

Implementing and Evaluating a Laughing Virtual Character

MAURIZIO MANCINI, DIBRIS, University of Genoa, Italy

BEATRICE BIANCARDI and FLORIAN PECUNE, CNRS-LTCl, Télécom-ParisTech, France

GIOVANNA VARNI, ISIR, Université Pierre et Marie Curie-Paris 6, CNRS UMR 7222, Paris, France

YU DING and CATHERINE PELACHAUD, CNRS-LTCl, Télécom-ParisTech, France

GUALTIERO VOLPE and ANTONIO CAMURRI, DIBRIS, University of Genoa, Italy

Laughter is a social signal capable of facilitating interaction in groups of people: it communicates interest, helps to improve creativity, and facilitates sociability. This article focuses on: endowing virtual characters with computational models of laughter synthesis, based on an expressivity-copying paradigm; evaluating how the physically co-presence of the laughing character impacts on the user's perception of an audio stimulus and mood. We adopt music as a means to stimulate laughter. Results show that the character presence influences the user's perception of music and mood. Expressivity-copying has an influence on the user's perception of music, but does not have any significant impact on mood.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**; *Graphical user interfaces*;

Additional Key Words and Phrases: HCI, virtual character, laughter, system, evaluation

ACM Reference Format:

Maurizio Mancini, Beatrice Biancardi, Florian Pecune, Giovanna Varni, Yu Ding, Catherine Pelachaud, Gualtiero Volpe, and Antonio Camurri, 2017. Implementing and evaluating a laughing virtual character. *ACM Trans. Internet Technol.* 17, 1, Article 3 (February 2017), 22 pages.

DOI: <http://dx.doi.org/10.1145/2998571>

1. INTRODUCTION

Virtual characters are a particular kind of computer interface exhibiting a human-like aspect and capable of human-like behavior. They can act as assistants, companions, and even establish a long-term rapport in humans' everyday life [Zhao et al. 2014]. In the last decade, such characters have been widely studied by the Human-Computer Interaction (HCI) community [Beale and Creed 2009] and they have been endowed both with affective [Marsella et al. 2010] and social [DeVault et al. 2014] capabilities.

This work was partially performed within the Labex SMART (ANR-11-LABX-65) supported by French state funds managed by the ANR within the Investissements d'Avenir programme under reference ANR-11-IDEX-0004-02. It has also been partially funded by: the French National Research Agency projects MOCA and IMPRESSIONS; the European Union's Horizon 2020 research and innovation programme under grant agreements No 645378 ARIA-VALUSPA and 645553 DANCE; the European Union's 7th Framework Programme under grant agreement No 270780 ILHAIRE.

Authors' addresses: M. Mancini, G. Volpe, and A. Camurri, Università degli Studi di Genova, Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria dei Sistemi (DIBRIS), Laboratorio Casa Paganini - InfoMus, Via All'Opera Pia, 13 - 16145 Genova, Italia; B. Biancardi, F. Pecune, G. Varni, and C. Pelachaud, Institut Systèmes Intelligents et de Robotique, Université Pierre et Marie Curie, ISIR - CNRS UMR 7222, 4 Place Jussieu, 75005 Paris, France; Y. Ding, Computer Graphics and Interactive Media Lab, University of Houston US.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2017 ACM 1533-5399/2017/02-ART3 \$15.00

DOI: <http://dx.doi.org/10.1145/2998571>

Laughter is an affective and social signal capable of facilitating interaction in groups of people. It communicates, for example, interest and reduces the sense of threat in a group [Grammer 1990; Owren and Bachorowski 2003]. Moreover, it can help to improve creativity [Hughes and Avey 2009] and to facilitate sociability [Dunbar 2008]. Communicative/social functions of laughter have been investigated in several works (e.g., Chapman [1983]). Pleasure can be spontaneously communicated by laughing [Provine 2001] and embarrassment can be masked with a fake laugh [Huber and Ruch 2007]. Laughter can also act as social indicator of in-group belonging [Adelswärd 1989]; it can work as speech regulator during conversation [Provine 2001]; it can also be used to elicit laughter in interlocutors, as it is very contagious [Provine 2001].

The relevance of laughter was recognized by the HCI community. The ILHAIRE Project¹ investigated how to incorporate laughter into Human-Avatar Interactions and addressed topics such as: automated detection of a user's laugh [Mancini et al. 2012; Griffin et al. 2015], influence of laughing virtual characters on a user's perception of funniness [Ding et al. 2014b], and laughter visual and acoustic synthesis.

This article illustrates an evaluation study in which a virtual character and a user jointly listen to specific kinds of music stimuli eliciting laughter. Laughter is captured by its multimodal components: voice, facial expression, and body movements. We aim to evaluate whether and how the co-presence of the laughing virtual character influences a user's perception of music stimuli and a user's mood. We also aim to measure how participants perceive the character's social presence and believability. A previous study showed that virtual characters with laughter capabilities can influence the user's perception of a joke told either by the user or by the agent [Ding et al. 2014b]. Music enables designing experiments where participants have no constraint on their visual channel, in contrast with experiments adopting stimuli like audiovisual recordings to elicit laughter. In our experiment, we free the participants' visual channel, i.e., participants sit on the side of the virtual agent and do not have to stare at the agent during the interaction. Moreover, music can have a strong power to elicit laughter: recent work of Schickele, who is also known by the name of his alter-ego character, "P.D.Q. Bach," consists of hundreds of humorist classical pieces. David Huron [2004] carried out studies on the Humor-Evoking Devices in Schickele's compositions.

The main contribution of the work presented in this article is to show that a virtual character exhibiting laughter and expressivity-copying behavior (that is, detecting the user's movement expressivity and dynamically adapting its expressivity to it) can improve the user's experience in terms of: (1) perception of music stimuli and (2) the user's mood. Compared to the existing work of Niewiadomski et al. [2013], (a) we focus on the perception of funniness of audio instead of video stimuli and (b) we exploit the expressivity-copying paradigm instead of detecting and copying the user's laughter timing (i.e., we copy *how* and not *when* the user laughs).

The study presented in this article is a follow-up of a previous study that provided weak results, reported in Pecune et al. [2015]. By slightly re-conceiving the technical setup and the analysis methodology, we ran another study where we focused more on the relevance of the physical presence of the virtual character. Moreover, the implementation of the character's copying behavior has been improved. The mapping of the detected user's body motion onto the virtual character's animation was more carefully calibrated. The copying behavior offers a smoother animation blend. These changes are described in the next paragraphs.

The first part of this article describes the virtual character we employed in our experiment. The user's movement expressivity is measured in real time by extracting the user's overall energy of movement and body leaning amplitude. The character's

¹<http://www.ilhaire.eu>.

animation is generated following two different strategies: the *copying* behavior one, in which the character's expressivity follows the user's one in real time; the *non-copying* behavior one, in which the character's expressivity is pre-defined. We designed a graphical tool that allows us to modulate the virtual character's body movements according to the user's. As demonstrated by Prepin et al. [2013], this approach allows us to establish a social coupling between the interaction participants by enhancing their reciprocal sense of engagement.

The second part of the article illustrates an evaluation study involving 28 participants in which each participant listens to music stimuli with or without the physical presence of a laughing virtual character performing copying vs. non-copying behavior. In particular, the following conditions have been tested out: (*no.Character*) the participants listen to music stimuli alone; (*yes.Character, no.Copying*) the participants listen to music stimuli with the character and the character does not copy participant's laughter expressivity; (*yes.Character, yes.Copying*) the participants listen to music stimuli with the character and the character copies participant's laughter expressivity (i.e., body energy and leaning amplitude) in real time. We first test the difference, in terms of the user's perception of music and the user's mood, between condition (*no.Character*) and condition (*yes.Character*). Then, we compared, in terms of the same dependent variables, condition (*yes.Character, no.Copying*) and (*yes.Character, yes.Copying*) to observe the difference between the presence of a virtual character performing copying vs. non-copying behavior. Finally, only for conditions (*yes.Character, no.Copying*) and (*yes.Character, yes.Copying*), we check the user's experience in terms of character's believability and social presence.

2. STATE OF THE ART

A truly interactive machine should exhibit behavioral adaptation toward the user's behavior in order to establish a social communication channel [Broekema 2011]: behavior mimicry, synchrony, and copying characterize human-human interaction and should be replicated to guarantee an effective human-machine interaction. Indeed, as argued by Nagaoka et al. [2007], nonverbal synchrony is a signal of the user's attitude toward other users or toward the interaction; Chartrand and Bargh [1994] affirmed that mimicry is exhibited to gain social influence. Bailenson and Yee [2005] confirmed such an effect (i.e., to gain social influence) in a study where a virtual character copied head movement of a human participant.

Some existing works [Castellano et al. 2012] focused on the *expressivity-copying* paradigm in virtual characters: a virtual character detects the user's expressive qualities (e.g., movement speed, amplitude, energy, and so on) and adapts its expressivity accordingly. For example, Castellano et al. [2012] investigated the mapping between the expressive characteristics of motion (e.g., velocity, acceleration, fluidity) that contribute to convey affective content in humans and the character's ones. An evaluation study demonstrates that users perceive the same emotion when it is expressed by the actors or by the virtual character performing expressivity-copying, independently from the shape of the user's gesture.

In our work, we aim to apply the *expressivity-copying* paradigm as a form of mimicry between the user and a laughing virtual character and to observe the influence of the character's presence and behavior on the user's mood and perception of funniness of music stimuli. Prepin and colleagues [Prepin et al. 2013] tested various conditions of mimicry of smile. They conducted perceptual studies with two virtual agents mimicking, or not, each other's smile, and with or without "snowball" effect; that is, with or without entrainment of the smile of one agent over the other agent. They found the agent smiling back to the other agent's smile is perceived more cooperative and engaging. In particular, when agents mimic and reinforce each others' smiles, the agents are evaluated as sharing greater mutual understanding, attention, interest, agreement,

and pleasantness. Ding et al. [2014b] studied the influence of a laughing virtual character on the user's perception of funniness of a joke. By listening to a joke told by a virtual character that laughs vs. a non-laughing character, the user evaluates the first joke funnier than the second one. Hofmann et al. [2015] have investigated the elicitation of amusement by a virtual character. Participants watch movies either alone or with the company of a virtual character. The virtual companion was either laughing at a prefixed time or at the same time as the human participants. The authors found that the participants laughed more when viewing movies with the virtual character; but they found no significant differences between the pre-scripted laughing agent and the dynamically adaptive laughing agent.

Other works attempted to model and implement laughter in virtual characters and robots. Urbain and colleagues [2012] built a multimodal interactive character system, the Laugh Machine, able to detect a user's laughter and to laugh back at the right moment. Niewiadomski et al. [2013] evaluated the impact of the Laugh Machine on the user, showing that it increases the user's perception of the funniness level of the experience (e.g., watching funny videos). Becker-Asano and colleagues [2009] investigated how to integrate laughter in humanoid robots. In their study, they found that the social effect of laughter depends on: (i) the situational context (that is, the task and the verbal/nonverbal behavior of the robot); (ii) the outer appearance of the humanoid robot; (iii) the interaction dynamics.

Compared to the aforementioned works, we use audio instead of video as stimulus, like in Urbain et al. [2012], and we do not aim to create a machine to detect the user's laughter as Niewiadomski [2013] did. Instead, our character laughs at predefined moments, performing the real-time copying of the user's expressivity. In this sense, our work is similar to that of Castellano et al. [2012] and Hofmann et al. [2015], but complementary: we ground the user-character interaction on laughter, a social signal that is not addressed in Castellano et al. [2012]; our character laughs at prefixed times, like it happens in Hofmann et al. [2015], but we test the influence on the user's perception of an expressivity-copying character vs. a non-copying one. Our work is also incremental with respect to Ding et al. [2014b]: we use audio stimuli instead of spoken text and we add expressivity-copying between the character's and the user's behavior. Similarly to the work presented by Prepin et al. [2013], we aim to evaluate the social presence of a copying vs. non-copying character. Our character is not interactive: it always laughs at prefixed times. This is why we propose to evaluate only its believability, that is, if a copying character is perceived more believable than a non-copying one.

3. ARCHITECTURE

We now describe the architecture of the laughing virtual character we exploited in the evaluation study presented in Section 4. The architecture is based on two main modules that have been integrated: the *Movement Features Extraction* module and the *Laughter Animation* module (see Figure 1).

As illustrated in Figure 1, the Movement Features Extraction module continuously extracts and communicates the user's energy of movement and body leaning to the Laughter Animation module. The user's body leaning is directly mapped onto the character's body leaning: if the user leans forward, the character leans forward as well. The user's body energy influences all the character's body movements: a high energy increases the amplitude of the character's movements, whereas a small energy reduces this amplitude. The Movement Features Extraction continuously sends messages containing such information to the Laughter Animation.

In the evaluation described in Section 4, the virtual character and the user listen to funny music stimuli together. The music stimuli are pieces created ad hoc to evoke a funny scenario (e.g., a piano player is continuously interrupted by a mosquito, see

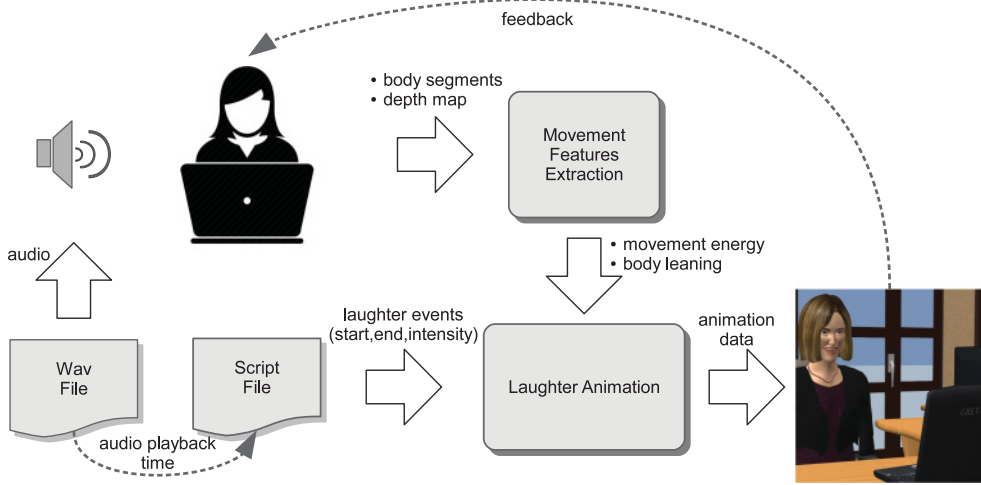


Fig. 1. The architecture of our laughing virtual character: the user, sitting beside a screen showing the virtual character (see Figure 4 for a detailed description of the experiment setup), listens to an audio stimulus (top left); the user's movement features (movement energy and body leaning) are extracted (top right); the audio stimuli triggers laughter events stored in a script file (bottom left); the virtual character's laughter animation depends on both the user's movements features and scripted laughter events (bottom right).

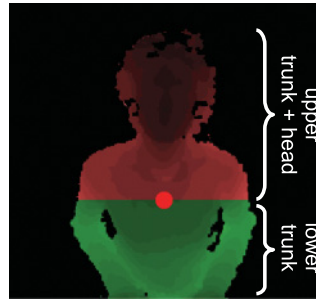


Fig. 2. Front/back body leaning: it is computed as the ratio between the mean color of the upper (including head) and lower body halves.

Section 4.4). Since automatically determining when music stimuli should trigger laughter is not in the scope of the presented work, we stored such information into Script Files containing start, duration, and intensity of each laughter-eliciting episode in audio. Details on the content of these files are provided in Section 4.4.2

3.1. Movement Features Extraction

The Movement Features Extraction module is grounded on the EyesWeb XMI platform² for real-time analysis and processing of synchronized multimodal data streams [Piana et al. 2013]. The platform consists of a visual programming interface enabling fast prototyping of applications by connecting input, processing, and output modules. EyesWeb includes libraries for gathering input for several sensors (video cameras, RGB-D sensors, microphones, and physiological sensors), for real processing of such data with particular reference to computing full-body movement features (e.g., energy,

²http://www.infomus.org/eyesweb_eng.php.

contraction, impulsivity, fluidity, rigidity, and so on), and for generation of audio and video output.

Extraction of movement features in EyesWeb follows the conceptual multi-layered framework presented in Camurri et al. [2016]. At the *Physical signals* layer, a Kinect sensor (640×480 @ 30 fps) performs reliable extraction of the user's upper-body silhouette in 3D (i.e., depth-map information). At the *Low-level signals* layer, we compute time-series of measures describing the movement being performed. In particular, we analyze upper-body movements only, focusing on body movements that are deemed important indicators of laughter. Ruch and Ekman [2001], for example, observed that laughter is often accompanied by one or more (i.e., occurring at the same time) of the following body behaviors: "rhythmic patterns," "rock violently sideways, or more often back and forth," "nervous tremor ... over the body," "twitch or tremble convulsively." Becker-Asano et al. [2009] observed that laughing users "moved their heads backward to the left and lifted their arms resembling an open-hand gesture." Markaki et al. [2010] analyzed laughter in professional (virtual) meetings: the user laughs "accompanying the joke's escalation in an embodied manner, moving her trunk and laughing with her mouth wide open" and "even throwing her head back." We decided to focus on movement energy and body leaning. According to the aforementioned studies, both high-energy (e.g., "rock violently," "tremble convulsively") and leaning (e.g., "rock ... back and forth," "moving her trunk") often accompany laughter. Moreover, they are particularly suited to be computed in real time.

Body movement energy is computed as the kinetic translation of a user's body segments, tracked by Kinect, and their percentage mass as referred by Winter [1990]. In particular, the full-body kinetic energy E_{FB} is equal to:

$$E_{FB} = \frac{1}{2} \sum_{i=0}^n m_i v_i^2, \quad (1)$$

where m_i is the mass of the i th user's body segment (e.g., head, right/left shoulder, right/left elbow, and so on) and v_i is the velocity of the i th segment, computed as the difference of the position of the segment at the current Kinect frame and the position at the previous frame divided by the sampling period.

Body leaning extraction is carried out starting from the grayscale depth map video captured by Kinect. In such an image, each pixel is a 16bit value (that is, a value in the range $[0, 65535]$), indicating the distance of that pixel from the sensor. EyesWeb uses the Microsoft Kinect SDK³ to get the grayscale depth map and to extract the user's silhouette from the background image. Body leaning is estimated by computing the mean color of the depth image of the user. We split the user's silhouette in two halves (lower and upper trunk), separated by the horizontal line passing through the user's barycenter. For both halves, we compute the mean color of the pixels belonging to each area. Finally, body leaning is the ratio between these two values. Instead of using this approach based on the depth image of the user, we could exploit the position of the user's body segments tracked by Kinect. To compute body leaning, we could compare the position of the user's shoulders vs. hips. However, in our setup, the user is sitting at her desk, thus, her hips are occluded by the desk, making it impossible to compute body leaning in such a way.

3.2. Laughter Animation

Virtual Interactive Behavior (VIB) is a platform which supports the creation of socio-emotional virtual characters [Pecune et al. 2014]. VIB has a modular and extensible architecture, where each module represents a character's functionality. We use one of

³<http://www.microsoft.com/en-us/kinectforwindows>.

the virtual characters provided by the VIB platform to carry out the experiment. We are aware that the character's rendering is not up to the current rendering standards. However, we are interested into the behavior of the agent rather than its realism. Although slightly outdated, VIB is still successfully exploited in many research projects.

In this section, we briefly describe the two most important modules that were used during the evaluation. First, we describe how facial expressions and upper body laughter animations are generated by our platform. Then, we explain how we blend the animation's parameters previously generated with the detected the user's parameters detected to obtain the animation displayed by the character.

We rely on our previous works to generate laughter animation [Ding et al. 2014a, 2014b]. In these works, data-driven approaches are applied to infer laughter motion (output signal) according to laughter audio (input signal). These approaches are first used to capture the temporal relationship between input and output signals recorded in a human audiovisual dataset; then, the relationship is rendered into output during the synthesis step. These data-driven approaches are applied on corpora collected by Urbain et al. [2010] and Ding et al. [2014a]. In Urbain et al. [2010], laughter audio and facial expressions were recorded; in Ding et al. [2014a], laughter audio and body motions, including head and upper torso motions, were captured. In both corpora, laughter audio has been segmented into 12 laughter pseudo-phonemes according to human hearing perception [Urbain et al. 2013]. So, laughter data have been segmented into sequences of laughter phonemes and their durations. Laughter prosodic features (such as energy and pitch) have been extracted using Praat [Boersma and Weenink 2001]. For the 12 laughter phonemes, 12 audiovisual sub-datasets have been built. Each sub-dataset contains segment motions as well as prosodic features, which are annotated with the same phoneme. Here, we describe briefly the three data-driven approaches, which are, respectively, used to infer lip motion, upper facial expression motion, and upper body (head and upper torso) motion.

Lip Motion: 12 parametric Gaussian models (PGMs) are used to model laughter lip motion. Each PGM is trained on an audiovisual sub-dataset. It is a Gaussian distribution whose mean vector standing for the lip motion depends on the speech features. It means that a PGM captures the mapping relationship between lip motion and prosodic features. Then, during the synthesis step, for any given input laughter phoneme and prosodic features, the mapping relationship is used to generate lip motions. More details can be found in Ding et al. [2014b, 2015].

Upper Facial Expression: Upper facial expression involves eyebrow, eyelid, and cheek. A unified framework is designed to synthesize the motion of these signals. The underlying idea is to select the segment motion from an audiovisual sub-dataset annotated by the input phoneme, and to concatenate the selected adjacent segment motions. Two criteria, called duration cost and continuity cost, are defined to select the motion segment from a sub-dataset. Duration cost is defined as the difference between input phoneme duration and the candidate segment motion duration; continuity cost is expressed as the distance between the end position of the selected last segment and the beginning position of the candidate segment. The candidate with the least weighted sum of the two costs is selected as output. Finally, the synthesized motion is obtained by concatenating and interpolating the selected samples between two successive segments. Details can be seen in Ding et al. [2014b]. None of the corpora we use have data on eye gaze motion. To avoid still perception, eyes move with a slight random movement.

Upper Body Motion: Upper body motions involve head and upper torso rotation motions. An animation generator, called Coupled Parametric Transition Loop HMM (CPTLHMM), is developed to calculate head and upper torso rotation motions. Such a framework is able to model laughter shaking-like motions and to introduce prosodic influence on the motion through transition probability parameterization. It can also

capture the dependencies between head and upper torso rotation motions. During the synthesis step, the head and upper torso motion is randomly inferred at each time frame, according to state transition probability distribution. The transition probability distribution is defined by prosodic features (input signals) at each time frame. Such framework is capable of generating shaking-like motion, considering the dependence between head and upper torso motions. More details can be found in Ding et al. [2014a].

3.3. Blending Laughter Animation with the User's Parameters

In the evaluation presented in Section 4, a virtual character is able to adapt its laughter body movement according to the user's behavior. We refer to this condition as the *yes_Copying* level of the **Copying** factor (see Section 4.3 for details). As explained earlier, the user's body leaning and his body movement energy influence, in real time, the animation of the character. While there is a direct mapping between the user's and the character's body leaning, the user's body movement energy has an overall influence on the laughter animation. If the user does not laugh at all and/or does not move at all, the character's laughter behavior will be inhibited, and the character will stay still. On the other hand, a user laughing out loud will lead the character to move even more.

To modify the generated animation of the character on the fly, we designed a graphical tool allowing us to manipulate different inputs and to blend them to obtain Body Animation Parameters (BAPs) [Pandzic and Forcheimer (Eds) 2002] for our character. This tool allows us to easily visualize and edit the influence of the user's body parameters on the output animation on the fly. We created a graph where the nodes are the inputs and outputs and the arcs represent the links between the input nodes and the output nodes. Like in neural networks, nodes have their own activation values, while arcs have their own weights. We used a similar method in the work of Charles et al. [2015] to blend the character's facial expressions with the user's detected level of empathy. Here, we blend two types of inputs: (1) BAPs generated by our laughter animation model and (2) the user's parameters (body leaning and movement energy). These inputs are respectively represented in Figure 3 by regular and dashed circles. The dotted circles represent the final BAPs that will be played by the character. The activation value for the node representing the user's parameters (dashed nodes) ranges from 0 (the user does not move at all) to 1 (the user moves a lot). For the BAPs nodes (regular and dotted nodes), the activation value ranges from $-\pi$ to π in order to be interpreted by our animation system, as specified by the MPEG-4 norm. For instance, if the value of the node *vc1 tilt* (cerebral vertebra along the neck) is set to $-\pi$, the character's head will be tilted backwards with an angle of $-\pi$. The weight of the arc linking the BAPs generated by our laughter animation model to the output BAPs depends on the level of the **Copying** factor (*yes_Copying* or *no_Copying*, see Section 4.3 for details) and the user's movement features.

The output nodes activation value is computed based on the activation function defined by McCulloch and Pitts [1943]. The final output activation value is a weighted sum of the inputs where $W_{i,j}$ represents the weights of the links, a_i represents the activation values of the input nodes, and θ_j represents the initial activation value of the output node:

$$a_j = \sum_i^n W_{i,j} * a_i + \theta_j \quad (2)$$

For the *no_Copying* level, that is, when the character does not adapt its behavior, the user's parameters nodes have no influence on the output animation. The weights linking the input BAPs nodes to the output BAPs nodes ($W_{2,3}$, $W_{5,6}$, and $W_{7,8}$ in Figure 4) are set to 1 and the laughter sequence is selected according to the laughter

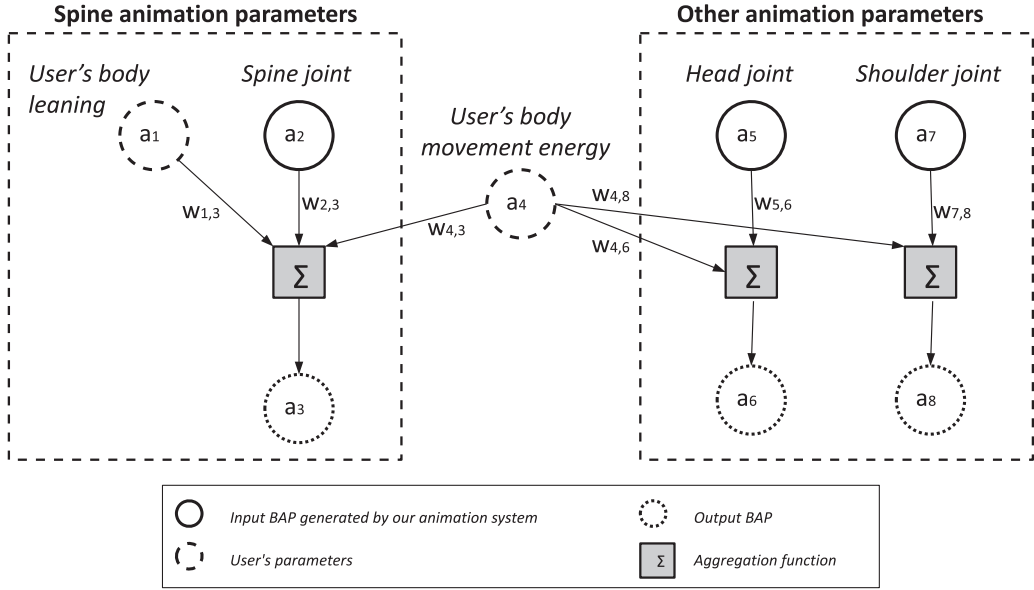


Fig. 3. Graph representing the mapping between the user's movement features and virtual character's animation parameters. The input nodes are represented as regular (character's input parameters) and dashed circles (user's features). The output nodes are represented as dotted circles (character's output parameters). The weight $W_{i,j}$ on an arc represents a factor in $[0, 1]$ determining how much the value of node i influences the value of node j .

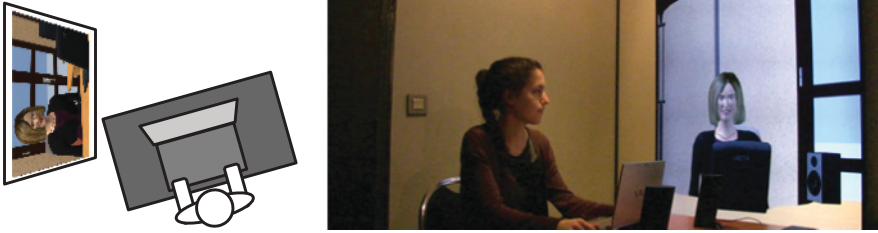


Fig. 4. Evaluation scenario: the participant is sitting at her desk with her laptop; the virtual character is projected on a screen on the wall on the left side of the participant's desk.

intensity sent from the script file as shown in Figure 1. Then, the generated animation parameters from our synthesis model are simply transmitted to and displayed by the character. For the *yes_Copying* level, however, the value of the output BAPs nodes corresponds to the weighted sum of the input BAPs nodes generated by our model and the user's body parameters. The user's body leaning influences the spine animation parameters: if the user leans forward while laughing, then the character does the same. The user's body movement energy influences not only the spine animation parameters, but also all the other body animation parameters generated by our laughter synthesis model: if the user moves with large movements, the amplitude of the character's movements is increased. On the other hand, if the user does not move at all or perform small motions, then the amplitude of the character's laughter motions decreases; it could decrease up to extremely low amplitude making the character appear to stay still. To maintain a natural laughter animation during the *yes_Copying* level, the influence of the user's body parameters on the final animation parameters (represented by the weights $W_{1,3}$, $W_{4,3}$, $W_{4,6}$, and $W_{4,8}$ in Figure 3) was calibrated empirically.

4. EVALUATION

We now describe the evaluation of a laughing character on the user's perception of the interaction. The evaluation aims to measure, through questionnaires, whether and how participants are influenced in terms of music perception, mood, and virtual character's believability in several conditions (e.g., when the character performs copying vs. non-copying behavior).

4.1. Scenario

The scenario aimed to recall an open space office, where the participant sat on his/her desk with a laptop and two loudspeakers. On the wall at the left of the participant's desk, the virtual character was projected on a screen in human size, sitting at its desk beside the user. The user's desk is positioned at nearly 90 degrees respect to the screen, as depicted in Figure 4. From her position, the participant could easily focus on both the laptop and the character. On her laptop, the participant could read introductory instructions and fill in questionnaires during the evaluation.

It is well known that, although people laugh a lot in their everyday interactions, laughter is a very difficult behavior to reproduce on demand, especially in laboratory settings [McKeown et al. 2013]. Thus, in our introductory instructions, the participant's attention was focused on the following *fake* research goal: "We are designing a virtual character that should exhibit laughter capabilities. We ask you to teach the character to laugh as a consequence of a music laugh-eliciting stimulus, that is, let yourself go and laugh when you feel the music is funny."

Using a fake story to hide true goals of an experiment is a common procedure already exploited in psychology (e.g., Platt et al. [2012]). We chose this kind of approach in order to reduce the effect for which a participant attending an experiment changes her behavior to fulfill the expectations of the researcher [French and Sutton 2010]. Moreover, since the purpose of our study was not to analyze the absolute ratings of the participants, but rather their relative difference between conditions (see Section 4.7), any influence of the fake story would apply to all the conditions, thus it would not be relevant in the context of our analyses.

Before starting the evaluation, participants were asked permission to video record the evaluation session and to use it for research purposes. We also informed them that their participation would remain totally anonymous.

4.2. Hypotheses

The evaluation study aimed to assess the following research hypotheses:

- H1** (baseline) - "The physical presence of a laughing virtual character positively affects: [**H1music**] the user's perception of the music stimulus along different dimensions (see Section 4.5.1) and [**H1mood**] the quality of the user's mood, in terms of different feelings (see Section 4.5.2)."
- H2** - "The physical presence of a *copying* laughing virtual character positively affects: [**H2music**] the user's perception of the music stimulus along different dimensions, (see Section 4.5.1) and [**H1mood**] the quality of the user's mood, in terms of different feelings (see Section 4.5.2)."
- H3** - "A *copying* laughter virtual character is perceived as more socially present and believable compared to a *non-copying* one."

The first hypothesis sets our ground truth: the physical presence of a laughing virtual character contributes to improve the perception of a music stimulus and impacts positively on the user's mood.

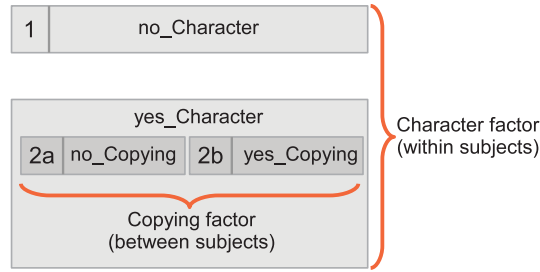


Fig. 5. Factorial design: *Character* is the within-subjects factor and *Copying* is the between-subjects factor. *Copying* is available only for the *yes_Character* level: it is related to the virtual character's behavior. Each experimental group experienced the *no_Character* level first [1], and either the *yes_Character*, *no_Copying* [2a] or *yes_Character*, *yes_Copying* [2b] second.

In addition, we check whether the physical presence of a character performing copying behavior has an influence on the user's perception of the music and her mood (H2).

Finally, in H3, we address the differences in the perception of the social presence and believability of a virtual character with different levels of copying behavior (copying vs. non-copying).

4.3. Factorial Design and Experimental Protocol

Figure 5 summarizes the experimental design we followed. The factors we manipulated during the evaluation study were:

- Character* (within subjects) - It concerns the presence of the laughing virtual character during the listening phase. It has 2 levels:
 - no_Character*: the participant listens to the music stimuli alone; that is, the virtual character does not appear on the wall on the left side of the user's desk (screen turned off);
 - yes_Character*: the participant listens to the music stimuli in the presence of the virtual character, which is displayed on her left;
- Copying* (between subjects) - It is about the different behaviors exhibited by the character during laughter. During the evaluation, the character, when present, **always laughed at prefixed times**. This factor has 2 levels:
 - yes_Copying*: the character's body expressivity (i.e., body movement energy and body leaning amplitude) follows the participant's movement expressivity.
 - no_Copying*: the character's body expressivity is pre-fixed, that is, it is not influenced by the participant's movement; the character's body expressivity is computed offline starting from the level of funniness annotated by human raters (see Section 4.4 for details).

All participants experienced both levels of *Character* factor, always in the following order: *no_Character* first, *yes_Character* second. That is, they listened to the same piece of music at first without the presence of the virtual character, and then with the presence of the character. We made this choice for one reason: following the experiment's *fake* goal (see Section 4.1), the participants were instructed to laugh when they felt that music was funny; so, by listening to music alone first, they could get familiar with the particular type of audio stimulus we exploited in the experiment.

During the second listening, that is, when experiencing the *yes_Character* level, the virtual character could copy or not the participant's movement expressivity. So, we divided participants into two independent groups, each one of them experiencing **only**

one level of the Copying factor, either the *yes.Copying* or the *no.Copying* level. The two independent groups of participants experiencing different levels of the *Copying* factor are:

- Group1*: the first group, for which the character **copied** the participant's movement expressivity; this group experienced the *yes.Copying* level;
- Group2*: the second group, for which the character **did not copy** the participant's movement expressivity; this group experienced the *no.Copying* level.

At the end of each listening (both, the one *without* and the one *with* the character), three questionnaires were administered to the participant (see Section 4.5 and the Appendix). The experiment lasted about 25 minutes for each participant, including initial instructions, hardware calibration, experiment, and debriefing.

4.4. Audio Stimuli and Laughter Scripts

For the evaluation, we designed and synthesized five ad hoc audio-only stimuli S_1, \dots, S_5 , exploiting existing research on classical music pieces that could elicit laughter in an audience. We ground on previous work of P. Schickele, or “P.D.Q. Bach,” who composed hundreds of humorist classical pieces.⁴ Schickele's compositions have been evaluated by Huron [2004] by counting and isolating audience laughter events occurring during the music performance.

Each one of our stimuli S_1, \dots, S_5 consists of a 2 minutes length audio piece in which the music content evokes a funny narration. In the following, we list the *Humor-Evoking Devices*, defined by Huron [2004], we exploited for creating our stimuli:

- Incongruous sounds*: they are sounds that do not match the expected music context; for example, during a quiet piano performance, a mosquito enters the scene and engages in a music fight with the player;
- Metric Disruptions*: a beat is eliminated or added to a measure; for example, while the Symphony No. 9, Op. 125 in D minor of L.W. Beethoven is played by a pianist, an extra beat is added at the end of a measure, creating a rhythmic pattern which is not part of the original composition;
- Implausible Delays*: they occur when a music phrase does not end at the expected time; for example, a crescendo performed by an orchestra evoking the conclusion of a music piece is repeated for an unexpected number of times or is played after an exaggerated long delay;
- Excessive Repetition*: it consists of the repetition of a music passage for a number of times out of the music norm; for example, the ring-tone of a mobile phone is repeated continuously while becoming faster and faster;
- Incongruous Quotation*: a well-known music piece is mixed with another music piece of a non-congruent style; for example, a quick cavalry charge phrase is mixed with a slow and quiet classical music piece.

4.4.1. Audio Stimuli Selection. In order to determine the two funniest stimuli among the five synthesized ones S_1, \dots, S_5 , four researchers with long-time experience in playing musical instruments were asked to continuously rate the funniness value of each stimulus. They provided their rating by regulating, through a mouse, the value of a slider going from 0 (not funny at all) to 1 (very funny) with 1,000 steps. The funniness value was sampled at 10Hz and stored in a file. The inter-rater's agreement was computed for each stimulus by using the square root of the Mean Squared Deviation (MSD). We defined a satisfactory agreement—an agreement for

⁴<http://www.schickele.com>.

which the square root of MSD is less than or equal to 0.05 (i.e., 5% of mean deviation among the ratings). With this assumption, the agreement was satisfactory for each stimulus: *agreement* – $S1 = 0.02$, *agreement* – $S2 = 0.003$, *agreement* – $S3 = 0.05$, *agreement* – $S4 = 0.02$, *agreement* – $S5 = 0.003$. Then, the Mean Funniness Curve (*MFC*) across raters was computed and normalized by the maximum funniness value among the participants. Finally, the area under the curve *AUC* of each stimulus was computed and normalized by the stimulus length to obtain a single descriptive value of the funniness. The two stimuli exhibiting the higher level of funniness ($S1$, $S3$) have been selected for the experiment. The difference between the funniness level of the two stimuli was negligible: $AUC - S1 = 0.053$, $AUC - S3 = 0.050$.

4.4.2. Laughter Scripts. As we illustrated in Section 4.3, in our experiment, the virtual character always laughs at prefixed times, exhibiting one of the following behaviors: expressivity-copying (the *yes_Copying* level) Vs. non-copying behaviors (the *no_Copying* level). We created script files containing the time at which the character has to start and stop laughing, and, for the *no_Copying* level only, the laughter intensity. For each of the 2 funniest music pieces described in the previous section, starting from the corresponding Mean Funniness Curves, we computed:

- Laughter event timing* - a laughter event starts when the normalized level of funniness *MFC* exceeds a threshold of 0.4; the event lasts until *MFC* returns below the same threshold;
- Laughter event intensity* - the laughter event intensity is the mean value of *MFC* on the interval between the laughter event's start and end.

4.5. Questionnaires

The evaluation of the perception of music stimulus, user's mood, and character's social presence was assessed by three questionnaires built from well-established questionnaires. The appendix at the end of the article reports the questions and items of the three questionnaires.

Each questionnaire was displayed on a different web page on the participant's desktop and no blank answers were allowed. Participants were asked to rate their level of agreement on a 5-point Likert scale. In order to avoid any misunderstanding about the meaning of the items, the questionnaires were translated into the participant's mother tongue, and each item was accompanied by a brief definition to avoid ambiguities.

4.5.1. Questionnaire Q1 - Music Perception. This questionnaire, mainly inspired by Ruch and Rath's work [Ruch and Rath 1993], aimed at assessing participant's perception of the music. In our questionnaire, we exploited the nine adjectives used by Ruch and Rath, referring, in their case, to the perceptive qualities of images (i.e., *witty*, *childish*, *aggressive*, *original*, *tasteless*, *subtle*, *embarrassing*, *funny*, *simple*). Then, we added three additional items adapted from Praid and Paiva [2009] (see Questionnaire 1, questions q10-q12 in the Appendix), in order to assess music likeability in an indirect way. The major changes from the original questionnaires concerned the substitution of the term "game" with "music," and the inversion of the polarity of one sentence from negative to positive.

4.5.2. Questionnaire Q2 - Participant's Mood. This questionnaire aimed at assessing the participant's mood. It includes eight adjectives used by Ruch and Rath [1993]: *excited*, *bored*, *activated*, *indignant*, *puzzled*, *angered*, *amused*, *unstimulated*.

4.5.3. Questionnaire Q3 - Character's Social Presence. This questionnaire aimed at assessing the social presence and the believability of the virtual character. It includes 12

items: q1-q4 were adapted from the “Temple Presence Inventory” by Lombard [2009], q8-q9 from Bailenson et al. [2003], and six were created by us to better address our research goals (see Questionnaire 3, questions q5-q7 and q10-q12 in the Appendix). The main adaptation we made concerned the use of assertions instead of questions, and the substitution of very general terms with those related to our study, for instance, substituting the name of our virtual character instead of “*the person*,” or “*office*” instead of “*place*.” Some example sentences we used in the questionnaire are: “*I felt as I was sharing the same office with the character*,” “*during the experiment, I could see the facial expressions of [character’s name]*,” and, “*during the experiment, I felt I could interact with [character’s name]*.”

4.6. Participants

We collected data from 28 participants (9 males, 19 females, mean age 25); all the participants were Italian, except one of them, who was Romanian, but with a high level of comprehension of the Italian written and spoken language; 46.43% had a high school degree, 28.57% had a bachelor’s degree, the 14.29% had a master’s degree; 42.86% were musicians. We checked for any influence of being musicians, but we did not find any difference in their ratings with respect to those of non-musicians (all p -values > 0.1). Thus, we will not consider it as a factor during the rest of the article.

All the participants were volunteers and they all signed a consent form.

4.7. Analysis and Results

Q1 and Q2 were analyzed by performing the t-test for each item separately. Questionnaire Q3, since all the items referred to the same dimension (Cronbach’s Alpha = 0.85, denoting a good internal consistency), was analyzed after pre-processing data by performing the following operations on them: first of all, the ratings of items referring to negative items have been inverted by applying a function which transformed the participant’s answer $a \in \{1, 2, 3, 4, 5\}$ into $\bar{a} = 6 - a$. Then, for each participant, we considered the sum of her scores over all the items.

4.7.1. Hypothesis 1.

H1music. To verify hypothesis H1fun, we questioned participants about the perception of the music stimulus without and with the physical presence of the laughing virtual character. In order to verify that there is a statistically significant difference between the two levels of the Character factor, a two-tailed paired-samples t-test was carried out for each Q1 item to compare the ratings given by all the participants in *no_Character* and in *yes_Character* condition: the results of two-tailed t-tests are shown in Table I.

A significant difference was found for *childish* ($t(27) = -2.83, p = 0.008$) and *funny* ($t(27) = 2.59, p = 0.02$); in particular:

- Music was rated as **less childish** in *yes_Character* condition ($M = 2.61, SD = 1.29$) with respect to *no_Character* condition ($M = 2.89, SD = 1.26$), see Figure 6(a). That is, the presence of the virtual character during the listening decreased participant’s perception of music childishness.
- Music was rated as **funnier** in *yes_Character* condition ($M = 3.35, SD = 1.16$) with respect to *no_Character* condition ($M = 3.0, SD = 1.25$), see Figure 6(b). That is, the presence of the virtual character during the listening increased the perceived funniness of the music.

Table I. t and p Values of Two-Tailed Paired T-Tests for Each Item of Q1

Items	<i>no_Character</i> vs. <i>yes_Character</i>	
	$t(27)$	p-value
Witty	0.63	0.54
Childish	-2.83	** 0.008
Aggressive	-1.16	0.26
Original	-0.94	0.36
Tasteless	-0.63	0.54
Subtle	-0.83	0.42
Embarrassing	-0.24	0.81
Funny	2.59	* 0.02
Simple	-1.29	0.21
Suggest to a friend	0.57	0.57
Listen again	-0.62	0.54
Longer duration	0.81	0.42

Within-subjects factor: character; $n=28$. Significance Levels: * for $p < 0.05$, ** for $p < 0.01$.

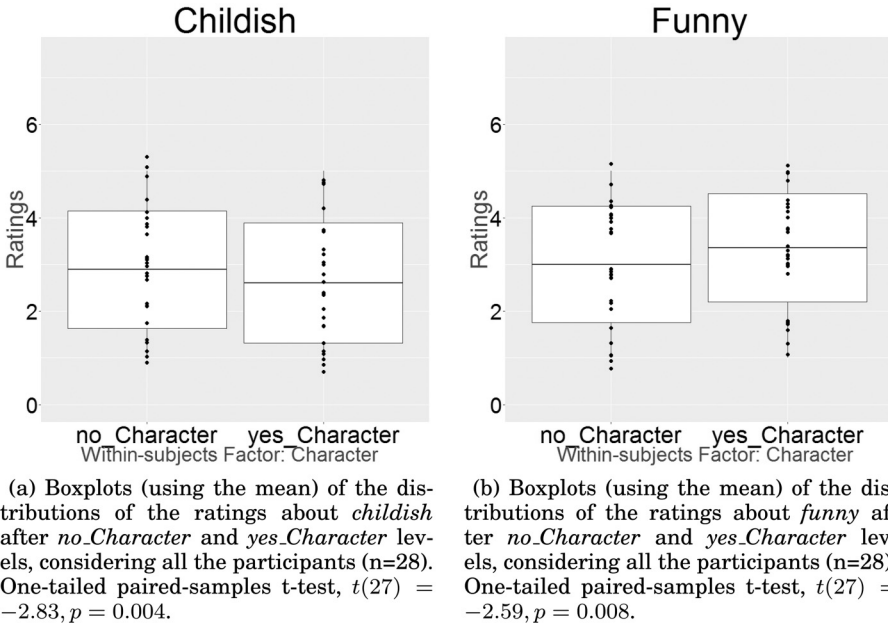


Fig. 6. Boxplots for significant adjectives of hypothesis H1music.

H1mood. Concerning hypothesis H1mood, we questioned participants about their mood after listening to music stimuli without and with the physical presence of the laughing virtual character. A two-tailed paired-samples t -test was carried out for each Q2 item to compare the ratings given by all the participants in *no_Character* and in *yes_Character* condition: the results of two-tailed t -tests are shown in Table II.

A significant difference was found for *activated* ($t(27) = 2.29$, $p = 0.02$) and *puzzled* ($t(27) = -4.26$, $p = 0.0002$); in particular:

Table II. t and p Values of Two-Tailed Paired T-Tests for Each Item of Q2

Items	<i>no_Character</i> vs. <i>yes_Character</i>	
	$t(27)$	p-value
Exhilarated	1.65	0.11
Bored	-1.61	0.12
Activated	2.29	* 0.02
Indignant	-0.44	0.66
Puzzled	-4.26	*** 0.0002
Angered	-1.44	0.16
Amused	0.72	0.48
Unstimulated	-1.14	0.26

Within-subjects factor: character; $n=28$. Significance levels: * for $p < 0.05$, *** for $p < 0.001$.

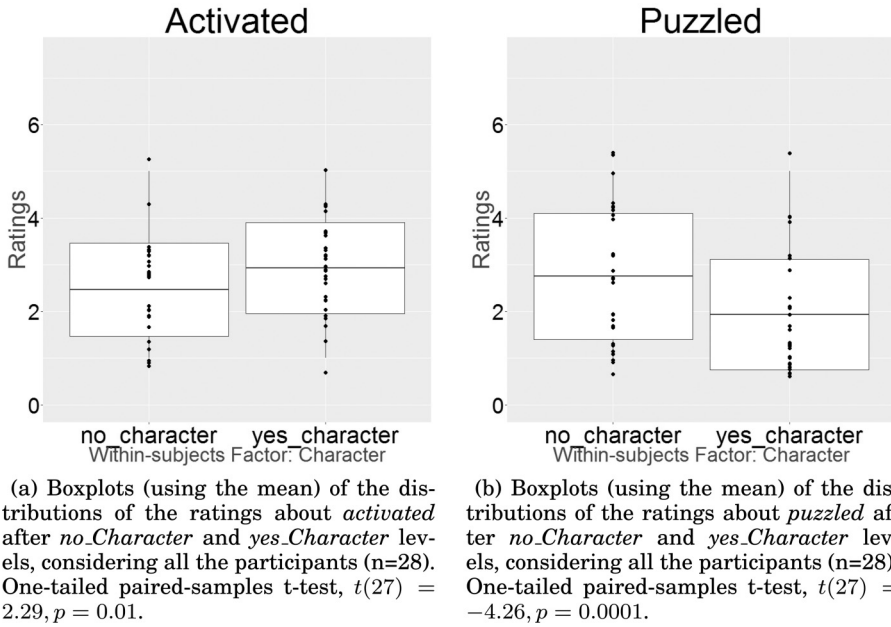


Fig. 7. Boxplots for significant adjectives of hypothesis H1mood.

- Users felt **more activated** in *yes_Character* condition ($M = 2.93, SD = 0.98$) with respect to *no_Character* condition ($M = 2.46, SD = 0.999$), see Figure 7(a).
- Users felt **less puzzled** in *yes_Character* condition ($M = 1.93, SD = 1.18$) with respect to *no_Character* condition ($M = 2.75, SD = 1.35$), see Figure 7(b).

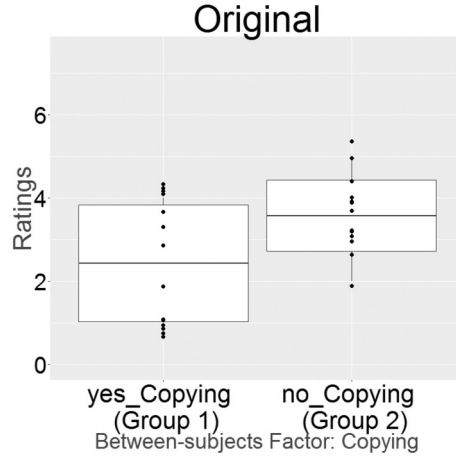
4.7.2. Hypothesis 2. In order to investigate the influence of Factor *Copying*, we compared Group 1 and Group 2 by considering only participant' ratings after the second listening, that is when the virtual character was present.

H2music. In order to investigate the effect of virtual character behavior on participant's perception of the music stimulus, a two-tailed Welch t -test was conducted to

Table III. t and p Values of Two-Tailed Welch T-Tests for Each Item of Q1

Items	<i>yes_Copying</i> vs. <i>no_Copying</i>	
	t (Welch t-test)	p-value
Witty	0.96	0.35
Childish	1.03	0.31
Aggressive	0.38	0.71
Original	2.61	* 0.02
Tasteless	0.65	0.52
Subtle	1.05	0.31
Embarrassing	0.6	0.55
Funny	0.98	0.34
Simple	1.6	0.12
Suggest to a friend	0.15	0.88
Listen again	-0.62	0.54
Longer duration	0.37	0.71

Between-subjects factor: copying; $n_1=n_2=14$. Significance Levels: * for $p < 0.05$.



(a) Boxplots (using the mean) of the distributions of the ratings about *original* in *yes.Character* condition, for Group 1 and Group 2 ($n_1=n_2=14$). One-tailed Welch t-test, $t(21.48) = 2.61, p = 0.008$.

Fig. 8. Boxplot for significant adjective of hypothesis H2music.

compare the ratings of Group 1 vs. Group 2 for each item of Q1. The results of the t-test are shown in Table III.

A significant difference was found for *original* ($t(21.48) = 2.61, p = 0.02$); in particular:

—Music was rated as **less original** in *yes_Copying* condition ($M = 2.43, SD = 1.4$) with respect to *no_Copying* condition ($M = 3.57, SD = 0.85$), see Figure 8. That is, when the agent copied the participant's movement expression, the music was perceived as less original than when the agent's behavior was pre-fixed.

Table IV. *t* and *p* Values of Two-Tailed Welch T-Tests for Each Item of Q2

Items	<i>yes_Copying</i> vs. <i>no_Copying</i>	
	<i>t</i> (Welch t-test)	<i>p</i> -value
Exhilarated	0.93	0.36
Bored	0.94	0.35
Activated	0.77	0.45
Indignant	0.91	0.37
Puzzled	1.65	0.11
Angered	0	1
Amused	0.49	0.63
Unstimulated	1.19	0.24

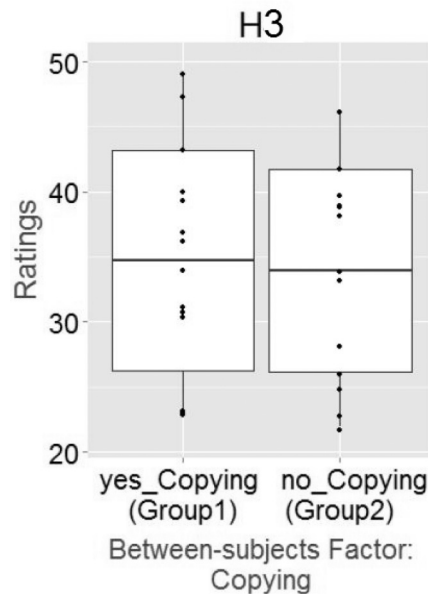
Between-subjects factor: copying; $n_1=n_2=14$.

Fig. 9. Boxplot (using the mean) for hypothesis H3. Distributions of the ratings about character's social presence and believability given by Group1 and Group2 ($n_1=n_2=14$). One-tailed Welch t-test, $t(25.82) = -0.25$, $p = 0.8$.

H2mood. In order to investigate the effect of virtual character behavior on a user's mood, a two-tailed Welch t-test was conducted to compare the ratings of Group 1 vs. Group 2 for each item of Q2. The results of the t-test are shown in Table IV.

In this case, no significant difference between the two groups was found for any of the adjectives of the questionnaire.

4.7.3. Hypothesis 3. This hypothesis is aimed to confirm that a character performing copying behavior is perceived more socially present and believable than a character not performing any copying behavior. In this case, we conducted a two-sample Welch t-test on the participants' answers to Questionnaire Q3 for the *yes_Copying* level vs. the *no_Copying* level, considering only the *yes_Character* level. No significant difference between them was found, contradicting the hypothesis: $t(25.82) = -0.25$, $p = 0.8$ (see Figure 9).

5. DISCUSSION AND CONCLUSIONS

Evaluation results show an effect of the physical presence of the virtual agent on participant's perception of the music stimulus. In particular, the same piece of music was rated by participants with statistically different degrees on two dimensions: *Childishness* and *Funniness*.

Despite using auditory stimuli instead of visual ones (images from booklets), our results are in line with the findings of [Ruch and Rath 1993], where *funny* was found to belong to a positive evaluation of humor, and *childish* to belong to a negative evaluation.

In addition, even though no statistical significance was reached, the ratings of other adjectives are in line with Ruch and Rath's findings, i.e., *witty* increased after the second listening ($M(\text{no_Character}) = 3.10$, $M(\text{yes_Character}) = 3.18$), while *simple* decreased ($M(\text{no_Character}) = 3.07$, $M(\text{yes_Character}) = 2.8$).

Regarding the influences on participants' mood, the presence of a character increased their level of activation, decreasing, at the same time, their level of puzzlement. This is an expected and strong result, as it confirms that virtual characters are capable of engaging users, increasing the quality of their interaction with computers.

Conversely, copying vs. non-copying virtual characters exhibited very limited differences in our study. Only one dimension, i.e., originality of music, was significantly affected by the copying vs. non-copying behavior of the character: music seems less original when the character exhibits copying behavior. This is not what we expected; it could be explained by the misalignment between the character's and music's expressivity in the non-copying condition, which could have induced a sort of feeling of "incoherence" in the participants. This point should be further investigated.

No significant effects could be observed on the participant's mood when watching a copying vs. non-copying character.

Finally, the type of behavior displayed by the virtual character does not seem to affect its social presence and believability, from the participant's point of view.

On the whole, these results confirm the positive role of the presence of a virtual character during an interaction, in our case, the listening of a music stimulus, and a limited importance for a virtual character to copy participant's movement expressivity to improve its laughing capabilities.

In our evaluation, we did not consider the reaction time of the virtual character as a variable: the modulation of the character's expressivity is instantaneous. Synchrony is a crucial parameter in social interaction and its relevance is well known to provide humans with a more natural and believable interaction. In future experiments, we plan to manipulate the character's reaction time in order to investigate how the perception of the music stimulus and character's believability may vary with respect to the presented results.

Another element that we aim to improve is the copying behavior of the virtual character. In addition to dynamically adapting the character's body movements, we aim to adapt the character's laughter sounds. Gaze behavior of the virtual character needs to be addressed carefully. In the present study, gaze behavior is not modeled as such. However, gazing at one partner during an interaction, and more specifically, during laughter, may enhance the impression of engagement of the participants in the interaction [Kovarsky et al. 2009]. We aim to study how these future capabilities, copying behavior and gazing, may affect the social skill of the virtual character and to measure which effect it may have on the perception of music funniness and participant's mood valence.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the article. Authors are thankful to C. Canepa from Casa Paganini - InfoMus for providing the original music stimuli exploited in the presented evaluation.

REFERENCES

- V. Adelswärd. 1989. Laughter and dialogue: The social significance of laughter in institutional discourse. *Nordic Journal of Linguistics* 12, 02 (1989), 107–136.
- J. N. Bailenson and N. Yee. 2005. Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science* 16, 10 (2005), 814–819.
- J. N. Bailenson, J. Blascovich, A. Beall, and J. M. Loomis. 2003. Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin* 29, 7 (2003), 819–833.
- R. Beale and C. Creed. 2009. Affective interaction: How emotional agents affect users. *International Journal of Human-Computer Studies* 67, 9 (2009), 755–776.
- C. Becker-Asano and H. Ishiguro. 2009. Laughter in social robotics - no laughing matter. In *Proceedings of the International Workshop on Social Intelligence Design*. 287–300.
- C. Becker-Asano, T. Kanda, C. Ishi, and H. Ishiguro. 2009. How about laughter? Perceived naturalness of two laughing humanoid robots. In *Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction*. IEEE, 1–6.
- P. Boersma and D. Weenink. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5, 9/10 (2001), 341–345.
- C. Broekema. 2011. Motion intensity matching in interaction with a virtual agent. In *Proceedings of the 14th Twente Student Conference on IT*.
- Antonio Camurri, Gualtiero Volpe, Stefano Piana, Maurizio Mancini, Radoslaw Niewiadomski, Nicola Ferrari, and Corrado Canepa. 2016. The dancer in the eye: Towards a multi-layered computational framework of qualities in movement. In *Proceedings of the 3rd International Symposium on Movement and Computing (MOCO'16)*. ACM, New York, NY, Article 6, 7 pages.
- G. Castellano, M. Mancini, C. Peters, and P. W. McOwan. 2012. Expressive copying behavior for social agents: A perceptual analysis. *IEEE Transactions on Systems Man and Cybernetics-Part A-Systems and Humans* 42, 3 (2012), 776.
- A. Chapman. 1983. Humor and laughter in social interaction and some implications for humor research. In *Handbook of Humor Research*, Vol. 1, P. E. McGhee and J. H. Goldstein (Eds.). 135–157.
- F. Charles, F. Pecune, G. Aranyi, C. Pelachaud, and M. Cavazza. 2015. ECA control using a single affective user dimension. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM, 183–190.
- T. L. Chartrand and J. A. Bargh. 1994. The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology* 76 (1994), 893–910.
- D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet, and others. 2014. SimSensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1061–1068.
- Y. Ding, J. Huang, N. Fourati, T. Artières, and C. Pelachaud. 2014a. Upper body animation synthesis for a laughing character. In *Intelligent Virtual Agents*. Lecture Notes in Computer Science, Vol. 8637. 164–173.
- Y. Ding and C. Pelachaud. 2015. Lip animation synthesis: A unified framework for speaking and laughing virtual agent. In *Auditory-Visual Speech Processing (AVSP'15)*. 78–83.
- Y. Ding, K. Prepin, J. Huang, C. Pelachaud, and T. Artières. 2014b. Laughter animation synthesis. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems (AAMAS'14)*. 773–780.
- R. I. M. Dunbar. 2008. Mind the gap: Or why humans are not just great apes. In *Proceedings of the British Academy*, Vol. 154.
- D. P. French and S. Sutton. 2010. Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? *British Journal of Health Psychology* 15, 3 (2010), 453–468.
- K. Grammer. 1990. Strangers meet: Laughter and nonverbal signs of interest in opposite-sex encounters. *Journal of Nonverbal Behavior* 14, 4 (1990), 209–236.
- H. Griffin, M. Aung, B. Romera-Paredes, C. McLoughlin, G. McKeown, W. Curran, and N. Berthouze. 2015. Perception and automatic recognition of laughter from whole-body motion: Continuous and categorical perspectives. *IEEE Transactions on Affective Computing* PP, 99 (2015).
- J. Hofmann, T. Platt, W. Ruch, R. Niewiadomski, and J. Urbain. 2015. The influence of a virtual companion on amusement when watching funny films. *Motivation and Emotion* 39, 3 (2015), 434–447.
- T. Huber and W. Ruch. 2007. Laughter as a uniform category? A historic analysis of different types of laughter. In *Proceedings of the 10th Congress of the Swiss Society of Psychology*.

- L. W. Hughes and J. B. Avey. 2009. Transforming with levity: Humor, leadership, and follower attitudes. *Leadership & Organization Development Journal* 30, 6 (2009), 540–562.
- D. Huron. 2004. Music-engendered laughter: An analysis of humor devices in PDQ Bach. In *Proceedings of International Conference on Music Perception and Cognition*. 700–704.
- Dana Kovarsky, Maura Curran, and Nicole Zobel Nichols. 2009. Laughter and communicative engagement in interaction. In *Seminars in Speech and Language*, Vol. 30. © Thieme Medical Publishers, 027–036.
- M. Lombard, T. B. Ditton, and L. Weinstein. 2009. Measuring presence: The temple presence inventory. In *Proceedings of the 12th Annual International Workshop on Presence*. 1–15.
- M. Mancini, G. Varni, D. Glowinski, and G. Volpe. 2012. Computing and evaluating the body laughter index. In *Human Behavior Understanding*, A. A. Salah, J. Ruiz-del Solar, C. Meriçli, and P.-Y. Oudeyer (Eds.). Lecture Notes in Computer Science, Vol. 7559. Springer, Berlin, 90–98.
- V. Markaki, S. Merlino, L. Mondada, and F. Oloff. 2010. Laughter in professional meetings: The organization of an emergent ethnic joke. *Journal of Pragmatics* 42, 6 (2010), 1526–1542.
- S. Marsella, J. Gratch, and P. Petta. 2010. Computational models of emotion. *A Blueprint for Affective Computing—A Sourcebook and Manual*. 21–46.
- W. S. McCulloch and W. Pitts. 1943. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics* 5, 4 (1943), 115–133.
- G. McKeown, W. Curran, C. McLoughlin, H. J. Griffin, and N. Bianchi-Berthouze. 2013. Laughter induction techniques suitable for generating motion capture data of laughter associated body movements. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 1–5.
- C. Nagaoka, M. Komori, and S. Yoshikawa. 2007. *Embodied Synchrony in Conversation*. 331–351.
- R. Niewiadomski, J. Hofmann, J. Urbain, T. Platt, J. Wagner, B. Piot, H. Çakmak, S. Pammi, T. Baur, S. Dupont, and others. 2013. Laugh-aware virtual agent and its impact on user amusement. In *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems*. 619–626.
- M. J. Owren and J.-A. Bachorowski. 2003. Reconsidering the evolution of nonlinguistic communication: The case of laughter. *Journal of Nonverbal Behavior* 27 (2003), 183–200. Issue 3.
- I. S. Pandzic and R. Forcheimer (Eds.). 2002. *MPEG4 Facial Animation - The Standard, Implementations and Applications*. John Wiley & Sons.
- F. Pecune, A. Cafaro, M. Chollet, P. Philippe, and C. Pelachaud. 2014. Suggestions for extending SAIBA with the VIB platform. In *Proceedings of the Workshop on Architectures and Standards for Intelligent Virtual Agents at IVA*. 336–342.
- F. Pecune, M. Mancini, B. Biancardi, G. Varni, Y. Ding, C. Pelachaud, G. Volpe, and A. Camurri. 2015. Laughing with a virtual agent. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems (AAMAS'15)*. 1817–1818.
- S. Piana, M. Mancini, A. Camurri, G. Varni, and G. Volpe. 2013. Automated analysis of non-verbal expressive gesture. In *Human Aspects in Ambient Intelligence*, T. Bosse, D. J. Cook, M. Neerinx, and F. Sadri (Eds.). Vol. 8. Atlantis Press, 41–54.
- T. Platt, J. Hofmann, W. Ruch, R. Niewiadomski, and J. Urbain. 2012. Experimental standards in research on AI and humor when considering psychology. In *Proceedings of the AAAI Fall Symposium: Artificial Intelligence of Humor*.
- R. Prada and A. Paiva. 2009. Teaming up humans with autonomous synthetic characters. *Artificial Intelligence* 173, 1 (2009), 80–103.
- K. Prepin, M. Ochs, and C. Pelachaud. 2013. Beyond backchannels: Co-construction of dyadic stance by reciprocal reinforcement of smiles between virtual agents. In *Proceedings of COGSCI 2013 the Annual Meeting of the Cognitive Science Society*.
- R. R. Provine. 2001. *Laughter: A Scientific Investigation*. Penguin.
- W. Ruch and P. Ekman. 2001. The expressive pattern of laughter. In *Emotion, Qualia and Consciousness*, A. Kaszniak (Ed.). World Scientific Publishers, Tokyo, 426–443.
- W. Ruch and S. Rath. 1993. The nature of humor appreciation: Toward an integration of perception of stimulus properties and affective experience. *Humor* 6 (1993), 363–363.
- J. Urbain, E. Bevacqua, T. Dutoit, A. Moinet, R. Niewiadomski, C. Pelachaud, B. Picart, J. Tilmanne, and J. Wagner. 2010. The AVLaughterCycle database. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'10)*. 19–21.
- J. Urbain, H. Çakmak, and T. Dutoit. 2013. Automatic phonetic transcription of laughter and its application to laughter synthesis. In *Proceedings of ACII*. 153–158.

- J. Urbain, R. Niewiadomski, J. Hofmann, E. Bantegnie, T. Baur, N. Berthouze, H. Cakmak, R. Cruz, S. Dupont, M. Geist, H. Griffin, F. Lingenfelser, M. Mancini, M. Miranda, G. McKeown, S. Pammi, O. Pietquin, B. Piot, T. Platt, W. Ruch, A. Sharma, G. Volpe, and J. Wagner. 2012. Laugh machine. In *Proceedings of the 8th International Summer Workshop on Multimodal Interfaces, eNTERFACE 12*. 13–34.
- D. A. Winter. 1990. *Biomechanics and Motor Control of Human Movement*. John Wiley & Sons, Inc., Toronto.
- R. Zhao, A. Papangelis, and J. Cassell. 2014. Towards a dyadic computational model of rapport management for human-virtual agent interaction. In *Intelligent Virtual Agents*. Springer, 514–527.

Received December 2015; revised September 2016; accepted September 2016