eNTERFACE 2013 project proposal

Laugh when you're winning

Principal investigators: members of the ILHAIRE Consortium (www.ilhaire.eu). In particular: Harry Griffin (University College London), Johannes Wagner (University of Augsburg), Maurizio Mancini (University of Genoa)

Abstract: Computer-based characters play an ever-increasing role in human-computer interaction, not only for entertainment but also for education, as assistants and potentially in healthcare. Such emotionally complex interactions demand avatars that can detect and synthesise emotional displays. Laughter is a ubiquitous and complex but under-investigated emotional expression. *Laugh when you're winning* will build on the work in *AVLaughterCycle* (eNTERFACE'09) and *Laugh Machine* (eNTERFACE'12) to deliver in three main areas:

First, the project will generate an avatar that is an active participant in social games, such as the yes/no game. The avatar capabilities developed for game playing will have many applications beyond simple entertainment. The complex human-avatar interaction of a game also demands considerable behavioural naturalness for the avatar to be a credible, trustworthy character.

Second, the avatar will respond to user laughter in a highly customised way by producing laughter of its own. Laughter detection and analysis among the speech, noise and body movements that occur in social games will be achieved through multimodal laughter detection and analysis of audio, video, body movements (Kinect) and respiration. Laughter decisions will integrate output from a module that will drive affiliative behaviour e.g., mimicry, in response to the detected parameters e.g., type and intensity, of users' laughter. Laughter synthesis will use a generation module enabling synthesis of new laughter sequences (as opposed to a finite repertoire of re-synthesized human laughs).

Thirdly, we will assess users' perception of avatar behaviour in the direct interaction involved in a social game. The level of emotional response displayed by the avatar will be varied (none, responsive, responsive with mimicry). Measures of users' personality will be analysed alongside short-term measures e.g., user laughter, and long-term measures of engagement e.g., mood, trust in the avatar. This spectrum of measures will test the applicability of an emotionally sensitive avatar and how its behaviour can be optimised to appeal to the greatest number of users and avoid adverse perceptions such as a malicious, sarcastic or unnatural avatar.

1. Project's objectives

Despite its ubiquity and importance in social interaction, laughter remains sparsely investigated in HCI. In order to integrate laughter detection, analysis and reactive synthesis into a real-life human-avatar interaction scenario this project aims to develop an avatar that is:

1. **Capable of playing at least one social game.** The avatar will participate in games with a single user. Suggestions for the games to be used are the yes/no game (in which players must respond to questions without saying "yes" or "no"), rock-paper-scissors, and Mad Libs (in which random words are inserted into a pre-scripted story). These games have simple rules and require only the production of one of a limited number of pre-scripted actions, questions or responses. Evaluation of the game-playing capabilities of the avatar will therefore be based on its adherence to the game rules and its correct production of the required actions. However, the games' simplicity does not preclude complex social interaction. For example, the user may show signs of enjoyment or frustration during the game and breaks between rounds of the game are often filled with laughter and conversation within a group. In order to be a credible participant the avatar must therefore be sensitive to some emotional displays.

2. C**apable of complex and reactive laughter behaviour.** The system will detect users' laughter, even in the potentially noisy and active environment of a social game. In order to achieve this, data from several modalities (video, audio, body-movements (from Kinect), and respiration) will be used. Laughter decisions will be influenced by a module which generates mimicry or other affiliative laughter behaviour in response to the precise characteristics of the user's laughter. Laughter synthesis will be driven from high-level information such as a real-time laughter intensity value. Acoustic laughter synthesis will use a novel HMM model. Facial laughter animation will be driven by pseudo-phoneme information and laughter intensities. Body animation will be driven by laughter intensity and also by the mimicry module. A more complex challenge to the system's capability will be for it to act as an emotionally expressive observer of a game played between two users who are not face-to-face with the avatar.

Having developed an avatar with the above capabilities we will use the avatar to investigate users' emotional experiences and cognitions in human-avatar interactions. For this, mood measurements pre and post the social games can be obtained to measure the change in cheerfulness, seriousness and bad mood in response to the exhilarating interaction. Furthermore, validated questionnaires which have been developed in former eNTERFACE projects can be applied to measure the believability, naturalness and perceived capabilities of the avatar.

2. Background information

The concept of a game playing robot has long intrigued humans, with examples, albeit fake, such as the Mechanical Turk in the 18th century (Poe, 1836). Games are complex social interactions and the possibility of victory or defeat can make them emotionally charged. The importance of emotional competence (the ability to detect and synthesise emotional displays) has therefore been recognised in more recent human-avatar/robot systems. Leite et al. (2012) describe an empathic chess-playing robot that detected its opponent's emotional valence. More children reported that the robot recognised and shared their feelings when it displayed adaptive emotional expressions during games.

Laughter often occurs in games due to their social context and energetic, exhilarating nature. Recognising and generating laughter during games is therefore vital to an avatar being an engaging, emotionally convincing game companion. In addition, a trend for gamification - "the use of game attributes to drive game-like behavior in a non-gaming context" (Owens, 2012) – may increase emotional expressions, such as laughter, in serious or mundane tasks. Thus an emotionally competent avatar developed for a game situation may well have value in situations such as education, exercise or rehabilitation

Previous laughter detection and response systems have generally used a limited human-avatar interaction. Fukushima et al. (2010) built a system that enhanced users' laughter activity during video watching. It comprised small dolls that shook and played pre-recorded laughter sounds in response to users' laughter. *AVLaughterCycle* aimed to create an engaging laughter-driven interaction loop between a human and the agent (Urbain et al., 2010). The system detected and responded to human laughs in real time by recording the user's laugh and choosing an acoustically similar laugh from an audio-visual laughter database. The *Laugh Machine* project endowed a virtual agent with the ability to laugh with a user as a fellow audience member watching a funny video (Urbain et al., in press). The closer interaction of a game scenario, proposed here, demands precise laughter detection and analysis and highly natural synthesised laughter. The social effect of robot laughter also depends on contextual factors such as the task, verbal and nonverbal behaviours beside laughter and the user's cultural background (Becker-Asano & Ishiguro, 2009; Becker-Asano et al., 2009). In addition social context and emotional valence have been shown to influence mimicry (Bourgeois & Hess, 2008). Therefore, in a game scenario with both positive and negative emotions, laughter and mimicry must be well-implemented in order to enhance rather than inhibit interaction.

Successful fusion of audiovisual cues for laughter detection is relatively recent (Petridis & Pantic, 2008). In the *Laugh Machine* project only audio cues were used, although sensors for multiple modalities were developed. Visual laughter synthesis has focussed on specific aspects of laughter expression e.g., respiration-related torso movements (Di Lorenzo et al., 2008) or facial expression Cosker and Edge (2009). Such systems may not be applicable for real-time, whole body animation. Until recently, work on the acoustic synthesis of laughter has been sparse and of limited success with low perceived naturalness (Lasarcyk & Trouvain, 2007; Sundaram & Narayanan, 2007). However, progress using HMM-based synthesis, which had been effectively applied to speech synthesis (Tokuda et al., 2002), have advanced the state-of-the-art (Urbain et al. 2012).

Synchronization and processing of different inputs (facial expressions, speech etc.) will be done using Social Signal Interpretation (SSI), an open framework to analyse and recognize human behaviour in real-time (Wagner 2011). Some parts of the processing - in particular the analysis of body movement - will be outsourced to EyesWeb (Camurri, 2000). The inferred user state will be distributed to the virtual agent using the SEMAINE architecture (Schroder, 2010), a distributed framework for real-time interactive agent-based systems e.g., GRETA (Niewiadomski et al., 2011), which coordinates different components facial expressions, speech etc. of the agent's behaviour both while it speaks and listens. SEMAINE is composed of a set of feature extractors, analysers and interpreters that continuously communicate with the SAIBA (Vilhjalmsson et al., 2007) compliant agent though a whiteboard message exchange system.

3. Detailed technical description

a. Technical description

Laugh when you're winning contains two overall workstreams: (i) development of the avatar game-playing capabilities, and (ii) development of avatar laughter detection, mimicry and synthesis capabilities. These workstreams will initially be separate, but will converge mid-project to produce a system that will be evaluated and generate psychological data in the final project phase.

Development of avatar game-playing capabilities

i. Formalising game rules

In order to participate in games the avatar's behaviour must be driven and constrained by the rules of the games in which it is participating. The structure and sequence of the games must therefore be formally described. This will include identifying games situations that requires action from the avatar and transitions from playing to more free interaction e.g., after one round of a game has been completed.

ii. Generating game stimuli

The games that will be used will be chosen for their relatively simple rules and structure and also because they can be played with only the production of a limited number of actions or responses. For example, the yes/no game, in which the avatar will play the role of the questioner, requires the avatar to rapidly ask questions for a limited time e.g., 1 minute, to try to elicit a yes or no response. These questions can be pre-recorded and selected for replay on the basis of output from a behavioural management module

iii. Behaviour management

This module will control the avatar's game-related responses based on the detected game-state at any time, for example whose turn it is and the overall score. Since the avatar will produce both game-related and laughter responses this module will have to work seamlessly with the dialog manager, which drives laughter behaviour. Ultimately, if possible, these two systems will be integrated into a single module.

Development of avatar laughter capabilities - Integration of existing modules

i. Multimodal laughter detection (video, audio, body movements, respiration)

The ILHAIRE project partners will provide initial working versions of the hardware and software required for detecting laughter features in the above modalities. Many of these modules were developed for *Laugh Machine* (eNTERFACE'12) but only audio was used for overall laughter detection. Therefore, one of the main aims in week 1 will be to integrate all modalities in order to increase the sensitivity and specificity of the laughter detection and increase its resilience to changes in the environment such as low light or background noise. In addition to the integration task, certain components will be improved and adapted to the specific requirements of the game scenario(s).

ii. Mimicry and non-mimicry affiliative behaviour

A key, novel feature of the avatar's behaviour will be its ability to adapt its laughter in response to the detected features e.g., type and intensity, of the user's laughter. This module will use the output from the integrated laughter detection system and apply a probabilistic model of what laughter features would be appropriate to respond to with mimicry or non-mimicry affiliative behaviour. The output from this module will be used to tune laughter synthesis, for example, by mimicking the rhythm of the user's laughter. A preliminary version of this model will be provided by the ILHAIRE partners and its parameters will be optimised during testing of the game scenarios, or in simpler pilot human-avatar interactions.

iii. Dialog manager

This component manages the agent's laughter reactions. Using the output from the multimodal detection and mimicry modules the dialog manager will determine most appropriate laughter response. If laughter is required it will output a high level description of the laugh (e.g., type and intensity). A working dialog manager will be provided by the ILHAIRE team which has been produced by analysing databases of human-human interaction and using particular reinforcement techniques to generalize to situations not encountered in the databases. Further optimisation of the dialog manager will be carried out during the project if required.

iv. Laughter synthesis

Audio synthesis will use Hidden Markov Models (HMMs), which have proven to be efficient for speech synthesis e.g., the free and widely used HMM-based Speech Synthesis System (HTS). The data used will come from the AVLaughter-Cycle database (AVLC), which contains around 1000 laughs from 24 subjects and includes phonetic transcriptions of the laughs.

Visual synthesis using Living Actor Laughter Synthesis is implemented using two different mechanisms, one for the face and another for the rest of the body. The facial animation is modelled from speech synthesis, where information about phonemes is sent by the Text to Speech engine along with the audio signal. For laughter, the visemes are composed of lip deformations but also movements of cheeks and of other parts of the face. Input data consist of pseudo-phoneme information and laughter intensities. The laughter visemes are chosen accordingly and their associated morphing function is applied with corresponding intensities.

The skeletal animation is based on Living Actor graph of animations where a target state is chosen from speech audio analysis and behaviour instructions. Transitions and loops are specifically added for laughter movements, along with a special mechanism to navigate through them depending on the type of laughter if available and its intensity.

Evaluation

Evaluation scenarios will be game sessions in which one or two users interact with the avatar. A single user will play multiple rounds of the chosen game with the avatar. In the more challenging case of two users, they will also play the game between themselves as dictated by turn-taking, while the avatar observes. The avatar will respond to users' laughter not only while the game is being played but also in rest periods between rounds during which there is often a substantial amount of laughter.

The avatar's laughter behaviour will vary according to the experimental condition:

i. No laughter

ii. Simple laughter response: user laughter is detected and the avatar laughs in response but using a fixed set of parameters.

iii. Complex laughter response: user laughter is detected and analysed. The characteristics of the avatar's laughter response are determined by the nature of the user's laughter e.g., mimicry: intense laughter eliciting similarly intense laughter. The optimal relationship between the user's and avatar's laughter will be determined by pre-evaluation pilot studies (WP3).

The effectiveness of the avatar as a game-playing character will be assessed by quantifying the ongoing flow of the game and the user's activities within it. For example, smooth, uninterrupted gameplay and user-avatar question-response or action-action patterns would be indicative of well-implemented behavioural control. Similarly, genuine user laughter in response to avatar laughter would be indicative of well-implemented laughter control as it would indicate that the avatar has laughed at a time and in a way that are perceived as natural or human-like. This is the type of measure that will be compared across the three experimental conditions.

Pre- and post-evaluation interviews will determine the effect of different avatar behaviour on the mood of the user. Validated questionnaires will the measure the believability, naturalness and perceived capabilities of the avatar. All of the measures will be analysed with respect to user factors such as sex, tendency to laugh, state and trait cheerfulness and gelotophobia. The perception of the avatar in relation to the outcome of the game will also be of interest since losing the game could make users more susceptible to negative perceptions of the avatar, for example, perceiving its laughter as mocking.

b. Resources needed

In order to maximise the presence of the avatar in a game-playing group it should be presented as close to life-size as possible. Therefore a back-projected screen or large display will be required in a separate room for system evaluation. Two powerful PCs with displays will also be required for real-time running of the system, sensors and recording equipment.

c. Project management

Members of the ILHAIRE team will coordinate the work according to the schedule laid out in section 4. We are seeking other project members with expertise in game design. Such members would play a key role in managing WP2 (formalising game structures etc.).

4.Work plan and implementation schedule

Work packages will be as follows:

- WP1: integration and synchronisation of core components provided by ILHAIRE for laughter detection. The components provided for laughter detection will be tested to overcome limitations in their performance.
- WP2: development of game-playing strategies, models and procedures. The avatar's game-playing behaviour must be established, including real-time control and the expression of pre-recorded questions and responses
- WP3: integration and optimisation of mimicry module in simple interaction. The mimicry module is an entirely novel technical aspect of this project and will need to be integrated with the laughter detection system to provide complex input to the laughter control system. Parameters to be optimised include the delay at which the avatar responds to laughter and the laughter type and intensity with which it responds.
- WP4: final integration, experimental design and feasibility testing for evaluation scenarios.
- WP5: evaluation in several game-based scenarios including user experience and psychological measures.
- WP6: analysis of evaluation and reporting

WP1 will start in week 1 with ILHAIRE partners demonstrating the capabilities of the individual modules for laughter detection. The modules will be integrated into a single multimodal platform not only for laughter detection, but also for laughter categorisation and intensity analysis.

WP2 will also start in week 1, initially with discussion of the games that the avatar could play in addition to the yes/no game. Requirements will include simplicity of rules and production of data useful for analysis. Formal descriptions of the rules for the chosen games will be drawn up and translated into behavioural controls. Speech detection systems will be tested to identify in-game user actions e.g., responses to questions. Where necessary, avatar questions, responses and actions will be recorded and annotated so that they can be activated by the behaviour manager in response to the user actions.

WP3 builds on WP1 as it uses the parameters of the detected laugh to drive its control of laughter synthesis. WP3 will therefore begin in week 2. The mimicry model will be applied to the avatar and the behaviour of the mimicry-driven avatar will be assessed in simplified interactions (e.g., simple question and answer) so that the parameters of the mimicry can be at least informally optimised prior to evaluation.

WP4 is a short work package in which the results of WPs 1-3 will be considered in finalising the experimental design and setup for evaluation and analysis. The detection, mimicry, management and synthesis modules will be integrated to produce a complete, evaluable system. This will take place at the beginning of week 3.

WP5 will begin in week 3. Users will interact with the avatar individually or in pairs. Measures of user experience e.g., trust, believability, naturalness and perception of avatar's abilities, will be recorded in conditions with varying levels of laughter response. Questionnaire-based assessment of facets of users' personality e.g., tendency to laugh, gelotophobia, will also be implemented.

WP6 analysis will immediately follow the evaluation in WP5 and will be summarised in the report prepared in week 4.

5. Benefits of the research

Development of a game-playing avatar has obvious applications in entertainment. Also the game-playing capabilities of the avatar will be useful in more serious situations that have game-like components such as repetitive goal-oriented tasks, rewards or penalties for achieving or failing to achieve a goal, collaborative or competitive action in groups. Such scenarios include educational and medical contexts e.g., physiotherapy for rehabilitation. The implementation of the laughter detection, mimicry and synthesis modules into a game scenario involving direct human-avatar interaction will be a stringent test of their effectiveness and it is expected that they will be improved substantially as a result.

The psychological data collected on the perception of the avatar, its laughter and perceived abilities will be vital in informing future iterations of avatar design. The avatar, in turn, will be a powerful psychological experimental tool allowing us to highlight the aspects of laughter behaviour, and mimicry in particular, that result in the perception of specific laughter characteristics. This is particularly relevant in light of the heterogeneity of laughter perception with respect to the observers' sex, personality and cultural background. It is possible that some users will have negative perceptions of avatar laughter in such a direct interaction. For example they may perceive the laughter as malicious, mocking or sarcastic or simply be uncomfortable with the idea of a computer that laughs in a realistically human way. In cases of gelotophobia the broader fear of laughter makes such a reaction almost inevitable. Such factors must therefore be taken into account in designing and installing avatars in roles where they will interact with the public.

The integration of multimodal signals for laughter detection will be a substantial technical achievement. Unimodal detection can be limited in its capabilities for laughter detection in adverse conditions, for example, audio detection of quiet laughter in a noisy environment or facial laughter detection in poor lighting. Exploiting multiple modalities will introduce redundancy and hence resilience into laughter detection. Similarly estimates of laughter intensity and type will be made more precise and reliable through the provision of individual estimates from multiple modalities. The degree to which different modalities and combinations of modalities effect this improvement will be assessed from the large corpus of human laughter which will be recorded during evaluation. The architecture developed for integration of the signals from multiple modalities will be of use in future work on laughter perception and human-avatar interaction.

6. Profile of the team

a. Project leaders

Johannes Wagner – PhD researcher at the University of Augsburg, Germany

Johannes Wagner graduated as a Master of Science in Informatics and Multimedia from the University of Augsburg, Germany, in 2007. Afterwards he joined the chair for Human Centered Multimedia of the same University. Among other projects, he is currently infolved in the European projects CEEDs and ILHAIRE. He is the core developer of the Social Signal Interpretation (SSI) toolkit, a general framework for real-time recognition of non-verbal behaviour in a multimodal approach.

Harry Griffin – Postdoc researcher at the University College London (UK)

Harry Griffin received his PhD in Psychology from the University of Cambridge (UK). His area of expertise is visual perception. He has worked on an EPSRC funded project on the perception of dynamic faces. He developed a method to generate an effective average photorealistic avatar onto which people's facial movements are mapped. His current research focuses on the perception of laughter from body movement and the factors that may affect the emotional contagion of laughter.

Maurizio Mancini – Postdoc researcher at the University of Genoa (Italy)

Maurizio Mancini received his PhD in Computer Science at University of Paris 8 (France) in 2008, working on expressive embodied conversational agents. Then he joined the InfoMus Lab at University of Genoa (Italy) as a post-doctoral researcher at the end of 2008. He currently works on the analysis of expressive gesture in human movement and he is still collaborating with researchers from other universities on expressive avatars.

b. Staff proposed by the leader

Thierry Dutoit – Full Professor at the University of Mons (Belgium)

Thierry Dutoit graduated as an electrical engineer and Ph.D. in 1988 and 1993 from the Faculté Polytechnique de Mons (now UMONS), Belgium, where he teaches Circuit Theory, Signal Processing, and Speech Processing. In 1995, he initiated the MBROLA project for free multilingual speech synthesis. Between 1996 and 1998, he spent 16 months at AT&T-Bell Labs, in Murray Hill (NJ) and Florham Park (NJ). He is the author of several books on Speech Synthesis and Applied Signal Processing, and he wrote or co wrote more than 20 journal papers, and more than 120 papers on speech processing, biomedical signal processing, and digital art technology. He has been an Associate Editor of the IEEE Transactions on Speech and Audio Processing (2003-2006) and the president of ISCA's SynSIG interest group on speech synthesis, from 2007 to 2010. In 2005, he initiated the eNTERFACE 4-weeks summer workshops on Multimodal Interfaces and was the organizer eNTERFACE'05 in Mons, Belgium. He was also part of the organizing committee of INTERSPEECH'07 in Antwerpen. T. Dutoit is a member of the IEEE Signal Processing and Biomedical Engineering societies, and is part of the Speech and Language Technical Committee of the IEEE since 2009. He is involved in collaborations between UMONS and ACAPELA-GROUP, a European company specialized in TTS products. Recently he founded the NUMEDIART Institute for Media Art Technology, of which he is the director.

Jérôme Urbain – PhD researcher at the University of Mons, Belgium

Jérôme Urbain graduated as an electrical engineer from the Faculté Polytechnique de Mons (FPMs), Belgium, in 2006. He then joined the Signal Processing and Circuit Theory (TCTS) Lab of the same University, where he carried works on sleep analysis, speech recognition, and is currently focusing his PhD research works on the acoustic aspects of laughter modelling, synthesis and recognition.

Hüseyin Cakmak - PhD student at the University of Mons, Belgium

Ir. Hüseyin Cakmak holds a double degree (T.I.M.E) in electrical engineering from UMONS (Mons/Belgium) and in aeronautical engineering from SUPAERO (Toulouse/France). During his master's thesis, he has been working on audio laughter synthesis using a HMM-based approach. The results obtained in a few months outperformed the state of the art in audio laughter synthesis. Those results were presented and used in the European FP7 ILHAIRE project, particularly during the Laugh Machine project (eNTERFACE 2012) to which he contributed for the audio synthesis part. After graduation, he has been chosen for a FRIA grant and now continues as a PhD Student in the field of audiovisual laughter synthesis based on 3D modelling and statistical approaches.

Radosław Niewiadomski – Postdoc researcher at Telecom ParisTech, France

Radosław Niewiadomski received the PhD degree in Computer Science in 2007 from the Universita degli Studi di Perugia, Italy. Currently he is a post-doc at the Telecom ParisTech. His research interests include embodied conversational agents, recognition and synthesis of emotions, interactive multimodal systems.

Jenny Hofmann - PhD researcher at the University of Zurich, Switzerland

Jenny Hofmann is a PhD candidate at the University of Zurich. She received a M.Sc. degree in psychology from Zurich. She studied the phenomenon of "laughing at oneself" experimentally and more recently the fear of being laughed at (i.e., gelotophobia), in a project founded by the Swiss National Foundation (SNF). She is a certified and experienced coder of the Facial Action Coding System (FACS).

Tracey Platt – PhD researcher at the University of Zurich, Switzerland

Tracey Platt is a PhD candidate at the University of Zurich. She received a M.Sc. degree in occupational psychology from Hull University, UK. She studied ridicule and teasing in relation to the fear of being laughed at (i.e., gelotophobia). Further research in gelotophobia involves different perspectives (emotions, cultural differences, aged populations, stutterers), partly in a project founded by the Swiss National Foundation (SNF). She is a certified FACS-coder.

Florian Lingenfelser – PhD researcher at the University of Augsburg, Germany

Florian Lingenfelser received his M.Sc. degree in Informatics and Multimedia from the University of Augsburg, Germany, in 2009. In 2010 he joined the chair for Human Centered Multimedia of the same University as PhD student. He is currently contributing to multimodal data fusion within the CEEDS project and developing the Social Signal Interpretation (SSI) framework.

Willibald Ruch – Professor for Personality Psychology and Assessment at the University of Zurich, Switzerland

Willibald Ruch is psychologist by training and full Professor of Personality Psychology and Assessment at the University of Zurich. He has worked on the expressive pattern of laughter, exhilaration and humour since 1978. He is specialized in conducting experiments on humour and laughter and has designed several questionnaire measurements on humour- related traits. He is the president of the International Society for Humour Studies (ISHS) and runs the International Summer School on Humour and Laughter Theory Research and Application.

Gary McKeown – Senior researcher at the Queen's University Belfast (UK)

Gary McKeown is a cognitive psychologist at the School of Psychology, Queen's University Belfast. He received his BA and PhD at QUB, his doctoral thesis concerned implicit learning in the control of complex systems. His research focuses on communication, with interest in risk perception and decision making in environmental and health settings. This led to an interest in emotion and in particular the inter-relationship of cognition and emotion. Recent research has focused on emotion, social signal processing and the cross-cultural emotion perception, most recently the SEMAINE project on emotion-sensitive avatars, where he has been responsible for recording large-scale data collections.

Nadia Berthouze – Senior Lecturer at the University College London (UK)

Nadia Berthouze is currently a Senior Lecturer in the UCL Interaction Centre (UCLIC) at the University College London. Her main expertise is the study of body movement in affective computing, and how it affects the design of systems for providing their users with a rich and positive affective experience. Her current research focuses on studying body movement as a medium to induce, recognize, and measure the quality of experience of humans interacting and engaging with/through whole-body technology. She is also investigating the various factors that affect such experience, including emotional contagion and cross-cultural differences.

Olivier Pietquin – Professor and head of Machine Learning and Interactive Systems group, Supélec, Metz (France)

Olivier Pietquin obtained an Electrical Engineering degree from the Faculty of Engineering, Mons (FPMs, Belgium) in June 1999 and a PhD degree in April 2004. In 2011, he received the Habilitation à Diriger des Recherches (French Tenure) from the University Paul Sabatier (Toulouse, France). He joined the FPMs Signal Processing department (TCTS Lab.) in September 1999. In 2001, he has been a visiting researcher at the Speech and Hearing lab of the University of Sheffield (UK). Between 2004 and 2005, he was a Marie-Curie Fellow at the Philips Research lab in Aachen (Germany). Now he is a Professor at the Metz campus of the Ecole Superieure d'Electricite (Supelec, France), and headed the "Information, Multimodality & Signal" (IMS) research group from 2006 to 2010 when the group joined the UMI 2958 (GeorgiaTech - CNRS) international lab. In 2012, he heads the Machine Learning and Interactive Systems group (MaLIS). From 2007 to 2011, he was also a member of the IADI INSERM research team (in biomedical signal processing). He is a full member of the UMI 2958 (GeorgiaTech - CNRS) since 2010 and coordinates the computer science department of this international lab. Since 2010, Olivier Pietquin sits at the IEEE Speech and Language Technical Committee and he is a Senior IEEE member since 2011. His research interests include spoken dialog systems evaluation, simulation and automatic optimisation, machine learning, speech and signal processing. He authored or co-authored more than 100 publications in these domains.

Bilal Piot – PhD student, Supélec, Metz (France)

Bilal Piot obtained an Engineering master degree from SUPELEC in 2010 and a Mathematical master degree in differential partial equations in 2010. In 2011, he obtained a master degree in applied mathematics and worked as a trading risk controller at BNP Paribas (London). He is currently a PhD student at SUPELEC and works in the inverse reinforcement learning field.

Émeline Bantegnie – R&D engineer at Cantoche, France

Émeline Bantegnie graduated in 2011 from ENSIMAG, one of top French engineering schools. She is part of Cantoche technical team that engages in research and development of the Living Actor™ software suite. Her main area of expertise is C++ and Java programming, and 3D computer graphics. She created the last version of Avatar Maker, a software programme used by Cantoche to create avatars. She also implemented new mechanisms for real time 3D animations of laughter.

c. Other researchers needed

We encourage researchers not involved in ILHAIRE to participate in this work and look forward to working with them. We are particularly looking for experts in the following domains:

i. Speech recognition

A crucial part of controlling the avatar's game and emotional responses will be detecting verbal responses from the user(s). For example, in the yes/no game the beginning and end of user's responses must be detected in order for the avatar to ask questions at the appropriate moment. Also, and more challengingly, the specific responses "yes" and "no" must be detected within the user's speech in order to mark the end of the game and trigger a suitable response e.g., laughter, in the avatar. The contribution of an expert in this field would be valuable as we could minimise the manual intervention required in this and similar scenarios by implementing automatic recognition and integrating it into the behavioural control module.

ii. Data streams in the network and/or Smartphone application design

The application design requires the simultaneous display of game-related information and an agent on a screen, as well as the capture of audio-visual data during the experiment. An expert in network communication and design of interactive (driven by network signals) web application would be helpful to manage the information streams. Also, we would have the possibility to port the laughing agent application to a Smartphone or Tablet if we are joined by an expert in the domain.

iii. Mixed reality (AR, holograms)

The believability of the agent will depend on its display as well as its attitude and communicative capabilities. It could be interesting to integrate the agent in a virtual environment. A specialist in augmented reality, 3D display (or holograms) would allow exploring these dimensions and increasing the impact of the application.

iv. Physiological sensors and processing of physiological signals

In addition to the modalities mentioned above, physiological sensors could be used during the experiments, to measure signals like heart rate, finger pulse pressure, etc. and analyse their evolution during the experiments. Such signals could also serve as input for inferring the user state and drive the dialog manager. A researcher with expertise on this type of data recording and processing physiological signals would be helpful for these developments.

v. Game design and AI

The games suggested in this proposal are relatively simple such that their implementation will be achievable within the limited timeframe of the project. The use of more complex games would be welcome as a wider test of the avatar's behavioural and emotional competence. We would therefore welcome researchers with an interest in game design, formal descriptions of games and game-behaviour, and the use of AI in games, for example, for controlling non-player characters.

7. References

Becker-Asano, C., Ishiguro, H. (2009), Laughter in Social Robotics – no laughing matter, 8th International workshop on Social Intelligence Design, Kyoto.

Becker-Asano, C., Kanda, T., Ishi, C., Ishiguro, H. (2009), How about laughter? Perceived naturalness of two laughing humanoid robots, Affective Computing and Intelligent Interaction, Amsterdam, 49-54.

Bourgeois, P. and U. Hess. (2008). The impact of social context on mimicry, Biological Psychology, 77(3), 343-352.

Camurri, A., Hashimoto, S., Ricchetti M., Ricci A., Suzuki, K., Trocca, R., Volpe, G. (2000). EyesWeb: Toward Gesture and Affect Recognition in Interactive Dance and Music Systems. Comput. Music J. 24, 1 (April 2000), 57-69.

Cosker, D., Edge, J. (2009). Laughing, Crying, Sneezing and Yawning: Automatic Voice Driven Animation of Non-Speech Articulations. In Proc. of Computer Animation and Social Agents (CASA09), pp. 21-24.

DiLorenzo, P. C., Zordan, V. B., Sanders, B. L. (2008). Laughing Out Loud : Control for modeling anatomically inspired laughter using audio. ACM Transactions on Graphics 27, 1.

Fukushima, S., Hashimoto, Y., Nozawa, T., Kajimoto, H. (2010). Laugh enhancer using laugh track synchronized with the user's laugh motion. In Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI '10), pp. 3613-3618.

Lasarcyk, E. and J. Trouvain. 2007. Imitating conversational laughter with an articulatory speech synthesis. In Proceedings of the Interdisciplinary Workshop on the Phonetics of Laughter, Saarbrücken, Germany, August 4-5, 43–48.

Leite I, Castellano G, Pereira A, Martinho C, Paiva A. 2012. Modelling empathic behaviour in a robotic game companion for children: an ethnographic study in real-world settings. Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction. :367–374

Niewiadomski, R., Bevacqua, E., Quoc Anh Le, Obaid, M., Looser, J., Pelachaud, C. (2011). Cross-media agent platform, 2011 Web3D ACM Conference, Paris, France, pp. 11-19.

Owens, M.D. (2012). It's all in the game: gamification, games and gambling. Gaming Law Review and Economics 16(3). DOI: 10.1089/glre.2012.1634

Poe E.A., (1836) Maelzel's Chess-Player, Southern Literary Messenger, 2:318-326

Schröder, M. (2010). The SEMAINE API: Towards a Standards-Based Framework for Building Emotion-Oriented Systems, In: Advances in Human-Computer Interaction, Volume 2010

Sundaram, S.and S. Narayanan. (2007). Automatic acoustic synthesis of human-like laughter. Journal of the Acoustical Society of America, 121(1):527–535.

Tokuda, K., Zen, H., Black, A., (2002). An HMM-based speech synthesis system applied to english. In: Proceedings of the 2002 IEEE Speech Synthesis Workshop. pp. 227-230. Santa Monica, California (September 2002)

Urbain, J., Niewiadomski, R., Bevacqua, E., Dutoit, T., Moinet, A., Pelachaud, C., Picart, B., Tilmanne, J., Wagner, J. (2010) AVLaughterCycle. Enabling a virtual agent to join in laughing with a conversational partner using a similarity-driven audiovisual laughter animation, Journal of Multimodal User Interfaces, vol. 4, n. 1, pp. 47-58.

Urbain, J., Cakmak, H., Dutoit, T. (2012), "Development of HMM-based acoustic laughter synthesis", Interdisciplinary Workshop on Laughter and other Non-Verbal Vocalisations in Speech, Dublin, Ireland, October 26-27.

Urbain, J., Niewiadomski, R., Hofmann, J., Bantegnie, E., Baur, T., Berthouze, N., Cakmak, H., Cruz, T., Dupont, S., Geist, M., Griffin, H., Lingenfelser, F., Mancini, M., Miranda, M., McKeown, G., Pammi S., Pietquin, O, Piot, B., Platt, T., Ruch, W., Sharma, A., Volpe, G., Wagner, J. (In press) Laugh Machine. Proc. eNTERFACE'12, Metz

Vilhjalmsson, H., Cantelmo, N., Cassell, J., Chafai, N.E., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A.N., Pelachaud, C., Ruttkay, Z., Thorisson, K.R., van Welbergen, H., van der Werf, R.J. (2007). The behavior markup language: Recent developments and challenges. In: Proc. of Intelligent Virtual Agents. (2007)

Wagner, J., Lingenfelser F., André, E. The Social Signal Interpretation Framework (SSI) for Real Time Signal Processing and Recognition. in Proceedings of Interspeech, 2011.