Proceedings of the 6th Workshop

Emotion and Computing –

Current Research and Future Impact

Dirk Reichardt (Editor)

Saarbrücken, Germany

September 24th, 2012

ISSN 1865-6374

6th Workshop

Emotion & Computing
Current Research and Future Impact

D. Reichardt

Computer Science Department
Baden-Württemberg Cooperative State University Stuttgart (DHBW)
Stuttgart, Germany
reichardt@dhbw-stuttgart.de

The workshop series “emotion and computing - current research and future impact” has been providing a platform for discussion of emotion related topics of computer science and AI since 2006.

This workshop intends to discuss the scientific methods considering their benefit for current and future applications. Especially when regarding the subject of emotion recognition, this also includes ethical aspects. We are looking back to 5 workshops with interesting talks and very interesting discussions on varying subjects.

This year we have 4 presentations of ongoing research work and – as a tradition in this workshop series – a moderated discussion on a selected subject in the field of emotion and computing.

We are looking forward to interesting presentations and fruitful discussions.

Dirk Reichardt

Organization and Scientific Committee

Prof. Dr. Dirk Reichardt, Baden-Wuerttemberg Cooperative State University Stuttgart

Dr. Christian Becker-Asano, Freiburg Institute for Advanced Studies

Dr. Patrick Gebhard, DFKI Saarbruecken

Prof. Dr. Nicola Henze, University of Hannover

Prof. Dr. Michael Kipp, Hochschule Augsburg

Prof. Dr. Paul Levi, University of Stuttgart

Prof. Dr. John-Jules Charles Meyer, University of Utrecht

Dipl.-Ing. Christian Peter, Fraunhofer-Institut fuer Graphische Datenverarbeitung, Rostock

Dr. Goetz Renner, Daimler AG, Customer Research Center

Prof. Dr. Michael M. Richter, University of Calgary

Dr.-Ing. Bjoern Schuller, TU Muenchen

Prof. Dr. David Suendermann, DHBW Stuttgart

Scientific Papers

Towards Robust Spontaneous Speech Recognition with Emotional Speech Adapted Acoustic Models

Bogdan Vlasenko, Dmytro Prylipko, and Andreas Wendemuth

Cognitive Systems, IESK & Center for Behavioral Brain Sciences,
Otto von Guericke University, D-39016 Magdeburg, Germany
bogdan.vlasenko@ovgu.de

Abstract. Speech signal in addition to the linguistic information contains additional information about the speaker: age, gender, social status, accent (foreign accent, dialects, etc.), emotional state, health etc. Some of these informational channels induce changes of the speech acoustic characteristics. This article presents evaluation of the ASR acoustic models (first trained on neutral, read speech) on acted and spontaneous emotional speech. In our research we used adaptation approaches to compensate the mismatch of acoustic characteristics between neutral speech samples and affective speech material. During experiments we observed that the affective-speech-adapted ASR acoustic models provide better emotional-speech-recognition performance. The improvements of affective speech recognition performance were 6.24% absolute (7.1% relative) for speaker-independent evaluations on the EMO-DB database and 7.08% absolute (25.43% relative) for cross-corpora evaluation on the VAM database.

Keywords: Emotional speech, adaptation, ASR

1 Introduction

The speech signal comprises not only linguistic content but also various additional information about the speaker: *age, gender, social status, accent, emotional state, health* etc. Characterization of the influence of some of these speech signal variations, together with related methods to improve automatic speech recognition (ASR) performance, is an important research field [2]. In order to deal with spontaneous speech we should not cut the above mentioned information channels from the input signal, but use them as an additional knowledge source and thus boost the performance.

Most of the ASR evaluations use an assumption that training and test datasets have similar acoustic characteristics (speaking rate, acoustic environment, vocal tracts variability, emotional speech, etc.). However, in real-life applications, this is usually not the case. The acoustic characteristics mismatch may significantly decrease the recognition performance compared to the ASR systems

build on data with matched acoustic characteristics. To compensate the mismatch of acoustic characteristics between test and training datasets, adaptation techniques are usually applied.

In our previous research [12] we characterized acoustical difference between emotional and neutral speech. We have shown a significant difference between vowel triangles form and their position in F1-/F2-dimensional space for emotionally colored and neutral speech samples. This difference illustrates why ASR models trained on neutral speech are not able to provide a reliable performance for affective speech recognition.

Acoustic models' adaptation techniques applied for ASR models are usually employed to compensate the mismatch of acoustic characteristics between various speaker, acoustic channels and noisy environments. Acoustic models' adaptation towards affective speech is a less popular adaptation concept.

In our research, we used adaptation approaches to compensate the mismatch of acoustic characteristics between neutral speech samples and affective speech material. We used acted affective speech samples from the popular public available database EMO-DB [4] to adapt acoustic models trained on emotionally neutral speech samples from The Kiel Corpus of Read Speech [8]. Adaptation on emotional speech yield a significant gain in emotional speech recognition performance over the basic ASR models trained on neutral speech samples from the Kiel database. Leave-one-speaker-out evaluations on the EMO-DB database showed 6.24% absolute (7.1% relative) accuracy improvement and cross-corpora evaluations on the VAM database which contains spontaneous emotions showed 7.08% absolute (25.43% relative) improvement.

2 Corpora

For initial acoustic models training we used a part of The Kiel Corpus of Read Speech [8] which contains emotionally neutral German read speech samples. For our evaluation we used speech samples from 1041 utterances produced by 6 female and 1033 utterances spoken by 6 male speakers.

For affective speech we decided to use the popular studio recorded Berlin Emotional Speech Database (EMO-DB) [4] and The Vera am Mittag (VAM) corpus [6]. The EMO-DB contains acted emotional speech samples. 10 professional actors (5 male and 5 female) spoke 10 German sentences with emotionally neutral linguistic meaning (e.g. 'The rag is on the refrigerator'). We used EMO-DB samples to derive optimal adaptation parameters (hyper-parameter τ for MAP adaptation and number of regression class trees rc for MLLR adaptation). Throughout perception tests by 20 subjects, 494 phrases have been chosen that were classified as more than 60% natural and at least 80% clearly assignable. We used these preselected utterances for our evaluations.

The VAM database [6] consists of 12 hours of audio-visual recordings taken from a German TV talk show. The corpus contains 947 utterances with spontaneous emotions from 47 guests of the talk show which were recorded from unscripted, authentic discussions.

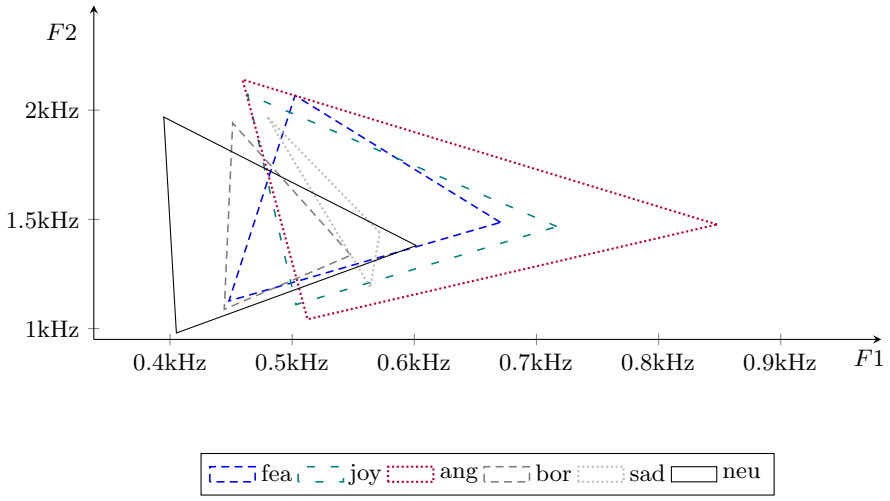

Fig. 1. Classical vowel triangle forms for male speaker's and various emotional states. The triangles are built on the mean positions of *i*, *a* and *o* vowels in the F1/F2 space. Estimated on speech material from the EMO-DB database

3 German phonetic pattern

We have used a simplified version of BAS SAM-PA [1] with a set of 39 phonemes (18 vowels and 21 consonants). Since VAM corpus does not provide such a lexicon, we created it by ourselves using two ways. The major part of the word transcriptions (1216 words) has been taken from other German corpora, namely Verbmobil [7] and SmartKom [11]. For the rest (688 words) we created transcriptions using a grapheme-to-phoneme conversion with *Sequitur* G2P converter [3]. The converter was trained on a joined lexicon based on SmartKom and Verbmobil lexicons (12460 German words at all). Prior to applying the G2P software to the missing VAM lexicon, we tested it on the constructed united lexicon, where 1% randomly selected words were moved into the test set. The phoneme error rate was 5.3% (56 from 1050), the word error rate was 29% (37 from 127).

In 1952, Paterson and Barney [10] created a classic plot of measured values of the first and second formant for 10 English vowels spoken by a wide range of male and female speakers. They pointed out that it is common to represent each vowel by a centroid in the formant space. To show the difference between acoustic characteristics of emotional speech samples we estimated these vowel triangles for selected EMO-DB's utterances. Instead of representing each vowel by a centroid, we represent each vowel by the mean value over utterances. As one can see from Fig. 1 the vowel triangles form and position are different for various emotional states of a speaker. This variability is one of the reasons why ASR models trained on neutral speech are not able to provide a reliable performance in affective speech recognition.

More detailed German phonetic pattern specification for male and female speakers can be found in [12].

4 Implemented adaptation techniques

One of the most popular adaptation technique applied within ASR systems are model-based transform: *Maximum Likelihood Linear Regression (MLLR)* and *Maximum a Posteriori (MAP)* [13].

The Maximum a Posteriori (MAP) approach (sometimes referred as the Bayesian adaptation) maximizes the *a posteriori* probability using a prior HMM parameter distribution.

The Maximum Likelihood Linear Regression (MLLR) is a widely used linear transformation method employed for speaker adaptation. It uses the Maximum Likelihood (ML) criterion to estimate a linear transformation which may be applied to adapt Gaussian parameters of HMMs.

5 Emotional adaptation of acoustic models

In this section we describe acoustic models specification, selection process to derive optimal adaptation parameters and provide experimental results for speaker independent (Leave-One-Speaker-Out, LOSO) evaluation on the EMO-DB corpus and cross-corpora evaluation on the VAM database.

5.1 Acoustic models training

The most robust and general acoustic technique in automatic speech recognition are hidden Markov models (HMM). For our evaluations we used the HTK toolkit [14] to create and test our German acoustic models. We applied continuous density HMM technique based on a multivariate Gaussian mixture model (GMM) with 32 mixture components.

We used *left-to-right* monophone models with three emitting states for acoustic modeling. Speech input is processed using a 25 ms Hamming window, with a frame rate of 10 ms. We employed 39-dimensional MFCC feature vectors (12 cepstral coefficients + log frame energy plus speed and acceleration coefficients).

5.2 Adaptation configuration

Two adaptation schemes have been tested: a basic MLLR and MAP. During the adaptation only the mean values of Gaussian mixture were updated because variance compensation provides only minor imrovement and requires additional computational overhead of non-diagonal Gaussian likelihood calculations [5]. Prior to adaptation and recognition on VAM, optimal parameters for each scheme should be determined. For MLLR a number of regression classes is important. MAP depends on the τ parameter (weight of the prior knowledge). The preliminary adaptation is performed on EMO-DB without having the actual data

Table 1. Optimal adaptation parameters selection. Basic models trained on Kiel, adapted and evaluated with LOSO on EMO-DB

Acoustic model	**Parameters**	**Word accuracy** [%]
Non-adapted basic		88.06
MLLR-adapted basic	rc=32	93.67
MAP-adapted basic	τ=2	94.30
EMO-DB trained		96.70

Table 2. Word-accuracy rates for each speaker presented in the EMO-DB database, received within LOSO evaluation of non-, MAP-, MLLR-adapted Kiel-trained acoustic models, original EMO-DB acoustic ASR models

Speaker ID	basic Kiel [%] non-adapted	MAP	MLLR	EMO-DB [%]
03	93.78	96.17	96.65	98.09
08	92.02	97.59	95.55	98.14
09	81.15	92.67	89.79	97.12
10	79.03	85.48	85.48	84.84
11	94.65	96.63	96.24	97.43
12	90.12	96.84	95.65	97.23
13	90.57	96.74	95.71	98.11
14	90.00	94.77	94.62	98.62
15	92.50	95.46	95.46	97.44
16	75.74	89.11	89.42	95.49

from the adaptation and validation corpus (VAM). Thus, the parameters are also determined with testing on EMO-DB in LOSO way. Acoustic ASR models presented in this section are evaluated with a bigram language model and a grammar scale factor $s = 10$.

As one can see from Table 1, HMM/GMM models trained on neutral speech samples from the Kiel dataset are not able to provide acceptable emotional-speech-recognition performance without adaptation on affective speech samples. For this configuration (72 words in lexicon, only 10 possible sentences) state-of-the-art recognition accuracy is higher than 95% [9]. However, MLLR-adapted basic models provide better recognition performance, which is close to the value achieved during the evaluation of EMO-DB-trained (*native*) acoustic models.

It can be seen from Table 2 that the acuuracies on German affective speech recognition for each speaker are approximately equal. Only for speaker 10 we obtained a comparably low affective-speech-recognition accuracy rate. Such a low performance can be explained by very specific vocal tract characteristics of the speaker 10.

For the MLLR, regression trees with $2, 4, 8, 16$ and 32 terminal nodes have been tested. Prior knowledge weight for MAP has been evaluated in range of $\tau = 2, ..., 20$. For MLLR adaptation the best emotion recognition performance

Table 3. Word-accuracy rates for cross-corpora emotion non-adapted acoustic ASR models. Trained and evaluated on the Kiel and VAM datasets correspondingly

Training set	Evaluation set	Word accuracy [%]
Kiel	VAM	27.84
VAM	VAM	42.75

Table 4. Word-accuracy rates for Kiel trained acoustic ASR models without and with pre-adaptation on EMO-DB samples, evaluated on the VAM database

Adaptation scheme	Word accuracy [%]
non-adapted	27.84
MLLR	34.92
MAP	33.54

on EMO-DB samples has beed obtained with 32 regression class trees ($rc = 32$). These configurations have been used further for adaptation and test on the VAM corpus. For MAP adaptation the best emotion recognition performance has been obtained with $\tau = 2$.

5.3 Experiments and results

Prior to adaptation we tested the baseline performance of the acoustic models trained on the Kiel corpus.

Except acoustic models, other components such as lexicon or language models have been taken from the test database. Training and testing on the VAM (single corpora mode) has been done in a speaker independent fashion using Leave-One-Speaker-Group-Out (LOSGO with 5 speaker groups at all) strategy.

The results presented in Table 4 show that training ASR models on neutral speech, and subsequent adaptation on affective speech samples, does have an impact on the recognition performance within emotional speech recognition. These results have been obtained after evaluations in a cross-corpora way. We used speech samples from the Kiel and Emo-DB databases for training and adaptation of acoustic ASR models. Finally, these acoustic models have been evaluated on the VAM database speech samples.

Also, we compared initial acoustic models with pre-adapted ones with unsupervised incremental adaptation (MLLR with HVite). The latter are acoustic models trained on the Kiel dataset and adapted with MLLR ($rc = 32$) on EMO-DB samples.

30 adaptation sentences were selected randomly from the whole VAM corpus. During the adaptation process they were fed to HVite sequentially. Other 917 sentences formed the test set. This difference in procedure is the reason why initial values of both curves depicted in Fig. 2 slightly differ from the values provided in Table 4. The transformations were applied after some number of frame occurrences (namely 800) which in our case corresponds to 6 sentences or 6.52 seconds of speech. One can see from Fig. 2, if we do not have at least

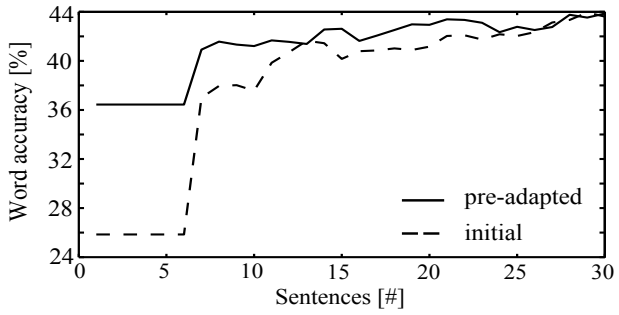

Fig. 2. Performance evolution during the incremental unsupervised adaptation on VAM database. Initial models trained on only Kiel samples, pre-adapted - trained on Kiel and adapted on EMO-DB samples.

25 sentences for unsupervised adaptation, pre-adapted acoustic models provide much better speech recognition performance.

6 Discussion and conclusions

The main issue of this research is to show that training ASR models on neutral speech, and subsequent adaptation on affective speech samples, does have an impact on the recognition performance within emotional speech recognition. It has been found that the adaptation on acted emotional speech samples favor a significant gain (about 25.43% relative improvement for word-accuracy rate) in spontaneous emotion speech recognition performance (34.92% with adapted models) over the basic ASR models trained on neutral speech samples from the Kiel database. In comparison with results presented for the EMO-DB database, speech recognition performance for the VAM database obtained with adapted models is relatively low. This result can be compared with a low word-accuracy rate of 42.75% obtained during speaker-independent LOSGO evaluation on the VAM database.

Comparison of these values to the state-of-the-art speech recognition performance is unfortunately hardly possible, due to the nature of the corpora. Both EMO-DB and VAM were designed for research in emotion recognition from speech, rather than for speech recognition. As it has been mentioned before, VAM is not even supplied with a lexicon. That is why most publications report on accuracies in emotion classication. For our best knowledge, there is no paper reporting on the accuracies of speech recognition on EMO-DB or VAM.

By using more accurate lexica (more expensive solution, generated by phonetic expert) and modeling non-linguistic cues presented in VAM speech samples we should be able to improve accuracy of speech recognition performances for both evaluation cases (*LOSGO* and *cross-corpora*). With more accurate lexica we can also switch to triphone models, which usually provide better speech recognition performances.

As a conclusion, we showed that acoustic models trained on read speech samples and adapted to acted emotional speech could provide better performance of spontaneous emotional speech recognition. We come to the conclusion, that implementing adaptation on affective speech samples can be an important issue towards robust spontaneous speech recognition.

References

1. Bavarian Archive for Speech Signals. Extended SAM-PA (PhonDat-Verbmobil). http://www.bas.uni-muenchen.de/Bas/BasSAMPA, 1996. Last accessed 08.08.2012.
2. M. Benzeghiba, R. Demori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, P. Laface, A. Mertins, and C. Ris. Automatic speech recognition and speech variability: A review. *Speech Communication*, 49(10-11):763–786, October 2007.
3. M. Bisani and H. Ney. Joint-Sequence Models for Grapheme-to-Phoneme Conversion. *Speech Communication*, 50(5):434–451, May 2008.
4. F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, and B. Weiss. A database of German emotional speech. In *Proceedings of European Conference on Speech Communication and Technology, EUROSPEECH*, pages 1517–1520, 2005.
5. M. Gales, D. Pye, and P. Woodland. Variance compensation within the MLLR framework for robust speech recognition and speaker adaptation. In *Proceedings of ICSLP*, pages 1832–1835. IEEE, 1996.
6. M. Grimm, K. Kroschel, and S. Narayanan. The Vera am Mittag German audio-visual emotional speech database. In *Proceedings of the IEEE International Conference on Multimedia and Expo, ICME*, pages 865–868, 2008.
7. W. Hess, K. Kohler, and H.-G. Tillman. The Phondat-Verbmobil speech corpus. In *Proc. of the Eurospeech 1995*, pages 863–866, Madrid, Spain, 1995.
8. K. J. Kohler. Labelled data bank of spoken standard German - the Kiel Corpus of read and spontaneous speech. In *Proceedings of International Conference on Spoken Language Processing*, pages 1938–1941, 1996.
9. D. Pallett. A Look at NIST's Benchmark ASR tests: Past, Present, and Future, 2003.
10. G. E. Peterson. Control Methods Used in a Study of the Vowels. *The Journal of the Acoustical Society of America*, 23(1):148, 1951.
11. F. Schiel, S. Steininger, and U. Turk. The Smartkom multimodal corpus at BAS. In *Proc. of the Language Resources and Evaluation (LREC)*, 2002.
12. B. Vlasenko, D. Prylipko, D. Philippou-Hübner, and A. Wendemuth. Vowels formants analysis allows straightforward detection of high arousal acted and spontaneous emotions. In *Proceedings of the Interspeech 2011*, Florence, Italy.
13. P. Woodland. Speaker adaptation for continuous density HMMs: A review. In *ISCA Tutorial and Research Workshop (ITRW) on Adaptation Methods for Speech Recognition*, pages 11–19, Antipolis, France, 2001.
14. S. Young, G. Evermann, M. Gales, H. T., D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland. *The HTK Book (for HTK Version 3.4)*. Cambridge University Engineering Department, 2009.

Emotion Sensitive Active Surfaces

Larissa Müller[1], Arne Bernin[1,4], Svenja Keune[2], and Florian Vogt[1,3]

[1] Department Informatik, University of Applied Sciences (HAW) Hamburg, Germany
[2] Department Design, University of Applied Sciences Hamburg (HAW) , Germany
[3] Innovations Kontakt Stelle (IKS) Hamburg , TuTech Innovation GmbH, Germany
[4] School of Computing, University of the West of Scotland (UWS), UK

Abstract. In this work we introduce a new user interface design that maps sensed emotions to an abstract device. This abstract display allows the exploring of emotional expressions avoiding anthropomorphisms caused by human or animal like designs. The presented interface senses emotions by facial video analysis. The device is implemented as an actuated surface that expresses emotions by changing its physical attributes. This change is triggered by sensed emotions via a behavioral mapping. The surface design, including texture, form of expression and dynamics, is a result of an iterative design process involving interdisciplinary participants. Finally, results of informal evaluations of the device during audience exhibits are presented.

1 Introduction

Computer interfaces have evolved over many years into various shapes and forms. The main focus of interface development has been on rational control paradigms using mental models such as *tools* or *request response behaviors*. In contrast, humans communicate on an emotional as well as a rational level. Emotional interactions with computers have not yet become widespread in computer interfaces, although research is developing this field in areas such as computer linguistics, affective computing, and human robotics. Building computer interfaces based on human to human communication is challenging and complex, since humans simultaneously process on both levels and integrate different modalities. Our approach is to disambiguate these two levels of processing with a *divide and conquer* strategy by building an interface that suppresses the rational level and emphasizes the emotional level.

For this purpose we chose an abstract surface design, since associations of anthropomorphism are less pronounced and there are no direct links to humans or animals. Otherwise, anthropomorphism could lead to misunderstandings and false expectations. The research by Gaver et al. [6] has found cases of frustration when user expectations are not met by the abilities of computer interfaces; this is also supported by the work of Höök [7]. In our user interface design we exclude any rational controls such as buttons or position sensors and deploy emotion sensors based on facial expression.

A purely emotional interface allows us to evaluate interaction patterns such as the *emotion mirror*, an interface which mirrors the emotion displayed by the user. *Provoking/reacting* is another pattern where an expressed emotion is mapped to a different displayed emotion thereby creating a character behavior. The basis of this work is to investigate if an interface has the ability to: 1. express emotions, 2. engage in an emotional dialog, and 3. create an emotional binding with users.

1.1 Related Work

Physical user interfaces of computers and machines have greatly evolved since the introduction of personal computers and have become more diverse in form and function. In particular, robotic

interfaces have taken a lead role to include emotional qualities in interfaces. These are categorized by two properties: The similarity to living beings including human or animal and The type of interface both physical and virtual. The following table shows the categorization of related work:

Interface model	**Physical**	**Virtual**
Human	Geminoid HI-1 [11]	Max [4], facial animations [9, 3]
Animal	Haptic creature [13], Kismet [5]	Nintendodogs [1]

Table 1: Examples of different types of emotion based interfaces

While the works above translate the behavior and character of living beings, our active surface exhibits abstract behavior, which makes it free of preconceptions.

2 Concept

The design process in our work is based on Mitch Resnick's ideas of a *Lifelong Kindergarten* [12]. Our fast prototyping is achieved in an iterative process using the steps of *imagine*, *create*, *play*, *share* and *reflect*. This fast prototyping offers the possibility of creative idea development in teams. Our short design and implementation cycle also includes informal feedback given by observers and focus groups. This enables us to perform integrated testing of the concepts of the emotional interface and avoids that observers misinterpret or do not perceive the displayed emotions.

We applied this design method with an interdisciplinary team consisting of computer scientists and designers. Multiple design evaluations with reflection and feedback from visitors of exhibitions have led to a number of designs. The final results are presented in this work.

2.1 Physical Design Concept

The active surface consists of a sense and a display module. The display module can change surface properties to provoke reactions from visitors. The behavioral patterns are triggered based on the presence and facial expressions shown by visitors.

The surface design consists of repeating basic elements, constituting colonies or swarm configuration. This concept was inspired by living organisms and natural elements, each having limited complexity but together having the ability to exhibit complex and engaging behavior.

Our final design resulted in an active surface which is made of fabric covered shell elements, as shown in Fig. 1. A group of these base elements are connected to servo motors to display movement. These servo motors are controlled by a microcontroller, connected to the displaying module.

The active surface works with four degrees of freedom to express emotions: speed of movements, intensity of movements, size, and position of the moving area. Since the different areas of the surface do not symbolize special meanings such as face or body parts, we discarded the position of the moving parts as a degree of freedom; instead we combined the size and position of movement, reducing to three degrees of freedom.

Fig. 1: Lomelia interface: A group of shells is connected to one of ten servo motors. To optimize detection of facial expressions, a frontal image is required, therefor a camera is mounted next to the pink flower acting as an eye catcher.

2.2 Software Architecture

The software design consists of two modules, loosely coupled through a message broker, implementing a publish-subscribe pattern. One module performs the emotion sensing, while the second module controls the active surface. Loose coupling was chosen for maximum flexibility in extending the setup with additional input and output. In order to sense emotions, a camera is mounted on top of the surface, taking pictures of visitors' faces in front of it.

These pictures are algorithmically searched for facial expressions and are processed by the SHORE library of the Fraunhofer Institute for Integrated Circuits (IIS) [8] to produce emotion estimates. If an emotion is sensed, it is sent to the message broker. The control module receives the published message and reacts according to a predefined mapping of emotions (see Fig. 2). It creates different surface movements.

2.3 Emotion mapping

We apply a state-based emotional model consisting of the emotions *angry*, *sad*, *surprised* and *happy*. These emotions are determined by the SHORE library. The movements of the shells display the

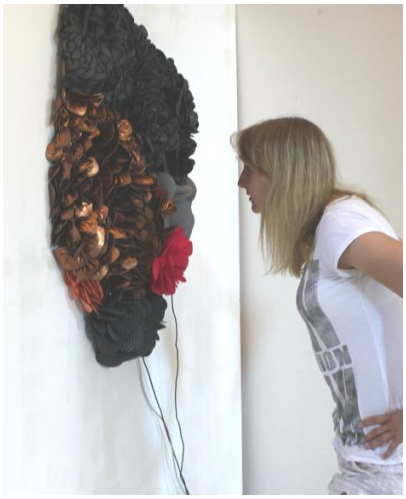

Fig. 2: Communication with our active surface interface

following emotions: *happy*, *angry*, *relaxed*, *excited*. The current mapping of emotions is shown in Table 2.

A salient part of the surface is the red flower in the center. This surface part provides an engaging focus by displaying happiness.

SHORE	**Lomelia**	**Expression**
happy	happy	rhythmic dance movement
sad	angry	jitter movement
angry	relaxed	breathing movement
suprised	excited	erratic movement

Table 2: Mapping of emotions: - recognized emotions of the interface (SHORE). These emotions are mapped to states of the active surface (Lomelia), which triggers particular movements (Expression).

2.4 Informal Evaluation

Informal evaluations took place during the DMY Berlin 2012 (see Fig. 3, [2]), an international design festival, with more than 2000 exhibition visitors , as well as during the Open House 2012 of the HAW Hamburg with more then 1000 exhibition guests. We observed the visitor interaction with our surface prototypes at the exhibits. In interviews the observers answered questions about their expectations and experiences regarding the emotional dialogue and engagements.

Recognizing the emotional state *sad* leads the surface interface to display *angry* which may trigger a feeling of discomfort in some observers. The visitors do not normally associate intention.

Fig. 3: Visitors at the DMY 2012 interacting with three different versions of the active surface

The fast movements were in most cases distinguished as *happiness* instead of *angry*. Recognition Rates of the SHORE Library varied for the different emotions, *happy* was recognized with the highest rate. *Surprised* and *angry* had a lower recognition rate, followed by *sad*.

A smiling person in front of the surface interface triggers a happy response. In most cases, observers recognized the interface as happy. Some of them compared the interface with a dog wagging its tail. Others made associations with coming home and finding a friend to share their happiness. Looking angry into the camera provoked a relaxed response from the interface. Slow movements were associated with a relaxed state. A surprised facial expression triggers a very fast movement. The aim was to show excited emotions. In most cases different movements while smiling and looking angry led to a lot of excitement. People laughed and had a lot of fun to trigger the emotions of the interface. They spend significant time playing with the interface, Lomelia. Some visitors were shocked by the movement. They did not expect the surface to react when they moved closer. Some screamed in surprise.

The acceptance of our idea to establish an emotional dialogue between an observer and an interactive textile surface interface was very high amongst exhibition visitors. The enthusiasm of the audiance was supported by the receipt of the audience award at the DMY Berlin 2012.

3 Summary

With this work we present our design of an emotion sensitive active device. We developed a physical abstract surface interface with the ability to sense the emotional states of observers and react by expressing emotions. We focused exclusively on the emotional aspects of the interaction, leaving rational input aside.

Our preliminary findings are based on observation and informal interviews from two exhibitions. Visitors were able to recognize different displayed emotions of our abstract surface. Furthermore many visitors engaged in an emotional dialogue by playing with Lomelia several times in `provoke response cycles`. In several instances we observed personal bindings, since visitors returned multiple times to interact with the active surface and shared their experience with friends and family. While the sensing module could be improved to include a wider range of emotion enabling an en-

riched dialogue, we think emotional interfaces have further far reaching computer applications in entertainment and computer interfaces.

4 Acknowledgement

We thank Kai von Luck, Renata Brink, Gunter Klemke, Birgit Wendholt, Franziska Hübler and the members of the EmotionLab at HAW Hamburg for their feedback and their support. In addition, we thank the Fraunhofer Institute of Integrated Circuits IIS for providing access to the SHORE library.

References

[1] Learning to care for a real pet whilst interacting with a virtual one? The educational value of games like Nintendogs (2008)
[2] Dmy - international platform for architecture interior and product design. http://dmy-berlin.com/ (May 2012)
[3] Albrecht, I., Haber, J., Seidel, H.P.: Automatic Generation of Non-Verbal Facial Expressions from Speech. pp. 283–293 (2002)
[4] Becker, C., Prendinger, H., Ishizuka, M., Wachsmuth, I.: Evaluating affective feedback of the 3d agent max in a competitive cards game. In: International Conference of Affective Computing and Intelligent Interaction. pp. 466–473. Springer-Verlag (2005)
[5] Breazeal, C.L.: Sociable machines: expressive social exchange between humans and robots. Ph.D. thesis (2000), aAI0801833
[6] Gaver, W., Dunne, A., Hooker, B., Kitchen, S., Walker, B.: The Presence Project. Interaction Design Research Department, Royal College of Art, London (2001)
[7] Höök, K.: User-centred design and evaluation of affective interfaces. In: Ruttkay, Z., Pelachaud, C. (eds.) From brows to trust, chap. User-centred design and evaluation of affective interfaces, pp. 127–160. Kluwer Academic Publishers, Norwell, MA, USA (2004), http://dl.acm.org/citation.cfm?id=1138317.1138323
[8] Kueblbeck, C., Ernst, A.: Face detection and tracking in video sequences using the modified census transformation. Journal on Image and Vision Computing 24(6), 564–572 (2006)
[9] Lee, Y., Terzopoulos, D., Waters, K.: Realistic modeling for facial animation. pp. 55–62 (1995)
[10] Picard, R.W.: Affective computing. MIT Press, Cambridge, MA, USA (1997)
[11] von der Pütten, A.M., Krämer, N.C., Becker-Asano, C., Ishiguro, H.: An android in the field. In: Proceedings of the 6th international conference on Human-robot interaction. pp. 283–284. HRI '11, ACM, New York, NY, USA (2011), http://doi.acm.org/10.1145/1957656.1957772
[12] Resnick, M.: All i really need to know (about creative thinking) i learned (by studying how children learn) in kindergarten. In: Proceedings ACM SIGCHI Conference on Creativity & Cognition (C&C-07). pp. 1–6. ACM (2007), http://doi.acm.org/10.1145/1254960.1254961
[13] Yohanan, S., MacLean, K.E.: A tool to study affective touch. In: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems. pp. 4153–4158. CHI EA '09, ACM, New York, NY, USA (2009), http://doi.acm.org/10.1145/1520340.1520632

Designing Interface Agents: Beyond Realism, Resolution, and the Uncanny Valley

Eva Krumhuber[1], Marc Hall[2], John Hodgson[2], Arvid Kappas[1]

[1] School of Humanities and Social Sciences,
Jacobs University Bremen,
Campus Ring 1, 28759 Bremen,
Germany
[2] Department of Computing, Engineering and Technology,
University of Sunderland
St Peter's Way, Sunderland, SR6 0DD,
United Kingdom
{e.krumhuber, a.kappas}@jacobs-university.de
{marc.hall, john.hodgson}@sunderland.ac.uk

Abstract. Previous attempts in designing interface agents have been concerned mainly with producing highly realistic-looking animations with emotions that are clearly recognizable. We argue that the choice of visual representation requires consideration of purpose-related psychological processes (i.e., theory of mind) in users. In an evaluation study, four synthetic characters ranging in appearance from non-human to very human (blob, cat, cartoon, human) were evaluated with respect to dispositional traits, mental states, as well as emotions. Results showed that the type of synthetic character strongly influenced what judgment was made. Whilst the blob and cat characters were well liked, attributions of intelligence, mind and complex emotions were found to be reserved more for the human-like counterparts. The findings suggest that independently of questions of realism and clarity of emotional signs, the design of interface agents should be based on attributions the type of character elicits and the function the character is to serve in a particular application.

Keywords: interface agent, visual appearance, emotion, uncanny valley, theory of mind

1 Introduction

The visual appearance of computer agents and avatars is a topic of particular interest in the fields of computer science and AI. For the user, the anthropomorphic embodiment of a program for interaction appears much more tangible than the “black box” or a computational device displaying printed text on a screen. As such, the personal nature of its appearance allows for it to be more approachable and life-like, thereby making it an immediate source of interaction. This is not just a question of liking; but it changes the social relationship between users and agents/bots. From previous research we know that people attribute personality traits and human characteristics to interface agents similarly as they might do to other people [1]. In

this sense, users respond emotionally to it and treat it as a social agent [2]. Considering this personification process, attempts have been made to increase the humanness of agents and avatars by adding human-like attributes. The ultimate goal of such endeavors consists for many developers in the creation of synthetic digital humans with photorealistic faces that exhibit life-like behavior [3], [4]. However, for practical reasons approaches are limited with respect to the type of realism that can be achieved [5]. Specifically, the design of such embodiments is driven by system constraints regarding the spatial and temporal resolution devices presently afford, as well as conceptual considerations. For example, anthropomorphic representations with high fidelity may lead to alienation as a consequence of the "uncanny valley effect" [6], [7]. Applying Mori's hypotheses which stem from a context in robotics to virtual agents one could argue that, if computer agents resemble humans too closely without making people fully believe that they are real, feelings of unpleasantness and uncanniness are triggered. In consequence, specific interactions might fail and users might try to avoid the "creepy" agent. To circumvent such pitfalls researchers have consequently chosen to implement the visual metaphor for their agents as cartoon-like humanoid characters, or animals, or animated objects, such as robots.

With regard to what these agents show, there has been a particular interest in the emotional expression. Emotions reveal much of a character's personality and influence the type and quality of interaction. For example, when users see a smiling agent they expect to have more enjoyable interactions compared to a non-smiling one [8]. The criterion here is typically whether an expression is recognized; this means whether a particular label, such as "happiness" is attributed to an entity when the designer intended to communicate this state. This concerns mainly how perceivers decode facial emotions, but it is not directly based on information on how senders would have encoded the expression in real life [9]. Furthermore, to maximize recognition, expressions are often not designed with ecological validity in mind. Thus, expressions correspond to stereotypical masks that are simplified in the type and quality of appearance. Mostly, these depict the six basic emotions (anger, fear, disgust, surprise, happiness, sadness) [10] and are displayed in a pure/exaggerated form [11]. Such expressions are well recognized because they function as clear representations of stereotypical emotion categories but they do not correspond to ecologically valid displays [12]. Furthermore, they do not necessarily capture the complexity of emotion attribution in the sense of what emotional states people really infer from the display [13].

Recent research in psychology may contribute in complicating the matter. Apparently, there is an interaction between how human we consider something and what mental and specifically emotional capacities we assume that "thing" to have [14], [15]. In other words, if something is less than human, we might not believe that it has the same mind a human has. Basic emotions, such as anger or happiness are easily attributed to animals, but more refined emotions, such as guilt or shame require more mind than we attribute to most animals [16]. Similarly, animated objects such as robots may remind of machines or automata and consequently lack emotions, cognitive flexibility and mind in the eye of the beholder. So what happens if agents range in appearance from highly anthropomorphic to cartoon-like or akin to animal,

perhaps to escape the uncanny valley? Could it be that the type of representation affects perceivers in ways how (affectively) smart these beings are thought of? To elucidate such questions we conducted an evaluation study in which different types of visual agent representations – from non-human to very human - were presented. Depending on how closely the characters resemble humans, it was predicted that perceivers would make different attributions of dispositional traits, mental states and emotions. Furthermore, we investigated the effects of movement on characters' evaluation. Since Mori [6], [7] made different predictions concerning the slope of the uncanny valley for static and moving displays, attributions should change as a function of the display condition.

2 Evaluation Study

The study involved forty participants (21 men, 19 women) aged between 18-35 years (M = 20.33, SD = 2.96) who participated on a voluntary basis from Cardiff University, UK. All were students or staff at the university and received £7.00 for their participation. Participants were presented with either static or dynamic displays of four embodied characters that differed in their degree of humanness: blob, cat, cartoon, and human (see Fig. 1). In the static condition, images of the characters in a neutral position were shown for 5 s. In the dynamic condition, each character consecutively displayed three types of movement – idle, bow, wave – which lasted about 10 s. All characters were displayed on blue background with an image size of 490 x 270 pixels.

Fig. 1. Four embodied characters – blob, cat, cartoon, human - from non-human to very human in a neutral position.

Participants were tested individually on a PC workstation. After signing an informed consent form, they were told that they would see several animated characters that they should rate on a number of dimensions. It was made clear that there were no right or wrong answers. Rather, they should indicate their first impression. Using MediaLab 2008 (Empirisoft) software, participants could initiate the stimulus sequence by using the mouse to click a start button on the computer screen. Each stimulus randomly appeared for 5 s (images) or 10 s (videos) and was prefaced by a rating dimension that

was displayed throughout the stimulus presentation. After the stimulus disappeared, participants were instructed to respond to the rating scale.

To allow for a varied nature in perception, we included a number of attributes that targeted dispositional traits, mental states as well as basic and social emotions. The following questions were answered on 7-point Likert-scales ranging from (1) *not at all* to (7) *very much*:

- How likeable is the character?
- How trustworthy is the character?
- How intelligent is the character?
- How engaging is the character?
- To what degree does the character have a mind on its own?
- To what degree can the character experience anger?
- To what degree can the character experience shame?

These questions were posed in random order, with one question per stimulus presentation.

3 Results

A multivariate analysis of variance (MANOVA) with condition (static, dynamic) and sex of participant (male, female) as between-subjects factors, and stimulus character (blob, cat, cartoon, human) as within-subjects factor was conducted on the seven dependent variables: likeable, trustworthy, intelligent, engaging, mind, anger, and shame. For all univariate analyses, a Greenhouse-Geisser adjustment to degrees of freedom was applied. There were no significant effects associated with sex of participant, $F(7, 30) = 0.98$, $p = .461$, and this factor was dropped in all further analyses. As expected, the multivariate main effect of stimulus character was highly significant, $F(7, 110) = 13.88$, $p = .000$. Univariate tests showed significant effects for nearly all variables: likeable, $F(2.56, 97.31) = 10.42$, $p = .000$, trustworthy, $F(2.87, 109.11) = 6.68$, $p = .000$, intelligent, $F(2.79, 106.21) = 14.97$, $p = .000$, engaging, $F(2.54, 96.41) = 1.80$, $p = .160$, mind, $F(2.75, 104.65) = 6.96$, $p = .000$, anger, $F(2.93, 111.32) = 15.98$, $p = .000$, and shame, $F(2.74, 104.32) = 5.26$, $p = .003$.

As can be seen in Fig. 2, for ratings of intelligence, the blob scored lowest and significantly different from the other characters (ps < .001). This was similar for attributions of mind in which the blob received lowest ratings which differed significantly from those of the cartoon ($p = .037$) and human character ($p = .005$). Furthermore, participants attributed less mind to the cat in comparison to the human character ($p = .003$). With respect to perceptions of anger, the cartoon character was judged to be most capable of experiencing anger with ratings significantly different from all other characters (ps < .01). Additionally, it was also perceived as least likeable and trustworthy, with ratings significantly lower than those of the remaining characters (ps < .05). The human character was perceived more capable to experience anger than the blob ($p = .05$). For ratings of shame, the human character scored significantly higher than both the blob ($p = .004$) and cat character ($p = .010$).

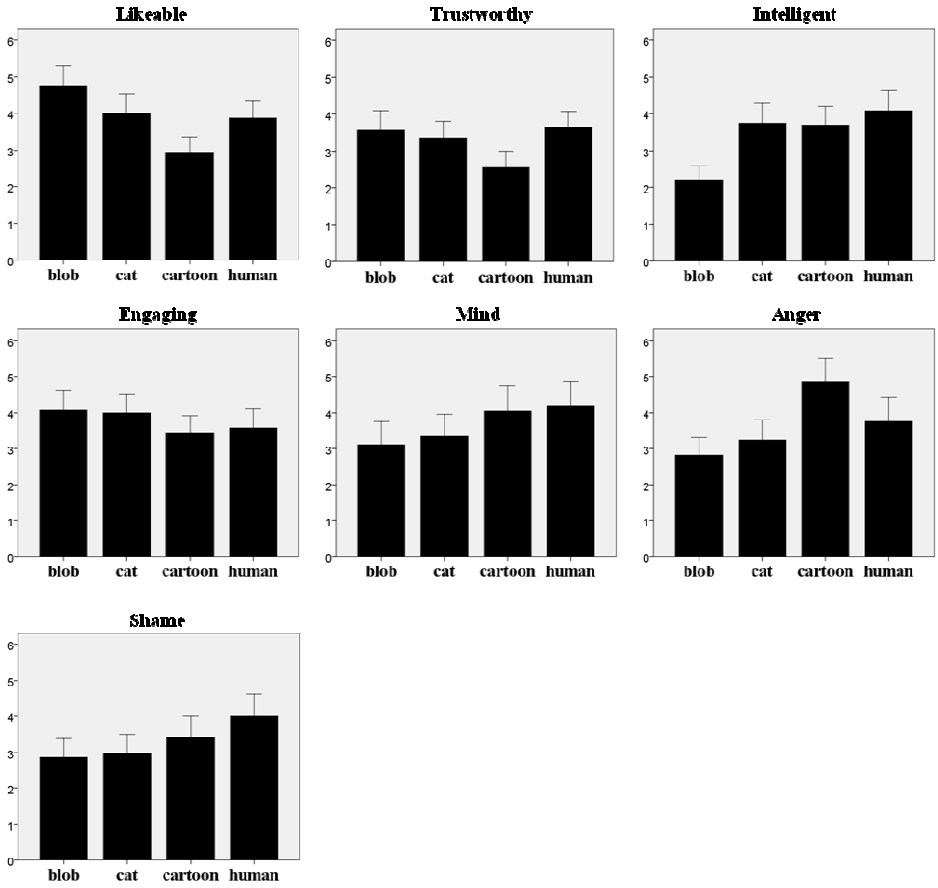

Fig. 2. Mean ratings of the four characters for the seven dependent measures. Error bars represent standard errors.

The multivariate main effect of condition was not significant, $F(7, 32) = 1.20$, $p = .331$. However, there was a significant interaction between condition and character, $F(7, 110) = 3.83$, $p = .001$. In univariate terms this interaction was significant only for ratings of likeability, $F(2.56, 97.31) = 4.57$, $p = .007$, and trustworthiness, $F(2.87, 109.11) = 3.03$, $p = .035$. Post-hoc comparisons showed that the cat character was perceived to be more likeable in the dynamic than in the static condition ($M_{dynamic} = 4.67$ vs. $M_{static} = 3.26$, $p = .006$). In contrast, trustworthiness ratings of the blob were significantly higher in the static than in the dynamic condition ($M_{dynamic} = 3.09$ vs. $M_{static} = 4.10$, $p = .042$).

4 Discussion

Results showed that the attribution of dispositional traits, mental states, as well as basic and social emotions differed depending on the type of computer agent. Overall, the blob was well liked, but ratings of intelligence and mind were lowest for this type of character. Given that it was the most non-human and object-like looking, participants might have ascribed less mental capacities which are usually reserved for humans [14], [15]. This is also reflected by the finding that the cat as a living, but non-human being was seen to possess less mind than the human character. Thus, the human appearance seems to play a crucial role in what kind of attributions people make. If something is less than human we might not perceive it as having the same mind as a human. Moreover, such lower perceived ability to reason and mentalize is interlinked with how emotionally smart those characters are seen. Specifically, refined emotions such as shame require more mind than what is attributed to objects and most animals. Respective ratings of the present study corroborate that notion. Both the blob and cat character were judged as being least capable to experience shame. In comparison, ratings of shame and anger were highest for the human and cartoon character, indicating that participants perceived them as being most capable to experience complex emotional and mental states.

For the proposed relation between human resemblance and perceiver's affinity, Mori [6], [7] had made slightly different predictions for moving and static displays. In the current study, attributions of likeability and trustworthiness were moderated by the type of display condition. Interestingly, this effect occurred for the two characters being furthest away from the human endpoint (i.e, blob and cat). Given that bowing and waving were chosen as representation of dynamic displays, it is feasible that these typical human movements exerted their influence particularly in how non-human characters were perceived. This is an intriguing finding as it suggests that the slope of the uncanny valley may not only be sensitive to the presence of motion, but also to the type of movement and how closely it represents human-like behavior.

5 Conclusion

The findings have important implications for the design of anthropomorphic characters in the field of computer science and AI. Previous efforts have focused largely on issues such as realism and emotional clarity. In that context, attempts have been made in producing highly realistic-looking animations with emotions that are easily recognizable [3], [11]. We argue that the design of agents is not just an issue of realism but requires consideration of purpose-related psychological processes in users. There is more to designing an agent than optimizing for the practical constraints of a particular implementation and avoiding the uncanny valley. It does make a difference whether an agent looks like a human, or an animal. It would appear that likeability is an important point, but if the blob is likeable but stupid, it would not be a good idea to use the blob to provide feedback in a serious matter. If the cartoon character is intelligent, but not trustworthy, you would not want to use such a

representation in a sales-type interaction. In other words, depending on the function that a particular agent has, the choice of visual representation should take into account issues such as what types of inferences regarding the cognitive and emotional intelligence it invites. Here a closer collaboration of psychologists and computer scientists and engineers can be particularly promising. It would be interesting to what degree such effects persist over longer periods of interaction, or to what degree users of different ages (e.g, children) or from different cultural background are susceptible to such effects. More research is needed regarding these issues.

In psychology there is much research regarding Theory of Mind – this relates to the capacity of humans to imagine the thoughts and feelings of other humans [17]. When designing artificial interactants, whether embodied in the shape of robots, or virtual in the shape of agents, we must also consider the Theory of Mind the users are going to employ as a function of the design choices the engineers make [18], [19]. This study provides a pointer towards the type of evaluation studies that might be helpful in this context, but it is only a starting point towards the development of a systematic attempt to clarify criteria for development of artificial entities that can realize the communicative intent of its designers.

Acknowledgments. This work has been conducted within the European Commission project eCUTE – Education in Cultural Understanding, Technologically-Enhanced (FP7-ICT-2009.4.2). We thank Tony Manstead for his help with data collection.

References

1. Dehn, D.M., Van Mulken, S.: The Impact of Animated Interface Agents: A Review of Empirical Research. Int. J. Hum-Comput. St. 52, 1--22 (2000)
2. Reeves, B., Nass, C.: The Media Equation. Cambridge University Press, New York (1996)
3. Takács, B., Kiss, B.: The Virtual Human Interface: A Photorealistic Digital Human. IEEE Comput. Graph. 23, 38--45 (2003)
4. Alexander, O., Rogers, M., Lambeth, W., Jen-Yuan Chiang, Wan-Chun Ma, Chuan-Chang Wang, Debevec, P.: The Digital Emily Project: Achieving a Photorealistic Digital Actor. IEEE Comput. Graph. 30, 20--31 (2010)
5. Donath, J.: Virtually Trustworthy. Science 317, 53--54 (2007)
6. Mori, M.: Bukimi No Tani. The Uncanny Valley (K. F. MacDorman & T. Minato, Trans.). Energy 7, 33--35 (1970)
7. Mori, M.: The Uncanny Valley. (K. F. Macdorman & N. Kageki, Trans.). IEEE Robot. Autom. Mag. 19, 98--100 (2012)
8. Sproull, L., Subramani, M., Kiesler, S., Walker, J.H., Waters, K.: When the Interface Is a Face. Hum-Comput. Interact. 11, 97--124 (1996)
9. Russell, J.A., Bachorowski, J.A., Fernández-Dols, J.M.: Facial and Vocal Expression of Emotion. Annu. Rev. Psychol. 54, 329--349 (2003)
10. Ekman, P.: All Emotions are Basic. In: Ekman, P., Davidson, R.J. (eds.): The Nature of Emotion: Fundamental Questions, pp. 15--19. Oxford University Press, New York (1994)
11. Kerlow, I.V.: The Art of 3D Computer Animation and Effects (3rd ed.). John Wiley & Sons, New Jersey (2004)

12. Kappas, A.: Smile When You Read This, Whether You Like It or Not: Conceptual Challenges to Affect Detection. IEEE T. Affect. Comput. 1, 38--41 (2010)
13. Krumhuber, E., Kappas, A.: Moving smiles: The Role of Dynamic Components for the Perception of the Genuineness of Smiles. J. Nonverbal. Behav. 29, 3--24 (2005)
14. Haslam, N., Loughnan, S., Kashima, Y., Bain, P.: Attributing and Denying Humanness to Others. Eur. Rev. Soc. Psychol. 19, 55--85 (2008)
15. Waytz, A., Epley, N., Cacioppo, J.T.: Social Cognition Unbound: Insights into Anthropomorphism and Dehumanization. Curr. Dir. Psychol. Sci. 19, 58--62 (2010)
16. Bilewicz, M., Imhoff, R., Drogosz, M.: The Humanity of What We Eat: Conceptions of Human Uniqueness Among Vegetarians and Omnivores. Eur. J. Soc. Psychol. 41, 201--209 (2011)
17. Frith U, Frith, C.D.: Development and Neurophysiology of Mentalizing. Philos. T. Roy. Soc. B. 358, 459--73 (2003)
18. von der Pütten, A., Krämer, N.: A Survey on Robot Appearances. In: 7th Annual ACM/IEEE International Conference on Human-Robot Interaction, pp. 267--268, ACM Press, New York (2012)
19. Hegel, F., Krach, S., Kircher, T., Wrede, B., Sagerer, G.: Theory of Mind (ToM) on Robots: A Functional Neuroimaging Study. In: 3rd ACM/IEEE International Conference on Human-Robot Interaction, pp. 335--342, ACM Press, New York (2008)

Robust EEG time series transient detection with a momentary frequency estimator for the indication of an emotional change

Gernot Heisenberg, Ramesh Kumar Natarajan, Yashar Abbasalizadeh Rezaei, Nicolas Simon, and Wolfgang Heiden

Institute of Visual Computing
Department of Computer Science
Bonn-Rhein-Sieg University of Applied Sciences, Germany
{gernot.heisenberg,ramesh.natarajan,yashar.abbasalizadeh,nicolas.simon, wolfgang.heiden}@h-brs.de
http://www.emotion-computing.org
http://vc.inf.h-brs.de

Abstract. This paper describes adaptive time frequency analysis of EEG signals, both in theory as well as in practice. A momentary frequency estimation algorithm is discussed and applied to EEG time series of test persons performing a concentration experiment. The motivation for deriving and implementing a time frequency estimator is the assumption that an emotional change implies a transient in the measured EEG time series, which again are superimposed by biological white noise as well as artifacts. It will be shown how accurately and robustly the estimator detects the transient even under such complicated conditions.

Keywords: momentary frequency, emotion computing, EEG, time series processing, adaptive filters, affective computing, brain computer interfaces

1 Introduction

In empiric sciences, one often has to deal with measured data sets, trying to interpret the underlying nature of an observed global system by the interactions of its partial systems. However, in most scientific areas it is simply not possible to separate the partial systems either experimentally nor theoretically such that a separate investigation becomes feasible. When there is no chance to divide the overall system into partial ones and instead the only possible interpretation is observing the global appearance and reaction of the system, time series processing often becomes the mean of choice in order to solve this inverse problem. In the optimal case it becomes possible to create a model out of the measured data sets. With this model one may again conclude the underlying process which is responsible for the production of the measured data in the sense of at least sufficient, in the optimal case, necessary conditions for their creation. Unfortunately this is not possible with EEG data sets. Here it is rather possible to classify

the processes by extracting features from the data sets. Moreover, EEG data sets are usually very noisy and transient and are created by an underlying non-stationary and stochastic dynamics. Hence, for the analysis of human emotional stress states stochastic analysis methods are one set of tools of choice. However, turning an experimental person from one emotional state into another, e.g. from being relaxed into being stressed, implies to be able to focus with the methods on the transient itself even though the transient is heavily superimposed with noise. The aim of this paper is to present a detection of this transient during an on-the-fly analysis. The momentary frequency estimation algorithm proposed by Grieszbach et al. [1] was considered and integrated into the openVibe [2] software framework.

2 Related Work

In Non-Invasive Brain-Computer Interface (BCI), to effectively interpret the brain signals to meaningful features for further application in various fields, a systematic approach with four different phases is to be used [3]:

- Signal Acquisition
- Signal Pre-processing
- Feature Extraction
- Signal Classification

Many Researchers have contributed to the BCI field especially detecting people's emotional state by adapting different techniques and algorithms for each of the above mentioned phases with a reasonable success rate. The emotion-recognition system developed by Kwang-Eun Ko et al., [4] uses relative power values and a Bayesian network. In this system, the power spectrum of the EEG signals is analyzed by applying a Fast Fourier transform (FFT) and by decomposing the signals into five bands of Alpha, Beta, Gamma and Theta for further classifications. The relative power values of the prominent frequency band is calculated by dividing the current absolute band to the sum of the frequency ranges. The same method is adapted also by Kwang Shin Park et al. [5]. The emotion-recognition system implemented by Hosseini [6] uses the thermodymanic property called entropy, that calculates the amount of disorder in the system by which each emotion state is estimated using approximate entropy and wavelet entropy. Four different emotions (disgust, happiness, surprise and fear) are recognized by two different lifting based wavelet transforms (LBWT) and then classifying them by Fuzzy C-Means (FCM) clustering [7]. The signal features of an emotional response due to various tastes have been extracted by Common Spatial Pattern (CSP) [8]. In the systems [4], [5], [6], [8], [9], the EEG signals are separated into various frequency bands using band-pass filters except for [7] which uses Average Mean Reference (AMR) for filtering. All the systems ignored the delta range $[0 - 4Hz]$ to reduce eye blinks and other physiological artifacts but the major drawback in ignoring the delta band causes the loss of valuable data for monitoring sleep waves in an adult. Usually, EEG signals are prone to

more noise since data is recorded non-invasively using electrodes. In addition, the problem using band-pass filters for distinguishing various frequency bands is that the results are not convincing if the signal-to-noise ratio is high and this leads to heavy preprocessing of the EEG signals before meaningful features can be extracted. Moreover, emotional change in a normal human is more gradual and continuous and so band-pass filters cannot be used in real-time scenarios (e.g. monitoring pilots) where there is subtle change of the emotional state. An emotion-recognition system which can track these subtle emotional changes using the same technique for both, signal pre-processing and feature extraction, can improve the processing speed and can also provide an intuitive real-time emotion monitoring. In order to address the above mentioned problem, a frequency estimation algorithm which shows robustness and also allows for real time emotion monitoring is introduced and its applicability is explored using EEG data.

3 Estimation of Momentary Signal Parameters

A classical example for an estimation algorithm is the estimation of the mean value M_n which is based on a sequence of independent random values $\{\xi_n\}_{n=0,1,2,\dots}$. If x_i will denote the realization of the random value ξ_i, then

$$M_n = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

is a consistent estimation for $E(\xi_i)$. M_n can be easily turned into a recursive order. It is

$$\begin{aligned} M_0 &= x_0 \\ M_{n+1} &= M_n - \frac{1}{n+1}(M_n - x_{n+1}) \qquad n = 0, 1, 2, \dots \end{aligned} \tag{2}$$

In this form the algorithm is already real-time capable since it needs only the previous estimation value M_n, the current data value x_{n+1} and the current time $n+1$. With this, the continuously recursive computation of the times series is possible. Put into a more general form, the estimation procedure S looks like

$$\begin{aligned} S_0 &= s_0 \\ S_{n+1} &= S_n - c_n K(S_n, x_{x+1}) \qquad n = 0, 1, 2, \dots \end{aligned} \tag{3}$$

where $X = \{x_i\}_{i=0,1,2,\dots}$ denotes the measured data points, K is a correction term for the estimation which itself depends on S_n, the momentary data point x_{i+1} and the adaptation constant c_n. Comparing this with equation 2, the adaption constant turns into $c_n = \frac{1}{n+1}$ and following the conditions

$$\sum_{n=0}^{\infty} c_n = \infty, \quad \sum_{n=0}^{\infty} c_n^2 < \infty \tag{4}$$

the estimation procedure converges with the probability of 1 against the constant expectation value of a stationary time series.

3.1 Adaptive Time Frequency Analysis

Before introducing the concept of a momentary frequency estimation a more general definition of frequency has to be provided. Normally, frequency is the parameter of sinusoidal periodical functions, with which the number of periods per time unit is characterized [1]. Thus, for this set of functions, frequency is also defined by half the number of zero crossings. Now, with this definition the frequency characterization can be applied to our EEG times series. Since it is evident that the EEG time series do not oscillate around zero but around a finite value the momentary mean of the time series has to be estimated too. The schematic representation of the concept of this estimation is shown in figure 1. $\mathbf{L}(X)$ denotes the time shift operator, $\mathbf{M}^c(X)$ denotes the momentary mean

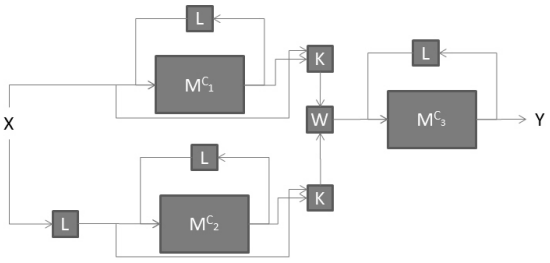

Fig. 1. Schematic for the momentary, adaptive frequency estimation.

operator. In addition, the following comparison operators are used: $\mathbf{K}(X,Y)$ (bigger) and $\mathbf{W}(X,Y)$ (unequal). $\mathbf{M}_1^c(X)$ estimates the momentary mean value of the signal whereas $\mathbf{M}_2^c(X)$ estimates the momentary mean value of the direct past. After that both mean values get interpreted by the comparison operators whether or not a "zero crossing" has occurred which is basically a crossing through the momentary mean value or not.
The adaption constant from equation 3 just needs to fulfill the condition $0 < c < 1$ in order to guarantee a convergence to the expectation value. Higher values of c will result in a faster adaptation to the time series mean value after a transient has occurred. However, this is accompanied by quite high variation. Vice versa, a smaller c means slower but smoother adaptation. In order to understand this have a look at equation 2 which is

$$\begin{aligned} M_0 &= m_0 \\ M_{n+1} &= M_n - c(M_n - x_{n+1}) \qquad n = 0, 1, 2, ... \end{aligned} \tag{5}$$

Putting it into the form

$$M_{n+1} - M_n = -c(M_n - x_{n+1}) \qquad n = 0, 1, 2, ... \tag{6}$$

and dividing both sides by the sampling frequency δt,

$$\frac{M_{n+1} - M_n}{\delta t} = -\frac{c}{\delta t}(M_n - x_{n+1}) \qquad n = 0, 1, 2, ... \tag{7}$$

it becomes obvious that equation 7 is the discretized form of

$$\frac{\partial M}{\partial t} = -\frac{1}{\tau}(M - x) \tag{8}$$

where τ denotes the adaption time. Now it becomes evident, that if the adaption time is chosen to be large in comparison to the sampling frequency δt, the adaption constant $c = \frac{\delta t}{\tau}$ will become small and vice versa.

4 Application

Before applying the momentary frequency algorithm to real measured EEG time series, two examples for precomputed transients are shown. Fig. 2 shows the momentary frequency estimation for a computed signal containing two transients. The first transient appears when the signal changes from pure white noise into

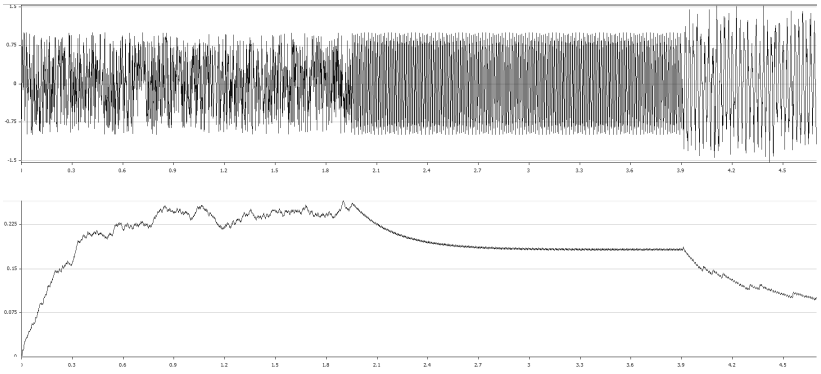

Fig. 2. This figure shows the momentary frequency estimation for a signal containing two transients, from white noise to a pure sinusoidal wave and then to a sine wave with a different frequency superimposed with white noise.

a pure sinusoidal oscillation. The estimator is able to detect the frequency jump and estimates the frequency of the sine wave. The third part of the signal is a superposition of white noise and another sinusoidal wave with a different frequency. The estimator detects this frequency jump as well and adapts very quickly. The adaption time for computations was set to $c = 0.008$.

Fig. 3 shows again a computed signal containing three different parts. Now, the difference to fig. 2 is that the third part is a superposition of the same sinusoidal wave of part two with the white noise from the first part of the signal. The estimation though is robust against the white noise and shows the same frequency. Just the variance of the signal is slightly increased which was to be expected.

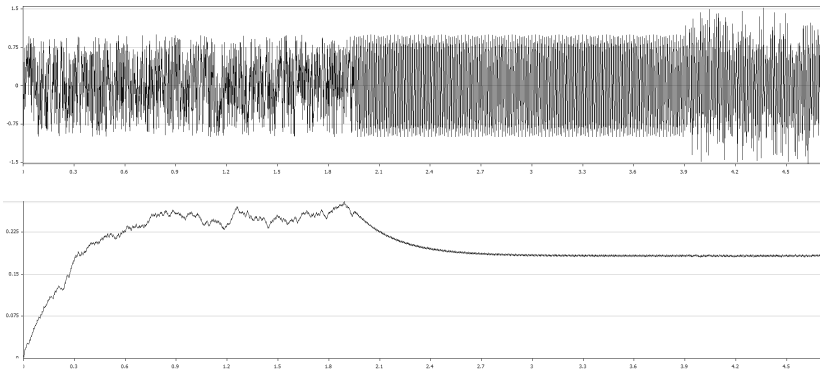

Fig. 3. This figure shows the momentary frequency estimation for a signal containing two transients again. This time the third part is a superposition of the white noise from part 1 and the sine from part two. As a result the white noise is filtered entirely.

The experiment

In order to stimulate five healthy male probands between 25-30 years old, the following experiment was carried out. The task was to play a challenging number table concentration game (`http://www.salticid.com/concentration.htm`) in a web browser. For data acquisition a non-invasive brain computer interface with 14 channels, P3 and P4 referenced, was used. For recording and processing the data stream openVibe was taken and the momentary frequency algorithm had been implemented as a box module. After $300s$ of high concentration the EEG

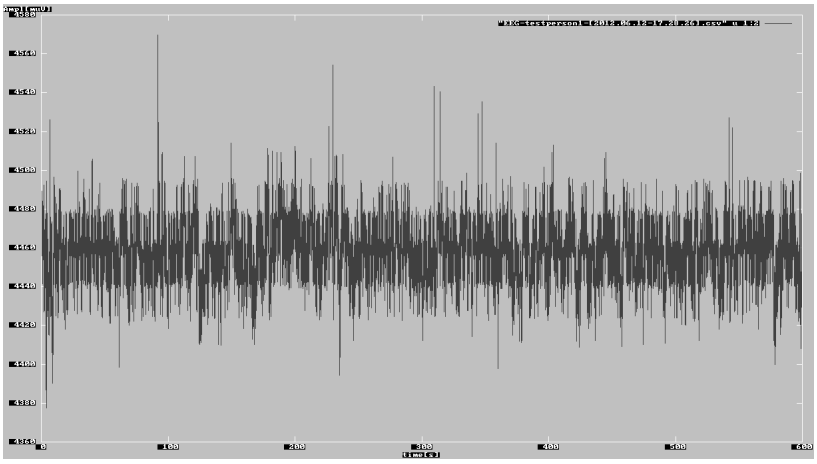

Fig. 4. $600s$ EEG data from the T7 electrode.

data showed a strong activity in the beta band which was to be expected. The

momentary frequency estimator oscillated around $18Hz$ which indicated strong mental activity due to the concentration. Fig 4 shows the recording of the EEG at the T7 electrode. After $300s$ the subjects were exposed to a $80s$ long sequence of images, lasting for about $10s$ each. The images were showing relaxing content such as beaches, holiday situations etc.. During this interval the concentration game was paused and the entire focus was on the visual consumption of the image sequence. However, the *raw* data in fig 4 does not show any significant change after the relaxing image sequence stimulus had been sent to the probands. In comparison, fig. 5 shows the momentary frequency together with the moving

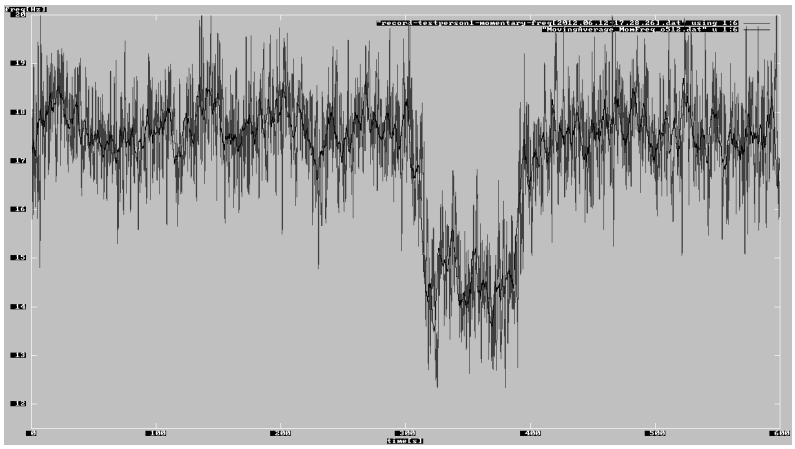

Fig. 5. Momentary frequency plotted together with its moving average of order 512 for $600s$ EEG data recorded by T7.

average of order 512 for the entire $600s$. The momentary frequency estimator was possible to detect a significant decrease down to approx. $15Hz$. After the image sequence was finished the subjects continued playing the concentration game. The momentary frequency increased quickly and reached the previous $18Hz$ again.

5 Conclusions and Future Work

This paper described adaptive time frequency analysis of EEG signals, both in theory as well as in practice. In particular, the momentary frequency estimation algorithm developed by Grieszbach et al. [1] has been derived and integrated into the openVibe software framework from INRIA and was applied to different computed and measured time series. The momentary frequency has proven its ability to detect transients in times series numerically correct (see fig. 3). Even white noise added to the signal has been filtered robustly (see fig. 2). In order to obtain the same result with a band-pass filtering approach the algorithm would looked

like: (1) recording the signal, (2) computing a power spectrum and looking for the main frequency contributions, (3) filtering the signal with a small frequency windowing of about 1-2 Hz width and (4) matching the result with the applied stimulus. Now it becomes evident, that the clear advantage of the momentary frequency estimation compared to band-pass filtering is that the unknown frequency of the transient is robustly computed without any preprocessing, the time is delivered implicitly when the transient appears, and most importantly, the algorithm allows for on-the-fly processing and does not need any storage of the signal. Since our future work will focus on carrying out further transient inducing experiments for detecting ERD (event related desynchronization) and ERP (event related potentials) and since we want to combine this with machine learning algorithms, the momentary frequency algorithm is the mean of choice.
This work is being supported financially by the Institute of Visual Computing at Bonn-Rhein-Sieg University of Applied Sciences. In addition, special thanks to all the volunteering probands of the concentration experiments.

References

1. Grieszbach, G., Schack, B., Putsche, P., Bareshova, E., Bolten, J.: Dynamic description of stochastic signal by adaptive momentary power and momentary frequency estimation and its application in analysis of biological signals. Medical and Biological Engineering and Computing **32** (1994) 632–637
2. Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., Bertrand, O., Lécuyer, A.: OpenViBE: An Open-Source Software Platform to Design, Test and Use Brain-Computer Interfaces in Real and Virtual Environments. Presence: Teleoperators and Virtual Environments / Presence Teleoperators and Virtual Environments **19**(1) (April 2010) 35–53 Département Images et Signal Département Images et Signal.
3. Sejnowski, T.J., Dornhege, G., Millán, J.d.R., Hinterberger, T., McFarland, D.J., Müller, K.R.: Toward Brain-Computer Interfacing (Neural Information Processing). The MIT Press (2007)
4. Ko, K.E., Yang, H.C., Sim, K.B.: Emotion recognition using EEG signals with relative power values and Bayesian network. International Journal of Control, Automation and Systems **7**(5) (October 2009) 865–870
5. Kwang Shin Park, Hyun Choi, Kuem Ju Lee, Jae Yun Lee, K.O.A., Kim, E.J.: Emotion recognition based on the asymmetric left and right activation. International Journal of Medicine and Medical Sciences **3(6)**(June) (2011) 201 – 209
6. Hosseini, S.A.: Emotion recognition method using entropy analysis of EEG signals. International Journal of Image **3**(5) (2011) 30
7. Murugappan, M., Rizon, M., Nagarajan, R., Yaacob, S., Zunaidi, I., Hazry, D.: Lifting scheme for human emotion recognition using EEG. In: 2008 International Symposium on Information Technology, IEEE (2008) 1–7
8. Park, C., Looney, D., Mandic, D.P.: Estimating human response to taste using EEG. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference **2011** (January 2011) 6331–4
9. Mikhail, M., El-Ayat, K., El Kaliouby, R., Coan, J., Allen, J.J.B.: Emotion detection using noisy EEG data. In: Proceedings of the 1st Augmented Human International Conference on - AH '10, New York, New York, USA, ACM Press (April 2010) 1–7