Toni Vanhala and
Jenni Anttonen (Eds.)

Proceedings of EHTI'08: The First Finnish Symposium on Emotions and Human-Technology Interaction

DEPARTMENT OF COMPUTER SCIENCES
UNIVERSITY OF TAMPERE

D-2008-3

TAMPERE 2008

UNIVERSITY OF TAMPERE
DEPARTMENT OF COMPUTER SCIENCES
SERIES OF PUBLICATIONS D
D-2008-3, MAY 2008

Toni Vanhala and
Jenni Anttonen (Eds.)

Proceedings of EHTI'08: The First Finnish Symposium on Emotions and Human-Technology Interaction

DEPARTMENT OF COMPUTER SCIENCES
FIN-33014 UNIVERSITY OF TAMPERE

ISBN 978-951-44-7383-8
ISSN 1795-4274

Preface

This book contains the proceedings of the First Finnish Symposium on Emotions and Human-Technology Interaction (EHTI'08) held in Tampere, Finland, May 30, 2008.

The event was organized as a meeting point for researchers studying emotions in the context of human-technology interaction in Finland. We were pleasantly surprised at the quality and quantity of this research which you can judge from the papers in these proceedings. We were even further motivated by the many researchers who expressed their support and enthusiasm for this kind of an event. It seems it was truly a time for a get-together.

We thank all the people who contributed to the success of the event. Antti Salovaara, the head of SIGCHI Finland, was the first to propose organizing this kind of an event. We thank him for the great idea! Erkki Mäkinen from the Department of Computer Sciences at the University of Tampere helped us in preparing the proceedings and publishing them in collaboration with the department. The sponsors made it possible to organize the event in a pleasant venue with good food and drinks, and disseminate the results as paper[1] and bits. Finally, our special thanks to all who contributed their time and effort to make this event a success by writing up their research and presenting it to all participants.

We hope that the readers of these proceedings will find them inspiring and a useful reference to emotion and technology related research in Finland. Now, we have a map of who and where we are, so let us collaborate!

Jenni Anttonen
Toni Vanhala
Organizers of EHTI'08

[1]You may use identical page numbering to reference the printed version:
T. Vanhala and J. Anttonen (Eds.) Proceedings of EHTI'08: The First Finnish Symposium on Emotions and Human-Technology Interaction. University of Tampere, Department of Computer Sciences, A-2008-3. ISBN 978-951-44-7347-0. ISSN 1459-6903.

Organization

EHTI'08 was organized by

Jenni Antonen[1,2] Toni Vanhala[2]
Jenni.Anttonen@cs.uta.fi *Toni.Vanhala@cs.uta.fi*

[1]SIGCHI Finland
[2]Research Group for Emotions, Sociality, and Computing (ESC)
Tampere Unit for Computer-Human Interaction (TAUCHI)
Department of Computer Sciences
University of Tampere

We thank our sponsors

SIGCHI Finland promotes human-computer interaction (HCI) research and teaching in Finland. SIGCHI Finland aims to bring together people with an interest in HCI by organizing seminars and events on different HCI-related themes. The organization is a member of the international ACM SIGCHI.

Tampere Unit for Computer-Human Interaction (TAUCHI) is a part of the Department of Computer Sciences at the University of Tampere. TAUCHI carries out multidisciplinary research and development in human-technology interaction harmonizing the potential of technology with human abilities, needs, and limitations.

The Department of Computer Sciences studies and teaches two main subjects at the University of Tampere. The research and education in Computer Science includes topics such as software development, algorithm analysis, data analysis, conceptual modelling, database design, computer security, health informatics, and electronic commerce. Interactive Technology focuses on multidisciplinary studies of human-technology interaction. TAUCHI organizes the research and teaching of this subject.

Contents

Foreword

In the wide spectrum of emotion research community the significance of functional human emotion systems for all human behavior is currently well understood. We know that emotions modulate and regulate our interactions, motivations, and cognitive processes continuously and significantly in many ways. To state a long research story in brief I am arguing, as before, that emotions are a necessary precursor for all human rational behavior! A quite recent extension in emotion research has been the incorporation of research dealing with human-technology interaction. This relates to the notion of bringing machines from passive tools closer to the process that can be called communication.

From the perspective of using computing technology as tools we know, but rarely notice, the fact that current computers and user interfaces actually function quite well. They are effective in computing and relatively easy to use. From the perspective of interacting with technology that has a high computing power we know that new research problems have developed in parallel with the improvement of technology itself. Indeed, one of the recently recognized problems relates to the dramatic emotional states that working with computers can evoke in users. An early example of this is a study called “Rage Against the Machine” which found that users swear, consider causing damage, deliberately pull out the plug, and even actually kick their computers while working with them. So, working with computers can evoke intense negative emotional reactions.

Looking a bit deeper these examples also imply that people actually treat computers as if they were social entities which they would like to threaten and punish in order to make them operate better in the future. According to the notion of “Computers Are Social Actors” people have so strong tendency towards social interaction that we treat the computers as if they had human abilities to understand us. At the same time we as users forget that machines have basically no means to observe, analyze, interpret, and respond to us in any even close to intelligent ways. Thus, from this starting point the huge amount of research and technology development needed in the area of affective computing exposes itself.

There are a multitude of approaches to affective computing internationally. These approaches cover aspects from basic research questions like how pupil size variation could be used as some kind of indication of user’s emotional state to a computer to the building emotionally educative and entertaining systems. I think that the multitude of these approaches is nicely reflected in this national symposium called “The first Finnish Symposium for Emotions and Human-Technology Interaction”. The 11 submissions for this first symposium cover a wide range of activities varying from laboratory studies investigating basic human reactions to authentic and synthetic stimuli, to the construction of interactive systems that evoke, perceive, and adapt to emotions. The common goal is the knowledge and understanding of the emotional and social processes that are present when interacting with technology. The submissions also represent the diversity of emotion related processes in terms of the possible modalities which include speech, facial expression, subjective experiences, physiology, and bodily expression of emotion. In sum, there is quite a lot of interest and activities in this field for our small country.

I hope that this symposium will act as pleasantly arousing sparkle for future exchange in this important area connecting people and technology.

Veikko Surakka
Professor in Interactive Technology, PhD
Research Group for Emotions, Sociality, and Computing
Tampere Unit for Computer-Human Interaction
Department of Computer Sciences
University of Tampere
Finland

Neurocognition of emotion and social perception

Jari Kätsyri, Mikko Viinikainen, Mikko Sams

Cognitive Science and Technology Research Group
Department of Biomedical Engineering and Computational Science (BECS)
Helsinki University of Technology (TKK)
Firstname.Lastname@tkk.fi

Abstract. This article gives an overview of current emotion and social perception research in the Department of Biomedical Engineering and Computational Science in Helsinki University of Technology (TKK)

Keywords: Emotional and social interaction, neural mechanisms of emotions, emotional auditory processing, emotional facial expressions, autism spectrum disorders, eye gaze tracking, pupillary mechanisms.

1 Introduction

Department of Biomedical Engineering and Computational Science (BECS) at the Helsinki University of Technology [1] conducts interdisciplinary research by combining experimental and computational methods to study living and artificial complex systems ranging from molecular interactions to information technology systems and human cognition. Our neurocognitive research relates to active hearing, and emotional and social interaction. An overview of the latter is given in this article.

2 Valence and arousal processing in the brain

A systemic view on emotions considers that there are rudimentarily separate dimensions that span the emotional space. The most important emotion dimensions are considered to be valence, which denotes the unpleasantness vs. pleasantness dimension, and arousal, which ranges from soporific to highly arousing [9]. Here we describe three functional magnetic resonance imaging (fMRI) experiments related to the processing of emotional valence and arousal in the human brain.

2.1 Emotional pictures

In an experiment [10] conducted in collaboration with NBM (see acknowledgments), we used emotionally evocative pictures from the International Affective Picture System (IAPS) [7] as stimuli. These pictures were presented in blocks of nine pictures using 100-ms presentation time and 2-s inter-stimulus interval (ISI). In multiple brain regions, blood oxygen level-dependent (BOLD) signals correlated positively with valence ratings for unpleasant pictures and negatively with valence ratings for pleasant pictures. Across all pictures the

T. Vanhala and J. Anttonen (Eds.) *Proceedings of EHTI'08: The First Finnish Symposium on Emotions and Human-Technology Interaction.* Tampere, Finland, May 30, 2008. ISBN 978-951-44-7383-8.

correlation was of inverted U-shape. Typical emotion processing areas, dorsomedial prefrontal cortex (DMPFC), anterior cingulate cortex (ACC), dorsolateral prefrontal cortex (DLPFC) and insula showed significant bilateral inverted U-shape correlations with valence. The activation patterns in insula region are portrayed in Fig. 1. For arousal, unpleasant pictures did not show significant correlations, but with pleasant pictures there was significant positive correlation with arousal in bilateral lateral sulcus and midbrain tegmentum.

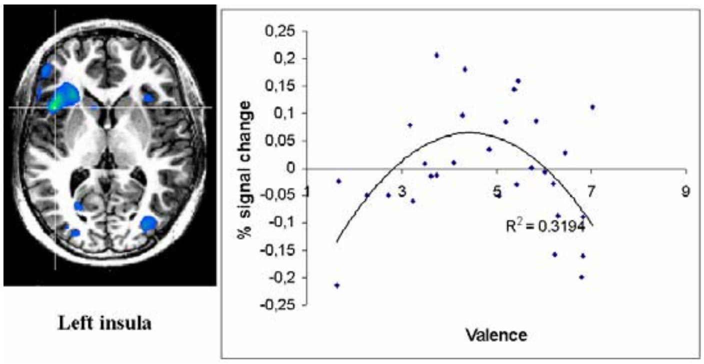


Figure 1. Correlation analysis for fMRI experiment using IAPS pictures as stimuli. Statistical parametric map from second-order nonlinear correlation has been shown in the left-hand side brain image, $p < 0.001$ uncorrected. Activations are seen in bilateral insula, but especially in the left hemisphere. The cross is centered on the highest activation peak in left insula and right-hand side diagram shows the inverted U-shape behavior in this region between percent BOLD signal change and stimulus valence.

2.2 Emotional facial expressions

In another unpublished experiment, we used acted emotional facial expression stimuli selected from TKK facial expression collection (described in [5]) as well as recordings specifically performed for this study. These still photographs were presented in blocks of eight pictures using 10-ms presentation time and 2-s ISI, and valence and arousal values were varied between the blocks. For negative facial expressions, valence and BOLD signal correlated positively in bilateral intraparietal sulcus/supramarginal gyrus (IS/SG) and bilateral middle frontal gyrus, and for positive facial expressions the correlation in these areas was negative. Thus, across all facial expressions, the activations on these regions were of inverted U-shape. For arousal, positive linear correlation across all pictures was observed in left inferior frontal gyrus (IFG), which is known to be part of the human mirror neuron system.

2.3 Emotional auditory stimuli

Presently, we are conducting an fMRI experiment where we use emotional auditory stimuli from the International Affective Digital Sounds (IADS) collection [2]. Selected auditory stimuli have been divided into blocks of three 6-s sounds so that valence is varied parametrically and arousal is kept approximately constant between the blocks. Only valence is studied due to small number of feasible stimuli. Our preliminary evaluation of IADS collection with 6 subjects suggested that some of the stimuli are considered unnatural (c. 20% of all stimuli) and some are perceived as both pleasant and unpleasant (c. 15% of non-neutral stimuli).

3 Face perception in Asperger Syndrome

Asperger Syndrome (AS) is an autistic spectrum disorder with severe social interaction difficulties [8], with difficulties in perceiving and interpreting faces. Our studies relate to emotional facial expression recognition and audiovisual speech perception in AS.

3.1 Recognition of basic emotions from facial expressions

In our previous research [6], conducted in collaboration with HUCH (see acknowledgements), we have analyzed the recognition of emotional facial expressions in AS. The recognition of four basic emotions from facial expressions with different levels of visual details was compared in 20 adult individuals with AS and their age- and sex-matched controls. The results showed typical recognition of basic emotions from unmodified facial expressions in AS, suggesting no difficulties in recognizing simple emotional states from faces. On the other hand, the performance of participants with AS was more degraded when accurate visual details were removed. This result is compatible with a 'weak central coherence' hypothesis of autistic disorders suggesting cognitive-perceptual reliance on details with a deficit in perceiving global shapes.

3.2 Audiovisual integration in speech perception

We are currently studying audiovisual speech perception in AS in collaboration with CBRU (see acknowledgements). We have conducted a study with 16 adult individuals with AS and their matched controls as participants. Audiovisual speech integration was studied by utilising McGurk stimuli, i.e. audiovisual speech stimuli with conflicting auditory (syllable /apa/) and visual (/aka/ and /ata/) components. In addition to behavioral measurements, eye gaze patterns were tracked while participants performed the task. Our preliminary analysis suggests weaker audiovisual integration in AS group (Fig. 2) caused by less accurate recognition of purely visual speech. Furthermore, preliminary results suggest that participants with AS focused less on mouth area than control participants.

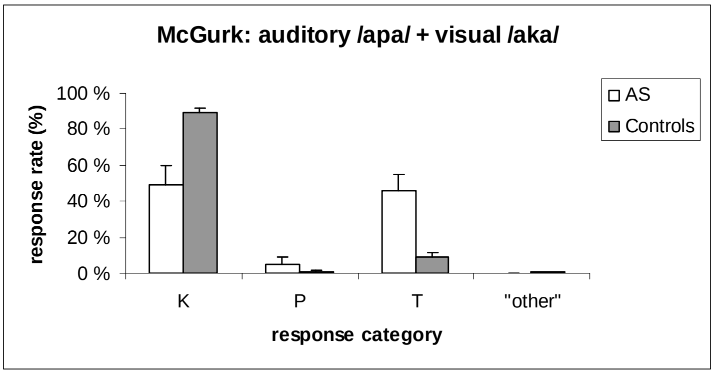


Figure 2. Preliminary results from an audiovisual speech perception study. Two groups were studied: participants with Asperger syndrome and their controls. Response rates are shown for different response categories when evaluating one of the presented incongruent audiovisual stimuli.

4 Pupillary contagion

Seminal pupillary studies have indicated that manipulating pupil size in images changes observers' evaluations of faces even when they are unaware of the modifications [4]. Furthermore, a recent study [3] has found that an observer's own pupil size mirrors those of presented faces. Curiously, this effect was present only when participants looked at sad faces but not when they observed happy, angry or neutral ones. In collaboration with MT and ICN (see acknowledgments), we have replicated this study with 23 participants. Sad faces and faces posing the basic emotions not evaluated in the original study (disgust, fear and surprise) were used as stimuli. In addition to facial expressions posed by human actors, our stimuli included animated facial expressions developed in MT. Pupil sizes were recorded with gaze tracking equipment while participants observed the stimuli, and the valence and arousal of stimuli was evaluated in a separate task. We are currently analyzing the obtained data.

Acknowledgments. We want to thank our collaborators: **NBM**: Laboratory of Neural Bases of Mind, Institute of Psychology, Russian Academy of Sciences, **HUCH**: Division of Child Neurology, Department of Paediatric and Adolescent Medicine, Helsinki University Central Hospital, **CBRU**: Cognitive Brain Research Unit, Department of Psychology, University of Helsinki, **MT**: Department of Media Technology, Helsinki University of Technology and **ICN**: Institute of Cognitive Neuroscience, University College London.

Jari Kätsyri, Ph.D. I work as a teaching researcher in TKK until the end of 2008. My research interests include face and emotional facial expression perception and the neural bases of emotional processing. I have participated in teaching emotion research, perceptual psychology and the philosophy of cognition. My undergraduate background comes from computer and cognitive sciences.

Mikko Viinikainen, M.Sc. (technology). I am a graduate student in Cognitive Science doing my doctoral thesis on the neural mechanisms of emotions within systemic framework. This research is performed mostly using fMRI. My background is in Engineering Physics and Mathematics, and I have worked earlier on some computational issues in brain research. Nowadays I work, however, mainly on the experimental side of neuroscience.

Mikko Sams, Professor (cognitive neuroscience). In addition to the research described in this paper, I am interested in the neurocognition of active hearing, i.e. how auditory system adapts to current processing needs. My other interests include music, literature, arts and food.

References

1. Department of Biomedical Engineering and Computational Science. URL: http://www.becs.tkk.fi/en/research/
2. Bradley, M.M. and Lang, P.J. (2007). The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual. *Technical report B-3*, University of Florida, Gainesville, Florida.
3. Harrison, N.A., Singer, T., Rotshtein, P., Dola, R.J., and Critchley, H.D. (2006). Pupillary contagion: central mechanisms engaged in sadness processing. *Social Cognitive and Affective Neuroscience Advance*, 1(1) 5-17.
4. Hess, E.H. and Petrovich, S.B.(1987). Pupillary behavior in communication, in *Nonverbal behavior and communication*, A.W. Siegman and S. Feldstein, Editors, Erlbaum: Hillsdale, NJ, 327-348.
5. Kätsyri, J. (2008). Human Recognition of Basic Emotions from Dynamic Posed and Animated Facial Expressions. Ph.D. thesis, Helsinki University of Technology.
6. Kätsyri, J., Saalasti, S., Tiippana, K., von Wendt, L., and Sams, M. Impaired recognition of facial emotions from low spatial frequencies in Asperger Syndrome. *Neuropsychologia*, In Press.
7. Lang, P.J., Bradley, M.M., and Cuthbert, B.N. (2005). International affective picture system (IAPS): Instruction manual and affective ratings. *Technical Report A-6*, The Center for Research in Psychophysiology, University of Florida.
8. Nieminen-von Wendt, T. (2004). Asperger Syndrome: Clinical, Neuroimaging and Genetical Findings, in Department of Child Neurology. University of Helsinki.
9. Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6) 1161-1178.
10. Viinikainen, M., Jääskeläinen, I.P., Alexandrov, Y., Balk, M.H., Autti, T., and Sams, M. Submitted.

Psychophysiology of emotions in digital games

Matias Kivikangas, Simo Järvelä, Niklas Ravaja

Center for Knowledge and Innovation Research
matias.kivikangas@hse.fi

Abstract. CKIR uses psychophysiological measurement methods to study the gaming experience and emotions during digital game play as a part of the FUGA EU project. Results include the emotional reactions to game events such as dying and killing and effects of single and multiplayer game modes.

Keywords: digital games, emotion, psychophysiology, facial EMG, skin conductance, user experience, game experience

1. Introduction

Center for Knowledge and Innovation Research is a multidisciplinary research unit at Helsinki School of Economics. In addition to studying several topics on economics, CKIR hosts a prominent research group doing world-class research on the psychophysiology of media experiences. The Emotion Team is led by Dr. Niklas Ravaja and it studies emotions, digital games, user experiences, and visual perception.

Our areas of expertise are experimental psychology and the psychophysiological methods (e.g., facial EMG, skin conductance, EEG, and cardiac measures). Currently our main research focus is digital game experiences, especially their emotional aspects, and developing a solid research methodology to examine the game experience comprehensively and with high temporal resolution as a part of the FUGA project. Physiological measures have been successfully used to study emotional reactions for decades in basic psychological research, and they have also proved their usefulness in media research [3, 4, 5, 6]. Studies conducted at CKIR extend that field to digital games, which offer a unique challenge to the methodology: researching an interactive and multimodal experience requires quite a different experimental control and framework of interpretation than studying more traditional stimuli.

2. Psychophysiological methods

Psychophysiological research uses physiological signals to study psychological phenomena. Psychophysiological methods are particularly useful for examining emotions, because they don't rely on subject's memory or interpretation of wordings, they are immune to respondent biases such as social desirability or wish to please the researcher, their reliability is independent from the observer, and they can be used automatically and continuously without disturbing the subject's natural behavior. As they can be used to study processes occurring during the experience with high temporal resolution, they enable event-related and long-term

T. Vanhala and J. Anttonen (Eds.) *Proceedings of EHTI'08: The First Finnish Symposium on Emotions and Human-Technology Interaction.* Tampere, Finland, May 30, 2008. ISBN 978-951-44-7383-8.

data analyses, and time series analyses [3]. Combined with other methods (e.g., self-report and observational data), the media user experiences can be studied with exceptional precision.

In psychophysiological emotion research it is suggested that all emotions can be located in two-dimensional space of arousal and valence, or bodily activation and emotional tone [2]. It is well established that arousal is connected to electrodermal activity (EDA, or skin conductance), and valence to facial electromyography (EMG) [3]. Cardiac measures such as heart rate have also been used for indexing arousal, as well as attention and other constructs. In addition, EEG research has established many useful indices for, e.g., attention and approach/withdrawal behavior. [1]

3. Projects

Our research group has been involved in many research projects in recent years that are closely related to emotions, digital games and other media, and human-computer-interaction.

3.1. FUGA

The Fun of Gaming: Measuring the Human Experience of Media Enjoyment (FUGA) is a three-year project (May 2006 - April 2009) funded by the European Union. The project is coordinated by CKIR, and it unifies six top-class European partners from Finland, Germany, Sweden, and the Netherlands to test and establish a set of scientific digital game experience measures. The operational goals of FUGA include the establishment of the construct validity, reliability, and predictive validity of the game experience measures that are based on the different measurement techniques. In addition to our own approach of psychophysiological methods, FUGA partners use brain imaging (fMRI), behavioral measures, implicit measures, eye-tracking, and a questionnaire created in the project specifically to probe dimensions of digital game experience. A further goal of the project is to develop a prototype of an emotionally adaptive game.

We have focused on developing sensitive, reliable and valid psychophysiological measurement methods to assess the game experience in several different setups. A novel technique was to use a specifically created modification of a digital game (Half-Life 2) and a system that enables automatized reviewing of the game from the video that had recorded the exact moments when events of interest occurred in the game play. This way we could have the subject assess the game events without interrupting the game, and also compare the emotional reactions caused by the actual game play with those caused by the subsequently reviewed video of the game play. Some research questions in these studies include: do the facial expression of a game character during intensive game play cause the same reaction as the facial expressions do in traditional picture viewing task, how does the congruence or incongruence of the facial expression and the affiliation of a game character affect the emotional responses to the character, and how does the enemy type affect the emotional response to an enemy in a game.

In addition to laboratory experiments, we have also been using a completely portable measurement device and handheld consoles to conduct experiments at cafeterias, metro stations, and subjects' homes in an effort to test how the gaming environment might affect the subjects' emotional responses. In other experiments we have studied the emotional effect of single and multiplayer modes, and of the subjects playing the game on the same side and against each other [see also 5].

3.2. PASION

The PASION (Psychologically Augmented Social Interaction Over Networks) is an European Union project under IST program scheduled to run for four years (January 2006 - December 2009). Current communication technologies are ineffective in conveying social, non-verbal and contextual information required for effective group interaction thus resulting in poor performance. The main goal of PASION is to achieve a more efficient, effective group interaction in mediated environment by understanding and tracking group behaviour, providing appropriate feedback services and by developing technological substitutes for traditional ways of conveying social, non-verbal and implicit information. PASION is focused on business and leisure oriented applications, such as collaborative work and social gaming.

The main task of CKIR in PASION is the development of a collaborative knowledge work application. Application is targeted to make users to be aware of the state of the entire work group and each other to coordinate and cooperate more effectively. The application consists of mobile and desktop parts; a Flash application for the mobile phone and a plug-in for Outlook e-mail client. The application collects and visualises social, emotional, and contextual data. Social data consist of e-mail and instant messaging traffic; from these data various social network analysis indexes are calculated. Emotional information is obtained from psychophysiological sensors, facial image analysis, or inputted manually by the user. Contextual data include location information and activity descriptions.

Along application development CKIR has carried out basic research on the effect of emotional cues while carrying out knowledge work tasks. In 2007 a psychophysiological laboratory experiment on emotional contagion while performing routine and creative knowledge work tasks was conducted.

3.3. Others

In the past CKIR has participated in several other projects involving digital media and game studies.

The Mobile Content Communities (MC2) project studied the social meaning and impact of new communication technology for communities that are interested in mobile gaming. The results of the MC2 project include evaluated and tested scenarios of mobile community gaming, template-based design tools that allow people to create their own games and game-related content, new open source tools to empower the community activity, and company-specific case studies to help the industry partners to benefit from community-created content.

In Presence: Measurement, Effects, Conditions (MEC) project multimethod approaches were used to measure presence. Questionnaires, distraction-paradigm, fMRI analysis, eye-tracking and psychophysiology were used. Based on the data the project built a model of spatial presence and its determinants.

4. Results

Unfortunately the most of the results of the FUGA studies are not ready for publication, as analyzing the data and the reporting process are not done at the time of writing. The results will be published in journals and conference papers, and the references can be found from the project web page [http://project.hkkk.fi/fuga/]. Some results from the earlier studies are presented.

4.1. Emotional reactions to game kills and deaths

One of the most widely noted results has been the counterintuitive emotional responses when player character kills an opponent or is killed himself. Instead of joy resulting from victory and success, wounding and killing the opponent elicited high-arousal negative affect, and despite the fact that the wounding and death of the player’s own character could be considered a defeat, it increased some aspect of positive emotion [6]. The negative reaction to killing and wounding of opponent is surprising, but in line with the research showing that negative emotional responses are elicited when people passively perceive images of victims of violence [2], although other explanations could also be provided. The positive reaction to the player characters own death, on the other hand, could be explained, for example, by transient relief from engagement as the action is paused when killed, or by positively experienced challenge.

Similar counterintuitive results were found, when the emotional responses to a bowling game were studied: Not only putatively positive game events, but also putatively negative events that involved active participation by the player elicited positive emotional responses in terms of facial EMG activity. In contrast, passive reception of negative feedback elicited low-arousal negative affect [4].

4.2. Effects of opponent relationship

Our other study concerned the relationship between the players in a multiplayer game. The subjects played a FPS game against a stranger or a friend, and also in single-player mode. The assumption of many gamers that it is more fun to play against a friend than a stranger, and more fun to play a multiplayer game than a single-player game, gained support from both psychophysiological measurements and self-reports; also spatial presence followed the same pattern [5]. More multiplayer experiments have been done in FUGA project.

5. Future prospects

In future we are planning several different projects and studies involving time series analyses of media experiences, more comprehensive examination of game events, and gamer personality profiles. Our aim is to develop the methodology so that it can be applied fruitfully across the wide field of media experience research.

Acknowledgements. The authors' research is supported by European Community NEST project 28765: “The Fun of Gaming: Measuring the Human Experience of Media Enjoyment.” FUGA Website: http://project.hkkk.fi/fuga/

J Matias Kivikangas, MA (psych). I am a researcher at Center for Knowledge and Innovation Research (CKIR), and a PhD student in psychology at the University of Helsinki, studying the psychophysiology of digital game experiences in FUGA project. My research interests include the methodology and use of psychophysiological methods in the context of digital games, how different people experience game enjoyment and different game features contribute to it, and how game research could help game design. Additionally I am interested in game design and role-playing theory.

Simo Järvelä, BBA, BA. I am an undergraduate student in cognitive science at the University of Helsinki and the Usability School, and a research assistant at CKIR. At the moment I’m involved in the FUGA project and researching digital games and user experience using psychophysiological methods. My other interests include role-playing theory, philosophy of mind and alternate states of consciousness.

Niklas Ravaja, PhD, Docent. Dr. Niklas Ravaja is the Director of Research at CKIR and he acts as the Coordinator and Principal Investigator of the FUGA project.

References

1. Cacioppo, J.T., Tassinary, L.G. & Berntson, G.G. (Eds.) (2000). *Handbook of psychophysiology* (2nd ed). New York: Cambridge University Press.
2. Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30*, 261–273.
3. Ravaja, N. (2004). Contributions of psychophysiology to media research: Review and recommendations. *Media Psychology, 6*, 193–235.
4. Ravaja, N., Saari, T., Laarni, J., Kallinen, K., Salminen, M., Holopainen, J., & Järvinen, A. (2005). The psychophysiology of video gaming: Phasic emotional responses to game events [CD-ROM]. In *Proceedings of DiGRA 2005 Conference*. Burnaby, British Columbia, Canada: Simon Fraser University.
5. Ravaja, N., Saari, T., Turpeinen, M., Laarni, J., Salminen, M.and Kivikangas, J.M. (2006). Spatial Presence and Emotions during Video Game Playing: Does It Matter with Whom You Play? *Presence, 15*, 381–392.
6. Ravaja, N., Turpeinen, M., Saari, T., Puttonen, S. and Keltikangas-Järvinen, L. (2008). The Psychophysiology of James Bond: Phasic Emotional Responses to Violent Video Game Events. *Emotion, 8*, 114–120.

Designing emotionally adaptive gaming

Kai Kuikkaniemi[1], Toni Laitinen[1], Ilkka Kosunen[1]

[1]Helsinki Institute for Information Technology (HIIT)
Firstname.Lastname@hiit.fi

Abstract. This paper describes the approach of HIIT's DCC-research group (Digital Content Communities) on emotionally adaptive gaming (or biosignal adaptive gaming). The paper includes short discussion about the terminology and some design ideas. Then we introduce the games we have been building, and the research experiment we are currently implementing. Finally we will discuss shortly about our findings and research interests.

Keywords: affective computing, applied gaming, emotional adaptation, adaptive systems, psychophysiology, biosignals, biofeedback

1 Emotionally adaptive gaming

Two years ago we got funding to start building emotionally adaptive games in a EU-project called FUGA. We had history of building experimental games, and we had some basic understanding on measuring psychophysiological signals. However, emotionally adaptive games were in practice a new domain for us. During the last two years we have understood that building an emotionally adaptive game is far from straightforward game development. First of all emotionally adaptive gaming as such is a problematic phrase, and it might be helpful to use some other terminology. Second, there are many different approaches to building emotionally adaptive games. Hence, defining the goals for emotional adaptation is probably the key issue in this kind of game development. Each signal has distinct "behavior" and usability. Player can learn to manipulate any signal if the reward is high enough (even EEG or EKG). Hence, it is important to be aware of what the gaming patterns are in practice, and how rookies and experts behave with the game.

1.1 Concepts and terminology

Emotionally adaptive gaming means in concept level that game measures some user signals (psychophysiological, voice, gestures, behavioral), interprets emotions from these signals and reacts accordingly – as simple as that. However, as we do not have perfect and absolute emotional model, not to mention explicit and absolute ways of measuring the emotion accurately and in real-time, we need to fine-tune the concept. In reality the use of for example voice or psychophysiological data in a game adaptation is handicapped by the fact that when player is aware of the adaptation she is able to control the signals, at least at some level. If player is not aware of the adaptation, there are many practical and ethical problems, which are crucial in our opinion. Term affective gaming is not any better as emotions are inherent part of the definition of affective computing. Biosignal adaptive gaming, or biofeedback gaming are in our opinion better phrases.

T. Vanhala and J. Anttonen (Eds.) *Proceedings of EHTI'08: The First Finnish Symposium on Emotions and Human-Technology Interaction.* Tampere, Finland, May 30, 2008. ISBN 978-951-44-7383-8.

Our experience for building the prototype games (chapter 2) has led us to conclude that GSR (Galvanic Skin Response) and respiration are probably the easiest signals to start with. These signals can produce relative measures of arousal. However, it sounds bit too overwhelming to state GSR adaptation as emotional adaptation, when the easiest way to increase GSR levels is to breathe intensely. Intense breathing is easy to manipulate and it is not accurate measure of any emotion. Anyways, in this paper we are using the term emotionally adaptive games, as it is closer to the topic of EHTI and the title of our work in FUGA-project, and because it is not incorrect to use the term. When we are talking about emotionally adaptive, biosignal adaptive or biofeedback gaming, we are talking about the same thing, and we are meaning games that utilize in some ways player’s biosignals and other complementary technologies such as gestures, voice and image recognition, behavioral data and accelerometers.

1.2 Approaches on emotionally adaptive gaming

It is hard to summarize all the lessons we have learned related to the various approaches on emotionally adaptive gaming. This is work in progress, and we hope we can have opportunity to give a comprehensive dissemination on our research results. Biofeedback is a technology that can be utilized practically in every current gaming type (multiplayer, FPS, strategy, roleplaying, simulators, casual), and it is possible that emerging gaming types (like pervasive-, edu- and exergaming) will be especially benefitted from biofeedback. Furthermore, biofeedback can open doors for new gaming types like therapy gaming.

Below is a short table of different game elements that can be configured based on biosignals. Configuring the elements causes distinct emotionally adaptive gaming patterns, and different gameplay experience.

Game element	**Details**
Player character (attributes)	Speed, size, shaking, ability to recognize by NPCs, can make gaming more immersive when avatar body qualities is linked with players body
Player character (outlook)	Especially in multiplayer gaming, body language in virtual environment
Game-world mechanics	Gravitation, lightning, sound, almost anything in game world can be linked with biosignals. The key question is not what makes sense but what makes a good game!
Narrative	Interactive narrative in the game can select paths based on players biosignals
Controls	Biosignals can be used as a new control interface or complementary control parameters (e.g. fine tune sensitivity) with current interfaces
Emotions visualization	Make emotions transparent, usable for example in social gaming and therapy gaming, not neccesarily avatar representation

Some of the elements in the list probably overlap with others and the list is not anyways comprehensive, hence there are definitively distinct and notable game elements, which can adapt to biosignals that are missing from our list.

2 HIIT's EMOGames

In HIIT we have been developing the emotionally adaptive games in an environment, which includes three components: biosignal capturing hardware (Varioport), biosignal analysis software (EMOEngine) and the prototype games. The following chapter defines our prototype games and our current game experiment. EMOEngine is analysis platform, which controls and listens hardware through serial port, and transmits analyzed data forward to games. EMOEngine also stores psychophysiological data for our experiments.

2.1 EMOTetris

EMOTetris was our first game prototype. It is fairly simple variation of the famous game of Tetris, in which the speed of falling objects is linked to player's arousal measured from GSR. We have also variation of Tetris in which the movement of the objects is controlled with facial muscles based on EMG sensors.

2.2 EMOShooter

EMOShooter is our second emotionally adaptive game prototype. Actually it is game platform. It is built on top of graphics engine (OGRE 3D) and physics engine (ODE). Because we did not use any existing game engine we do not have limitation of other games. In EMOShooter platform we have possibility to modify pratically any game world, player avatar, avatar outlook, or control parameter. EMOShooter is FPS game platform capable for displaying videotextures and applicable for multiplayer gaming. We have been testing various adaptation patterns with EMOShooter by utilizing EMG, GSR and respiration adaptations.

2.3 FPS-Experiment

Our current efforts concentrate on FPS-Experiment, which is based on EMOShooter platform. In the experiment we are researching the effect of biosignal adaptation on game enjoyment. In the experiment GSR and respiration signals are connected to player avatar attributes in simple FPS-stage. Goal of the game is to kill cube-enemies either with sniper or machine gun. Adaptation includes changes in rate of fire, recoil, movement speed and shaking. Movement speed and rate of fire are generally positive adaptations and shaking and recoil are mostly negative. This combination maintains game balance nicely.

Signals are calibrated by using dynamic range (basically a variation of dynamic signal normalization algorithm), which has memory of few seconds (depending on signal). Dynamic range is easy to use and effective calibration mechanism. It also helped us to eliminate some problematic game patterns. The respiration adaptation is connected to amplitude not frequency, so in some cases it would be desirable to hold breath for quite some time. Dynamic calibration with short memory makes holding breath useless after few seconds, and forces player to find rhythm in gaming instead of forcing the player to control her biosignals. Relative change seems to be more practical than absolute values in this kind of gaming.

The experiment is divided in two phases. In first phase player plays 8 stages (2 x GSR, respiration, and two types of signal simulations) in random order without knowledge of biosignal adaptation. After 8 stages player is introduced the concept of biosignal adaptation and she has opportunity to learn to control biosignals. Then she plays again the stages with biosignal adaptation. In the experiment we want to explore the change between simulations and biosignals, and how gaming changes before and after the knowledge of biosignal adaptation. Currently we are in the middle of subject data collection. We gather information from behavioral game log, EMG, GSR, respiration, video capture, interviews and questionnaires after each stage.

3 Key issues for building emotionally adaptive games

There are few fundamental issues that must be solved while building emotionally adaptive games. For all of these issues there are some generic rules and then the case-by-case varying solutions. The issues may sound obvious, but we have found out that the solutions are much more than trivial. The key issues in our opinion are emotionally adaptive gaming patterns, adaptation algorithms and signal processing, and sensors in the game context. This list also shows some of the key issues we are concentrating in our research. The complexity of applying these issues arises from the inherent complexity of building games. Even a simple game includes many components starting from controls, game rules to AI and graphics. Good games are composed of delicate synthesis of the components, which creates pleasant game balance and challenge for players. Creating game balance and challenge, which utilize emotional adaptation patterns and new game sensors increases the complexity. Furthermore, the design of the game must include deep understanding of player's emotions and motivations during the game play. This is hard, especially when players have so different personalities and tastes.

3.1 Emotionally adaptive gaming pattern

Good game has a balance. It is important that game is challenging and rewarding. Many times the game goal is for example to achieve maximum amount of points or be fastest etc.. Or elements in the game are designed to make the achieving the goal challenging and rewarding in a right balance. If the goal of the game is not to teach player to control her biosignals, the utilization of biosignals must also obey the game goal. For this reason the game designer must consider that a good game balance cannot require player to keep smiling, be very highly aroused or hold breath for all the time in order to achieve maximum points. In order to achieve the goal player should find the right rhythm or balance of playing the game and controlling her emotions (or biosignals).

3.2 Adaptation algorithm and biosignal processing

Biosignals must be analyzed before they can be used in emotionally adaptive gaming. Traditionally psychophysiology signal analysis has been done in retrospect to the data by using various statistical tools. In adaptive gaming case the analysis must take place in real-time. This limits the possibility of using some of the statistical algorithms. In practice we must use time-series analysis with short sample windows. In practice EKG, EEG and EMG require always extensive data processing. GSR and respiration can be almost used as such to create some adaptation signal, but even in these cases also the sensitivity, smoothing, mapping and range must be defined somehow. Furthermore, artifact elimination is important. Finding absolute calibration for biosignals is hard due the nature of human body and psychophysiology. Calibration should be personal and varies in time – a lot. This is why the calibration should be part of the game pattern and adaptation algorithm. For example dynamic range algorithm eliminates the need for calibration when using respiration and GSR.

3.3 Sensors in the game context

Usability of psychophysiological recording devices is currently poor. Respiration, pulse and GSR are probably the easiest. But they are even too complicated and time consuming for normal home video gaming. New piezo-electrodes appear promising. The usability with sensors without paste is increased significantly. Also voice and image recognition based signal detection is potential due the better usability. In case of popular new games such as Wii Sports or Guitar Hero the new kinds of interfaces and sensors have an in-game story. Also in case of emotional adaptation the design of the game should include the physical design of the sensors, e.g. "Detective hat" for EEG sensors or "Sniper-gloves" for GSR sensors. Hence, the sensors could be designed as part of the game story.

Acknowledgments. Our work is conducted under the supervision of professor Marko Turpeinen and professor Timo Saari. Our work is done in close collaboration with docent Niklas Ravaja and researcher Matias Kivikangas from Center of Knowledge and Innovation Research (CKIR). Thanks to CKIR for teaching us on how to measure biosignals, and lending us some devices.

Kai Kuikkaniemi, M.Sc. I am a PhD student working in the Helsinki Institute for Information (HIIT) Digital Content Communities (DCC) research group. I have been working in HIIT for five years in several different projects. Many of the projects are involved in gaming. Topics vary from mobile gaming, business models to emotional adaptation. However, my current PhD topic is related to cinema experience, and I am actually a PhD student in University of Arts and Design. I did my master's thesis in TKK department of Industrial Engineering and Management.

Toni Laitinen. I'm a computer science student at Helsinki University of Technology. Currently I'm working at Helsinki Institute for Information Technology (HIIT) and doing my Master's Thesis on issues related to biofeedback and affective feedback in context of digital gaming. At the moment I'm working on two projects that are both focused on issues related to affective computing in digital media. In FUGA project I'm developing our biofeedback game platform and in CALLAS I'm developing an interactive multimodal application that can be used for improvised storytelling.

Ilkka Kosunen. I'm computer science student in University of Helsinki, and have been working for 1.5 years in Helsinki Institute for Information Technology(HIIT) doing research on online machine recognition of human emotions from various psychophysiological signals and the use of this information in biofeedback applications, such as games and assisted learning systems. I'm interested both on the more theoretical side of machine recognition/learning of human emotions as well as more abstract side of how to use this information in HCI-context.

Affective pictures and emotion analysis of facial expressions with local binary pattern operator: Preliminary results

Seppo J. Laukka[1], Antti Rantanen[1], Guoying Zhao[2], Matti Taini[2], Janne Heikkilä[2]

[1]Learning Research Laboratory, University of Oulu, Finland
seppo.laukka@oulu.fi
antti.rantanen@oulu.fi
[2] Machine Vision Group, Infotech Oulu and Department of Electrical and Information Engineering, University of Oulu, Finland
gyzhao@ee.oulu.fi
mattitai@mail.student.oulu.fi
jth@ee.oulu.fi

Abstract. In this paper we describe a setup and preliminary results of an experiment, where machine vision has been used for emotion analysis based on facial expressions. In the experiment a set of IAPS pictures was shown to the subjects and their responses were measured from a video recording using a spatiotemporal local binary pattern descriptor. The facial expressions were divided into three categories: pleasant, neutral and unpleasant. The classification results obtained are encouraging.

Keywords: facial expression, affective pictures, emotion, local binary pattern operator

1 Introduction

Emotion is very important in communication. Facial expression reveals lots of information of emotion. The measurements of facial expressions have been developed since 1940 [6]. Nowadays one of the most well known fully automatic facial expression method is Facial Action Coding System (FACS) which is based on state-of-the-art machine learning techniques [2,3]. We have utilized the spatiotemporal local binary pattern operators to describe the appearance and motion of facial expression. The preliminary results from the subject-independent and subject-dependent experiments are quite promising.

2 Methods

2.1 Participants

A total of 5 right-handed, (native fluency in Finnish and no reported history of speech disorder) Finnish speaking, undergraduate female students from the University of Oulu participated in the experiment. The students participated voluntarily in the experiment during their psychology studies. Eyesight was tested using the Snellen card and anxiety was measured using the State Trait Anxiety Inventory (STAI) [8]. In addition, possible alexithymia was tested with TAS-20 [1], the validity of which has been tested in Finland too [4]. All subjects had normal eyesight (≥ 1.0), and no anxiety (STAI score < 35) nor alexithymia (TAS-

T. Vanhala and J. Anttonen (Eds.) *Proceedings of EHTI'08: The First Finnish Symposium on Emotions and Human-Technology Interaction.* Tampere, Finland, May 30, 2008. ISBN 978-951-44-7383-8.

20 score < 51) was found. After the explanation of the experimental protocol, the subjects signed their consent.

2.2 Apparatus

The IAPS pictures (International Affective Picture System) [5] were presented on the screen (17") of a computer with Intel Pentium 4 processor which was connected to a Tobii 1750 eye tracking system (Tobii Technologies AB, Sweden). The sample rate was 50 Hz and the spatial resolution was 0.25 degrees. The eye tracking system located every fixation point and measured the duration of fixation, the pupil size variation and the distance of the eye from the computer screen. The heart rate variations were measured using beat-to-beat RR-intervals with a Polar S810i heart rate monitoring system (Polar Oy, Finland). The facial expressions were recorded by one IEEE 1394 firewire camera (Sony DFW-VL500, Japan). In addition, the subject's speech was recorded a wireless microphone system (Sennheisr HSP2, Denmark) (Figure 1).

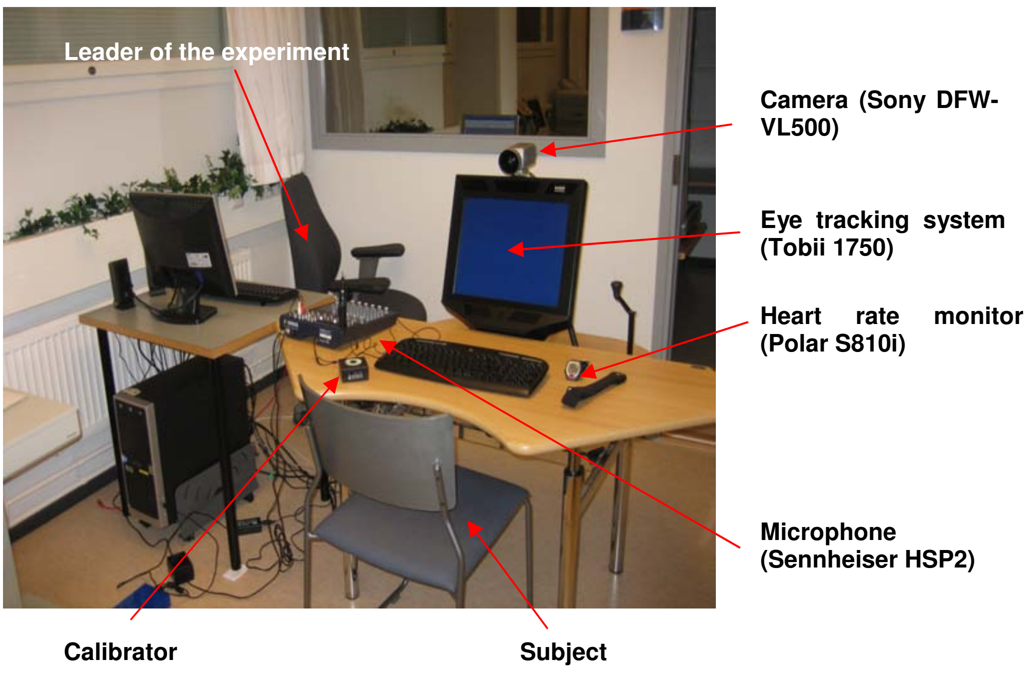


Figure 1. The experimental design.

2.3 Materials

A total number of 48 International Affective Pictures [5] were used in the experiment. The pictures were divided into three different groups; 16 pleasant, 16 neutral and 16 unpleasant pictures. The reliability of the valences and arousals of the pictures used in the experiment has been tested in Finland earlier [7]. The overall luminance levels of the pictures were adjusted with Adobe Photoshop 6.0 software.

2.4 Procedure

The subjects were interviewed and STAI (Form 2) and TAS-20 questionnaires were presented before the experiment. Subsequently, the subject was able to practise the experimental procedure from “the paper version” with the experimenter. Thereafter, the subject practised the procedure with the computer. Before the actual experiment, the subject rested 60 s, while the heart rate monitoring system, audio and camera systems were combined to the eye tracking system. The subject’s eye movements were also calibrated into the eye tracking system.

In the experiment the pictures were presented on the computer screen and the distance of the subject from the screen was 65 cm. At first, the subject had to look at the letter X for 30 s, which appeared in the middle of the screen. Sequentially either a pleasant or neutral or unpleasant picture appeared on the screen for 20 s in random order. Immediately after the 20 s, the SAM scale (Self-Assessment Manikin, Lang 1994) appeared. The subject’s task was to orally report the valency and arousal of the picture according to the SAM scale. After the report, the subject had to press the enter button in order to darken the screen. In this phase, the subject’s task was to orally report on what was seen, what was happening and what was going to happen in the picture for the experimenter who was sitting behind the computer screen. After the report, the subject had to press the enter button for the next picture to appear. After 48 pictures, the letter X appeared for 30 s. Finally, the STAI (Form 1) questionnaire was presented.

2.5 Data analysis

Our dataset in the experiments currently includes 240 video sequences from five subjects, each subject with 48 videos. Facial expression video information is utilized for three kinds of emotion recognition: pleasant, neutral, and unpleasant. There are two kinds of the ground truth: one is form pictures, the other is from the assessment given by the subject.

An effective spatiotemporal local binary pattern descriptor: LBP-TOP [9] is utilized to describe facial expressions, in which local information and its spatial locations should also be taken into account. A representation which consists of dividing the face image into several overlapping blocks was introduced. The LBP-TOP histograms in each block are computed and concatenated into a single histogram. All features extracted from each block volume are connected to represent the appearance and motion of the facial expression sequence.

After extracting the expression features, a support vector machine (SVM) classifier with one-against-one strategy was selected since it is well founded in statistical learning theory and has been successfully applied to various object detection tasks in computer vision.

3 Results

For evaluation, we separated the subjects randomly into n groups of roughly equal size and did a “leave-one-group-out” cross validation, which also could be called an “n-fold cross-validation” test scheme. Here we use 10-fold cross-validation. Every time leave one group out, and train on the other nine groups. This procedure is done for ten times. The overall results are from the average of the ten-time repetition. Table 1 lists the recognition results for each emotion and overall from all three emotions. The last row gives the results from the ground truth making by researchers’ observation. Because we found in the experiments that the assessment from pictures and subjects are not consistent to the observations based on facial expressions, that is to say, even though the assessment from subjects or the pictures shows one video should be that emotion, but the subjects did not make obvious corresponding facial expressions in their faces. This makes it is hard to recognize the emotions just from expressions.

Table 1. Emotion recognition results for subject-independent experiments.

Ground Truth	pleasant	neutral	unpleasant	10-fold cross-validation
Pictures	42.50	41.25	43.75	42.50
Subjects	57.53	65	68.66	63.75
Our observation	77.42	51.28	73.77	63.75

4 Discussion

It can be seen from the results the recognition rates from subjects’ assessment are better than those from pictures. It is reasonable, because for same pictures, different subjects would have different understanding and thus have different emotions. Even though the overall result from our observation assessment is same to that from subjects’ ground truth, the pleasant and unpleasant emotions obtain much better results. How to make a reliable ground truth is an interesting research in the future. As well, it can be seen that facial expression can provide quite much information for emotion recognition. If combining with the other modalities, like voice, we believe it would improve the recognition performance, especially for those subjects who do not usually show their emotions in faces.

Acknowledgments We would like to thank Satu Savilaakso (BEd) for technical help of data analyzing.

References

1. Bagby, M., Parker, J., and Taylor, G. (1994). The Twenty-Item Toronto Alexithymia Scale-I. Item selection and cross-validation of the factor structure. *J. Psychosom Res.* 38:23-32.
2. Bartlett, M.S., Movellan, J.R., Littlewort, G., Braathen, B., Frank, M.G., and Sejnowski, T.J. (2005). In: What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS), Second Edition. Ekman, Paul (Ed); Rosenberg, Erika L. (Ed); pp. 393-426. New York, NY, US: Oxford University Press, 2005. xxi, 639 pp.
3. Cohn, J.F., Ambadar, Z., Ekman, P. (2007). *Observer-based* measurement *of* facial expression *with the* Facial *Action Coding System.* In: Handbook of emotion elicitation and assessment. Coan, James A. (Ed); Allen, John J. B. (Ed); pp. 203-221. New York, NY, US: Oxford University Press, 2007. viii, 483 pp.
4. Joukamaa, M., Miettunen, J., Kokkonen, P., Koskinen, M., Julkunen, J., Kauhanen, J., Jokelainen, J., Veijola, J., Läksy, K., and Järvelin, M.R. (2001). Psychometric properties of the Finnish 20-item Toronto Alexithymia Scale. *Nord J Psychiatry* 55:123-7.
5. Lang, P.J., Bradley, M.M., and Cuthbert, B.N. (2005). International affective picture system (IAPS): Affective rating of pictures and instruction manual. *Technical Report A-6*. University of Florida, Gainesville, FL.
6. Lynn, J.G. (1940). An apparatus and method stimulating, recording and measuring facial expression. *Journal of Experimental Psychology*, Vol. 27(1), Jul 1940. pp. 81-88.
7. Nummenmaa, L., Hyönä, J., and Calvo, M.G. (2006). Eye Movement Assessment of Selective Attentional Capture by Emotonal Pictures. *Emotion* Vol 6 (2): 257-268.
8. Spielberger, C.D., Gorsuch, R., and Lushene, R. (1970). The State-Trait Anxiety Inventory (STAI) test manual. Palo Alto, CA: Consulting Psychologists Press.
9. Zhao G. and Pietikäinen M. (2007). Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6), 915-928.

From emotions to intentions: neurocognitive mechanisms of face perception

Laura Pönkänen, Jari K. Hietanen

Human Information Processing Laboratory, Department of Psychology,
University of Tampere
http://www.uta.fi/laitokset/psyk/hip
Laura.Ponkanen@uta.fi
Jari.Hietanen@uta.fi

Abstract. Human faces convey important visual information that is used actively when interacting with other people. We have currently been studying affective and cognitive processes related to the perception of face and gaze information. In particular, we have been investigating whether the processing differs when viewing the faces live or as pictures on a computer screen. Our recent studies have shown that live faces, holding the potential for interpersonal communication, elicit intensified affective and motivational responses when compared to pictures of faces on a computer screen.

Keywords: face perception, social communication, motivation, electroencephalography, event-related potential, frontal asymmetry.

1 Affective processing of facial information

1.1 Introduction

A face is a powerful stimulus in social interaction; it is capable of conveying signals about other persons' emotions, motivation, and attention direction, to mention only a few examples. In the course of evolution, rapid evaluation of affective-motivational significance of other people's faces has become a crucial skill in face-to-face interaction. Thus, not surprisingly, human adults are highly adept at recognizing facial information [1], and several behavioural, neurophysiological, and neuroimaging studies [2–4] have shown that the visual processing of facial information is specialized compared to the processing of other objects.

Recently our research group has been studying several aspects of facial information processing: especially gaze direction and eye contact. We have been particularly interested in affective and cognitive processing of facial information in different visual and social contexts. Although faces are typically perceived in the context of human interaction, face processing is commonly studied by displaying faces on a computer screen. However, when presenting the faces on a computer screen, the viewing context lacks the social dimension. Thus, one of the most interesting research questions has concerned whether the physiological and

T. Vanhala and J. Anttonen (Eds.) *Proceedings of EHTI'08: The First Finnish Symposium on Emotions and Human-Technology Interaction.* Tampere, Finland, May 30, 2008. ISBN 978-951-44-7383-8.

behavioural responses to facial stimuli differ depending on whether participants are viewing faces live or as pictures on a computer screen.

1.2 General methodology

To examine the differences between live and picture stimuli, we have presented facial stimuli both on a computer screen and live through an electronic shutter. In the live condition, a panel containing a window with a liquid crystal (LC) shutter is placed between the experimenter and the participant. A voltage sensitive LC-shutter can be made opaque or transparent within a millisecond range, thus it allows presentation of live stimuli with a strict control of timing.

In our studies, we have been using event-related potentials (ERPs) analyzed from the electroencephalographic (EEG) activity in order to study the neurocognitive processes involved in perception of facial information. In addition, we have been measuring frontal asymmetry in the EEG activity and skin conductance responses to investigate motivational-affective reactions. These latter mentioned measures are related to the approach-withdrawal action tendencies [5] and to the arousal component [6] inherent in affective reactions. We have also used questionnaire methods such as participants' self-ratings of arousal and valence to different types of facial stimuli.

2 Studying faces in context

2.1 Facial images vs. live faces

Both realistic (e.g., photographs) and schematic pictures of faces generally elicit the same pattern of face-sensitive ERP responses [7]. However, neither of them holds the potential for interpersonal communication, and it is known that rapid evaluation of affective-motivational significance is crucial in a real face- to-face interaction.

In our recent study [8] we examined the question of whether the presentation mode (live vs. picture) has an influence on the affective-motivational modulations on ERP responses. The participants of the study were shown a human face, a dummy face, and a control object live and as pictures on a computer screen. By including the dummy face we controlled for the possibility that the ERP differences between real faces and pictures of real faces might reflect the fact that live faces are 3D. In the live condition, both the human and the dummy face were 3D, but only the human face had a potential for social interaction.

We found that the ERP-responses, especially a component known as early posterior negativity that is suggested to reflect the affective-motivational processing of visual information, reliably discriminated between the human and the dummy face. Interestingly, this pattern of results was evident in the live condition only, but not when faces were presented as pictures on a computer screen. The evidence from several studies shows that affective-motivational factors can intensify visual processing reflected already in the early stage of stimulus encoding and selection regardless of the valence (i.e., threat or reward) of the stimulus [9,10]. The human face was also subjectively rated as more pleasant than the dummy face, but again, only when seen live. This possibly reflects the affective-motivational processing that facing a live person elicits. In sum, our study proposes that seeing a live human face elicits affective processes and intensifies early visual processing more than seeing a picture of a face, thus possibly preparing the organism for successful planning of appropriate reactions to social stimuli.

2.2 Affective-motivational processing of eye-gaze

Eye gaze is known to be an important factor in regulating social interaction and expression of intimacy and social goals [11]. Recent evidence suggests that direct and averted gaze can signal the sender's motivational tendencies to approach or avoid. The systems regulating basic motivational-emotional responses comprise of sets of neural networks regulating (i) the *direction* of the responses, i.e., the motivational tendency to approach or avoid the source of stimulation and (ii) the energy used for these responses, i.e. the *intensity* of the motivational tendency [12].

A recent study [13] conducted in our lab examined, whether seeing another person's direct vs. averted gaze has an influence on the observer's neural approach-avoidance responses. To test this hypothesis, we measured hemispheric asymmetry in the frontal EEG (direction of motivation) and skin conductance responses (arousal) to another person's direct and averted gaze. The relatively stronger activation of the left than the right frontal cortex has been associated to the activation of the approach-related motivational system, whereas the relatively greater activation of the right frontal cortex has been associated with the activation of the avoidance system [14]. As in our study presented above [8], it was also investigated whether it would make a difference if the participants were looking at the face of a real person or a picture of a face.

It was found that another person's direct gaze elicited a relative left-sided frontal EEG-activation (approach tendency), whereas averted gaze activated right-sided asymmetry (avoidance tendency). In addition, skin conductance responses were larger to faces in general than to control object (a radio), and to faces with direct relative to averted gaze, indicating that faces with direct gaze elicited more intense activation in the autonomic nervous system and strength of the motivational tendencies than did the control stimuli. However, all these effects were observed only when participants were facing a real person, not when looking at a picture of a face. This finding was suggested to reflect the motivational responses to gaze direction being activated in the presence of another person.

2.3 Discussion and future directions

The results of the studies described above indicate that a live face as a potentially interacting stimulus is processed differently than an inanimate picture of a face already at the early processing stages. In addition, seeing the eye-direction of a real person activates physiological responses related to affect and motivation differently than seeing the same stimuli as pictures. As action-readiness is an essential element of social interaction, it should be taken into account when studying the emotional or motivational processing of facial stimuli presented on a computer screen, and also when designing computer programs that aim at simulating face-to-face social situations. It is possible that some effects on physiological responses may not be observed at all when face processing is investigated using pictures of faces, or the effects are diminished as compared to effects elicited by real faces. Further research is needed to explore more detailed the specific neurocognitive mechanisms involved in looking at a living face versus looking at a picture of a face.

Acknowledgments. The research project of Laura Pönkänen and Jari Hietanen are supported by the NEURO -program of the Academy of Finland (Project no. 1111850).

Laura Pönkänen, MA. I am a PhD student working in the Human Information Processing Laboratory at the University of Tampere. My research topic covers face perception, especially the cognitive and affective processing of gaze direction. Particularly, I have investigated whether the processing of facial information differs when faces are viewed live or on a computer screen. I am also studying how the affective and cognitive processing of gaze and eye-contact develops within the first year of life (NEURO-program of the Academy of Finland *The development of the social brain: An affective view*).

Jari Hietanen, Ph.D., professor of psychology. I conduct research in the Human Information Processing Laboratory at the University of Tampere. My field of expertise is cognitive and neurocognitive mechanisms of face perception. Especially, my research has been targeted on gaze perception and perception of facial expression of emotions. The methodology we use extends from behavioral (response time and performance accuracy) to various physiological (electroencephalography, functional magnetic resonance imaging, skin conductance, facial electromyography, heart rate) measurements.

References

1. Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral & Cognitive Neuroscience Reviews 1*, 21 –62.
2. Carmel, D., and Bentin, S. (2002). Domain specificity versus expertise: Factors influencing distinct processing of faces. *Cognition*, 1–29.
3. Haxby, J.V., Hoffman, E., and Gobbini, M.I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry,* 59–67.
4. Itier, R. J, Latinus, M., and Taylor M. (2006). Face, eye and object early processing: What is the face specificity? *Neuroimage,* 667–676.
5. Davidson, R. J. (1993). The neuropsychology of emotion and affective style. In: M Lewis and J. M. Haviland (Eds.) *Handbook of Emotions* (pp. 143-166). New York: Guilford.
6. Dawson, M., Schell, A., and Fillon, D. (1990). The electrodermal system. In J. T. Cacioppo and L. G. Tassinary (Eds.), *Principles of psychophysiology* (pp. 295-324). New York: Cambridge University Press.

7. Sagiv, N. and Bentin, S. (2001). Structural encoding of human and schematic faces: Holistic and part-based processing. *Journal of Cognitive Neuroscience*, 937–951.
8. Pönkänen, L. M., Hietanen, J.K., Peltola, M. J., Kauppinen, P. K., Haapalainen, A. and Leppänen, J. M. (2008). Facing a real person: an event-related potential study. *Neuroreport*, 49–501.
9. Balconi M. and Pozzoli U. (2003). Face-selective processing and the effect of pleasant and unpleasant emotional expressions on ERP correlates. *International Journal of Psychophysiol*ogy, 67–74.
10. Werheid K, Schacht A, and Sommer W. Facial attractiveness modulates early and late event-related potentials. *Biological Psychology*, 100–108.
11. Kleinke, C. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, 78-100.
12. Lang, P. J., Bradley, M. M., and Curhbert, B. N. (1990). Emotion, attention and the startle reflex. *Psychological Review*, 37–395.
13. Hietanen, J. K., Leppänen, J. M., Peltola, M. J., Linna-aho, K. & Ruuhiala, H. J. (in press). Seeing direct and averted gaze activates the approach–avoidance motivational brain systems. *Neuropsychologia.*
14. Harmon-Jones, E. (2004). Contributions from research on anger and cognitive dissonance to understanding the motivational functions of asymmetrical frontal brain activity. *Biological Psychology*, 51 –76.

Research Group for Emotions, Sociality, and Computing

Veikko Surakka, Jenni Anttonen, Toni Vanhala, Yulia Gizatdinova, Mirja Ilves, Jani Lylykangas, Katri Salminen, Jouni Vaaramo

Research Group for Emotions, Sociality, and Computing (ESC)
Tampere Unit for Computer-Human Interaction (TAUCHI)
Department of Computer Sciences
University of Tampere
firstname.lastname@cs.uta.fi

Abstract. In the Research Group for Emotions, Sociality, and Computing, we conduct basic and applied research on the use of emotionally and socially meaningful signals and processes in human-technology interaction. The main focus of our research is on developing perceptual and expressive intelligence through experimental and constructive research. Our research methods include several types of measures covering the experiential, behavioral, and physiological components of emotion. We see that emotions are a crucial part of all human motivational behavior. Thus, our results in the context of emotions and technology have the potential to significantly facilitate human-technology interaction.

Keywords: affective computing, psychophysiology, wireless physiology, non-invasive measurement, facial EMG, EEG, ECG, heart rate, pupil measurements, synthesized speech, embodied virtual agents, haptics.

1 Introduction

The Research Group for Emotions, Sociality, and Computing is a part of the Tampere Unit for Computer-Human Interaction (TAUCHI) at the department of Computer Sciences at the University of Tampere. The research group was founded in 1999 when Veikko Surakka joined the TAUCHI unit. Thereafter, the group has grown steadily and currently, in addition to the group leader, we have 7 researchers with varying backgrounds from interactive technology and psychology, to computer science and physics.

We do primarily experimental and constructive empirical research that is most often carried out in the laboratory environment. Our research covers both basic research on human processing of social and emotional information and applied research, mostly in the form of developing new interaction techniques and testing new prototypes of technologies for emotion-related measurements. Our laboratory is well equipped with technologies for measuring psychophysiological, neurophysiological, behavioral, and subjective emotion-related measurements. We also provide teaching that is related to the themes of our research group.

The core philosophy of our research is to study how emotionally and socially meaningful signals and processes can be used in human-technology interaction. Our studies are motivated by the crucial meaning of social and emotional

T. Vanhala and J. Anttonen (Eds.) *Proceedings of EHTI'08: The First Finnish Symposium on Emotions and Human-Technology Interaction.* Tampere, Finland, May 30, 2008. ISBN 978-951-44-7383-8.

processes in all human behavior. We see that emotions are an important factor also in improving the quality of human-technology interaction.

Fundamentally, we have two research lines: building perceptual intelligence into technological systems and studying the means of expressive intelligence [13]. The goal of perceptual intelligence is to have technological systems that can perceive emotionally meaningful signals from the user. For example, we have studied the classification of emotional information from expressive facial images and facial EMG data [4, 10]. Expressive intelligence, on the other hand, is aimed to provide technological systems with the ability to express emotional information in a way that is meaningful for the human user. Our results have shown that technology can be used to both evoke and regulate emotions in users, and emotional responses to technology can enhance and facilitate cognitive processes [3, 5, 8, 9, 11, 14].

2 Technology with perceptual intelligence

Interaction with technology evokes emotional responses in users. Our aim is to develop methods and technologies for perceiving signals that carry emotional information from the user. This involves both basic research on emotional responses and studying methods to perceive and process emotional information. On the level of basic research, we have studied emotional responses using response measures such as heart activity, facial expressions and facial muscle activations (electromyography, EMG), pupil size, brain activity (electroencephalography, EEG), subjective ratings of emotions, and some behavioral measures (e.g. body posture).

There is a need for methods and technologies that can be used as emotion-related input to a computer. For example, our studies analyzing expressive facial images have shown that facial landmarks, namely, eyes, eyebrows, nose, and mouth can be detected with sufficiently high rates from expressive images of high complexity. Further, we aim at improving previous landmark detection algorithms to work under more challenging conditions in real time applications [4]. We have also studied algorithms for classifying emotional reactions based on electrical facial muscle activity (i.e., EMG) and heart rate data [10, 15].

We have been involved in several projects that have developed different kinds of innovative measurement technologies. This work has aimed at developing ubiquitous, non-invasive, wireless and wearable sensors to measure emotion-related signals. One promising example is the electromechanical film (EMFi) chair that was developed together with engineering scientists. It is a regular looking office chair with embedded sensors for detecting changes in pressure. These changes are due to movements of the body and they can be further processed to reveal behavioral and physiological responses of the person sitting on the chair. For example, our studies have shown that the EMFi chair can be used to unobtrusively measure emotionally significant heart rate changes in response to emotional stimuli [1, 2].

Another interesting line of research focuses on bioelectrical and biomagnetic measurements. The currently running Face Interface project[1] aims at developing wireless, non-invasive technologies for measuring gaze direction and facial activity in different human-computer interaction tasks, for example, typing and web browsing. Our results so far include, for example, a new multimodal technique for controlling a computer interface and an easy-to-wear headband for measuring facial EMG [6, 12].

3 Technology with expressive intelligence

Building expressive intelligence involves studying the means of conveying emotional information to the user. We have accumulated experience with various types of emotionally expressive stimuli including pictures, sounds, music, synthetic speech, videos of human faces, haptic stimuli, and embodied agent designs [1, 2, 5, 7, 11]. For example, a recent study showed that even simple, artificial haptic stimulation can carry emotional information [11].

Our results support the view that similar social and emotional cues and processes are active in both human-human and human-computer communication. For example, we have found that positively worded synthesized speech can be used to evoke positive subjective experiences and significant physiological effects in terms of electrical facial activity [5]. As an example of proxemic cues in HCI, the distance to a virtual humanlike head was found to affect subjective experiences of dominance (i.e. from being controlled to being in control) [9].

One of the aims of our work is to understand how users' emotions can be regulated in a way that is beneficial to the user, for example, by evoking positive emotions in the user. According to our results, these evoked emotions may then facilitate the user's performance and behavior. For example, when positively worded interventions from a speech synthesizer were delivered after a pre-programmed mouse delay, they significantly improved problem solving performance and increased smiling behavior [8]. In a similar study, participants were lead to believe that they were given feedback based on their performance in solving mathematical tasks [3]. It was found that positive feedback from a speech synthesizer significantly facilitated performance, although the feedback was not actually related to the level of performance.

We have also studied methods for voluntarily regulating emotion-related physiological processes in human-technology interaction. Our results showed that computer guided voluntary facial activations significantly affected autonomous nervous system activity as measured by heart rate and heart rate variability [14]. In sum, our findings suggest that social, emotional, and cognitive processes can be regulated using both voluntary and involuntary, automatic responses to technology (e.g., synthesized feedback and instructions).

---

[1] The research in the Face Interface project is carried out by the Wireless User Interfaces (WUI) consortium that consists of four research groups from the University of Tampere and the Tampere University of Technology (see http://www.cs.uta.fi/hci/wui).

4 Future vision

One of the ultimate aims of affective computing is to build a socio-affective loop where social and emotional information is exchanged in a meaningful fashion between the user and the technology. Our approach to this task has been to investigate new kinds of perceptual and expressive capabilities for technology. In this process we have found several channels that show promise for intelligent perception and expression. Further, our results so far have shown that emotional and social communication with technology may significantly benefit the user in terms of inducing positive experiences and facilitating her or his performance. In a larger context, our results suggest that computers and other technology items may soon evolve from simple tools to social and emotional partners that help us in expressing and regulating our behavior and experiences.

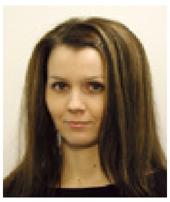
Acknowledgments. The research reported in this paper has been supported by the Academy of Finland, Graduate School in User-Centered Information Technology, the University of Tampere Foundation, Information Society Institute, and Finnish Funding Agency for Technology and Innovation (TEKES).

Veikko Surakka, PhD (psychology), professor in interactive technology. My research interests originate from the study of human emotions. Those interests have included, for example, facial and neural mechanisms of emotions, emotional contagion and mimicry mechanisms. Since 1999 I have worked in the area of human-computer interaction focusing on the use of social-emotional processes for user interface and communication development.

Jenni Anttonen, M.Sc. My doctoral studies concern measuring users' emotional responses by using physiological signals and exploring how that information could be used in HCI. My research interests also include studying the emotional aspects of user experience and how technology can contribute to social and emotional well-being of older adults.

Toni Vanhala, M.Sc. My doctoral thesis aims to construct a computer system for regulating emotions based on physiological and behavioral measures. The system is to be used for the treatment of mild social and emotional disorders using adaptive virtual stimulation. The Face Interface project benefits my thesis work by enabling new wireless and wearable measurments for this purpose.

Yulia Gizatdinova, M.Sc. My research interests include computer vision, image processing, pattern recognition, and human visual perception. In my doctoral work, I developed methods of automatic localization of facial landmarks from expressive images of high complexity. Recently, I have started new research on developing gaze tracking algorithms to estimate the focus of a user's attention.



Mirja Ilves, M.A. I received my Master of Arts degree in psychology from the University of Tampere in spring 2003. In the autumn of 2002 I started to work in ESC group as a psychology trainee and since then I have worked as a researcher. In my doctoral work I will concentrate emotional feedback and intervention in human computer interaction.

Jani Lylykangas, M.A. I am a psychologist and I have been working in ESC group since 2001. At the moment I am working in Tekes funded Mobile Haptics project. Previously I have participated in several research projects where I have conducted empirical research and surveys in the context of human-technology interaction and human emotions.

Katri Salminen. I'm a psychologist who has been working at TAUCHI from 2004. Currently I'm a post-grad student and working in Tekes funded Mobile Haptics project. My interests involve HCI in general, but especially emotions evoked by different user interfaces and feedback methods.

Jouni Vaaramo. I'm a research assistant in the ESC group. I am currently finishing my M.Sc. thesis that is focused on new multimodal techniques for typing text and browsing the web in the Face Interface project.

References

1. Anttonen, J. and Surakka, V. (2005). Emotions and heart rate while sitting on a chair. *Proeedings. of CHI'05*, 491–499.
2. Anttonen, J. and Surakka, V. (2007). Music, heart rate, and emotions in the context of stimulating technologies. *Proceedings of ACII 2007, LNCS*, 4738, 290-301.
3. Aula, A. and Surakka, V. (2002). Auditory emotional feedback facilitates human-computer interaction. *Proceedings of HCI 2002*, 337–349.
4. Gizatdinova, Y. and Surakka, V. (2006). Feature-based detection of facial landmarks from neutral and expressive facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligance*, 28, 1, 135–139.
5. Ilves, M. and Surakka, V. (2004). Subjective and physiological responses to emotional content of synthesized speech. *Proeedings. of CASA 2004*, 19–26.
6. Nöjd, N.; Puurtinen, M.; Niemenlehto, P.; Vehkaoja, A.; Verho, J.; Vanhala, T.; Hyttinen, J.; Juhola, M.; Lekkala, J. & Surakka, V. (2005). Wireless wearable EMG and EOG measurement system for psychophysiological applications, *Proceedings of NBCBME 2005*, 144-145
7. Partala, T., and Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, 59, 1-2, 185-198.
8. Partala, T., and Surakka, V. (2004). The effects of affective interventions in human-computer interaction. *Interacting with Computers*, 16, 2, 295-309.
9. Partala, T., Surakka, V., and Lahti, J. (2004). Affective effects of agent proximity in conversational systems. Proceedings of NordiCHI 2004, 353-356.
10. Partala, T., Surakka, V., and Vanhala, T. (2006). Real-time estimation of emotional experiences from facial expressions. *Interacting with Computers*, 18,2, 208–226.
11. Salminen, K., Surakka, V., Lylykangas, J., Raisamo, J., Saarinen, R., Raisamo, R., Rantala, J., and Evreinov, G. (2008). Emotional and behavioral responses to haptic stimulation. *Proceedings of CHI'08*, 1555-1562.
12. Surakka, V., Illi, M., and Isokoski, P. (2004). Gazing and frowning as a new human-computer interaction technique. *ACM Transactions on Applied Perception*, 1, 1, 40–56.
13. Surakka, V. and Vanhala, T. Emotions in human-computer interaction. A book chapter to appear in A. Kappas (Ed.) (2008). *Multi-channel communication on the Internet*. New York: Cambridge University Press (Accepted).
14. Vanhala, T. and Surakka, V. (2007a). Facial Activation Control Effect (FACE). *Proceedings of ACII 2007, Lecture Notes in Computer Science*, 4738, 278-289.
15. Vanhala, T. and Surakka, V. (2007b). Recognizing the Effects of Voluntary Facial Activations Using Heart Rate Patterns. *Proceedings of the 11th WSEAS Int. Conf. on Computers*, 628-632.

Interactive emotional embodied experience

Tapio Takala, Inger Ekman, Antti Poikola, Meeri Mäkäräinen

Department of Media Technology
Helsinki University of Technology
Firstname.Lastname@tml.hut.fi

Abstract. In the context of experiential applications such as games and interactive art, we investigate emotional man-machine communication by various non-verbal cues, including gaze tracking and pupil size, gestures and other bodily motion, as well as acoustical manipulation of sound.

Keywords: multimodality, post-WIMP interaction, motion, eye tracking, pupils, facial expression, gesture, human animation, emotional sound.

1 Introduction

Since early 90's there has been active research at Helsinki University of Technology (TKK) on virtual environments. Our spearhead focus has been virtual acoustics combined with high-quality visuals and interactive control. Currently the Art & Magic group consists of researchers interested more broadly in new media and unconventional interaction technologies, with a special inclination towards artistic applications. Our research method is often constructive, building prototypes of interactive systems, with which the users' behavior and experience are tested. In order to make emotionally engaging applications, such as games or art installations, not only traditional usability but the whole user experience is important.

Bodily motion is a research theme approached by character animation on one hand, and human motion tracking on the other. An early showcase was the virtual orchestra, an animated band conducted with motion of a baton. Musical emotions of the conductor are supposed to be communicated through the baton, which was partly confirmed in our tests (Ilmonen and Takala, 2005). Later several embodied game applications have been built, allowing free motion of the player through camera-based tracking, or utilizing physiological measures, such as heart rate, in interaction. However, emotions have been in our direct focus only recently.

2 Ongoing research

2.1 Pupil Based Interaction

We investigate consciously controlled pupil size as an input modality. Pupil size is affected by various processes, e.g., physical activation, strong emotional experiences and cognitive effort. Our approach to pupil-based interaction is basically a combination of biofeedback with affective computing. Continuous feedback allows users to know their pupil size at any time, allowing them to map

T. Vanhala and J. Anttonen (Eds.) *Proceedings of EHTI'08: The First Finnish Symposium on Emotions and Human-Technology Interaction.* Tampere, Finland, May 30, 2008. ISBN 978-951-44-7383-8.

this information with what they do and how they feel. Effectively, we reverse the traditional setting of Affective Computing: instead of monitoring the user's emotions with the intention to appropriate the system's responses, we use system responses to guide users to change how they feel. This is, at the same time, continuing on a line of research on unconventional human-computer interfaces, such as using brainwaves, skin conductivity (SC) and heart rate (HR) for controlling a computer.

The results we have so far indicate that knowing the relationship between actions and pupil size, with a little training, people learn to use self-invoked emotions to indirectly control pupil size. We have investigated how seven specific activities influence pupil size and the results show statistically significant increases in pupillary size variation for physical activity, self-induced pain, positive emotions and the changing of gaze focus point between near and far points. Another experiment has confirmed that subjects can produce pupil dilation and constriction on demand during paced tasks. Whereas pupil manipulation is possible with bodily activation methods (such as muscle effort and self inflicted pain), most interesting aspect for future research in our view is that pupils provide a communication channel for expression of emotional activity. When given free choice of methods, most subjects spontaneously chose positive thoughts or smiling as one component of their pupil control and, furthermore, used these methods successfully to manipulate their pupil size. This specific aspect is significant in utilizing pupil manipulation in the interface.

Originally, we set out to use pupil feedback as a way of modeling willpower in the user interface. We conceptualized willpower as a combination of concentration and cognitive load. However, user tests now indicate a more promising approach than cognitive load is excitement, especially positive excited 'hype'. We continue on this investigation and test different approaches to control in an experimental game application using various forms of eyes-only interaction.

2.2 Emotional and non-verbal cues in social interaction

Information in different forms is transmitted between participants in a situation of social interaction. Beside the verbal content many subtle emotional and non-verbal cues, including gaze behavior, paralanguage (non-verbal elements of speech), facial expressions, changes in breathing, body posture and gestures, are received and interpreted by the participants. Most of these cues are result of spontaneous behaviors and in many cases they are also interpreted unconsciously by the other party. We believe that the experience of the social situation is significantly influenced by these non-verbal cues.

We are interested in the process of forming an impression of the other party's personality and emotional state in a social situation. Specific questions are, how the impression of the other party is formed in a situation of social interaction, and which non-verbal cues are important in forming the impression of the other party in social interaction?

We aim at developing believable interactive agents capable of transferring different emotions and characteristics of personality in nonverbal communication. We will make behavioral studies of direct human-human interaction in order to identify the meaningful non-verbal cues and form behavioral models for natural nonverbal communication. The models will be implemented for an interactive agent and experimentally validated in situations of social interaction.

2.3 Emotional Sound

The conventional use of sound in static audiovisual media (such as film and television) has shown the expressive power of sound, not only as a means of transferring information but in defining mood and guiding emotions. Whereas much of the effect is linked to the identity of a sound (what), higher level modifiers like spatialization (how) heavily influence meaning and emotional tone (e.g. Västfjäll, Larsson & Kleiner 2002).

The goal of our research in this area is to find, implement and experimentally validate such higher level modifiers. We are now identifying and mapping features influencing semantic and emotional meaning of sound reproduction and applying them in real-time virtual acoustics. We also investigate methods of deliberate exaggeration as a tool of influencing sound meaning.

3 How to Handle Emotions in the User Interface?

Our research is done in the spirit of "post-WIMP" Human-Computer Interaction (Jacob et al. 2007). This paradigm views user perceptions as based on action rather than cognition (Dourish 2002). The approach highlights the importance of considering the user experience on a general level instead of taking a narrow focus as suggested by concepts like *usability*.

Emotions are part of all human experience and emotions arise in interaction with computer whether or not an interface is designed with an explicit purpose to cause emotional experiences. Experience can be thought of as a snap-shot of the user's state at any moment. The holistic *experience* is influenced and interpreted in the context of various processes. Emotions present one way of looking at and labeling some of these sensations and their (optional) related cognitive interpretations. However, we consider a rather liberal array of background factors, and also include sensations with less clear relations to outer events than the typical list of emotions (like hunger and sleepiness).

3.1 Emotions are embodied

Emotions are crucially part of how people make sense of the world. The interplay between emotion and cognition is strong, and the influence goes in both directions (Damasio, 1994). Also, while emotions make us react bodily, bodily actions can also induce emotions (Zajonc, 1985). Because emotions influence how we think and behave, emotions can be detected by looking at human behavior. Facial gestures are commonly used in emotion detection, but emotion can be approached via less obvious measurements like the movement style of a conductor's stick. On the other hand, since behavior influences emotion, we can design types of interaction that make people experience emotions not only because the information mapping works smoothly or poorly but because the actions that are required in the interaction are emotional in themselves, like having interaction work by making a smiling face or thinking really angry thoughts.

Taking emotions into account in the user interfaces means designing for expressiveness. Interaction should consider the whole experience including embodied aspects of use. Interaction can encourage, even challenge, users to have emotions and express them strongly.

3.2 Emotion is interaction

One challenge is to design communication channels that allow conveying various types of cues implicitly, using modalities that are sensitive to emotional expression.

- Use input modalities that are sensitive to emotionally salient cues.
- Identify and preserve emotional cues and convey them to the recipient.
- Present emotional cues in a form that allows the recipient to react to them without having to be consciously aware of it.

A common approach in Affective Computing (Picard, 1997) is to try and identify the user's emotions and modify the program response accordingly, in order to influence user mood and behavior. Here, the important part is not identifying and labeling emotions, but making sure emotionally salient information is allowed space in the communication channel.

We believe that emotions are not passive responses to stimulus events – emotional expression is rooted in two-way communication. Communication also teaches people how to show, or not to show, emotions to express what they want in a certain context. People do not automatically display to everyone what they feel, but use emotional expression purposefully and also hide and fake emotions. Any interface that changes its responses according to user emotions will influence how (and whether) emotions are expressed, not only how the user feels.

3.3 Emotion is fun – and horrible, sad, shocking

People do not seek only positive emotional experiences. While a restriction to positive emotions can be motivated in some studies, stressful, sad or scary events also have their place in the interface. People are rather good at using media (music, film, games) purposefully to regulate and guide their emotions. Emotion research should not forget that negative emotions are as important in these contents as positive ones.

Tapio Takala studied physics, electronics and computer science at Helsinki University of technology, receiving his Dr.Tech in 1987 on modeling in computer aided design. Since 1995 he has been professor of interactive digital media, currently as head of the Dept. of Media Technology. His research interests include computer graphics and interactive technologies, animation and 3D modeling, virtual reality, sound and music, with an inclination to artistic applications and creativity. Current research focus is in embodied and emotional “non-WIMP” interaction.

Inger Ekman received her MSc from University of Tampere in 2003. She is a doctoral student at University of Tampere and a member of the KIT graduate school. Her research topics include sound design techniques for interactive environments and game design. In her current research, she is looking into techniques for influencing sound semantics and emotional content in real-time interactive sound environments.

Meeri Mäkäräinen received her MSc. from Helsinki University of Technology in 2006. She is a doctoral student at Helsinki University of Technology and member of the UCIT graduate school. She is also studying for a MA degree at the University of Art&Design. Her research focuses on animation of emotional gestures and especially facial expressions. This work includes developing models for emotional and expressive behavior of autonomous agents.

Antti Poikola is a MSc student at Helsinki University of Technology, expected to graduate in 2008. His research interests are emotional and non-verbal cues in social interaction.

References

1. Damasio, Antonio R. (1994). Descartes’ Error: emotion, reason and the human brain. Putnam’s Sons.
2. Dourish, P. (2001). Where the Action Is. MIT Press.
3. Ilmonen, T; Takala, T (2005). Detecting Emotional Content from the Motion of an Orchestra Conductor. The 6th Intl. Workshop on Gesture in HCI and Simulation, Vannes, France.
4. Jacob, R.J.K.; Girouard, A.; Hirshfield, L.; Horn, M.; Shaer, O.; Solovey, E.; and Zigelbaum, J. Reality-based interaction: unifying the new generation of interaction styles. Proc. CHI '07 extended abstracts, 2465-2470.
5. Picard, R. (1997). Affective Computing. MIT Press.
6. Västfjäll, D; Larsson, P.; and Kleiner, M. (2002) Emotion and Auditory Virtual Environments: Affect-Based Judgments of Music Reproduced with Virtual Reverberation Times. *CyberPsychology & Behavior*, Vol. 5 (1): 19-32.
7. Zajonc, R.B. (1985). Emotion and facial efference: a theory reclaimed. *Science*. Apr 5;228 (4695): 15-21.

Pleasure and enjoyment in digital games

Jari Takatalo[1], Jukka Häkkinen[2], Jari Lipsanen[1], Miikka Lehtonen[1], Jyrki Kaistinen[1] and Göte Nyman[1]

[1]Department of Psychology, University of Helsinki, Finland
P.O. Box 9, FIN-00014 University of Helsinki
firstname.lastname@helsinki.fi
[2]Nokia Research Center, Ruoholahti, Finland
P.O. Box 407, 00045 Nokia Group
jukka.hakkinen@nokia.com

Abstract. Digital games are played for the experiences they offer. Usually the quality of this experience is considered positive and fun. Here we analyze the difference between two different empirically derived fun measures. The study is theoretically grounded on the key concepts of flow and cognitive theories of emotion. Structural equation modeling was used to test these theories in an empirical dataset (n=2182). The results deepen previous findings of how cognitive evaluation affects emotional outcomes. Pleasurable gaming experience is received when one feels competent, whereas enjoyment requires also arousing challenges. A need to understand psychological processes in order to create desired gaming experiences is discussed.

Keywords: emotion, cognition, digital games, user experience in games.

1 Introduction

Fun is considered key motivation to play digital games. Several factors (e.g., challenge, sociality) are assumed to contribute fun gaming experience [11]. However, the number of fun dimensions and their components is not well understood [7]. In psychology, fun, enjoyment, and pleasure are all seen as agreeable affective reactions. However, there are studies, which make distinction between pleasure and enjoyment, stating that simplistic hedonic models are not enough to account for all behaviors [5]. For example, a rock climber may have unpleasant, yet exhilarating and highly motivating experiences while climbing. There is also empirical evidence that pleasure (hedonic valence) and enjoyment can be distinguish in a digital gaming context [16]. Whether an experience is pleasurable or enjoyable requires analysis of the person-situation relationship more deeply (e.g., abilities and evaluations) [5]. In such analysis the close relationship between cognition and affect must be included.

Cognitive theories of emotions [6,10] suggest that cognitive interpretations and appraisals of events in the world are essential parts of emotions. There are various appraisal features and components, such as the effort anticipated in a situation, perceived obstacles, and the sense of control, all of which shape the emotions attached to these events [6]. In addition, these theories suggest that cognition and emotion form combinations, which in turn affect oncoming evaluations and

T. Vanhala and J. Anttonen (Eds.) *Proceedings of EHTI'08: The First Finnish Symposium on Emotions and Human-Technology Interaction.* Tampere, Finland, May 30, 2008. ISBN 978-951-44-7383-8.

emotions [9]. The theory of flow [5] shares the core idea of these theories. In flow, the cognitive evaluation is made between evaluated challenges of the situation and the skills one possess. A different ratio between these two is likely to lead to various emotional outcomes. Also cognitive-affective combinations (e.g., challenge-arousal) leading to positive flow -experiences have been empirically found among the Internet users [12].

In this study we disclose a specific cognitive-affective process underneath distinct pleasure and enjoyment dimensions. The results show how cognitive evaluations of competence (skill) and challenge are combined with emotions (arousal and control) to form super-ordinate concepts, which in turn shape both pleasure and enjoyment. These findings are in line with cognitive theories of emotion [9]. Disclosing the complex cognitive-affective process deepens the findings such as the assumption that challenge leads to fun in games.

2 Methods

2.1 Origin and collection of the dataset

The data have been collected from both laboratory experiments and an Internet survey using the EVE-Experience Questionnaire (EVEQ-GP) [15,17]. In the field of behavioral sciences the use of questionnaires has proven to be a valid way of assessing various mental phenomena [2,4,8,13]. Both the paper and pencil and the online versions of the EVEQ-GP were composed of 180 items (1-7 Likert-scale and semantic differentials) measuring different experiential aspects obtained from being and performing in the game world. The instructions for completing the EVEQ-GP encourage participants to reflect on their subjective gaming experience of one particular gaming session; moreover, the instructions emphasized that the questionnaire was to be completed immediately after a gaming session. Thus, the gaming experience was operationalized as a situated experience stemming from one particular game. The method enables the player to report, within pre-set multidimensional boundaries, how it felt to interact with a digital game. Also included were 27 background questions. In the present study different laboratory experiments and an Internet survey were included in large dataset.

2.2 Description of the dataset

The data consist of 2182 subjects (1972 males, 210 females), who filled in the questionnaire. The mean age of the respondents was 21.5 years (SD=6.0). The average time of playing was 127 minutes (SD=111) and the average size of the display used was 19.2” (SD=4.4). 33% of the respondents played daily, 29.6% played at least every other day, and 24.5% played often but not as often as every other day.

Most of the games played (31.5%) before the questionnaire was filled in were first-person shooters (FPS) either online (15.0%) or offline (16.5%). The second most popular genre (15.0%) was massive multiplayer online role-playing games

(MMORPG), and the third (13.1%) was single role-playing games (RPG) (13.1%). The most popular single game played was *World of Warcraft* (n=265), which is a MMORPG. Altogether the data included approximately 320 different games, giving a broad scope to the psychology of digital games. Since *Pelit* magazine focuses on PC games, 85.2% of the games were played with a PC and 14.8% with a console.

2.3 The scales used

In this study, six measurement scales composed of 42 questionnaire items were used. More information concerning the scale composition is provided by our previous studies [14,16]. C*hallenge* scale (5 items, Cronbach's α=.69) is composed of items measuring how challenging and demanding the game felt. *Competence* scale is a combination of items measuring user skills and positive feeling of effectiveness (11 items, α=.87). *Control* (5 items, α=.71) measures the feeling of being in control and independent, whereas *arousal* (5 items, α=.71) being active and stimulated instead of passive and unaroused. Hedonic *valence* is the bipolar pleasure /displeasure scale (10 items, α=.86) that is mainly composed of semantic differentials (e.g., happy/sad). Also items measuring boredom and anxiety were negatively associated to valence, emphasizing the bipolarity of the scale. *Enjoyment* scale (7 items, α=.83) includes aspects of pleasantness and agreeableness. Playing was also evaluated exciting and somehow special (e.g.," I will recommend it to my friends" and "I had peak experiences while playing"). A model was constructed to study the relationships between the cognitive-affective scales and disclose the difference between valence and enjoyment in gaming context. Structural equation modeling (SEM) was used to test the model with AMOS 7 software. SEM offers an assessment of the fit between a model and observed data.

3 Results

We included two latent cognitive-affective constructs for competence-control and challenge-arousal into the model. Since cognitive theories of emotion [9] suggest that new emotions arise from such appraisal combinations, we studied how the two latent cognitive-affective constructs were related to valence and enjoyment. The model fit was tested with a Maximum likelihood estimation method. Overall goodness-of-fit of the model was reasonable with root-mean-error of approximation (RMSEA) equal to 0.052. RMSEA values about or below 0.05 indicate a close fit of the model in relation to degrees of freedom, and values below 0.08 indicate a reasonable fit [3]. Also comparative fit index (CFI=0.959), goodness of fit index (GFI=0.993), and adjusted goodness of fit index (AGFI=0.978) were above the suggested minimum value of 0.9 [1], indicating a good model fit. Although, the chi-square test was significant ($\chi^2(7)= 47.77$, $p<0.000$) it is known that this tests gets exceedingly imprecise with larger samples [3]. All standardized path coefficients were significant. Estimated correlation between latent competence and challenge structures was significant at $p<0.05$, all the others were significant at $p<0.001$ (see Figure 1.).

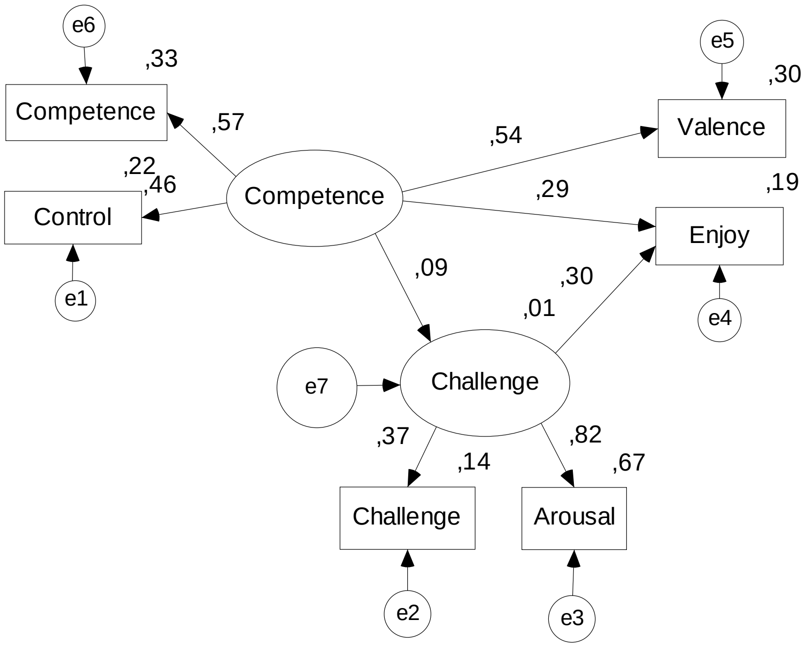


Figure 1. Structural equation model of competence, challenge and emotions.

4 Discussion

Digital games are played for the experiences they offer. Typically the quality of this experience is considered positive and fun. This study shows the difference between two empirically derived measures of fun, that is, pleasure and enjoyment [16]. Structural equation model based on cognitive theories of emotion [6,10] and theory of flow [5] was conducted. The results indicated that being competent and in control in the game leads to pleasurable experience. It was also shown that in addition to being competent and in control, enjoyable gaming experience requires challenges and elevated arousal. This combination makes enjoyment more intensive and stronger emotion as compared to pleasure. Enjoyment being more complex emotion compared to pleasure is also suggest by Csikzentmihalyi [5]. Our results also support previous theorizations and empirical findings that cognitive activities can be combined with emotions to form latent constructs [9,12].

In its current form the model shows how game mechanics are cognitively evaluated and emotionally experienced. In future studies a motivational component should be included in the model in order to understand more about the measured emotions [9]. Also other relationships besides linear should be considered to understand, for example, “bad challenges”. Considering player psychology in more depth in the game design phase will lead to more controlled and desired gaming experiences.

Acknowledgments. We thank Antti Hulsi, Heikki Särkelä and Jeppe Komulainen for help in collecting the data. This work has been supported by the User Centered Information Technology graduate school.

Jari Takatalo, Lic A. (Psyc.). I am a PhD student working in the Psychology of Evolving Media and Technology (POEM) research group at the University of Helsinki. My doctoral studies concern understanding and measuring user experience in digital environments such as games.

Jukka Häkkinen, PhD. I'm a PhD in experimental psychology working in both POEM and Nokia Research Center in Helsinki. My current main research focus is in defining subjective image quality of stereoscopic contents.

Miikka Lehtonen, I am a M.A. (Psyc.) student working in POEM. My studies concern social issues in digital gaming.

Jari Lipsanen, M.A. (Psyc.). I am a PhD student working at the department of psychology. My doctoral studies concern psychometrics.

Jyrki Kaistinen, M.A. (Psyc.). I am a PhD student working in POEM. My studies concern ergonomics and distributed virtual collaboration.

Göte Nyman, professor. I am the leader of POEM and have specific interest in the subjective methodology development and technological application issues within the scope of POEM.

References

1. Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin,* 107, 2, 238-46.
2. Breakwell, G. M. (2006). *Research Methods in Psychology.* Sage Publications.
3. Browne, M.W. and Cudeck, R. (1993). Alternative ways of assessing model Fit. In: Bollen K.A. and Long J.S., (Eds.), Testing structural equation models. Sage Publications, 136-162.
4. Couper, M. (2000). Web surveys: a review of issues and approaches. *Public opinion quarterly,* 64, 4, 464-94.
5. Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety.* Jossey-Bass Publishers.
6. Ellsworth, P. C. and Smith, C. A. (1988). From appraisal to emotion: Differences among unpleasant feelings. *Motivation and Emotion,* 12, 3, 271-302.
7. Hocking C. (2006). The last 3D revolution (Keyote speech in Futureplay 2006). Available at: http://clicknothing.typepad.com/Design/hockingc_FPLAY06_Revolution.zip. Accessed January, 2007.
8. Labaw, P. J. (1980). *Advanced Questionnaire Design.* Abt Books.
9. Lazarus, R. S. (1991). Cognition and motivation in emotion. *The American Psychologist,* 46, 4, 352-67.
10. Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *The American Psychologist,* 46, 8, 819-34.
11. Lazzaro N. (2004). Why We Play Games: Four Keys to More Emotion Without Story. Available at: http://www.xeodesign.com/whyweplaygames/xeodesign_whyweplaygames.pdf. Accessed November, 2007.
12. Novak, T. P., Hoffman, D. L. and Yung, Y. F. (2000). Measuring the Customer Experience in Online Environments: A Structural Modeling Approach. *Marketing Science,* 19, 1, 22-42.
13. Rust, J. and Golombok, S. (1999). *Modern Psychometrics: The Science of Psychological Assessment.* Routledge.
14. Takatalo, J., Häkkinen, J., Särkelä, H., Komulainen, J. and Nyman, G. (2006). Involvement and presence in digital gaming. *In Proc. NordiCHI 2006,* 393-6.
15. Takatalo, J. (2002). Presence and flow in virtual environments: an explorative study. *Master's thesis,* University of Helsinki.
16. Takatalo, J., Häkkinen, J., Lehtonen, M., Komulainen, J., Kaistinen, J. and Nyman, G. (Submitted). Quality of the user experience in digital gaming.
17. Takatalo, J., Häkkinen, J., Särkelä, H., Komulainen, J. and Nyman, G. (2004). The experiential dimensions of two different digital games. *In Proc. PRESENCE 2004,* 274-278.

Automatic discrimination of emotion in spoken Finnish: research utilizing the *MediaTeam Speech Corpus*

Juhani Toivanen[1], Eero Väyrynen[2], Tapio Seppänen[3]

[1]Academy of Finland and University of Oulu
juhani.toivanen@ee.oulu.fi
[2]University of Oulu
eero.vayrynen@ee.oulu.fi
[3]University of Oulu
tapio.seppanen@ee.oulu.fi

Abstract. Research on the automatic discrimination of emotion in spoken Finnish is described. The research is based on a large emotional speech database, the *MediaTeam Speech Corpus*. A speech analysis software, the *f0Tool*, is described: the algorithm can analyze some 45 acoustic/prosodic parameters in a speech sample fully automatically. The parameters can be used in experiments on the automatic discrimination of basic emotions in Finnish, with potentially important application fields.

Keywords: automatic discrimination of emotion, spoken Finnish, prosody, speech corpus

1 Introduction

It is well known that emotions play an important role in social interaction, both regulating and displaying behavior. It is equally well known that, in the phonetic sciences, the vocal expression of emotion has been studied over a long period of time, and the vocal parameters of the basic emotions are now understood relatively well. A common view is that pitch (fundamental frequency, f0), is, perhaps, the most important parameter for the discrimination/recognition of emotion; other relevant parameters include energy, duration and speaking rate.

Speech scientists and engineers are now taking an increased interest in the role of emotion in speech communication; witness, for example, the ISCA Workshop on Voice and Emotion (Newcastle, Northern Ireland). It seems that the study of the vocal expression of emotion is steadily reaching a level of maturity where the main focus is on important applications, particularly those involving human-computer interaction in various forms. In the area of information retrieval, the potential of prosodic/acoustic cues in signaling different affective speaker-states is receiving more attention in the form of prosodic data mining. Potential, and/or already existing, applications include hi-tech call-center environments and automated telephone systems with a possibility to localize customer discontent automatically, more powerful data mining solutions (e.g. content-based information retrieval for audio databases), and even medical applications of different kinds. For example, a "hot spotter" that could localize potentially escalating communicative break-downs in routine patient-caretaker interaction,

T. Vanhala and J. Anttonen (Eds.) *Proceedings of EHTI'08: The First Finnish Symposium on Emotions and Human-Technology Interaction.* Tampere, Finland, May 30, 2008. ISBN 978-951-44-7383-8.

making an alarm when the emotionally symptomatic speech behavior changes abruptly, would be a (highly) useful application: the successful system would monitor patient messages and prioritize them with respect to their emotional content. Entertainment applications, including robotic pets that “understand” human emotional behavior, are currently becoming increasingly available (and usable).

The human and/or automatic classification/discrimination of emotional content from speech has been studied for a number of languages, including minor languages such as Finnish [1]. Our aim is to approach the issue of the vocal expression of emotion from a viewpoint of automatic classification of spoken Finnish. A relevant application, for our research group, would be a search engine capable of browsing Finnish audio files according to their emotional content (and such a search engine would utilize, first and foremost, prosodic information although e.g. lexical features would also be of relevance). To work towards this goal, experiments on the automatic classification of spoken affect are necessary, and this activity forms the core of the research carried out utilizing the *MediaTeam Speech Corpus*.

2 Speech data

2.1 Speech data set I

The *MediaTeam Speech Corpus*, consisting of two data sets, is an extensive emotional speech database for Finnish. Fourteen professional actors (8 men, 6 women, Oulu City Theater) produced the speech data; the speakers were paid for their services. The speakers were aged between 26 and 50 (average age was 39), without any known hearing or speaking disorders. All were speakers of the same northern variety of Finnish. The speakers read out a phonetically rich Finnish passage of some 120 words simulating three basic emotions, in addition to “neutral”: “sadness”, “anger”, and “happiness/joy”. The text was emotionally completely neutral, representing matter-of-fact newspaper prose. The speakers were encouraged to take their time to prepare for each emotional state, and they could retake the reading (as many times as they wished) if they were not satisfied with the first rendition.

The recording was made in an anechoic chamber using a high quality condenser microphone and a DAT recorder to obtain a 48 kHz, 16-bit recording. The microphone was placed at a distance of 50 cm from subject’s lips. Calibration for intensity measurements was obtained from a standard voice source (93.6 dB). The data was stored in a PC as *wav* format files. The collected speech material was used in the experiments unedited as regards the content: the speech samples occasionally contained brief semi-conventionalized expressions of affect (e.g. sighs and laughter). These “affect bursts”, clearly identifiable by human listeners, were *not* deleted from the files.

2.2 Speech data set II

The speech database was enlarged with new material including a total of 450 simulated emotional speech samples (5x10x9). The samples were produced in five emotional contexts, all repeated 10 times by 9 professional actors, five men and four women, aged between 26 and 45 (some of whom had participated in the first round of data productions sessions). The samples were recorded in collaboration with the Helsinki University of Technology (HUT) in an anechoic room at the University of Oulu. The speech data consisted of multiple repetitions of the following emotions in Finnish speech: "neutral", "sadness", "joy", "anger", and "tenderness". The subjects read out a passage of some 80 words from a Finnish novel admitting several emotional interpretations but not containing, semantically/lexically, any obvious emotional content. Each speaker produced, in a random order, ten renditions of each emotional state (however, an emotional state was never repeated in the next rendition). The subjects were given a full interpretative freedom as to how, vocally, the emotions should be portrayed.

The technical procedure (concerning equipment, recording arrangements, etc.) was similar to that used in the first data collection round.

2.3 Emotions

The particular emotions were chosen in order to obtain speech data on basic emotional states. It can be argued that at least "anger", "sadness", and "joy" represent the very core of basic emotions in human affective behavior, and the "neutral" emotion is an appropriate baseline level. Regarding the second speech database, the five particular emotions (including "tenderness") apparently represent both positive and negative valence and high and low psycho-physiological activity levels, the neutral emotion having a zero value for valence and activity.

3 Acoustic analysis

3.1 f0Tool

The acoustic analysis was carried out with the *f0Tool*, a speech analysis software implemented in the MATLAB language; the *f0Tool* is a cepstrum-based voiced/unvoiced segmentation and time domain f0 extraction algorithm using waveform-matching [2]. The algorithm features the following processing stages: voiced/unvoiced segmentation, f0 contour estimation and prosodic feature calculation. The algorithm is capable of analyzing large speech data sets in a single batch run or single file analysis with a GUI for visual inspection and verification purposes. The only input required by the f0Tool is an audio waveform file; no phonemic annotations are used. The performance of the f0Tool was tested on a very demanding data set with radio conversations involving Finnish fighter pilots. The reliability of the algorithm was found to be very good; see [1].

3.2 Parameters

From the speech data, more than 40 prosodic/acoustic parameters can be automatically extracted and analyzed with the f0Tool. The parameters are f0-related, intensity-related, temporal and spectral parameters generally known to correlate with emotions in speech. Some of the parameters are listed below; see [1] and [2] for a full list of parameters and the explanations.

Table 1. A sample of the prosodic parameters extracted and analyzed by the *f0Tool*.

Feature	Definition
Median	Median f0 (Hz)
frac Range	5%-95% f0 range (Hz)
Range	f0 range (Hz)
LF500	proportion of low frequency energy below 500 Hz
GD riseav	average f0 rise steepness (Hz/cycle)
GDmax	max steepness of f0 rise (Hz/cycle)
GDposav	average f0 rise during continuous voiced segment (Hz)
GDrisemax	max f0 rise during continuous voiced segment
MeanInt	mean RMS intensity (abs., dB)
fracIntRange	5%-95% intensity range (abs., dB)
MavLngth	average duration of voiced segments
ManLngth	average duration of unvoiced segments shorter than 300 ms
Max_Nlngth	max duration of unvoiced segments
Vratio	ratio of voiced segments

4 Status praesens

The automatic discrimination of emotion for the particular emotional label sets has been carried out, and human listening tests have been used to obtain baseline data or "truth" data for comparison. In general, the statistical classification experiments indicate that, with an optimal combination of prosodic feature vectors, automatic emotion discrimination close to human emotion recognition ability is attainable. The interested reader finds the references at http://www.mediateam.oulu.fi/research/lat/?lang=en.

Currently, the team's research interest is beginning to embrace some additional levels of emotion display in human communication. First of all, it must be pointed out that, in addition to automatically measurable prosodic features, more linguistically (phonologically) oriented parameters should be looked into in the research on the automatic discrimination of vocal affect. To this end, it will be necessary to annotate the speech data in terms of e.g. ToBI system to be able to describe the *intonational* or *phonological*, as opposed to *phonetic*, prosodic structure. Secondly, at some point, it may be necessary to investigate other bio-signal features of emotion display, e.g. heart rate variation, pupil size variation, etc. Research on these topics in the team is underway.

Acknowledgments. The authors gratefully acknowledge the support of Academy of Finland.

Juhani Toivanen, Docent. I am Academy Researcher and Docent in Applied Speech Research and Speech Technology. I am interested in the vocal communication of affect in Finnish and English. My background is in applied linguistics and phonetics, as well as second language acquisition research.

Eero Väyrynen, M.Sc., (Eng). I am a post graduate student and researcher with the MediaTeam Oulu research group. My research interests include digital signal processing and statistical pattern recognition with applications in various semantic concept recognition tasks (e.g. emotional state) using prosodic features of speech. My background is in electrical engineering (signal processing).

Tapio Seppänen, Professor. I received MSc degree in electrical engineering in 1985 and the Ph.D. degree in computer engineering in 1990 from the University of Oulu, Finland. Currently I am serving as a full professor at the same university. I teach and conduct research on multimedia signal processing and biomedical signal processing. My special interests include language technology, digital watermarking, pattern recognition applications and content-based multimedia retrieval.

References

1. Toivanen, J., Väyrynen, E., and Seppänen, T. (2004). Automatic discrimination of emotion from spoken Finnish. *Language and Speech*, 47 (4), 383–412.
2. Väyrynen, E. (2005). Automatic emotion recognition from speech. Master's Thesis. Department of Electrical and Information Engineering. University of Oulu.

Research on Service-mediated Emotion Computing and Communication

Changrong Yu[1], Jiehan Zhou[2]

[1]University of Oulu, Faculty of Humanities, English Philology
[2]University of Oulu, Department of Electrical and Information Engineering
changrong.yu@oulu.fi; jiehan.zhou@ee.oulu.fi

Abstract. Human emotions are motivated by desiring a service; instant service provision mediates human emotion behaviors; computer-based computing and communication is enabling instant emotion-aware service provisions by human-system interactions. Those assumptions are highlighted in this paper with our generalized term of service-mediated human emotion experience. An emotion-aware service computing and communication architecture is explored in line with emerging service-oriented computing and communication. A high level emotion elicitation scheme is studied and specified from the aspect of cognitive theories.

Keywords: emotion computing, emotion communication, service-oriented computing

1. Introduction

Humans experience emotional communication everyday and everywhere. Emotions are commonly grouped into negative emotions (e.g., fear, distress, disappointment, shame, being sorry, etc.) and positive emotions (e.g., joy, happiness, liking, satisfaction, etc.). Emotions and emotional intelligence have been studied by biologists [1][3], psychologists [2], anthropologists[4, 5][6], sociologists [4], social constructionists[7], social psychologists, and linguists[18-19].

This paper highlights that emotions result from the appraisal of an event, prospect, object, agent, arouse by desiring a service from the perspective of cognitive theories. The proper and instant service provision can interrupt and regulate human emotions, turning negative emotions into positive emotions. With the rapid advance of computer-based computing and communication technology, e.g., ambient intelligence [8], pervasive computing [9], ubiquitous computing [10, 11], etc., instant provision of emotion-aware services is being enabled in terms of instant multimedia messages, networked multimedia, mobile TV [12], wireless networks, IP multimedia subsystem,[13] sensor networks, etc. Although emotion-aware application scenarios have been frequently reported in well-known projects, e.g., homesickness service via instant communication, etc. in [14], emotion-aware service computing and communication are not the focus in these projects.

This paper is a continuation of our previous research [15, 16], and aims to explore an IT-enabled service mediated emotion architecture, facilitating applications development by leveraging cognitive emotion theory and instant service provision technology. Such an application detects the user's emotions from sensor data and other available context information. It acquires user's motivations or desiring

T. Vanhala and J. Anttonen (Eds.) *Proceedings of EHTI'08: The First Finnish Symposium on Emotions and Human-Technology Interaction.* Tampere, Finland, May 30, 2008. ISBN 978-951-44-7383-8.

services (identifies the intention of the emotion), and reasons a proper service that matches the user's emotions and other context.

The remainder of the paper is organized as follows. Section 2 presents a service-mediated emotion computing and communication architecture. Section 3 studies emotion categories based on emotion elicitation conditions in context of event, agent and object cognition. Section 4 discusses and concludes the paper.

2. Services-mediated emotion architecture

Our research is based on the cognitive emotion theory [17] and perceives human emotion communication by applying the following general sequences (Figure 1): human emotion communication starts by emotion generation and emotion action by the emotion producer. The observer of the emotion action creates an emotional response that in turn affects the emotion producer. Emotion generation produces various types of emotions; positive emotions like happiness, love, and kindness, and negative emotions like anger, fear, and depression. These emotions are motivated by the individual's intentions or desire for services. The emotion action, for example, voice rising, blood pressure change, crying, or smiling, makes the producer's emotions observable. The emotion responses, like making an instant conversation, a soulful kiss, a helpful call, or encouraging words, are triggered by emotion actions. The emotion responses could be enforced by emotion producers themselves inwardly or emotion observers outwardly. Positive responses nourish the emotion producer's positive emotions and help her/him to suppress negative emotions and to avoid foolish actions. In the context of emotion-aware service computing and communication, we use the term of services, instead of emotion responses, which is exactly what the observer or the computer system wants to provide for mediating the producer's emotion.

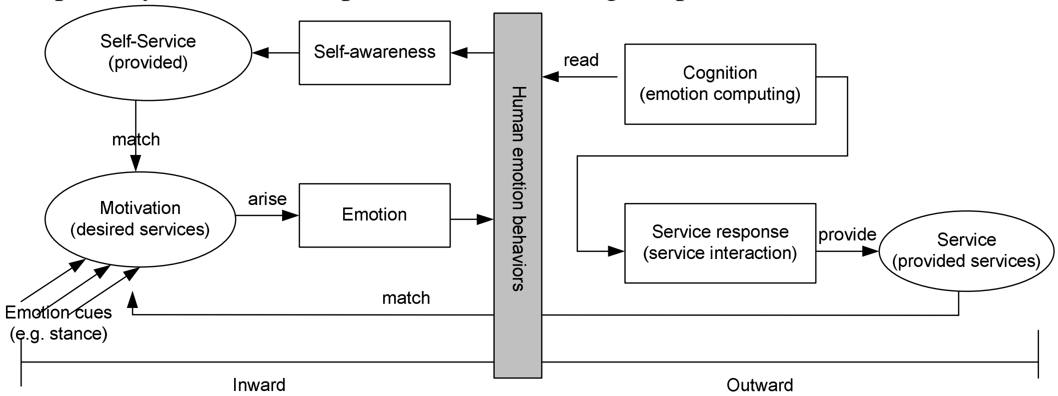


Figure 1. Human emotion computing and communication from a service perspective

To support service-mediated human emotion experience, an emotion-aware service computing and communication architecture is illustrated in Figure 2. This architecture integrates cognitive emotion modeling, emotion-aware service modeling and advanced IT computing and communication (e.g., ambient intelligence, affective computing, emotion-aware services).

Emotion modeling is responsible for emotion detection and emotion motivation acquisition. Emotion detection component detects positive and negative emotions that are represented by emotion actions through speech, face, and body behavior. Motivation acquisition identifies the intention of the emotion or desiring services.

Emotion-aware service modeling is responsible for meeting the identified emotion motivations by creating services, delivering services, and managing the delivered services. The service creation includes emotion-aware service composition and emotion-aware service development. Composition refers to assembling existing services, like in the case of a courage service, it could be created by composing a text message service, music service, and video service. Service development refers to creating new services responding to identified emotion motivation. Service delivery is responsible for supplying appropriate emotion services to the users via Internet or 3G mobile devices.

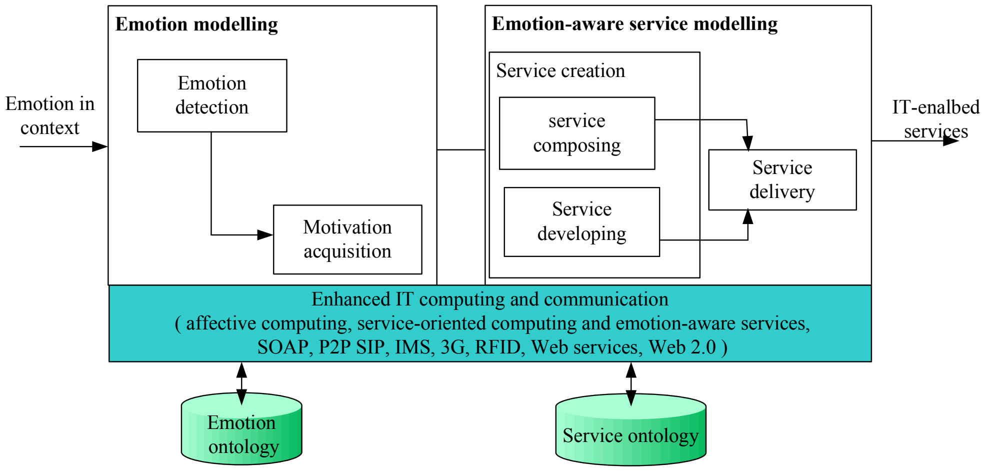


Figure 2. The service-mediated emotion architecture

3. Emotion specification

The cognitive emotion theory [17] rests on the assumption that emotions are valenced affective reactions to the world events, agents, and objects. Table 1 presents a scheme of emotion categories.

Event-based emotions involve reactions of the experiencing person to the implications of events for the person experiencing the emotion. This class is sub-categorized into other-aware and self-aware according to whether the person experiencing the emotions is reacting to the consequences of the focal event with respect only to himself, or with respect to some other person.

Agent-based emotions represent emotions having to do with people's reactions to the agency that they attribute to, namely, attribution emotions. Basically, these emotions are forms of the affective reactions of approving and disapproving of an agent's actions.

Object-based emotions result from reactions to objects, also called attraction emotions, which are all variations of the affective reactions of liking and disliking.

Table 1. Scheme of emotion categories based on [17]

Cognitive emotion categories				Descriptor	elicitation conditions
Event-based (desirable or	self-aware	prospect-relevant	pleased	Hope	(pleased about) the prospect of a desirable event

undesirable)			displeased	Fear	(displeased about) the prospect of an undesirable event
			pleased	satisfaction	(pleased about) the confirmation of the prospect of a desirable event
			displeased	fears-confirmed	(displeased about) the confirmation of the prospect of an undesirable event
			pleased	relief	(pleased about) the disconfirmation of the prospect of an undesirable event
			displeased	disappointment	(displeased about) the disconfirmation of the prospect of a desirable event
		prospect-irrelevant	pleased	contented	(pleased about) a desirable event
			displeased	distressed	(displeased about) an undesirable event of a loss
	other-aware	pleased	desirable	happy-for	(pleased about) an event presumed to be desirable for someone else
			undesirable	gloating	(pleased about) an event undesirable for someone else
		displeased	desirable	jealousy	(displeased about) an event desirable for someone else
			undesirable	compassion	(displeased about) an event presumed to be undesirable for someone else
Agent-based (approving or disapproving)	self-aware		praiseworthy	pride	(approving of) one's own praiseworthy action
			blameworthy	shame	(disapproving of) one's own blameworthy action
	other-aware		praiseworthy	appreciation	(approving of) someone else's praiseworthy action
			blameworthy	reproach	(disapproving of) someone else's blameworthy action
object-based	appealing			Liking	(liking) an appealing object
	unappealing			disliking, hate	(disliking) an unappealing object

4. Conclusion

In this paper, the process of emotion computing and communication is analyzed in a service-centric perspective first and IT-based emotion-aware service computing and communication architecture is designed. A scheme of reasonable emotion categories is studied and their elicitation conditions are specified. Three major tasks are identified in the further work: 1) study on identifying emotion-aware service categories. 2) study the structure of emotion in English conversation. 3) study empirical data occurring in daily emotion communication.

Acknowledgement. Thanks to Dr. Elise Kärkkäinen for her suggestive comments.

Changrong Yu, M.A. Ph.D student in Linguistics of English Philology, Faculty of Humanities, University of Oulu. Her research interests include emotion expression and analysis in English conversation.

Jiehan Zhou, PhD, a Post-doctoral research scientist at MediaTeam, Information Processing laboratory, University of Oulu. He obtained his PhD in manufacturing and automation from the Huazhong University of Science and Technology, Wuhan, China in 2000. His current research interests are related to emotion-aware computing and communication.

References

[1] C. Darwin, The Expression of Emotion in Man and Animals. Indy Publish, 1872,

[2] S. Freud, Beyond the Pleasure Principle. New York: Norton, 1975,

[3] W. James, "Psychological essay: What is an Emotion?" Mind, vol. 9, pp. 188-205, 1884.

[4] M. Lweis, J. M. Haviland and J. M. Haviland, Handbook of Emotion. New York: The Guilford PressLweis, M, 1993,

[5] C. Lutz, Unnatural Emotions: Everyday Sentiments on a Micronesian Atoll and their Challenge to Western Theory. Chicago: University of Chicago Press, 1988,

[6] R. Cornelius, The Science of Emotions. Upper Saddle River, NJ: Prentice-Hall, 1996,

[7] R. Plutchik, H. Kellerman and H. Kellerman, Emotion: Theory, Research and Experience. , vol. 1, New York: Academic PressPlutchik, R, 1980,

[8] Issarny, V., Sacchetti, D., Tartanoglu, F., Sailhan, F., Chibout, R., Levy, N., and Talamona,A., "Developing Ambient Intelligence Systems: A Solution based on Web Services," Autom. Software. Eng, vol. 12, 2005.

[9] R. Grimm, "One.world: Experiences with a Pervasive Computing Architecture," IEEE Pervasive Computing, vol. 3, pp. 22-30, 2004.

[10] J. Zhou and E. Niemela, "State-of-the art on metamodel driven multimedia over mobile ubiquitous computing environments," in Proceedings of the Communication, Internet and Information, 2005, pp. 278-286.

[11] J. Riekki, J. Huhtinen, P. Ala-Siuru, P. Alahuhta, J. Kaartinen and J. Roning, "Genie of the Net, an Agent Platform for Managing Services on Behalf of the User," Computer Communications Journal, Special Issue on Ubiquitous Computing, vol. 26, pp. 1188-1198, 2003.

[12] Europa, "An EU strategy for mobile TV – frequently asked questions. MEMO/07/298," July 2007.

[13] J. Rittwik and J. Serban, "From IPTV to mobile TV to IMS-TV?: Implications and standards for a network operator," in IPTV Workshop International World Wide Web Conference, 2006,

[14] ITEA-AMBIENCE, "Context-Aware Environments for Ambient Services," www.Itea-Office.org/public/project_leaflets/AMBIENCE_results_oct-04.Pdf, 2004.

[15] J. Zhou and P. Kallio, "Ambient emotion intelligence: From business-awareness to emotion-awareness," in The 17th International Conference on Systems Research, 2005, pp. 47-54.

[16] J. Zhou, C. Yu, J. Riekki and E. Kärkkäinen, "AmE Framework: a Model for Emotion-aware Ambient Intelligence, 2007," The Second International Conference on Affective Computing and Intelligent Interaction (ACII2007): Lisbon, Portugal., 2007.

[17] O. Andrew, C. L. Gerald and C. Allan, the Cognitive Structure of Emotions. Cambridge University, 1988

[18] W. Chafe, Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing. Chicago: University of Chicago Press. 1994

[19] N. L. Stein, R. S. Bernas, and &D. Calicchia, Conflict talk: Understanding and resolving arguments. In T. Giron (Ed.) Conversation: Cognitive, communicative and social perspectives.Amsterdam: John Benjamins. 1996

Facial expression recognition with spatiotemporal local descriptors

Guoying Zhao, Matti Pietikäinen

Machine Vision Group, Infotech Oulu and Department of Electrical and Information Engineering, P. O. Box 4500 FI-90014 University of Oulu, Finland
{gyzhao, mkp}@ee.oulu.fi

Abstract. This paper reviews our research on facial expression recognition using spatiotemporal descriptors, combining the appearance and motion together and describing the local transition features. Experiments show very promising results.

Keywords: facial expression recognition, spatiotemporal descriptors, emotion, local binary patterns.

1 Introduction

The human face plays a significant role in verbal and non-verbal communication. Fully automatic and real-time facial expression recognition whose goal is to determine the emotional state of the face, e.g. happiness, sadness, surprise, neutral, anger, fear, and disgust as shown in Fig. 1, regardless of the identity of the face, could find many applications, for example, in human-computer interaction (HCI), biometrics, telecommunications and psychological research. It has been argued that to truly achieve effective human-computer interaction, there is a need for the computer to be able to interact naturally with the user, similar to the way human-human interaction takes place. Therefore, there is a growing need to understand the emotions of the user. The most informative way for machine perception of emotions is through facial expressions in video. Most of the past research on facial expression recognition has been based on static images [1,2]. Psychological studies [3] have shown, however, that facial motion is fundamental for the recognition of facial expressions.

Surprise Sadness Fear Disgust Anger Happiness

Figure 1. Facial expressions.

Some research on using facial dynamics has been carried out; however, reliable segmentation of lips and other moving facial parts in natural environments has turned out to be a major problem.

In our earlier studies [4], we proposed a method for facial expression recognition from static images using a novel face representation based on local binary pattern

T. Vanhala and J. Anttonen (Eds.) *Proceedings of EHTI'08: The First Finnish Symposium on Emotions and Human-Technology Interaction.* Tampere, Finland, May 30, 2008. ISBN 978-951-44-7383-8.

(LBP) features [5,6], obtaining good results. In this approach, the facial expression image is divided into several regions from which the LBP features are extracted and concatenated into an enhanced feature vector to be used as an expression descriptor.

In this paper, we review our recent work on facial expression recognition using an effective spatiotemporal dynamic descriptor and present some experimental results.

2 Facial expressions with spatiotemporal local descriptors

A volume LBP operator (VLBP) [7] was developed which combines temporal and spatial information (i.e. motion and appearance). A simplified method based on concatenated LBP histograms computed from three orthogonal planes (LBP-TOP) [8] was also proposed.

Then, we extended these spatiotemporal descriptors to deal with non-traditional dynamic textures, such as facial expressions, in which local information and its spatial locations should also be taken into account. A representation which consists of dividing the face image into several overlapping blocks was introduced. The LBP-TOP (or VLBP) histograms in each block are computed and concatenated into a single histogram, as shown in Fig. 2. All features extracted from each block volume are connected to represent the appearance and motion of the facial expression sequence [9].

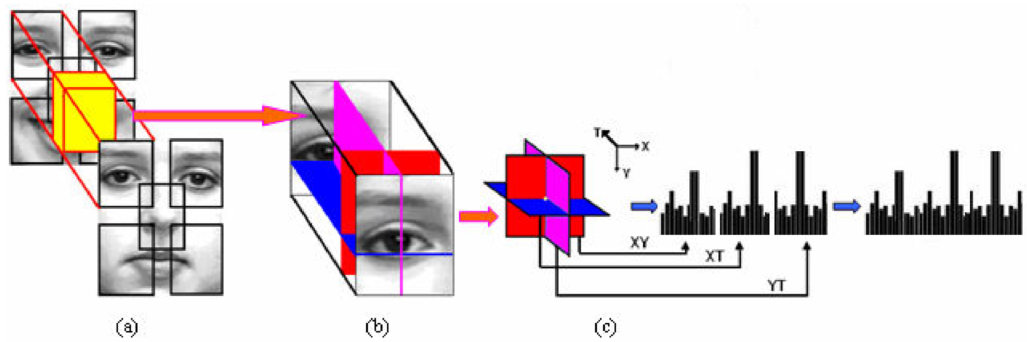


Figure 2. Spatiotemporal LBP based representation for facial expression recognition.

After extracting the expression features, a support vector machine (SVM) classifier was selected since it is well founded in statistical learning theory and has been successfully applied to various object detection tasks in computer vision. Since SVM is only used for separating two sets of points, the six-expression classification problem is decomposed into 15 two-class problems (happiness-surprise, anger-fear, sadness-disgust, etc.), then a voting scheme is used to accomplish recognition. Sometimes more than one class gets the highest number of votes. In this case, 1-NN template matching is applied to these classes to reach the final result.

We experimented with the Cohn-Kanade database [10] which consists of 100 university students with age ranging from 18 to 30 years. Sixty-five percent were female, 15 percent African-American, and three percent Asian or Latino. Subjects were instructed by an experimenter to perform a series of 23 facial displays that included single action units and combinations of action units, six of which were based on descriptions of prototypic emotions, anger, disgust, fear, joy, sadness, and surprise. For our study, 374 sequences from the dataset were selected from the database for basic emotional expression recognition. The selection criterion was that a sequence to be labeled is one of the six basic emotions. The sequences came from 97 subjects, with one to six emotions per subject. The positions of the two eyes in the first frame of each sequence were given manually and then these positions were used to determine the facial area for the whole sequence. The whole sequence was used to extract the proposed spatiotemporal LBP features.

A recognition result of 96.26% in ten-fold cross-validation was achieved with the combination of region-based LBP-TOP and VLBP descriptors, which is better than those obtained in earlier studies, even though we used in our experiments a simpler image preprocessing and a larger number of people (97) and sequences (374) than most of the others have used. Our approach was shown to be robust with respect to changes in illumination and errors in face alignment, and it does not require error prone segmentation of facial features such as lips.

To evaluate the performance in real-world environments, a simple web camera is used to capture videos with a resolution of 320 by 240 pixels. Subjects were about 0.5-1 meters in front of the camera. The face was detected automatically using boosted Haar-like features and the size of the cropped faces was around 40 by 50. Fig. 3 demonstrates results taken from an example image sequence, showing the detected faces and classified expressions for each case.

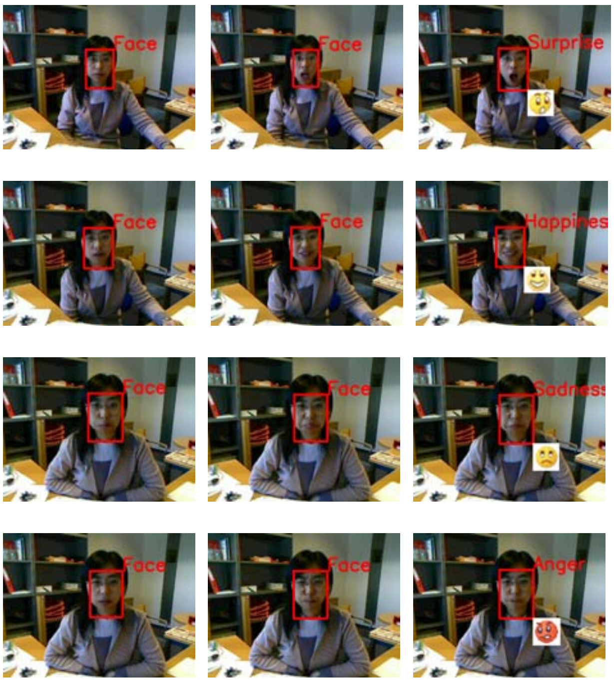


Fig. 3. Facial expression recognition from a web camera. From left to right are the start, middle and the end of the sequences.

More recently, spatiotemporal local binary patterns computed at multiple resolutions were proposed for describing dynamic events, combining static and dynamic information from different spatiotemporal resolutions [11]. Appearance and motion are the key components for visual analysis related to movements. AdaBoost algorithm is utilized for learning the principal appearance and motion from spatiotemporal descriptors derived from three orthogonal planes, providing important information about the locations and types of features for further analysis. In addition, learners are designed for selecting the most important features for each specific pair of different expression classes. The experiments carried out on facial expression recognition show the effectiveness of the approach.

We have also extended the LBP-TOP based approach for visual recognition of spoken phrases using mouth movements, obtaining very promising results [12].

3 Future work

In the future, we plan to investigate the use of near infrared imaging (NIR) techniques with an LBP-TOP based approach for facial expression recognition. Using NIR imaging with an LBP-like operator, the difficult problems caused by varying illumination can be handled well. Recognition of facial expressions even under very low lighting should be possible. Moreover, we plan to combine multiple modalities, i.e. facial expression and speech, to analyze the emotions of people and to make human-computer-interaction affective.

Guoying Zhao, Dr. Tech. I received the MS degree in computer science from North China University of Technology, in 2002 and the PhD degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, in 2005. Since July 2005, I have been a postdoctoral researcher in the Machine Vision Group at the University of Oulu, Finland. My research interests include dynamic texture recognition, facial expression recognition, human motion analysis, and person identification. I have authored more than 30 papers in journals and conferences.

Matti Pietikäinen, Dr. Tech. I received the doctor of science in technology degree from the University of Oulu, Finland, in 1982. In 1981, I established the Machine Vision Group at the University of Oulu (http://www.ee.oulu.fi/mvg). This group has achieved a highly respected position in its field, and its research results have been widely exploited in industry. Currently, I am a professor of information engineering, scientific director of Infotech Oulu Research Center, and leader of the Machine Vision Group at the University of Oulu. My research interests include texture-based computer vision, face analysis and their applications in human-computer interaction, person identification, and visual surveillance. I have authored over 200 papers in international journals, books and conferences.

References

1. Pantic M. and Rothkrantz L.L.M. (2000). Automatic analysis of facial expressions: The state of the art. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no.12, 1424-1455.
2. Fasel B. and Luettin J. (2003). Automatic facial expression analysis: a survey. *Pattern Recognition*, vol.36, 259-275.

3. Bassili J. (1979). Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology*, vol. 37, 2049-2059.
4. Feng X., Pietikäinen M. and Hadid A. (2005). Facial expression recognition with local binary patterns and linear programming. *Pattern Recognition and Image Analysis*, 15(2), 546-548.
5. Ojala T., Pietikäinen M. and Mäenpää T. (2002). Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 971 - 987.
6. Ahonen T., Hadid A. and Pietikäinen M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12), 2037-2041.
7. Zhao G. and Pietikäinen M. (2006). Dynamic texture recognition using volume local binary patterns. *Proc. ECCV 2006 Workshop on Dynamical Vision, In conjunction with ECCV*, Graz, Austria, LNCS 4358: 165-177.
8. Zhao G. and Pietikäinen M. (2006). Local binary pattern descriptors for dynamic texture recognition. *Proc. Int'l Conf. Pattern Recognition*, vol. 2, 211-214.
9. Zhao G. and Pietikäinen M. (2007). Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6), 915-928.
10. Kanade T., Cohn J.F., and Tian Y. (2000). Comprehensive database for facial expression analysis. *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, 46-53.
11. Zhao G. and Pietikäinen M. (2008). Principal appearance and motion from boosted spatiotemporal descriptors. Submitted.
12. Zhao .G, Pietikäinen M. and Hadid A. (2007). Local spatiotemporal descriptors for visual recognition of spoken phrases. *Proc. 2nd International Workshop on Human-Centered Multimedia (HCM2007)*, Augsburg, Germany, 57-65.

Index

Author Index