PREREQUISITES FOR AFFECTIVE SIGNAL PROCESSING (ASP) – PART II

Egon L. van den Broek
http://www.human-centeredcomputing.com/
vandenbroek@acm.org

Joris H. Janssen
User Experience Group, Philips Research Europe, High Tech Campus 34, 5656 AE Eindhoven, The Netherlands
joris.h.janssen@philips.com

Jennifer A. Healey
Future Technology Research, Intel Labs Santa Clara, Juliette Lane SC12-319 Santa Clara CA 95054, USA
jennifer.healey@intel.com

Marjolein D. van der Zwaag
User Experience Group, Philips Research Europe, High Tech Campus 34, 5656 AE Eindhoven, The Netherlands
marjolein.van.der.zwaag@philips.com

Keywords: affective signal processing, emotion, user identification, theoretical specification

Abstract: Last year, in van den Broek et al. (2009a), a start was made with defining prerequisites for affective signal processing (ASP). Four prerequisites were identified: validation (e.g., mapping of constructs on signals), triangulation, a physiology-driven approach, and contributions of the signal processing community. In parallel with this paper, in van den Broek et al. (2010) another set of two prerequisites is presented: integration of biosignals and physical characteristic. This paper continues this quest and defines two additional prerequisites: identification of users and theoretical specification. In addition, the second part of a review on the classification of emotions through ASP is presented; the first part can be found in van den Broek et al. (2009a).

1 INTRODUCTION

Almost half a century ago, Ulric Neisser (1963, p. 194) described three fundamental and interrelated characteristics of human thought that are conspicuously absent from existing or contemplated computer programs.

1. Human thinking always takes place in, and contributes to, a cumulative process of growth and development.
2. Human thinking begins in an intimate association with emotions and feelings which is never entirely lost.
3. Almost all human activity, including thinking, serves not one but a multiplicity of motives at the same time.

Nonetheless, artificial intelligence (AI) aimed at understanding human thought and developing computational and executable models of human thought without considering these three notions. Although, nowadays, a computer can beat the world’s best chess players, the general opinion is that AI has failed. We, and others, think that Ulric Neisser’s words are of vital importance and should be brought into AI practice.

In this paper, we will treat the second issue Neisser raised, that of emotions and feelings or, in other words, affect. Ever since Picard’s book Affective Computing (AC) 1997, this direction of research has received growing interest. However, as with AI, the results with AC are disappointing. This is explained by the fact that research relevant for AC is scattered over a broad range of sciences and lacks generalization and robustness.

To force a breakthrough in results on AC we propose to consider a set of prerequisites for affective signal processing (ASP), before starting with AC in practice. The first part of these prerequisites was introduced last year in van den Broek et al. (2009a). This set of prerequisites was, however, not complete. This paper introduces the second part of the prerequisites on ASP. In parallel, the third part of the prerequisites on ASP is introduced in van den Broek et al. (2010). Together, these three papers should form the foundation for more successful ASP and AC.
In the next section, we will briefly denote ASP and AC, including a review presented in Table 1. This table contains the second part of our survey on ASP and AC, complementary to the part presented in van den Broek et al. (2009a). Section 3 introduces two new prerequisites for successful ASP, complementary to both those introduced in van den Broek et al. (2009a) and in van den Broek et al. (2010). Finally, we draw conclusions and denote the prerequisites’ implications for applications on ASP.

2 AFFECTIVE SIGNAL PROCESSING (ASP)

ASP is often employed from three specialized areas of signal processing:

• movement analysis (Gunes and Piccardi, 2009),
• computer vision techniques (Gunes and Piccardi, 2009), and
• speech processing (Ververidis and Kotropoulos, 2006).

However, these signals still have their major disadvantages. In contrast, such issues have been resolved for biosignals in recent years: currently, high fidelity, cheap, and unobtrusive biosignal recordings are easy to obtain. Moreover, the recording devices can be easily integrated in various products (van den Broek and Westerink, 2009; Gamboa et al., 2009). Therefore, this paper focuses on biosignals. For an overview of the most commonly used biosignals and their features, we refer to van den Broek et al. (2009a).

The review in Table 1 illustrates both differences and similarities among research on AC, conducted over the last decade. 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 ranges from 1 to 50, although studies including > 20 participants are relatively rare; cf. Table 1. The number of features determined through ASP also ranges considerably; i.e., from 3 to 193. Only half of the studies applied feature selection/reduction, where this would be advisable in general.

For AC, a broad plethora of classifiers are used. The characteristics of the categories among which has to be discriminated is different from most other classification problems. The emotion classes used are typically ill defined, which makes it hard to compare studies. Moreover, the number of emotion categories (i.e., the classes) to be discriminated ranges considerably: from 2 to 8. Although these are small numbers in terms of pattern recognition and machine learning, the results are behind that of other classification problems. With AC recognition rates of 60%–80% are common, where in most other pattern recognition problems, recognition rates of > 90% and often > 95% are often reported. This illustrates the complex nature of AC and the need to consider prerequisites for ASP.

3 PREREQUISITES – PART II

In van den Broek et al. (2009a), the first set of prerequisites for ASP was introduced: validity, triangulation, a physiology-driven approach, and contributions from signal processing. One additional set of prerequisites is presented in van den Broek et al. (2010) and comprises: physical characteristics and integration of biosignals. Here we present a third set, which complements the two other, by discussing user identification and theoretical specification.

3.1 Tailored ASP – User identification

Throughout the field of AC, an ongoing debate is present on generic versus personal approaches to emotion recognition. Some research groups specialized in AC have moved from general AC to AC for specialized groups or individuals. For example, the group of Picard currently focuses on autism (Picard and Goodwin, 2008). In general, the identification of users has major implications for ASP. We propose three distinct categories, among which research in affective science could choose:

1. all: generic ASP; see also Table 1 and van den Broek et al. (2009c); van den Broek and Westerink (2009)
2. group: tailored ASP; e.g., Choi and Woo (2005); Sternbach and Tursky (1965)
3. individual: personalized ASP; e.g., Picard et al. (2001); Healey and Picard (2005)

Although attractive from a practical point of view, the category all will probably not solve the mysteries concerning affect. As is long known in neurology and psychology, special cases can help in improving ASP. For the categories group and individual, the following subdivision can be made:

1. Specific characteristics; e.g., autism (Picard and Goodwin, 2008)
2. Psychological traits; e.g., Personality (Krohne et al., 2002; van den Broek et al., 2009c) or empathic intelligence (Håkansson and Montgomery, 2003).
<table>
<thead>
<tr>
<th>Information source</th>
<th>Year</th>
<th>Signals</th>
<th>Participants</th>
<th>Number of Features</th>
<th>Selection / Reduction</th>
<th>Classifiers</th>
<th>Target</th>
<th>Classification Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picard et al.</td>
<td>2001</td>
<td>C,E,R,M</td>
<td>1</td>
<td>40</td>
<td>SFS, Fisher LDA</td>
<td>8 emotions</td>
<td>81%</td>
<td></td>
</tr>
<tr>
<td>Lisetti &amp; Nasoz</td>
<td>2002</td>
<td>C,E,S</td>
<td>10</td>
<td>3</td>
<td>kNN, LDA</td>
<td>5 emotions</td>
<td>85%</td>
<td></td>
</tr>
<tr>
<td>Kim et al.</td>
<td>2002</td>
<td>C,E,S</td>
<td>50</td>
<td>10</td>
<td>SVM</td>
<td>4 emotions</td>
<td>61%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SVM</td>
<td>3 emotions</td>
<td>55% / 78%</td>
<td></td>
</tr>
<tr>
<td>Kim et al.</td>
<td>2004</td>
<td>C,E,S</td>
<td>50</td>
<td>10</td>
<td>Fisher SVM</td>
<td>4 emotions</td>
<td>62%</td>
<td></td>
</tr>
<tr>
<td>Rani et al.</td>
<td>2004</td>
<td>C,E,M</td>
<td>1</td>
<td>6</td>
<td>FLS</td>
<td>2 anxiety levels</td>
<td>??%</td>
<td></td>
</tr>
<tr>
<td>Healey &amp; Picard</td>
<td>2005</td>
<td>C,E,R,M</td>
<td>9</td>
<td>22</td>
<td>Fisher LDA</td>
<td>3 stress levels</td>
<td>97%</td>
<td></td>
</tr>
<tr>
<td>Kim et al.</td>
<td>2005</td>
<td>C,E,R,M</td>
<td>3</td>
<td>26</td>
<td>SFS</td>
<td>4 emotions</td>
<td>53% / 74%</td>
<td></td>
</tr>
<tr>
<td>Lisetti &amp; Nasoz</td>
<td>2005</td>
<td>C,E,S</td>
<td>41</td>
<td>86</td>
<td>kNN, ANN (2x)</td>
<td>2 emotions (3 sets of)</td>
<td>92%</td>
<td></td>
</tr>
<tr>
<td>Liu et al.</td>
<td>2005</td>
<td>C,E,M</td>
<td>15</td>
<td>13(?)</td>
<td>kNN, RT, BN, SVM</td>
<td>5 emotions</td>
<td>86%</td>
<td></td>
</tr>
<tr>
<td>Liu et al.</td>
<td>2006</td>
<td>C,E,M,S</td>
<td>14</td>
<td>35</td>
<td>RT</td>
<td>3 anxiety levels</td>
<td>70%</td>
<td></td>
</tr>
<tr>
<td>Rainville et al.</td>
<td>2006</td>
<td>C,G,S,M,P</td>
<td>43</td>
<td>18</td>
<td>PCA</td>
<td>4 emotions</td>
<td>49%</td>
<td></td>
</tr>
<tr>
<td>Jones &amp; Troen</td>
<td>2007</td>
<td>C,E,R</td>
<td>13</td>
<td>11</td>
<td>ANN</td>
<td>5 arousal levels</td>
<td>31% / 62%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5 valence levels</td>
<td>26% / 57%</td>
<td></td>
</tr>
<tr>
<td>Yang &amp; Liu</td>
<td>2007</td>
<td>C,E,R,M</td>
<td>1</td>
<td>193</td>
<td>BPSO kNN</td>
<td>4 emotions</td>
<td>86%</td>
<td></td>
</tr>
<tr>
<td>Kim</td>
<td>2007</td>
<td>C,E,R,M,S</td>
<td>3</td>
<td>77</td>
<td>SBS</td>
<td>4 emotions</td>
<td>51%–71%</td>
<td></td>
</tr>
<tr>
<td>Cheng &amp; Liu</td>
<td>2008</td>
<td>M</td>
<td>1</td>
<td>14</td>
<td>DWT ANN</td>
<td>4 emotions</td>
<td>75%</td>
<td></td>
</tr>
<tr>
<td>Lichtenstein et al.</td>
<td>2008</td>
<td>C,E,R,M,S</td>
<td>40</td>
<td>5</td>
<td>SVM</td>
<td>5 emotions</td>
<td>47%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 levels of arousal</td>
<td>82%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 levels of valence</td>
<td>72%</td>
<td></td>
</tr>
<tr>
<td>Chanel et al.</td>
<td>2009</td>
<td>C,E,R</td>
<td>10</td>
<td>18</td>
<td>LDA, SVM, RVM</td>
<td>3 emotions</td>
<td>51%</td>
<td></td>
</tr>
<tr>
<td>Van den Broek et al.</td>
<td></td>
<td>E,M</td>
<td>21</td>
<td>10</td>
<td>ANOVA, PCA kNN, SVM, ANN</td>
<td>4 emotions</td>
<td>61%</td>
<td></td>
</tr>
</tbody>
</table>

*Signals: C: cardiovascular activity; E: electrodermal activity; R: respiration; M: electromyogram; S: skin temperature; and P: pupil diameter. Selection: PCA: Principal Component Analysis; SFS: Sequential Forward Selection; SBS: Sequential Backward Selection; BPSO: Binary Particle Swarm Optimization; DWT: Discrete Wavelet Transform; Fisher: Fisher projection; and ANOVA: ANalysis Of Variance. Classifiers: RT: Regression Tree; BN: Bayesian Network; ANN: Artificial Neural Network; SVM: Support Vector Machine; RVM: Relevance Vector Machines; LDA: Linear Discriminant Analysis; kNN: k-Nearest Neighbors; and FLS: Fuzzy Logic System. Note: * The authors present 100% successful classification for two emotion categories. However, this might indicate a questionable training and testing setup.
3. Demographics; e.g., age, sex, race (Sternbach and Tursky, 1965), or level of education (van den Broek et al., 2009c).
4. Activities; e.g., office work (Janssen et al., 2009), driving a car (Healey and Picard, 2005) or flying a plane, and running (Healey, 2009).

This subdivision is based current practice with ASP; however, possibly it should be altered.

So far, comparisons between research results on ASP are mostly made between results of either individuals or groups selected to resemble the general population; cf. Table 1. However, user-tailored approaches should be explored as well. In particular, experiences with specific groups can substantially contribute to the further development of ASP, as has been seen in other sciences; e.g., biology, psychology, and medicine. But also, individual biosignal response patterns should be taken into consideration, since people’s affective signals differ widely.

Having said that, the question remains, how to handle this striking variety between people. We present three approaches, which are on another level than usually adopted, but can tackle these problems:

Hybrid classification systems (Berzal et al., 2004).
Most often, such architectures incorporate both a (logic-based) reasoning system and a machine learning component. To the authors knowledge, so far, this approach has not been applied for ASP. It has, however, been applied successfully for speech-based emotion recognition (Schuller et al., 2004).

Multi-agent systems and multi-classifier systems.
Two approaches within this field could be of interest: 1) Multi-layered architectures, where each layer determines the possible classes to be processed or the classifiers to be chosen for the next layer and 2) An ensemble of classifiers, trained on the same or distinct biosignals and their features. Their outputs are collected into one compound classification, often determined through a voting scheme. For more information on this topic, we refer to Lam and Suen (1995) and Kuncheva (2002).

Biosignal signatures. Related to schemes that are used in forensics (Rogers, 2003), ASP could benefit from personalized profiles or schemes that tailor to a generic profile to people’s unique biosignal signatures. Moreover, this approach could be extended to incorporate context information, as is already done in forensics (Rogers, 2003). Biosignal signatures require advanced multi-modal data mining and knowledge discovery strategies, and is related to the baseline matrix as proposed by Picard et al. (2001).

Each of these approaches enable processing of multi-modal data, which allows to incorporate a range of characteristics. This makes them promising for ASP applications, also outside the scope of user identification.

3.2 Theoretical specification

Changes in biosignals relate to changes in many psychological constructs Cacioppo and Tassinary (1990). For ASP, it is important to distinguish between these different psychological constructs. This involves two different situations: Firstly, biosignals can have more or less equal response patterns but in different time frames; e.g., short time frames for emotions and longer time frames for moods. Second, biosignals can have the same response patterns in the same time frames but still relate to different psychological constructs; an increase in inspiration rate can imply increased positive moods but also increased task demands or mental effort. We propose three ways of dealing with this complexity: (1) specification of the relation between construct of interest and biosignal, (2) involving context information, and (3) using multiple classifier systems.

In the first place, a thorough description of the relation between the construct of interest and the biosignals is necessary. By doing this, distinguishing biosignal properties for the construct of interest can be found. For instance, when classifying emotions short-term changes are of interest, whereas when classifying moods only long-term changes are relevant. In addition, when trying to distinguish workload from mood, one should not be interested in skin conductance changes. Instead, one could look at heart rate variability as this typically reacts stronger to workload than mood.

Second, context provides a lot of information on the psychological constructs which might have been changed. For example, while driving a car workload is changing quickly depending on the road situation while your mood is likely to remain equal. On the other hand, during watching a television show changes in biosignals are more likely to come from affect induction than from changes in cognitive load, motivation, or memory. By inserting context information, e.g. captured by a camera, the probability that changes will occur in specific psychological constructs can be modeled into the system. Thereby, increasing the change of allocating changes in biosignals to changes in the correct psychological construct.

Finally, one can use multiple classifiers to make a classification of all separate psychological constructs. In turn, the construct that with the most certain classification can be selected as the influenced construct. Moreover, an extra classifier can be trained that receives it’s input from the separate classifiers and
makes a selection based on this information.

To conclude, these three ways can help to deal with the problem of the many-to-many relationship between psychology and physiology. Note that we have assumed that the psychological constructs are independent of each other, which is actually not the case. Nonetheless, treating them as if being independent of each other is necessary for practical purposes.

4 CONCLUSION

This paper provided the second part of prerequisites for ASP. For the first and third part, we refer to respectively van den Broek et al. (2009a) and van den Broek et al. (2010). Here, the prerequisites identification of users and theoretical specification are introduced. These prerequisites are complementary to those presented in the other two papers: validity, triangulation, the physiology-driven approach, and contributions of signal processing (van den Broek et al., 2009a) and physical characteristics and integration of biosignals (van den Broek et al., 2010). Moreover, the second part of a review on ASP has been presented; see Table 1, complementary to the review table presented in van den Broek et al. (2009a).

The review (see Table 1) and the prerequisites, illustrate the complexity and lack of success of AC. This urges us to emphasize that a step back should be made by looking at prerequisites for successful ASP to achieve true progress on a later stage, instead of running forward and ignoring the problems encountered in previous studies. We sincerely hope that the prerequisites can contribute to or even guide the promising future ASP provides us with.

ACKNOWLEDGMENTS

The authors would like to thank Joyce H.D.M. Westerink (Philips Research, Eindhoven, The Netherlands) for her comments on an earlier version of this paper. Furthermore, we would like to thank the anonymous reviewers, who provided us the opportunity to improve this paper.

REFERENCES


