# A virtual-agent head driven by musical performance

Maurizio Mancini, Roberto Bresin and Catherine Pelachaud

*Abstract*— **In this paper we present a system in which visual feedback of an acoustic source is given to the user using a graphical representation of an expressive virtual head. In this system we also included the notion of expressivity of the human behavior. We provide several mapping. On the input side, we have elaborated a mapping between values of acoustic cues and emotion as well as expressivity parameters. On the output side, we propose a mapping between these parameters and the behaviors of the virtual head. These mappings ensure a coherency between the acoustic source and the animation of the virtual head. After presenting some background information on expressivity of humans we introduce our model of expressivity. We explain how we have elaborated the mappings between the acoustic and the behavior cues. Then we describe the implementation of a working system that controls the behavior of a human-like head that varies depending on the emotional and acoustic characteristics of the musical execution. Finally we present the tests we conducted to validate our model of mapping between music performance emotions and expressivity parameters.**

*Index Terms*— **acoustic cues, music, emotion, virtual agent, expressivity**

## I. Introduction

WHAT happens when it is a computer listening to the music? In HCI applications affective communication plays an increasingly important role. It would be helpful if systems could express what they perceive and communicate it to the human user through visual and acoustic feedbacks.

Listening to music is an everyday experience. But why do we do it? For example one could do it for tuning her own mood. Research results show that we are not only able to recognize different emotional intentions used by musicians or speakers [1] but that we also feel these emotions. It has been found that when listening to music, people experience a change in bio-physical cues (such as blood pressure, etc.). This change may correspond to either the feeling of the emotion arising from listening the music or the recognition of the emotion evoked by the music [2].

Virtual agents with a human-like appearance and communication capabilities are being used in an increasing number of applications for their ability to convey complex information through verbal and nonverbal behaviors like voice, intonation, gaze, gesture, facial expressions, etc. Their capabilities are useful when being a presenter on the web [3], a pedagogical agent in tutoring systems [4], a talking head helping hearing-impaired people to “listen” to a telephone call by lipreading [5], a companion in interactive setting in public places such as museums [6], [7], or even a character in virtual storytelling systems [8]. The expressivity of behaviors, that is the way behaviors are executed, is also an integral part of the communication process as it can provide information on the state of an agent, such as current emotional state, mood, and personality [9].

In our work we implemented a system that gives to the user a visual feedback by moving and modifying the expression of a virtual agent’s head. The agent’s behavior (movement plus expression) explicitly visualizes the emotional intentions of the musical execution. It is meant to show the direct connection between body motion and expressivity in music performance [10].

In the next section we present the state of the art. Then we give some background information on expressivity for human behavior, voice and music execution. Perceptual tests on expressivity of body and vocal cues are also provided. Then, in Section V we introduce our real-time application for visual feedback of musical execution. We provide information on the mapping between acoustic cues and animation parameters. In section VI we describe the tests we conducted to validate our model of mapping between music performance emotions and expressivity parameters. Finally we conclude the paper.

## II. State of the art

Some previous works [11]–[13] have addressed the generation of synthetic human behavior depending on music (or sound) input. The works by Lee et al. [14] and by Cardle et al. [12] are mainly focused on adapting pre-calculated animations like walking or dancing to a given music input. These systems analyze the music and extract parameters such as *tempo*. Based on the values of the extracted parameters, the *rhythm* of the animation is changed. In the works by Cornwell et al. [15] and by Downie and Lefford [13] interaction between agents are modulated by music and sound. The emotive content of the acoustic source are positively correlated to the quality of the interaction between agents. For example a group of agents will tend to collaborate more if listening to a *happy* and *positive* piece of music [15]. [13] also underlines that music can help to *give life* to inanimate objects, increasing their credibility. Taylor et al. [16] developed a system that allows a user to adapt the way she plays a music instrument to the reaction of a virtual character. The user has to try to vary her execution to make virtual character reacts in some desired way.

Our work is most similar to DiPaola et al.’s work [11]. The authors emphasize that affective information can be delivered through several means (music, facial expression, body movement, etc) by translating the original message into the language used by each mean. So if music is the starting mean and facial expression is the output mean, the system elaborates the information coming from the music and translates it into facial expressions and head movements. As in this work [11], we view that the translation, that is the *mapping*, between cues from one mean to another one is of high relevance.

## III. Expressivity

Human individuals differ not only in their reasoning, their set of beliefs, goals, and their emotive states, but also in their way of expressing such information through the execution of specific behaviors. We refer to these behavioral differences with the term *expressivity*. In the sub-section III-A we present definitions of expressivity in human behavior from the perception studies point of view while sub-section III-B describes some work in voice and music expressivity.

### A. Expressivity in behavior

Many researchers (Johansson [17], Wallbott and Scherer [9], Gallaher [18], Ball and Breese [19], Pollick [20]) have investigated human motion characteristics and encoded them into categories. Some authors refer to body motion using dual categories such as slow/fast, small/expansive, weak/energetic, small/large, unpleasant/pleasant. Behavior expressivity has been correlated to energy in communication, to the relation between temporal/spatial characteristics of gestures, and/or to personality/emotion. For Wallbott [21] it is related to the notions of quality of the mental state (e.g. emotion) and of quantity (somehow linked to the intensity factor of the mental state). Behaviors encode not only content information (the What is communicating through a gesture shape for example) but also expressive information (the How it is communicating through the manner of execution of the gesture). There are evidence that some movement qualities are characteristic to emotions. These qualities are the spatial extension of the movement, its energy/power and the activity [21].

In a recent study by Dahl and Friberg (paper IV in [10]) a correlation between body motion and acoustic cues in expressive music performance has been observed. For example, in angry and happy performances faster movements of the body correspond to faster *tempi* in the performances, and larger amount of movement to louder sound level, in sad performances more fluent movements correspond to a more *legato* articulation. In a perceptual test done within the same study, subjects could recognize the intended emotions by rating muted video clips, each showing one of three different musicians, a marimba player, a saxophone player and a bassoon player, performing the same score with four emotional intentions (Fear, Anger, Happiness, Sadness). In particular this study highlights the importance of head movements in the communication of the emotional intentions of the player.

### B. Expressivity in voice and music

Sound is an important mean of communicating emotions. Much of the essence of speech and music concerns the communication of moods and emotions. Sound can be characterized in terms of a number of physical variables (*cues*): onset time, decay time, pitch, loudness, timbre, and tempo as well as the rate of change of these variables. Combinations of these cues can be used for describing the expression of emotion in sound.

In the last 30 years, Klaus Scherer and co-workers at the Department of Psychology, University of Geneva, have been conducting extensive work in the field of emotions and in particular in the area of perception of emotions in speech. They shed light on the multi-faceted problems related to emotions in vocal expression (for an overview see [22] [23]). Among other findings they clearly identified how cues are manipulated in the communication of emotions and specially during appraisal processes [24].

In the past decade, research in music communication has focused on the analysis and formalization of expressive communication [25] [26] [27]. Striking analogies between spoken and musical communication have been revealed, with respect to how emotions and moods are expressed using acoustic cues [1] [28]. It has been noticed that expressive rendering can help in marking more clearly the structure of the message being communicated [28]; and several acoustic cues involved in the communication of emotional expression have been identified [29] [1] [30] [31] [32] [33]. These cues can be combined in various ways for signaling the same emotion, thus affording robustness in communication via redundancy and variation. In a review of 101 papers on vocal expression and 41 on music performance [1], similarities were found between both channels in their use of acoustic cues for the communication of emotions.

## IV. Modeling expressivity

### A. Expressivity for virtual agents

*1) Behavior expressivity parameters:* In order to increase its credibility and life-likeness, a virtual agent should not only be able to show an emotional state but also to show it with a certain quality [34]; that is, the agent should be able to alter its way of expressing a given emotion through the application of some *modifications* on the quality of its movements. In our work we are more interested in what is visually perceived of a given behavior than in the internal reasons (for example mental state, personality, mood, etc) that have triggered that behavior. We base our model on perceptual studies. Starting from the results reported in [9] and [18], we have defined the expressivity of body movements over 5 *dimensions*. Four of these dimensions influence qualitatively the animation of our virtual agent [34]:

- *Overall Activity*: amount of activity (e.g., passive/static or animated/engaged). This parameter influences the number of single behaviors happening during the communication. For example, as this parameter increases, the number of head movements per unit of time will increase. It is a single float-valued ranging from 0 to 1 where a value of *zero* corresponds to *no activity*, and a value of one corresponds to *maximum activity*.
- *Spatial Extent*: amplitude of movements (e.g., expanded versus contracted). This parameter determines the quantity of physical displacement of the body parts involved in the communication process (e.g., amplitude of head rotations or arms opening). This attribute, like all the following, is a real number defined in the interval $[-1, 1]$, where *zero* corresponds to a *neutral* behavior, that is the behavior of our virtual agent without any expressivity control.

- *Temporal Extent*: duration of movements (e.g., quick versus sustained actions). This parameter modifies the speed of execution of behaviors. Low values produce very quick movements while higher values produce slower ones. For example low values produce very fast head rotations while higher values produce slower rotations.
- *Fluidity*: smoothness and continuity of movement (e.g., smooth, graceful versus sudden, jerky). Higher fluidity allows smooth and continuous execution of movements; while lower value creates a discontinuity in the movements. Figure 1 shows the same movement executed with different fluidity values.
- *Power*: dynamic properties of the movement (e.g., weak/relaxed versus strong/tense). Higher (resp. lower) values increase (resp. decrease) the acceleration of the muscles contraction, making movements become more (resp. less) powerful. Increasing this parameter will also produce movement *overshooting*. Figure 2 shows some examples of curves with different tensions.

*2) Perceptual tests for behavior expressivity:* To validate our expressivity model, we performed two perceptual tests [35]. In the first study we aimed to evaluate the implementation of each expressivity parameter while in the second study we aimed to understand if the set of expressivity parameters would allow us to model expressive behaviors. In both studies videos showing the same behaviors but with different expressivity parameters setting were created. Subjects had to evaluate the videos and find out, for the first study, which parameter has been modified and for the second study, which expressivity setting was the closest to reach a given movement quality. Both tests gave positive results. Subjects could perceive relatively well each expressivity parameter and which movement quality was intended [35].

### B. Automatic extraction of expressivity in music performance

CUEX (CUe EXtraction) is an algorithm developed at KTH and Uppsala University for extracting acoustical cues from an expressive music performance [36] [37]. Acoustical cues that can be extracted by CUEX are articulation (*legato* or *staccato*), local tempo (number of events in a given time window), sound level (dB), spectrum energy above 1000 Hz, attack speed (dB/s), musical tone, and *vibrato*. The CUEX algorithm has been validated by testing it on real monophonic expressive performances[1] played with electric guitar, piano, flute, violin, and saxophone. In average about 90% of tone onsets were correctly detected. CUEX has also been tested with voice and gave similar results.

Research in music performance has shown that musicians control acoustic cues for communicating emotions when playing [40] [41]. Particular combinations and relative values of the cues correspond to specific emotions. In Table I we present the use of acoustic cues by musicians when performing with happiness, anger, or sadness. Complete data have been reported by Juslin [41]. The acoustic cues extracted by CUEX can be mapped into a 2-dimensional space that represents the expressivity of the performance. The 2-dimensional space is defined by the axes pleasure-displeasure (valence) and degree of arousal (activity) as proposed by Russell [42]. In the present work, acoustic cues extracted by CUEX are mapped onto this 2-dimensional activity–valence space using a fuzzy logic approach [43]. For example, if a piece of music is played with *legato* articulation, soft sound level, and slow tempo, then it will be classified as "sad"; while it will be classified as "happy" if the performance is characterized by a more *staccato* articulation, louder sound level, and faster tempo. CUEX is implemented both in `Matlab` and `PD` [44]. The latter is a simplified version of the `Matlab` one, with less precision, but it runs in real-time. In this study we used the `PD` implementation.

## V. Visualization of expressivity: from acoustic cues to an animated virtual head

In this section we turn our attention to an explicit visual representation of expressivity in music performances. The system, called Music2Greta, has been realized by interfacing the output of the acoustic features extraction system CUEX described in section IV-B with the input of the Greta virtual agent [45], see figure 3. Both components communicate through a TCP/IP socket. Acoustic cues extracted by CUEX and the corresponding emotional content are transmitted to Greta in real-time. After receiving them, the module called *Acoustic params to expressivity* applies a mapping (section V-A) between the acoustic cues and the expressivity parameters of the Greta agent (section IV-A). At the same time the *Emotion blending* module generates the facial expression that has to be assumed by the Greta's face (section V-B).

### A. Mapping acoustic cues to expressivity parameters

The acoustical cues extracted by CUEX (that is sound level, tempo, articulation) are linearly mapped into the behavior expressivity parameters using a scaling factor to adapt their ranges of variation. The variation of each expressivity parameter is as follow:

- *Sound level*. The current sound level of the music performance is linearly mapped into the *Spatial Extent* and *Power* expressivity parameters. It influences the angle of rotation of head movements (*Spatial Extent*) as well as their acceleration and quantity of overshooting (*Power*).
- *Tempo*. This parameter represents the local *tempo* of the musical performance and influences *Temporal Extent* and *Overall Activity* expressivity parameters. It acts on the duration of head movements (*Temporal Extent*), and on the frequency of head movements (*Overall Activity*).
- *Articulation*. It reflects the style and the quantity of the articulation in the music performance, i.e. the amount of *staccato* or *legato*. It varies the *Fluidity* expressivity parameter. For example it acts on the continuity of head movements making them less continuous and less co-articulated as the articulation becomes more and more *staccato*.

[1] These performances were collected, and rated in listening tests in previous experiments [38] [39]. Listeners were able to identify the intended emotions in musicians' performances.

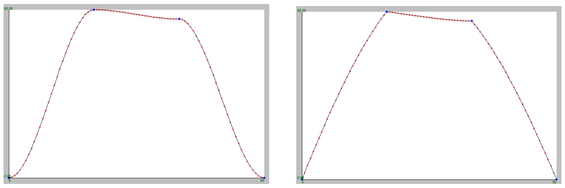

Fig. 1. Fluidity variation: left diagram represents high fluidity, right diagram represents low fluidity for the same behavior.

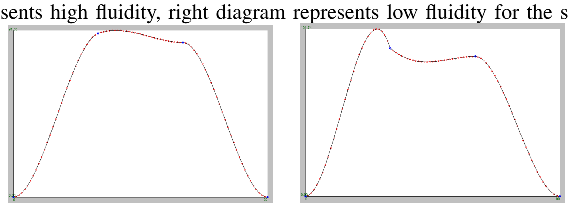

Fig. 2. Power variation: left diagram represents movement executed with low power; while the right diagram represents the same movement with high power.

TABLE I

MUSICIANS' USE OF ACOUSTIC CUES WHEN COMMUNICATING EMOTION IN MUSIC PERFORMANCE (FROM [41])

| **Emotion** | **Acoustic cues** | **Emotion** | **Acoustic cues** |
|---|---|---|---|
| Sadness | slow mean tempo | Anger | fast mean tempo |
| | large timing variations | | small tempo variability |
| | low sound level | | high sound level |
| | legato articulation | | staccato articulation |
| | small articulation variability | | spectral noise |
| | soft duration contrasts | | sharp duration contrasts |
| | dull timbre | | sharp timbre |
| | slow tone attacks | | abrupt tone attacks |
| | flat micro-intonation | | accent on unstable notes |
| | slow vibrato | | large vibrato extent |
| | final ritardando | | no ritardando |
| Happiness | fast mean tempo | | rising micro-intonation |
| | small tempo variability | | fast tone attacks |
| | small timing variations | | bright timbre |
| | high sound level | | sharp duration contrasts |
| | little sound level variability | | staccato articulation |
| | large articulation variability | | |

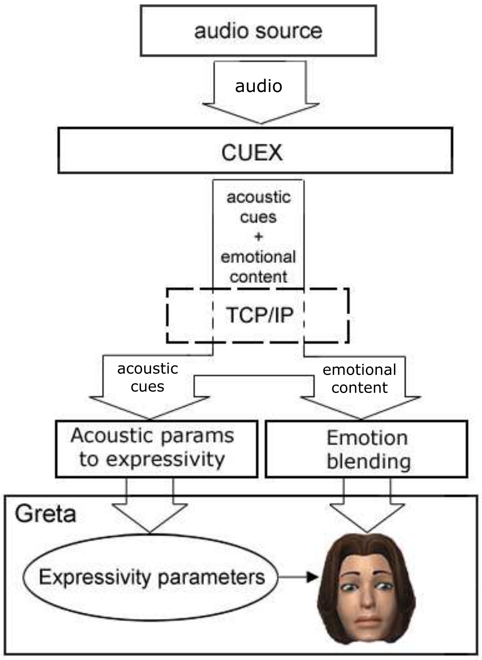

Fig. 3. Music2Greta architecture

We can notice that our mapping establishes a mimicry between sound quality and movement quality: loud sound is matched by large head movement; fast speed tempo by rapid movement; staccato performance by discontinuous movement, etc. These positively correlated relations between acoustic cues variation and behaviors quality variation have been noticed between pitch accent on emphatic word(s) and nonverbal behaviors [46].

### B. Mapping emotional intention onto the agent's face

The emotional intention recognized in the music performance by the CUEX system is mapped onto the facial expression to be displayed by the agent Greta. As described in section IV-B, the CUEX system determines the emotional content of a music performance in real-time. The result takes the form of a vector which coordinates are the amount of happiness, anger, or sadness contained in the actual performance. Thus it is possible that CUEX indicates that more than one emotion is present at the same time; that is, there is more than one emotion for which the level of arousal is greater than zero. In such a case, a blend of emotions is computed. The corresponding facial expression is obtained by applying the rules defined by P. Ekman and W. Friesen [47]. Two facial areas (*upper face* (eyes and eyebrows) and *lower face* (cheeks and mouth)) are considered [48], [49]. Facial expressions of blended emotions are computed by combining the expression shown on the upper face of one expression with the expression shown on the lower face of the other expression. Based on Ekman's research [47], expressions of negative emotions are mainly recognized from the upper face (e.g., frown of anger) while positive emotions are from the lower face (e.g., smile of happiness). We applied these finding and elaborated the following rules:

- if *anger* and *sadness* are present: the lower face shows anger (tense lips) and the upper face displays sadness (inner raise eyebrows);
- if *anger* and *happiness* are present: the lower face shows happiness (smile) and the upper face displays anger (frown);
- if *happiness* and *sadness* are present: the lower face shows happiness (smile) and the upper face displays sadness (inner raise eyebrow);
- if *anger*, *sadness* and *happiness* are all three present: the lower face shows happiness (smile) and the upper face displays sadness (inner raise eyebrow). Anger will be revealed through rapid head movements.

Emotions may have visible effect on the agent through two other aspects: skin color and head movements quality. For example, when the music performance becomes "angry" (faster attack and higher spectrum energy) the face becomes redder and leans toward the user; while for sadness emotion (slow attack and low spectrum energy) the face leans backward and becomes paler.

## VI. Testing the mapping between emotional intention and expressivity parameters

Our system is based on several mappings: from music performance to emotional intention, from emotional intention to facial expression and expressive head movement, and from expressivity parameters to animation. In earlier studies we have already conducted studies related to the first [31] and third mappings [34] (see section IV-A.2 and section IV-B). In the present tests our aim was to find out if the mapping between the emotional intention (as extracted from the acoustic cues) and the expressivity parameters of the virtual agent Greta were perceived by subjects.

### A. Experimental setup

Two groups of subjects took part at the tests. Group 1 consisted of researchers and doctoral students of music acoustics and speech technology at KTH, 3 females and 4 males, aged 25–46 (average 32), who played a musical instrument in average for 14 years. Group 2 was composed of researchers and doctoral students at the University of Paris8, 2 females and 4 males, aged 24–44 (average 32), who played a musical instrument in average for 5 years. In total, subjects were 13 and of 10 different nationalities.

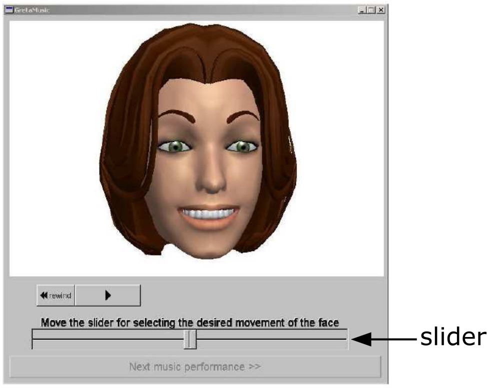
GretaMusic
rewind
Move the slider for selecting the desired movement of the face
Next music performance >>
slider

Fig. 5. Screen-shot of interface used for the tests: on top of the window there is the agent's head; below there is the slider used by the subjects for controlling the quality of the head movement.

As musical stimuli, we used performances of two melodies, Brahms' 1st theme of the *poco allegretto* 3rd movement Symphony Op.90 No.3, in C minor, and Haydn's theme from first movement of Quartet in F major for strings, Op. 74 No. 2. The two melodies were performed by a professional guitar player with three emotional intentions: anger, sadness, and happiness. The testing set of musical pieces consisted of $3 * 2 = 6$ stimuli (3 emotions, 2 melodies).

Participants were asked to sit in front of a PC where the test application was running. They were also instructed to read the explanation of the test procedure before starting the test.

For the purposes of this test we realized a slightly different version of the Music2Greta system. The test consisted in listening to the 6 pieces of music while watching the virtual agent's head moving on the screen and having the possibility of altering its movement quality (e.g., faster movement, lower amplitude) by moving a slider on the screen (figure 5 shows the interface used for the tests).

For each musical stimulus, the agent's face displayed the appropriate emotion. Each subject was instructed to change the position of the slider on the screen and see how the agent's

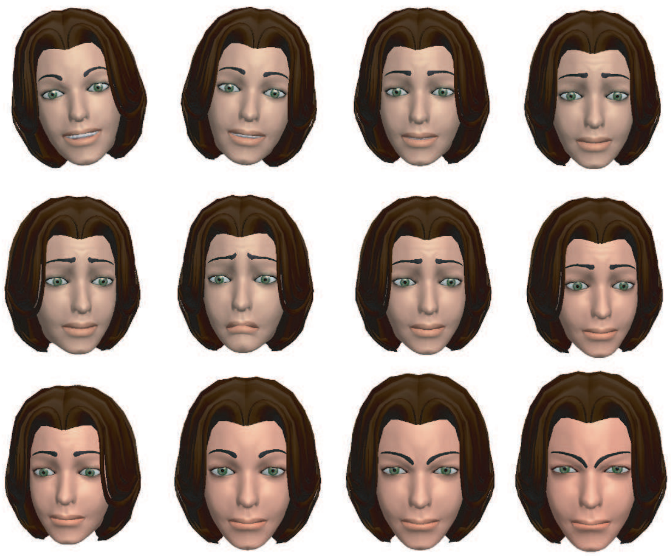

Fig. 4. This sequence shows an example of output of the Music2Greta feedback system. From left to right and top to bottom we can see the agent tilting her head on the side while displaying happiness. Then this facial expression fades to sadness as the head rotates downward. Finally its expression changes into anger with the skin reddening and the head leaning towards the user.

head changed its movement quality (the slider did not affect the emotion shown on the agent's face, which was fixed for each piece of music). When the subject found a good match between the musical stimulus and the agent's head movement she could go ahead to the next stimulus.

The slider value influenced the agent's head movement by altering the value of the 5 expressivity parameters described in IV-A. To do so, we had to create a mapping from a 1-dimension variable (the slider value) to a 5-dimension space (the expressivity parameters values).

At first, we had to decide which expressivity values should be considered as the *expected expressivity values* for each of the 3 emotions in our tests (anger, sadness, and happiness). We based our decision on perceptual studies conducted by Wallbott [21] and Gallaher [18]. We then associated this set of *expected expressivity values* to a randomly chosen value of the slider which could be anywhere along the range of variation of the slider. The position of the slider associated to the *expected expressivity values* was unknown to the subjects.

Let us give an example. Figure 6 shows a graph with the correspondence between the slider position ($X$ axis) and the 5 expressivity values ($Y$ axis). In this figure, the value $X = -0.3$ (that is $slider = -0.3$) corresponds to the pre-decided *expected expressivity values* (the dashed box) for the emotion for a given piece of music. Let us see the variation of just one of the expressivity parameters, e.g. Power ($PWR$). Position $X = 0$ in figure 6 corresponds to $PWR = 0.5$. By moving the slider towards the left, i.e. smaller $X$ values, the value of $PWR$ tends to 0, while by moving the slider towards the right $PWR$ tends to 1. Similarly the other parameters Overall Activity ($OAC$), Fluidity ($FLD$), Temporal Extent ($TMP$), and Spatial Extent ($SPC$) are simultaneously varied when adjusting the slider position.

At the end of the test, the subjects' choice, that is the final position $X$ of the slider, is compared to the value of $X = -0.3$ to check whether our set of *expected expressivity values*, $PWR$, $OAC$, $FLD$, $TMP$, and $SPC$, is correctly perceived by the subjects.

Subjects could listen to each musical stimulus as many times as they liked to, and they could constantly change the position of the slider, while watching the corresponding head's behavior on the screen. The order of the 6 musical stimuli was randomized for each subject, and the right and left extremes of the slider were randomly switched between subjects.

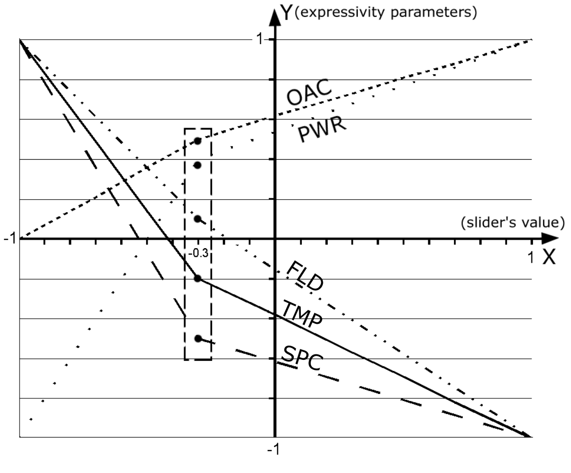

Fig. 6. Example of graph showing the correspondence between the value of the slider ($X$) and the values of the 5 expressivity parameters ($Y$). In the dashed box we highlighted the *expected expressivity values* for the given example. $PWR$ = Power, $OAC$ = Overall Activity, $FLD$ = Fluidity, $TMP$ = Temporal Extent, and $SPC$ = Spatial Extent.

### B. Results and discussion

Since the data collected for the two groups of subjects were not significantly different, they were pooled together in the analysis that follows. The emotional intention of the face and the performances (the independent variable) had a considerable effect on the listeners positioning of the slider, that controls Overall Activity, Spatial Extent, Temporal Extent, Fluidity and Power (the dependent variables). Main results for the angry, sad, and happy emotional intentions are plotted in Figure 7. As one can observe, there is an interaction of the tonality of the musical stimuli with the expressive movements chosen by the subjects. It is well know that minor tonality is often associated to sadness and major tonality to happiness. This can explain why in this experiment, for the emotional intentions sadness and happiness, subjects tend to choose similar but different expressive movements for the same facial expression depending on the musical stimulus. In particular when a facial expression was presented together with the Haydn's melody, in Major tonality, subjects tend to choose expressive movements with higher Overall Activity. For the emotional expression of anger, subjects preferred the same expressive movements for both melodies, but that differ from the expected ones.

The *expected expressivity values* were originally chosen for facial expressions that were not necessarily associated to music listening. This could partly explain the deviations from expected values observed in this experiment. Subjects' choices suggest that Greta's expressive movements should be controlled differently when providing feedback to expressive music than when speaking. Greta's expressivity settings, as given by subjects, were also influenced by the tonality, and the overall character of the musical piece.

## VII. Conclusion

In this paper, we have presented an application in which an animated virtual head is used as visual feedback on an expressive music performance. The acoustic parameters in the performance, such as tempo, sound level, and articulation, are extracted and analyzed. Their values are used to identify the emotional intention of the performer. The values of these parameters are also mapped onto the behavior expressivity parameters controlling the movement quality of the head. Finally the emotional intention and the expressivity parameters are given in input to an animated virtual agent's head that shows facial expressions of emotion, and expressive head movement. In this way we have a visualization of what we hear in a music performance, and generally in any audio stream including voice. A possible application could therefore be a virtual butler whose expressive behavior is driven by the acoustic input from the local, remote, or virtual environment. The butler would give real-time, silent, and informative feedback about the acoustic environment.

## Acknowledgment

The present work has been supported by the EU funded Human-Machine Interaction Network on Emotion Network of Excellence (HUMAINE http://emotion-research.net/) and COST 287 Action "Gesture CONtrolled Audio Systems" (ConGAS http://www.cost287.org). We are very thankful to Björn Hartmann for developing the gesture expressivity model and to Stéphanie Buisine for conducting the evaluation test of this model.

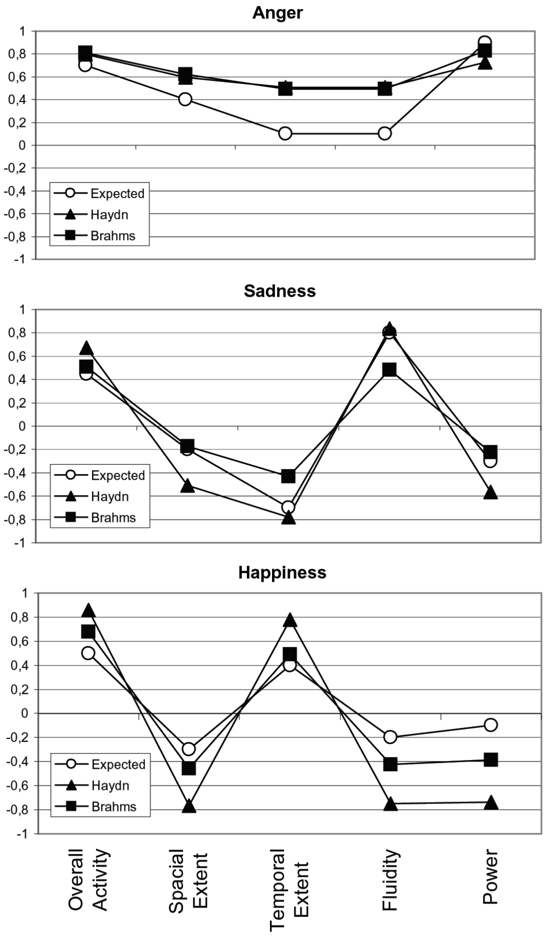

Fig. 7. Mean values for Overall Activity, Spatial Extent, Temporal Extent, Fluidity and Power as resulted from the test. Empty circles represent the expected values. Full triangles represent the mean values when the major melody was played. Full squares represent the mean values when the minor melody was played.

## References

[1] P. N. Juslin and P. Laukka, "Communication of emotions in vocal expression and music performance: Different channels, same code?" *Psychological Bulletin*, vol. 129, no. 5, pp. 770–814, 2003.

[2] C. L. Krumhansl, "An exploratory study of musical emotions and psychophysiology," *Canadian Journal of Experimental Psychology*, vol. 51, no. 4, pp. 336–352, 1997.

[3] H. Welbergen, A. Nijholt, D. Reidsma, and J. Zwiers, "Presenting in virtual worlds: Towards an architecture for a 3D presenter explaining 2d-presented information," in *Lecture Notes in Computer Science*, vol. 3814, 2005, pp. 203–212.

[4] W. L. Johnson, "Animated pedagogical agents for education training and edutainment," in *ICALT*, 2001, p. 501.

[5] J. Beskow, I. Karlsson, J. Kewley, and G. Salvi, "Synface - a talking head telephone for the hearing-impaired," in *Computers helping people with special needs - ICCHP 2004*, K. Miesenberger, J. Klaus, W. Zagler, and D. Burger, Eds., 2004, pp. 1178–1186.

[6] L. Chittaro, L. Ieronutti, and R. Ranon, "Navigating 3D virtual environments by following embodied agents: a proposal and its informal evaluation on a virtual museum application," *PsychNology Journal (Special issue on Human-Computer Interaction)*, vol. 2, no. 1, pp. 24–42, 2004.

[7] S. Kopp, L. Gesellensetter, N. Krmer, and I. Wachsmuth, "A conversational agent as museum guide – design and evaluation of a real-world application," in *Intelligent Virtual Agents*, P. et al., Ed. Springer-Verlag, 2005, pp. 329–343.

[8] E. Figa and P. Tarau, “The VISTA project: An agent architecture for virtual interactive storytelling,” in *TIDSE’2003*, N. Braun and U. Spierling, Eds., Darmstadt, Germany, 2003.

[9] H. G. Wallbott and K. R. Scherer, “Cues and channels in emotion recognition,” *Journal of Personality and Social Psychology*, vol. 51, no. 4, pp. 690–699, 1986.

[10] S. Dahl, “On the beat: Human movement and timing in the production and perception of music,” Ph.D. dissertation, Speech, Music and Hearing, KTH, Royal Institute of Technology, Stockholm, Sweden, 2005.

[11] S. DiPaola and A. Arya, “Affective communication remapping in musicface system,” in *Electronic Imaging & Visual Arts*, 2004.

[12] M. Cardle, L. Barthe, S. Brooks, and P. Robinson, “Music-driven motion editing: Local motion transformations guided by music,” *EGUK 2002 Eurographics UK Conference*, June 2002.

[13] M. Downie and N. Lefford, “Underscoring characters,” Massachusetts Institute of Technology, May 1999.

[14] H.-C. Lee and I.-K. Lee, “Automatic synchronization of background music and motion in computer animation,” in *Eurographics 2005 proceedings*, Dublin, Ireland, 2005.

[15] J. Cornwell and B. Silverman, “A demonstration of the pmf-extraction approach: Modeling the effects of sound on crowd behavior,” in *11th BRIMS, SISO*, May 2002.

[16] R. Taylor, D. Torres, and P. Boulanger, “Using music to interact with a virtual character,” in *The 2005 International Conference on New Interfaces for Musical Expression*.

[17] G. Johansson, “Visual perception of biological motion adn a model for its analysis,” *Perception and Psychophysics*, vol. 14, pp. 201–211, 1973.

[18] P. E. Gallaher, “Individual differences in nonverbal behavior: Dimensions of style,” *Journal of Personality and Social Psychology*, vol. 63, no. 1, pp. 133–145, 1992.

[19] G. Ball and J. Breese, “Emotion and personality in a conversational agent,” in *Embodied Conversational Characters*, S. P. J. Cassell, J. Sullivan and E. Churchill, Eds. Cambridge, MA: MITpress, 2000.

[20] F. E. Pollick, “The features people use to recognize human movement style,” in *Gesture-Based Communication in Human-Computer Interaction - GW 2003*, ser. LNAI, A. Camurri and G. Volpe, Eds. Springer, 2004, no. 2915, pp. 10–19.

[21] H. G. Wallbott, “Bodily expression of emotion,” *European Journal of Social Psychology*, vol. 28, pp. 879–896, 1998.

[22] K. R. Scherer, “Vocal communication of emotion: A review of research paradigms,” *Speech Communication*, vol. 40, pp. 227–256, 2003.

[23] K. R. Scherer, T. Johnstone, and G. Klasmeyer, “Vocal expression of emotion,” in *Handbook of the Affective Sciences*, R. J. Davidson, H. Goldsmith, and K. R. Scherer, Eds. New York and Oxford: Oxford University Press, 2003, pp. 433–456.

[24] K. R. Scherer, A. Schorr, and T. Johnstone, Eds., *Appraisal Processes in Emotion: Theory, Methods, Research*, ser. Series in Affective Science. Oxford University Press, 2001.

[25] A. Gabrielsson, “The performance of music,” in *The Psychology of Music*, D. Deutsch, Ed. San Diego: Academic Press, 1999, pp. 501–602.

[26] ——, “Music performance research at the millennium,” *Psychology of Music*, vol. 31, no. 3, pp. 221–272, 2003.

[27] A. Friberg and G. U. Battel, “Structural communication,” in *The Science and Psychology of Music Performance: Creative Strategies for Teaching and Learning*, R. Parncutt and G. E. McPherson, Eds. New York and Oxford: Oxford University Press, 2002, pp. 199–218.

[28] J. Sundberg, “Emotive transforms,” *Phonetica*, vol. 57, pp. 95–112, 2000.

[29] P. N. Juslin and P. Laukka, “Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion,” *Emotion*, vol. 1, no. 4, pp. 381–412, 2001.

[30] P. N. Juslin and J. A. Sloboda, Eds., *Music and emotion: Theory and Research*. New York and Oxford: Oxford University Press, 2001.

[31] P. Laukka, P. N. Juslin, and R. Bresin, “A dimensional approach to vocal expression of emotion,” *Cognition and Emotion*, vol. 19, no. 5, pp. 633–653, 2005.

[32] K. R. Scherer, “Emotion expression in speech and music,” in *Music, language, speech, and brain*, J. Sundberg, L. Nord, and R. Carlson, Eds. London: Macmillan, 1991, pp. 146–156.

[33] ——, “Expression of emotion in voice and music,” *Journal of Voice*, vol. 9, no. 3, pp. 235–48.

[34] B. Hartmann, M. Mancini, and C. Pelachaud, “Implementing expressive gesture synthesis for embodied conversational agents,” in *The 6th International Workshop on Gesture in Human-Computer Interaction and Simulation*, VALORIA, University of Bretagne Sud, France, 2005.

[35] B. Hartmann, M. Mancini, S. Buisine, and C. Pelachaud, “Design and evaluation of expressive gesture synthesis for embodied conversational agents,” in *Third International Joint Conference on Autonomous Agents & Multi-Agent Systems (AAMAS)*, Utretch, July 2005.

[36] A. Friberg, E. Schoonderwaldt, P. N. Juslin, and R. Bresin, “Automatic real-time extraction of musical expression,” in *International Computer Music Conference - ICMC 2002*. San Francisco: International Computer Music Association, 2002, pp. 365–367.

[37] A. Friberg, E. Schoonderwaldt, and P. N. Juslin, “Cuex: An algorithm for extracting expressive tone variables from audio recordings,” in *Acoustica united with Acta Acoustica*, in press.

[38] A. Gabrielsson and P. N. Juslin, “Emotional expression in music performance: Between the performer’s intention and the listener’s experience,” *Psychology of Music*, vol. 24, pp. 68–91, 1996.

[39] P. N. Juslin and E. Lindström, “Musical expression of emotions: Modeling composed and performed features,” in *Abstracts of the 5th ESCOM Conference*, 2003.

[40] A. Gabrielsson and P. N. Juslin, “Emotional expression in music,” in *Handbook of affective sciences*, H. H. Goldsmith, R. J. Davidson, and K. R. Scherer, Eds. New York: Oxford University Press, 2003, pp. 503–534.

[41] P. N. Juslin, “Communicating emotion in music performance: A review and a theoretical framework,” in *Music and emotion: Theory and research*, P. N. Juslin and J. A. Sloboda, Eds. New York: Oxford University Press, 2001, pp. 305–333.

[42] J. A. Russell, “A circumplex model of affect,” *Journal of Personality and Social Psycholog*, vol. 39, no. 6, pp. 1161–1178, 1980.

[43] A. Friberg, “A fuzzy analyzer of emotional expression in music performance and body motion,” in *Music and Music Science 2004*, J. Sundberg and W. Brunson, Eds., Stockholm, 2005.

[44] M. Puckette, “Pure data,” in *International Computer Music Conference - ICMC 1996*. San Francisco: International Computer Music Association, 1996, pp. 269–272.

[45] C. Pelachaud and M. Bilvi, “Computational model of believable conversational agents,” in *Communication in Multiagent Systems*, ser. Lecture Notes in Computer Science, M.-P. Huget, Ed. Springer-Verlag, 2003, vol. 2650, pp. 300–317.

[46] D. Bolinger, *Intonation and its Part*. Stanford University Press, 1986.

[47] P. Ekman and W. Friesen, *Unmasking the Face: A guide to recognizing emotions from facial clues*. Prentice-Hall, Inc., 1975.

[48] T. D. Bui, D. Heylen, M. Poel, and A. Nijholt, “Generation of facial expressions from emotion using a fuzzy rule based system,” in *Proceedings of 14th Australian Joint Conference on Artificial Intelligence (AI 2001)*, D. C. M. Stumptner and M. Brooks, Eds. Adelaide, Australia: Springer, 2003, pp. 83 – 94.

[49] M. Ochs, R. Niewiadomski, C. Pelachaud, and D. Sadek, “Intelligent expressions of emotions,” in *Affective Computing and Intelligent Interaction, First International Conference*, ser. Lecture Notes in Computer Science, J. Tao, T. Tan, and R. W. Picard, Eds., vol. 3784. Springer, 2005, pp. 707–714.

**Maurizio Mancini** IUT de Montreuil - Université de Paris 8 - 140 rue de la Nouvelle France 93100 Montreuil, France phone: +33 (0) 148703463 email: m.mancini(at)iut.univ-paris8.fr

**Roberto Bresin** KTH - Royal Institute of Technology, CSC - School of Computer Science and Communication, TMH - Department of Speech, Music and Hearing - Lindstedtsvägen 24 - 100 44 Stockholm, Sweden phone: +46 (8) 7907876 email: roberto(at)kth.se

**Catherine Pelachaud** IUT de Montreuil - Université de Paris 8 - 140 rue de la Nouvelle France 93100 Montreuil, France phone: +33 (0) 148703702 email: c.pelachaud(at)iut.univ-paris8.fr