Friberg, A. (2005a). pDM: an expressive sequencer with real-time control of the KTH music performance rules. Manuscript submitted for publication.

Friberg, A. (2005b). Home conducting – control the overall musical expression with gestures. Manuscript submitted for publication.

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(3), 1469-1484

Friberg, A., Schoonderwaldt, E., and Juslin, P. N. (2005). CUEX: An algorithm for extracting expressive tone variables from audio recordings. Accepted for publication in *Acoustica united with Acta Acoustica*.

Friberg, A., Schoonderwaldt, E., Juslin, P.N. and Bresin, R. (2002). Automatic Real-Time Extraction of Musical Expression. In *Proceedings of the International Computer Music Conference 2002* (pp. 365-367), San Francisco: International Computer Music Association,.

Gabrielsson, A., and 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). New York: Oxford University Press.

Hunt, A., Kirk, R. and Neighbour, M. (2004). Multiple Media Interfaces for Music Therapy. *IEEE Multimedia*, 11(3), 50-58.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1797-1813.

Juslin, P. N. (2001). Communication of emotion in music performance: A review and a theoretical framework. In P. N. Juslin & J. A. Sloboda (eds.), *Music and emotion: Theory and research* (pp. 309-337). New York: Oxford University Press.

Juslin, P. N. and Laukka, J. (2003). Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code? *Psychological Bulletin,* 129(5), 770-814.

Juslin, P. N. and Sloboda J. A., (Eds.) (2001). *Music and Emotion: Theory and Research*, Oxford: Oxford University Press.

Juslin, P. N., Friberg, A., and Bresin, R. (2002). Toward a computational model of expression in performance: The GERM model. *Musicae Scientiae special issue 2001-2002*, 63-122.

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

Mion, L. (2003) Application of Bayesian Networks to automatic recognition of expressive content of piano improvisations. In R. Bresin (ed.) *Proceedings of the Stockholm Music Acoustics Conference 2003*, Vol II (pp. 557-560).

Moeslund, T. B. and Granum, E. (2001). A Survey of Computer Vision-Based Human Motion Capture. *Computer Vision and Image Understanding*, 81, 231-268.

Niskanen, V. A. (2004). *Soft Computing Methods in Human Sciences*, Springer, Berlin.

Puckette, M. (1996). Pure Data. *Proceedings of the 1996 International Computer Music Conference* (pp. 269-272), San Francisco: International Computer music Association.

Rinman, M.-L., (forthcoming). Designing game, play and interaction in location-based game spaces. Doctoral thesis, KTH.

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

Seif El-Nasr, M., Yen, J. and Iorger, T. R. (2000). FLAME – Fuzzy logic adaptive mode of emotions. *Autonomous Agents and Multi-Agent Systems,* 3, 219-257.

Wanderley, M. and Battier, M. (eds.) (2000). *Trends in Gestural Control of Music*. IRCAM - Centre Pompidou.

Zimmerman, H.-J. (1996). *Fuzzy set theory – and its applications*, 3rd ed., Kluwer, Boston.