

PREREQUISITES FOR AFFECTIVE SIGNAL PROCESSING (ASP) *Part V - A Response to Comments and Suggestions*

Egon L. van den Broek

*Human-Media Interaction (HMI), Faculty of EEMCS, University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands
Karakter University Center; Radboud University Medical Center Nijmegen
P.O. Box 9101, 6500 HB Nijmegen, The Netherlands*

Joris H. Janssen

*Human Technology Interaction, Eindhoven University of Technology, Den Dolech 2, 5600 MB Eindhoven, The Netherlands
Body, Brain & Behaviour Group, Philips Research, High Tech Campus 34, 5656 AE Eindhoven, The Netherlands*

Marjolein D. van der Zwaag, Joyce H. D. M. Westerink

Body, Brain & Behaviour Group, Philips Research, High Tech Campus 34, 5656 AE Eindhoven, The Netherlands

Jennifer A. Healey

Future Technology Research, Intel Labs Santa Clara, Juliette Lane SC12-319, Santa Clara CA 95054, U.S.A.

Keywords: Affective signal processing, Emotion, Review, Temporal aspects, Prerequisites, Guidelines.

Abstract: In four papers, a set of eleven prerequisites for affective signal processing (ASP) were identified (van den Broek et al., 2010): validation, triangulation, a physiology-driven approach, contributions of the signal processing community, identification of users, theoretical specification, integration of biosignals, physical characteristics, historical perspective, temporal construction, and real-world baselines. Additionally, a review (in two parts) of affective computing was provided. Initiated by the reactions on these four papers, we now present: i) an extension of the review, ii) a post-hoc analysis based on the eleven prerequisites of Picard et al.(2001), and iii) a more detailed discussion and illustrations of temporal aspects with ASP.

1 INTRODUCTION

To align research on affective signal processing (ASP), a set of eleven prerequisites for ASP were proposed (van den Broek et al., 2010): validity, triangulation, a physiology-driven approach, signal processing contributions, physical characteristics, baselines, historical perspective, integration of biosignals, user identification, temporal construction, and theoretical specification. Since the publication of these papers, the authors have received many suggestions and comments following these papers. This article provides a response to the three most prominent reactions: i) an extension of the review on ASP, ii) the use of the prerequisites in practice, and iii) concerns on various temporal aspects of ASP.

The problem with a review is that it is impossible to be complete. However, one can always aim to achieve this goal as closely as possible. Therefore, this paper presents a review on ASP and affective computing (AC), complementary to the previous two (van den Broek et al., 2010); see Table 1. In addition, to show the use of the prerequisites in practice, we apply the prerequisites to post-hoc analyze the seminal work of (Picard et al., 2001); see Table 3. In Section 3, we elaborate more on temporal aspects of ASP and, more in general, of biosignals. Finally, we draw conclusions and denote the prerequisites' implications for applications on ASP.

Table 1: An overview of 18 studies on automatic classification of emotions, using biosignals / physiological signals.

information source	year	signals	part- cipants	number of features	selection / reduction	classifiers	target	classification result
Fernandez & Picard	1997	C, \mathcal{E}	24	5	B-W	HMM, Viterbi	frustration / not	63%
Healey & Picard	1998	$C, \mathcal{E}, \mathcal{R}, \mathcal{M}$	1	11	Fisher	QuadC, LinC	3 emotions anger / peacefulness 2 arousal levels 2 valence levels	87%-75% 99% 84% 66%
Healey & Picard	2000	$C, \mathcal{E}, \mathcal{R}, \mathcal{M}$	1	12	SFS	kNN	4 stress levels	87%
Takahashi & Tsukaguchi	2003	C, \mathcal{B}	10	12		NN, SVM	2 valence levels	62%
Rani et al.	2003	$C, \mathcal{E}, \mathcal{M}, \mathcal{S}$	1	18		FL, RT	3 anxiety levels	59%-91%
Herbelin et al.	2004	$C, \mathcal{E}, \mathcal{R}, \mathcal{M}, \mathcal{S}$	1	30	reg.LDA	kNN	5 emotions	24%
Takahashi	2004	$C, \mathcal{E}, \mathcal{B}$	12	18		SVM	5 emotions	42%
			12	18		SVM	3 emotions	67%
Zhou & Wang	2005	$C, \mathcal{E}, \mathcal{S}$, and others	32	?		kNN, NN	2 fear levels	92%
Rainville et al.	2006	C, \mathcal{R}	15	18	ANOVA, PCA	LDA	4 emotions	65%
			15	18	ANOVA, PCA	LDA	2 emotions	72%-83%
Liu et al.	2007	$C, \mathcal{E}, \mathcal{M}$	3	54		SVM	3 × 2 levels	85%/80%/84%
Villon et al.	2007	C, \mathcal{E}	40	28		regression model	5 emotions	63%-64%
Rani et al.	2007	$C, \mathcal{E}, \mathcal{M}, \mathcal{S}$	5	18		FL, RT	anxiety scale?	57%-95%
Hönig et al.	2007	$C, \mathcal{E}, \mathcal{R}, \mathcal{M}, \mathcal{S}$	24	4 × 50		LDA, GMM	2 levels of stress	94%-89%
Kreibig et al.	2007	$C, \mathcal{E}, \mathcal{R}, \mathcal{M}, \mathcal{O}$	34	23	ANOVA	PDA	fear, sadness, neutral	69%-85%
Liu et al.	2008	$C, \mathcal{E}, \mathcal{M}$	6	54		SVM	3 × 2 levels	81%
		$C, \mathcal{E}, \mathcal{M}$	6	54		SVM	3 levels	72%
Conn et al.	2008	$C, \mathcal{E}, \mathcal{M}, \mathcal{S}$	6	?		SVM, QV-learning	3 × 2 levels	83%
			6	?		SVM	3 behaviors	81%
Cheng et al.	2008	\mathcal{M}	1	12	DWT	NN, TM	4 emotions	75%
Benovoy et al.	2008	$C, \mathcal{E}, \mathcal{R}, \mathcal{S}$	1	225	SFS, Fisher	LDA, kNN, NN	4 emotions	90%

Signals: C : cardiovascular activity; \mathcal{E} : electrodermal activity; \mathcal{R} : respiration; \mathcal{M} : electromyogram; \mathcal{S} : skin temperature; \mathcal{O} : Expiratory pCO_2 .

Classifiers: HMM: Hidden Markov Model; RT: Regression Tree; NN: Artificial Neural Network; SVM: Support Vector Machine; LDA: (Fisher) Linear Discriminant Analysis; kNN: k-Nearest Neighbors; FL: Fuzzy Logic System; TM: Template Matching classifier; QuadC: Quadratic classifier; LinC: Linear classifier; Viterbi: Viterbi decoder

Selection: B-W: Baum-Welch re-estimation algorithm; PCA: Principal Component Analysis; SFS: Sequential Forward Selection; ANOVA: Analysis of Variance; DWT: Discrete Wavelet Transform; Fisher: Fisher projection; PDA: Predictive Discriminant Analysis.

2 REVIEWING AFFECTIVE SIGNAL PROCESSING (ASP)

ASP is mainly employed in four specialized areas of signal processing: movement analysis, computer vision techniques, speech processing, and biosignal processing (van den Broek et al., 2010). This article focusses on the last category, which has received very little attention compared to the other three.

Although studies on AC are sometimes claimed to be successful, their results are hardly brought to the market (cf. van den Broek, 2010b). The burden on ASP no longer lies in the recording and processing of biosignals. Nowadays, high fidelity, cheap, and unobtrusive biosignal recordings are easy to obtain and can even be easily integrated into various products. The problem lies in the lack of in depth understanding of the relation between biosignals and our emotions (Pi-

card, 2010; van den Broek, 2010a; van den Broek, 2010b). (Picard, 2010; van den Broek, 2010a; van den Broek, 2010b).

The review in Table 1 illustrates both the differences and the similarities between studies on AC. As this table shows, most studies recorded people's cardiovascular and electrodermal activity. However, differences between the studies prevail over the similarities. The number of participants varies from 1 to 40, with studies including > 15 participants being rare, see Table 1. The number of features extracted from the biosignals also varies considerably: from 5 to 225. Only half of the studies applied feature selection/reduction, where this would be advisable in general.

For AC, a plethora of classifiers are used. The characteristics of the categories among which has to be discriminated is different from those in most other classification problems. The emotion classes used are

Table 2: Physiological processes and delay in recording them through biosignals.

Physiological process	delay
Cardiovascular activity (except HR)	30 sec.
Heart Rate (HR)	1 sec.
Electrodermal Activity (EDA)	> 2-4 sec.
Skin temperature (ST)	> 10 sec.
Respiration	5 sec.
Muscle activity <i>through</i> EMG	instantly
Movements / Posture	instantly

typically ill defined, which makes it hard to compare studies. Also, the number of emotion categories (i.e., the classes) to be discriminated is small: from 2 to 5 (see Table 1) up to (sometimes) 8 (Picard et al., 2001; Picard, 2010). Although these are small numbers in terms of pattern recognition and machine learning, the results lie behind those of other classification problems. Moreover, it is unlikely that human’s affect can be described via discrete states. With AC, a large variety of recognition rates is present: 42%–94%; see also Table 1. In other pattern recognition problems, only recognition rates of > 90% are reported. Taken together, this all illustrates the complex nature of AC and the need to consider prerequisites for ASP.

To bring the prerequisites from theory to practice, we conducted a post-hoc analysis of the most influential article on affective computing, and with that on ASP, so far: (Picard et al., 2001). As is shown in Table 3, this revealed pros and cons of this study and provides valuable directives for future research.

3 TEMPORAL CONSTRUCTION

Among the questions the authors received on their prerequisites for ASP, a significant body concerned temporal aspects in ASP. In processing biosignals, temporal aspects are (indeed) of crucial importance. This importance exceeds the domain of affective computing and holds, in general, for biosignal processing. Therefore, in this section, we will elaborate on temporal aspects in biosignals and explain and illustrate why they should be taken into account with ASP.

First of all, people habituate; this is something ASP has to deal with. For this, it is necessary to track the number of stimuli that could trigger changes in emotional state. However, outside controlled (lab) environments, this requires a true understanding of both context and user, which is beyond science’s current state-of-the art (van den Broek, 2010a; van den Broek, 2010b). An initial approach could be to add a variable to current models that represents the time or amount

of stimuli that the participants have received so far. This should help to model a part of the habituation noise.

Second, another issue is the use of different time windows. For instance, we can look at parts of > 30 minutes but also at 5, 10, or 60 seconds, see also Figure 1. Figure 1a provides an EDA signal under investigation. On the right, Figure 1 presents three time windows of this signal that surround the event indicated in Figure 1 on the right. The distinct shapes of the signal within these three time windows perfectly illustrates the significant impact window length has on calculating a signal’s features, which is confirmed by their statistics as shown in Table 4. The challenge lies in the fact that we must estimate when the actual emotional event occurred and how long it persisted. This is problematic because there is always a lag between the changes in biosignals and user responses on the onset of the event; see also Table 2. In practice, time window selection can be done empirically; either manually or automatically; for example, by finding the nearest significant local minima or making assumptions about the start time and duration of the emotion.

Third, different psychological processes develop over different time scales. So, the time window to be used should depend on the psychological construct studied. Furthermore, the lag between the occurrence of an emotion and the accompanying physiological change differs per signal. For example, heart rate changes almost immediately, skin conductance takes 2 – 5 seconds, and skin temperature can take even longer to change; see Table 2. Hence, time windows should depend on the used signal.

Finally, physiological activity tends to move to a stable neutral state. So, the more extreme a physiological state is, the smaller reactions towards this extreme become given the same stimulus. Therefore, the pre-stimulus physiological level should always be taken into account. This is also known as the principle of initial values (van den Broek et al., 2010). The principle of initial values has shown to be a linear relationship; hence, it can be conveniently modeled (e.g., using linear regression).

4 CONCLUSIONS

This paper provides a response to the three main comments the authors received on their four prerequisites papers (van den Broek et al., 2010). First, following the comments that our review was limited, a review on ASP, complementary to the previous two (van den Broek et al., 2010), has been presented; see Ta-

Table 3: Post-hoc analysis of (Picard et al., 2001), using the complete set of eleven prerequisites (van den Broek et al., 2010).

Introduction One of the first cases of the application of ASP to the extended monitoring of an individual's emotional patterns was (Picard et al., 2001). This was a laboratory study in which a single subject used method acting (i.e., acting and feeling) to portray eight emotions every morning for approximately one month.

Validity The Clynes (1977) protocol for emotion generation of was adopted. This provided a detailed specification of eight emotions. The subjects kept a record describing the particular aspects of each emotion during each day's session. However, only an aggregate label was used to analyze each of the emotions as a group, which decreased the construct validity of the data, because the actual signal for each day is less accurately described. The construct validity of the expressed emotions would have been increased if there had been sufficient data to analyze different types of expression of each emotion. The construct validity is limited by the fact that the emotions were always expressed in the same order and that the length of each expression was three minutes. Because of this it is unclear to what extent steady state anger is being measured as opposed to the transition between "no emotion" and "anger". Ecological validity for this study is confined to lab conditions with seated subjects. Moreover, there was no automatic way to capture context that was not part of the experiment.

Triangulation Multiple biosignals as well as observation and introspection were used to measure the construct under investigation. This enabled a rich set of data.

Physiology driven approach The goal of this project was to develop a physiological self-monitoring system that was tailored to the individual. This paper shows how unique patterns can develop for each of the emotion states. However, these patterns may be individual-specific and as such can potentially achieve a higher degree of specificity than is generally the case.

Signal Processing Contributions This paper puts forward several unique signal processing contributions, including the idea of using features of the spectrum of respiration for identifying individual emotion patterns and exhaustively combining sets of features to find the optimal differentiating set for this individual. The Sequential Floating Forward Search (SFFS) method used in the paper is a standard feature selection method and could be applied to the same data set for a different individual, perhaps allowing the creation of different optimal subsets of sensors for different individuals.

Identification of Users This study used data from only one person. So, the models fit very well for this person but are probably over fitted with respect to other people. Moreover, the authors did not record any specifics about this user so no generalization to groups can be made. Therefore, it is impossible to determine which of the features are valid for only this user as opposed to which might be valid for all users or a group of users. Thus, this study poorly fulfills this prerequisite and would definitely have benefited from including more users.

Theoretical specification The theoretical specification of this study is limited. Although there are references to literature specifying the theoretical relationship between the signals chosen and various emotions, there is neither discussion on the reason for choosing particular features, nor an explanation for the success of certain subsets of features in differentiating between these emotional states, where other subsets fail.

Integration of biosignals Although this study uses advanced signal processing techniques to find patterns and evaluate them across multiple physiological features, it fails to take advantage of known relationships between biosignals and integrate them at the feature level. The models used had generic feature selection algorithms and do not represent a framework that theorizes a relationship between multiple biosignals, an appropriate model, data gathering, and model training. Instead, data was first gathered with no particular hypothesis about the relationship between features and an exhaustive search was conducted, using randomly selected groups of features to find the best result. Incorporating features combining respiration and heart rate variability, or accelerometer and EDA data might have resulted in a higher performance.

Physical characteristics Equipment and materials used are summarized. However, information on the type of electrodes used (wet or dry), the size of the EMG electrodes, and the position of the location of the EDA electrodes are omitted. In addition, the temperature and humidity of the environment are not reported, although it must be noted that the authors deal with this by normalization of the data per session per participant. This makes it hard to judge whether or not these results can be generalized to different situations, and how the models would work with small wearable sensor devices.

Historical perspective Previous work on signal processing and pattern recognition is discussed, in particular that in relation to affective computing. In contrast, little attention was given to the rich history of both psychophysiology and emotion research. However, in more recent work, (Picard, 2010) stresses the importance of this prerequisite.

Temporal Construction The subject was trained in acting and visualization techniques and could probably continue to evoke the emotion consistently for the three minute period. Considering the fact that emotions are generally much shorter than three minutes, the actor had to re-express each emotion several times within the three minutes. This will definitely have led to habituation and created noise in the biosignals. Moreover, there was no gap to allow the subject to fully recover between emotion states. With the analysis, the authors tried to minimize this effect by taking only the later part of each emotion period, but EDA recovery times can be quite long (>15 minutes); so, this may still represent a temporal transition. Instead, adjusting for the initial value with regression would probably have been a more successful solution.

Baselining Because this work studied only one user, baselining over different users was not necessary. However, they did find large inter-day differences and used baselining to normalize for those. Specifically, they tried two approaches that were both successful, which shows how important baselining is for ASP.

Conclusion By applying the prerequisites to a post-hoc analysis of the seminal work of Picard et al. (2001), we have illustrated their use. Some of the prerequisites were already applied successfully by the authors, while others were neglected. By taking into account all of the prerequisites, we are sure that, even post-hoc, the results can be improved.

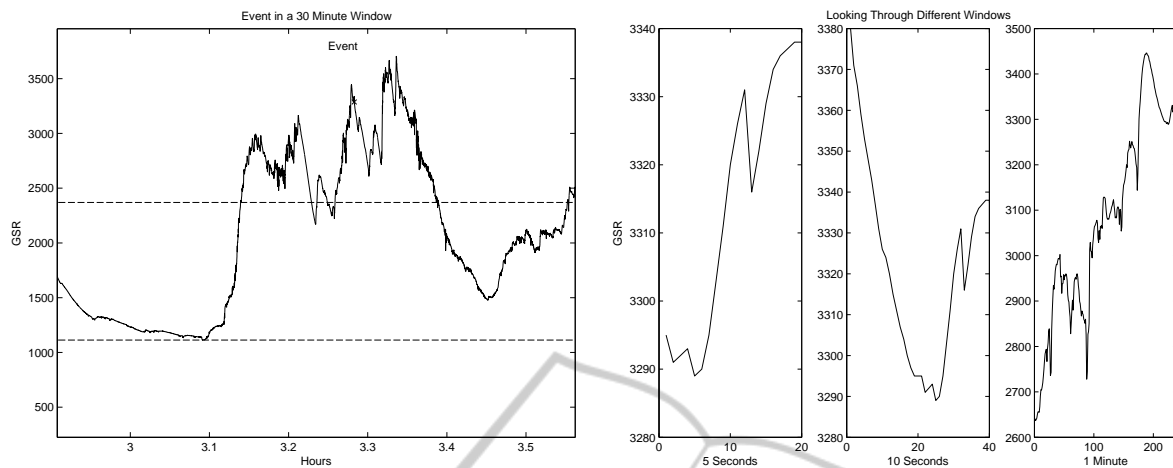


Figure 1: A 30 minutes time window of an EDA signal with three close-ups surrounding an 'event' (denoted as such).

ble 1. Second, a post-hoc analysis was conducted of the seminal work of (Picard et al., 2001), as provided in Table 3. This analysis illustrated the use of all eleven prerequisites in practice. Third, a more detailed discussion on temporal aspects of ASP was initiated, since the authors received many comments on this prerequisite. This extended elaboration on this prerequisite includes a concise overview of the lag of biosignals; see Table 2. In addition, the impact of choosing the time window was illustrated both by Figure 1 and by the accompanying statistics as provided in Table 4.

The additional review, the post-hoc analysis of Picard et al. (2001), and the additional discussion on temporal aspects of ASP all illustrate both the complexity and lack of success of AC. This makes one wonder whether or not affective computing can pay off its promises. The bottleneck is not the technique but our lack of understanding of affective signals (Picard, 2010; van den Broek, 2010a; van den Broek, 2010b). We hope that these prerequisites can initiate a first step towards rethinking ASP.

Table 4: Standard statistics on the three time windows of an EDA signal, as presented in Figure 1 (right).

Statistic	Time window		
	5 sec.	10 sec.	60 sec.
mean	3314	3312	3083
standard deviation	19	23	217
slope	43	-69	697

ACKNOWLEDGEMENTS

Egon L. van den Broek acknowledges the support of the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science. Further, we thank the three reviewers for their constructive suggestions on and Lynn Packwood for proof reading of a previous version of this article.

REFERENCES

- Benovoy, M., Cooperstock, J. R., and Deitcher, J. (2008). Biosignals analysis and its application in a performance setting: Towards the development of an emotional-imaging generator. In *Biosignals 2008: Proceedings of the first International Conference on Biomedical Electronics and Devices*, volume 1, pages 253–258, Funchal, Madeira, Portugal. INSTICC.
- Cheng, B. and Liu, G. (2008). Emotion recognition from surface EMG signal using wavelet transform and neural network. In *Proceedings of The 2nd International Conference on Bioinformatics and Biomedical Engineering (ICBBE)*, pages 1363–1366, Shanghai, China. Piscataway, NJ, USA: IEEE Press.
- Clynes, M. (1977). *Sentics: The Touch of the Emotions*. New York, NY, USA: Anchor Press/Doubleday.
- Conn, K., Liu, C., Sarkar, N., Stone, W., and Warren, Z. (2008). Affect-sensitive assistive intervention technologies for children with Autism: An individual-specific approach. In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication*, pages 442–447, Munich, Germany. Piscataway, NJ, USA: IEEE, Inc.
- Fernandez, R. and Picard, R. W. (1997). Signal processing for recognition of human frustration. In *Proceed-*

- ings of the *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 6 (Multimedia Signal Processing), pages 3773–3776, Seattle, WA, USA. Piscataway, NJ, USA: IEEE, Inc.
- Healey, J. and Picard, R. (2000). SmartCar: detecting driver stress. In Sanfeliu, A. and Villanueva, J. J., editors, *Proceedings of the 17th IEEE International Conference on Pattern Recognition (ICPR 2000)*, volume 4, pages 4218–4221, Barcelona, Spain.
- Healey, J. A. and Picard, R. W. (1998). Digital processing of affective signals. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP98)*, volume 6, pages 3749–3752, Seattle, WA, USA. IEEE.
- Herbelin, B., Benzaki, P., Riquier, F., Renault, O., and Thalmann, D. (2004). Using physiological measures for emotional assessment: A computer-aided tool for cognitive and behavioural therapy. In Sharkey, P., McCrindle, R., and Brown, D., editors, *Proceedings of the Fifth International Conference on Disability, Virtual Reality and Associated Technologies*, pages 307–314, Oxford, United Kingdom.
- Hönig, F., Batliner, A., and Nöth, E. (2007). Real-time recognition of the affective user state with physiological signals. In *Proceedings of the Doctoral Consortium of Affective Computing and Intelligent Interaction (ACII)*, pages 1–8, Lisbon, Portugal.
- Kreibig, S. D., Wilhelm, F. H., Roth, W. T., and Gross, J. J. (2007). Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films. *Psychophysiology*, 44(5):787–806.
- Liu, C., Conn, K., Sarkar, N., and Stone, W. (2007). Affect recognition in robot assisted rehabilitation of children with autism spectrum disorder. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1755–1760, Roma, Italy. Piscataway, NJ, USA: IEEE, Inc.
- Liu, C., Conn, K., Sarkar, N., and Stone, W. (2008). On-line affect detection and robot behavior adaptation for intervention of children with Autism. *IEEE Transactions on Robotics*, 24(4):883–896.
- Picard, R. W. (2010). Emotion research by the people, for the people. *Emotion Review*, 2(3):250–254.
- Picard, R. W., Vyzas, E., and Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1175–1191.
- Rainville, P., Bechara, A., Naqvi, N., and Damasio, A. R. (2006). Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International Journal of Psychophysiology*, 61(1):5–18.
- Rani, P., Sarkar, N., and Adams, J. (2007). Anxiety-based affective communication for implicit human-machine interaction. *Advanced Engineering Informatics*, 21(3):323–334.
- Rani, P., Sarkar, N., Smith, C. A., and Adams, J. (2003). Affective communication for implicit human-machine interaction. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, volume 5, pages 4896–4903, Washington, D.C., USA. Piscataway, NJ, USA: IEEE, Inc.
- Takahashi, K. (2004). Remarks on emotion recognition from bio-potential signals. In *Proceedings of the 2nd International Conference on Autonomous Robots and Agents*, pages 186–191, Palmerston North, New Zealand.
- Takahashi, K. and Tsukaguchi, A. (2003). Remarks on emotion recognition from multi-modal bio-potential signals. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 2, pages 1654–1659, Washington, D.C., USA. Piscataway, NJ, USA: IEEE, Inc.
- Van den Broek, E. L. (2010a). Beyond biometrics. *Procedia Computer Science*, 1(1):2505–2513.
- Van den Broek, E. L. (2010b). Robot nannies: Future or fiction? *Interaction Studies*, 11(2):274–282.
- Van den Broek et al., E. L. (2009/2010). Prerequisites for Affective Signal Processing (ASP) – Parts I–IV. In Fred, A., Filipe, J., and Gamboa, H., editors, *BioSTEC 2009/2010: Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies*, pages —, Porto, Portugal / Valencia, Spain. INSTICC Press.
- Villon, O. and Lisetti, C. L. (2007). A user model of psychophysiological measure of emotion. *Lecture Notes in Computer Science (User Modelling)*, 4511:319–323.
- Zhou, J. Z. and Wang, X. H. (2005). Multimodal affective user interface using wireless devices for emotion identification. In *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society*, pages 7155–7157, Shanghai, China. Piscataway, NJ, USA: IEEE, Inc.