

Unobtrusive Sensing of Emotions (USE)

Egon L. van den Broek^{a,*}, Marleen H. Schut^b, Joyce H. D. M. Westerink^c and Kees Tuinenbreijer^b

^a *Center for Telematics and Information Technology (CTIT), University of Twente, P.O. Box 217, 7500 AE, Enschede, The Netherlands*

^b *Philips Consumer Lifestyle Advanced Technology, High Tech Campus 37, 5656 AE, Eindhoven, The Netherlands*
E-mail: {marleen.schut,kees.tuinenbreijer}@philips.com

^c *User Experience Group, Philips Research, High Tech Campus 34, 5656 AE, Eindhoven, The Netherlands*
E-mail: joyce.westerink@philips.com

Abstract. Emotions are acknowledged as a crucial element for artificial intelligence; this is, as is illustrated, no different for Ambient Intelligence (AmI). Unobtrusive Sensing of Emotions (USE) is introduced to enrich AmI with empathic abilities. USE coins the combination of speech and the electrocardiogram (ECG) as a powerful and unique combination to unravel people's emotions. In a controlled study, 40 people watched film scenes, in either an office or a home-like setting. It is shown that, when people's gender is taken into account, both heart rate variability (derived from the ECG) and the standard deviation of the fundamental frequency of speech indicate people's experienced valence and arousal, in parallel. As such, both measures validate each other. Thus, through USE reliable cues can be derived that indicate the emotional state of people, in particular when also people's environment is taken into account. Since all this is crucial for both AI and true AmI, this study provides a first significant leap forward in making AmI a success.

Keywords: Ambient Intelligence (AmI), emotion, unobtrusive sensing, speech features, heart rate variability

When dealing with people, let us remember that we are not dealing with creatures of logic; we are dealing with creatures of emotion ... (Carnegie, 1936; p. 41)

Dale Carnegie (1888–1955)[†]

1. Introduction

On behalf of the EU's IST Advisory Group, Ducatel, Bogdanowicz, Scapolo, Leijten, and Burgelman [12] described "Scenarios for Ambient Intelligence in 2010". Two of their key notions will be assessed in this paper: emotion and unobtrusive measurements. Hereby, the lessons learned in Artificial Intelligence (AI), Cybernetics, psychophysiology, and other disciplines will be taken into account.

* Corresponding author. E-mail: vandenbroek@acm.org.

[†] For the interested reader, we refer to "Historical foundations of social effectiveness? Dale Carnegie's principles" [13], which illustrates the timeless significance of Carnegie's work.

AI pioneer Herbert A. Simon [42] was the first to denote the importance of emotion for AI. Minsky [32] confirmed this by stating:

The question is not whether intelligent machines can have emotions, but whether machines can be intelligent without emotions. (p. 163)

Nevertheless, in practice emotions were mostly ignored in the quest towards intelligent machines until Picard [36] introduced the field "affective computing". Since then, the importance of emotion for AI slowly became acknowledged; e.g., [33].

We stress that emotions are not only of crucial importance for true AI but are, at least, as important for Ambient Intelligence (AmI). This was already acknowledged by Emile Aarts [1]:

Ubiquitous-computing environments should exhibit some form of emotion to make them truly intelligent. To this end, the system's self-adaptive capabilities should detect user moods and react accordingly. (p. 14)

This paper describes the quest toward Unobtrusive Sensing of Emotions (USE) for AmI. The research respects the complexity of emotions as well as the current limitations of unobtrusive physiological measurements. Nevertheless, with the exploration of the speech and electrocardiogram (ECG) signals, we coin a unique combination to enable USE. We expect that features from these two signals hold promising features for unraveling people's emotional state.

First, we will introduce the constructs emotion (Section 2) and USE (Section 3) with its two physiological signals: speech and the ECG. Next, in Section 4, a study will be described, which explores the feasibility of using these two signals for USE. Last, in Section 5, the implications of this research for AmI will be discussed and future directives will be provided.

2. Emotion

A lengthy debate on the topic of emotion would be justified; however, this falls beyond the scope of the current paper. Hence, no overview of the various emotion theories and the levels on which emotions can be described will be provided. Instead, a thoroughly composed definition will be used as a starting point. In addition, the model for emotion applied in the research will be introduced.

Kleinginna and Kleinginna [22] compiled a list of more than 100 definitions of emotion. Regrettably, they had to conclude that *psychologists cannot agree on many distinguishing characteristics of emotions*. Therefore, they proposed a working definition: *Emotion is a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems, which can (a) give rise to affective experiences such as feelings of arousal and pleasure / displeasure; (b) generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labeling processes; (c) activate widespread physiological adjustments to the arousing conditions; and (d) lead to behavior that is often, but not always, expressive, goal directed, and adaptive*. In the current paper, we adopt this definition as working definition.

Kleinginna and Kleinginna [22] also address the influence of emotions on people's cognitive processes: issues (b) and (d). Hence, emotions by themselves should be taken into account; but, also their effect on cognitive processes (e.g., attention, visual perception, and memory) and, thereby, our functioning. This emphasizes the importance of taking emotions into

account in AmI. Moreover, Kleinginna and Kleinginna [22] address the influence of emotions on our physiology. This research exploits this through measuring unobtrusive physiological signals to unravel people's emotional state.

In line with the frequently adopted circumplex or valence-arousal model of emotions [23,31,37], the definition of Kleinginna and Kleinginna [22] distinguishes arousal and valence (i.e., pleasure / displeasure). The valence-arousal model denotes valence and arousal as two independent bipolar factors that describe emotions.

Although the valence-arousal model is successful, it suffers from two severe limitations. First, no emotions are identified with high scores, either positive or negative, on both the valence and the arousal axis [23]. Second, the model cannot handle mixed emotions; i.e., parallel experience of both positive and negative valence [6,48].

To enable the identification of mixed emotions and provide a suitable processing scheme, the valence-arousal model is sometimes extended; e.g., [6,48]. Such an extended valence-arousal model incorporates, instead of one bipolar valence dimension, two unipolar valence dimensions: one for positive and one for negative valence. Hence, the extended valence-arousal model incorporates three dimensions, instead of two. This approach is also adopted for the current research.

3. Unobtrusive Sensing of Emotions (USE)

Emotions contain a core affective state that can be defined as the simplest raw feeling that is consciously accessible; e.g., joy, sadness, frustration [40]. Such a core affective state is accompanied by both behavioral and physiological changes [16,22,31,45].

People's emotional state can be assessed by processing a range of their biosignals. When reviewing literature, it becomes apparent that these signals can be assigned to two groups:

1. A broad range of physiological measures signals. For recent overviews, we refer to [7,15,31,45].
2. Specialized areas of signal processing:
 - (a) speech processing [10,41,47,50]
 - (b) movement analysis [11,19]
 - (c) computer vision techniques [10,19,49,50]

These distinct measurement methods are seldomly combined; where, on the one hand, several physio-

logical measures are frequently combined and, on the other hand, speech processing, movement analysis, and computer vision are frequently combined. A recent study of Bailenson et al. [3] is an exception on this. They combined both computer vision and physiological measures. Their study illustrates the usefulness of this approach, providing better and more robust results.

Physiological measures are often obtrusive and, hence, disregarded for user-centered applications, as AmI is. However, wearable computing and wireless sensing technologies relieve this problem [17,20,26,27]. In contrast, speech and computer vision are unobtrusive but very noise sensitive. The audio recordings used for speech processing suffer from various types of noise. However, with no need for speech recognition, the remaining problem is binary: a speech signal or no speech signal, which makes it feasible. Computer vision techniques, although appealing, are only usable for emotion recognition in very stable environments; e.g., without occlusion, stable light sources, and the users sitting at a desk or in a couch [49].

Speech and physiological measures, in particular the ECG, are not yet combined to access the emotional state of users, although especially their combination is promising. A possible explanation is the lack of knowledge that exists on the application of this combination of measures for emotion measurement; cf. [10,11,19,41,47,50] and [7,15,31,45].

From features of both the speech and the ECG signal, we expect to extract cues on people's experienced valence and arousal. Since this study is (one of) the first to employ the combination of speech and ECG, we chose for a controlled study to assess their feasibility for AmI purposes. Before the study is described, each of the signals used are introduced.

3.1. The speech signal

Speech processing, speech dialogue, and speech synthesis can exhibit some form of intelligent, user-perceived behavior and, hence, are useful in designing AmI environments [1]. However, speech comprises another feature: emotion elicitation; e.g., [10,41,44,47].

The human speech signal can be characterized by various features and their accompanying parameters. However, no consensus exists on the features and parameters of speech that reflect the emotional state of the speaker. Most evidence exists for the variability (e.g., standard deviation; SD) of the fundamental fre-

quency (F0), energy of speech, and intensity of air pressure [10,41,44,47].

3.2. Electrocardiogram

The electrocardiogram (ECG) is an autonomic signal that cannot be controlled easily, as is the case with electrodermal activity. ECG can be measured directly from the chest. Alternatively, the periodic component of the blood flow in the finger or in an ear can be translated into the ECG. From the ECG, the heart rate (HR) can nowadays be easily obtained; e.g., [17]. Research identified features of HR as indicators for both experienced valence and arousal [2,8,31,34].

In addition to the HR, also the HR variability (HRV) can be determined from the ECG. The HRV is a frequently used variable in psychophysiological research; e.g., [28]. HRV decreases with an increase in mental effort, stress, and frustration [7,20,21,31]. Moreover, some indications have been found that HRV is also influenced by the valence of an event, object, or action [2,34,38,39].

4. Validation of USE

In this section, we introduce a controlled study to determine the feasibility of USE through speech and ECG. The scheme presented in Fig. 1 provides an overview of all information sources obtained throughout the validation of USE. Moreover, the scheme presents how these sources have been processed, as is also depicted in the forthcoming sections.

4.1. Method

4.1.1. Participants and design

40 volunteers (20 male, 20 female; average age 27; SD: 7.6) participated. None of them had hearing impairments or any known cardiovascular problems. All had normal or corrected to normal vision. The participants were ignorant to our research goals. All participants signed an informed consent.

The participants were divided into two groups of 20 each. One group of participants was assigned to an office environment, in which they took place in an office chair. The other group of participants was assigned to a living room environment, in which they sat on a couch. At both locations, the room was silent and darkened and a screen was placed in front of the participants.

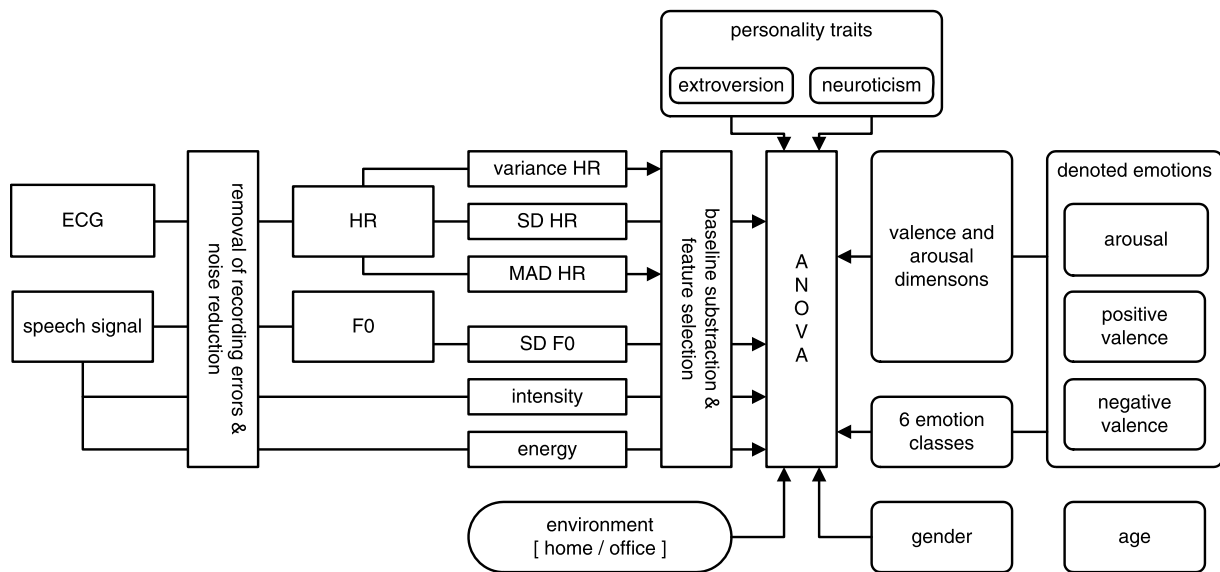


Fig. 1. The processing scheme of Unobtrusive Sensing of Emotions (USE). It shows how the physiological signals (i.e., speech and the ECG), the emotions as denoted by people, personality traits, people's gender, and the environment are all combined in one ANOVA. Age was determined but not processed. Note that the ANOVA can also be replaced by a classifier or an agent, as a module of an AmI; e.g., [46].

Explanation of the abbreviations: ECG: electrocardiogram; HR: heart rate; F0: fundamental frequency of pitch; SD: standard deviation; MAD: mean absolute deviation; and ANOVA: ANalysis Of VAriance.

4.1.2. Materials

To elicit an emotional response, the participants watched six scenes, adopted from [48]. The film scenes were presented on a 15.4 inch screen (1280×800 pixels, 60 Hz refresh rate; video card: ATI MOBILITY RADEON 9700). The films were presented in a random order.

During the study, speech utterances were recorded continuously by means of a Trust Multi Function Headset with microphone. The recording was performed in SoundForge 4.5.278 (sample rate: 44.100 Hz; sample size: 16 bit). Parallel with the speech recording, a continuous recording of the ECG was done through a modified Polar ECG measurement belt, which was connected to a data acquisition tool (NI USB-6008). Its output was recorded in a LabVIEW 7.1 program, with a sample rate of 200 Hz.

To be able to investigate possible influences of personality characteristics on the experienced emotions, all participants were asked to fill in a revised, short scale of the Eysenck Personality Questionnaire (EPQ-RSS) [14]. This questionnaire determined participants' personality traits extroversion and neuroticism.

4.1.3. Procedure

After instructions, the participants signed an informed consent, and the ECG measurement belt and

headset were positioned. Next, the participant read aloud a non-emotional story to a) verify whether or not the participant had understood the instructions, b) to test the equipment, and c) to determine their personal baseline for both the speech and the ECG signal.

Each of the six film scenes that were shown had a duration of 3 minutes and 18 seconds. After each scene, the participants had 30 seconds to describe the most emotional part of the scene, followed by a resting period of 60 seconds. During these 90 seconds (speaking and resting), a gray screen was shown. The experiment started and finished with displaying a gray screen during 90 seconds.

After the film scenes were shown, the participants rated them. This was done using 11 point Likert scales, ranging from 0 to 10. Since its introduction in 1932 [25], Likert scales have become a standard method for assessing people's subjective attitudes.

The Likert scales were embedded in a Digital Rating System (DRS). With each film scene, the DRS presented three Likert scales, one for each of the three dimensions positive affect, negative affect, and arousal; see also Section 2. The DRS displayed pictures of the film scenes in random order together with the Likert scales to jog the participants' memories.

4.2. Noise reduction

The reduction of noise consisted of two phases. First, recording errors were removed. Second, for both the speech and the ECG signal, preprocessing and noise reduction were applied.

4.2.1. Recording errors

The speech signal of two participants was not recorded due to technical problems. Of two other participants, the speech signal was too noisy. Of these four participants, the speech signals were excluded from further analyses.

With nine participants, either a significant amount of noise was present in their ECG or the signal was even completely absent. The ECG signals of these participants were omitted from further processing.

4.2.2. Speech signal

Some preprocessing of the speech signal was required before the features could be extracted from the signal. We started with the segmentation of the recorded speech signal in such a way that for each film scene separately, its speech signal was determined.

After the segmentation of the speech signal, the actual noise reduction was applied. The speech signal was noisy, due to technical inconveniences; e.g., the microphone placed too close to the mouth and, consequently, breathing is recorded. Moreover, noise due to participant's behavior (e.g., tapping on the table, coughing, scraping the throat, yawning) and speaking (e.g., silences, saying 'eh') needed to be removed.

Noise was removed from the speech signals in two subsequent sessions: 1) the silences were removed and 2) utterances such as 'eh' and noise due to participant's behavior were removed. This resulted in a 'clean' signal, as is also illustrated in Figs 2(a) and 2(b).

4.2.3. ECG

The output of the ECG measurement belt has a constant (baseline) value during the pause between two heart beats. Each new heart beat is characterized by a steep slope upwards, within one sample. To be more specific, a heart beat is characterized by a R-wave, which is an upward deflection, following the Q-wave, which is a downward deflection of the ECG arising from ventricular activation; see also Fig. 3. The vertical lines in this figure point out the R-waves. The HR is calculated from the intervals between the R-waves (R-R intervals) [28].

The measurement belt for the ECG signal appeared to be sensitive to movements of the participant. This

resulted in four types of noise that can be distinguished: 1) a heart beat that differs from the normal PQRS shape, see Fig. 3; 2) heart beats that succeeded too quickly; 3) missing heart beats in a sequence; and 4) no HR signal at all. The ECG signal was checked for all these types of noise and corrected where necessary.

4.3. Data reduction

This section describes how the questionnaires were processed. This includes both the personality questionnaires that were completed by the participants and the experienced emotions (i.e., subjective measurements). Additionally, the data reduction for both the speech signal and the ECG are described. See also Fig. 1 for an overview.

4.3.1. Personality questionnaires

To determine participants' personality traits extroversion and neuroticism the revised, short scale of the Eysenck Personality Questionnaire (EPQ-RSS) [14] were processed. This resulted in two binary indices for participants' personality traits.

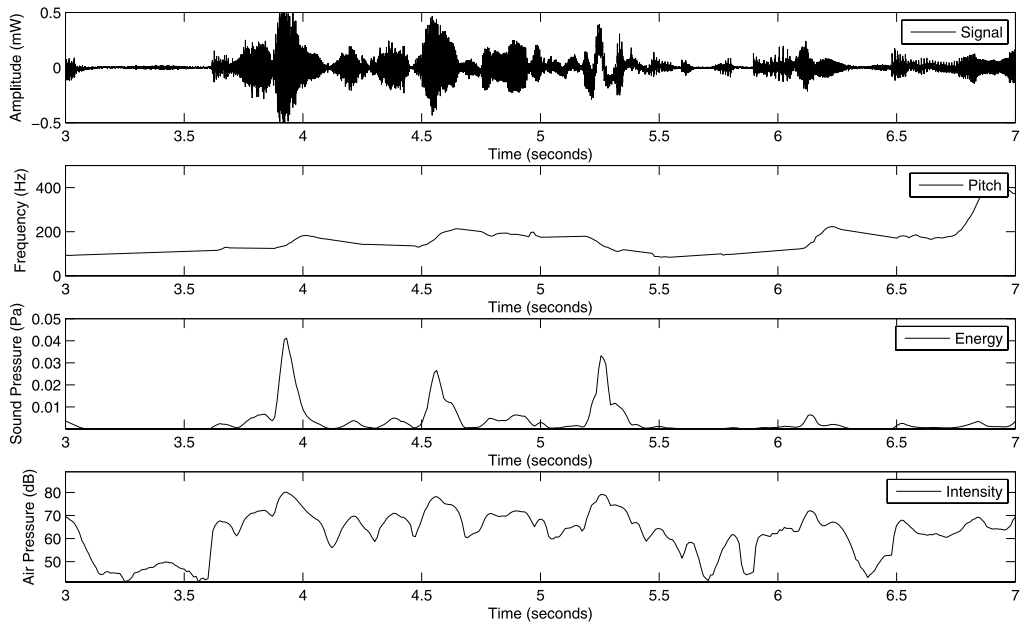
The two binary indices enabled us to denote all participants as either extrovert or introvert. In addition, through the second personality trait, all participants were categorized as being either neurotic or not neurotic.

4.3.2. Subjective measurements

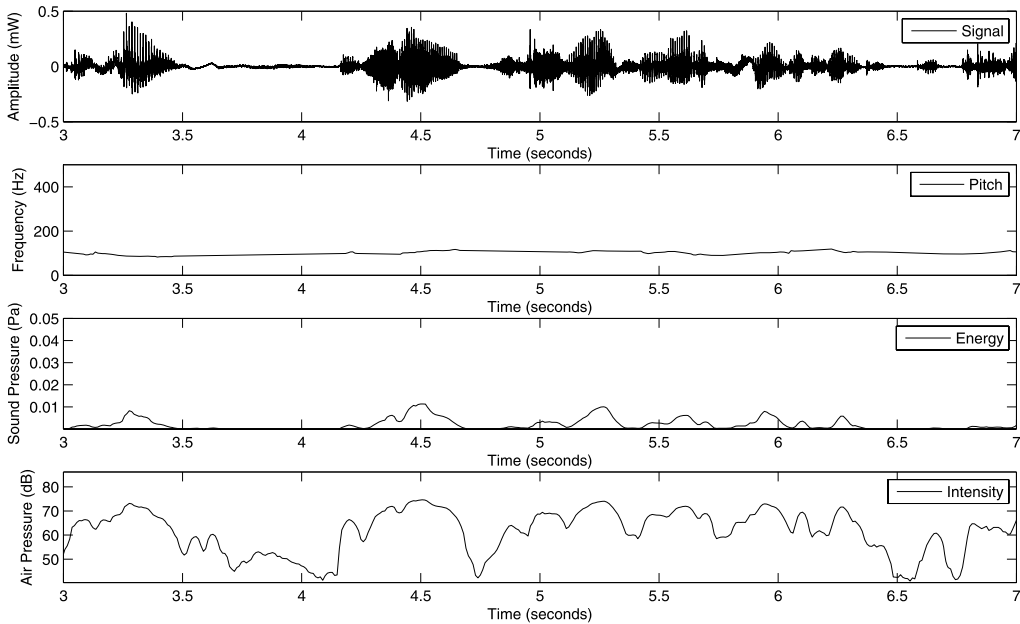
The ratings of the film scenes were provided by the participants after each of the experiments on three scales: positive valence, negative valence, and arousal, as denoted in Table 1. Combinations of these scales, allowed the creation of emotion categories, according to a valence-arousal model. See also Section 2.

For each film scene, the average ratings on each of the three scales over all participants were calculated. This resulted in a classification of the film scenes in two categories (i.e., high and low) for each of the three scales: positive, negative, and arousal. From these classifications, we derived three categories for valence: positive, negative and neutral. The category neutral denotes neither a positive valence nor a negative valence. In addition, two categories for arousal were derived: high arousal and low arousal. Together, these two categorized dimensions of the valence-arousal model depicted six emotion classes.

Each of the six emotion classes was represented in this research by one film fragment. The emotion



(a) A speech signal and its features of a person in a relaxed state.



(b) A speech signal and its features of a person in a sad, tensed state.

Fig. 2. Two samples of speech signals of the same person (an adult man) and their accompanying extracted fundamental frequencies of pitch (F0) (Hz), energy of speech (Pa), and intensity of air pressure (dB). In both cases, energy and intensity of speech show a similar behavior. The difference in variability of F0 between (a) and (b) indicates the difference in experienced emotions.

classes with the values on the three dimensions, their categorization in the valence and arousal categories, and their accompanying film fragment are denoted in Table 1.

4.3.3. Speech signal

Of each participant, the sound recorded during the study lasts approximately 25 minutes; however, only the parts in which the participants spoke are of inter-

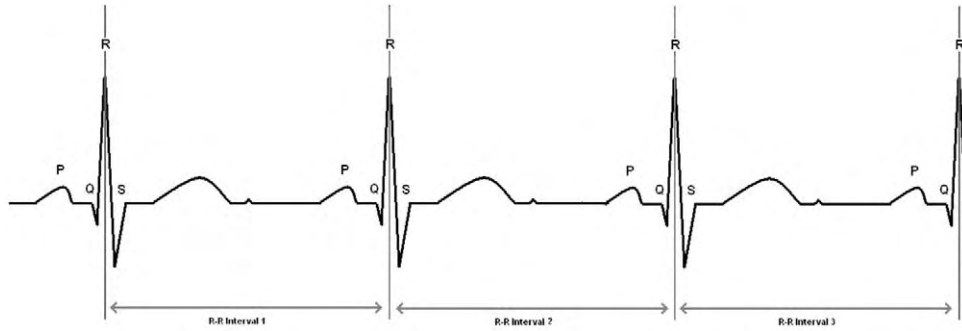


Fig. 3. A schematic representation of an electrocardiogram (ECG) denoting four R-waves, from which three R-R intervals can be determined. Subsequently, the heart rate and its variance (denoted as standard deviation (SD), variability, or mean absolute deviation (MAD)) can be determined.

est. Those parts in which the participants did not speak were removed from the sound signal.

In order to cope with interpersonal differences in speech, all data was normalized by subtracting a baseline from the original signal. Subsequently, the speech processing environment Praat 4.0.4 [5] was used to extract the required features: i.e., SD F0, energy of speech, and the intensity of air pressure (see also Figs 2(a) and 2(b)).

4.3.4. Electrocardiogram

The ECG signal was segmented into separate signals per stimulus, before it was processed. Next, the heart beats were identified; see also Fig. 3. This enabled the extraction of the features.

4.4. Feature extraction

From both the speech signal and the ECG signal a large number of features could be derived; e.g., see [10,41,47] and [2,28,45]. This research did, however, not aim to provide an extensive comparison of speech and ECG features. Instead, the use of the combination of these two signals was explored. Therefore, a limited set of features was extracted from both signals, as will be defined in the next subsections. See also Fig. 1.

4.4.1. Speech signal

In a variety of settings, several parameters derived from speech are investigated with respect to their use in the determination of the emotional state of people. Although no general consensus exists concerning the parameters to be used, much evidence exists for the SD F0 [10,41,44,47], the Energy of speech, and the Intensity of air pressure [41]; cf. Figs 2(a) and 2(b). They are useful for measuring experienced emotions.

For a domain $[0, T]$, the energy of speech is defined as:

$$\frac{1}{T} \int_0^T x^2(t) dt, \quad (1)$$

where $x(t)$ is the amplitude or sound pressure of the signal in Pa (Pascal) [5]. The following equation is its discrete equivalent:

$$\frac{1}{N} \sum_{i=0}^{N-1} x^2(t_i), \quad (2)$$

where N is the number of samples.

For a domain $[0, T]$, the intensity of air pressure in the speech signal is defined as:

$$10 \log_{10} \frac{1}{T P_0^2} \int_0^T x^2(t) dt, \quad (3)$$

where $P_0 = 2 \cdot 10^{-5}$ Pa is the auditory threshold [5]. The Intensity is computed over the discrete signal in the following manner:

$$10 \log_{10} \frac{1}{N P_0^2} \sum_{i=0}^{N-1} x^2(t_i). \quad (4)$$

It is expressed in dB (decibels) relative to the auditory threshold P_0 .

Both the Intensity and the Energy of speech are directly calculated over the clean speech signal. To determine the F0 of pitch from the clean speech signal, a fast Fourier transform has to be applied over the signal. Subsequently, its SD is calculated; see also Eq. (5). For a more detailed description of the processing scheme, we refer to [4].

Table 1

The six film scenes with the average ratings given by the participants on the positive valence, negative valence, and arousal Likert scales. From the positive and negative valence ratings, three valence categories are derived: neutral, positive, and negative. Using the scores on arousal, two arousal categories are determined: low and high

Film scene	Valence			Arousal	
	Positive	Negative	Category	Score	Category
Color bars	0.13	2.51	neutral	0.49	low
Final Destination	2.59	4.38	neutral	6.54	high
The bear	5.79	0.74	positive	3.49	low
Tarzan	7.31	0.26	positive	4.77	high
Pink flamingos	0.49	7.18	negative	6.00	low [‡]
Cry freedom	0.56	7.90	negative	7.69	high
Average	2.81	3.83		4.83	

[‡] This score is higher than average. Nevertheless, it is categorized as low. This is done for two reasons: 1) The experienced arousal is low relative to the other film scene with which a negative valence was experienced and 2) This categorization facilitated a balanced design, which enabled the preferred statistical analyses.

4.4.2. ECG signal: Heart rate variability

From the ECG signal, the intervals between the R-waves (R-R intervals) were determined; see also Fig. 3. Subsequently, the mean R-R interval was determined.

In literature, usually, the variability of a data set (e.g., a signal) is defined by the SD, variance, or the mean absolute deviation (MAD). To be able to determine the variability of the heart rate (HRV) from an ECG, the intervals between the R-waves (R-R intervals) of the ECG need to be identified. Two methods were applied for the calculation of the HRV, defined as follows:

The variance of the R-R intervals:

$$\sigma^2 = \frac{1}{\mathcal{R}} \sum_{i=0}^{\mathcal{R}-1} (\Delta_i - \bar{\Delta})^2, \quad (5)$$

with the SD of the R-R intervals is defined as its square root: σ . Δ_i denotes an R-R interval, $\bar{\Delta}$ denotes the average R-R interval, and \mathcal{R} denotes the number of R-R intervals.

The MAD of the R-R intervals:

$$\text{MAD} = \frac{1}{\mathcal{R}} \sum_{i=0}^{\mathcal{R}-1} |\Delta_i - \bar{\Delta}|. \quad (6)$$

Please note that various other measures are applied for the determination of the HRV. For more discussion on this topic and an extensive review, we refer to Chapter 3 of [28]. However, with these three measures we expected to have a good indication of the use of HRV for emotion detection.

4.5. Considerations with the analyses

As denoted in the Section 4.2, 13 corruptions of signals were detected of in total 11 participants. The recordings of two of these participants suffered from two types of noise. Through interpolation, corrections could have been made for the absence of this data. However, this would have decreased the reliability of the analyses done. Therefore, we chose to omit all data of participants of which problems were encountered with the recordings. This resulted in data of 29 participants that could be analyzed. Note that this has the disadvantage that the chance on finding significant results declined substantially.

Preliminary analyses of the ECG signal, using the two methods to determine the HRV, showed that the SD, the variance, and MAD (see Eqs (5) and (6)) provided similar results. However, the SD of the R-R intervals showed to be the most sensitive measure for HRV. Therefore, in the main analyses, variance and MAD of the R-R intervals as measures for HRV were excluded; see also Fig. 1. So, only the SD of the R-R intervals was processed with the forthcoming analyses. From this point on, the SD of the R-R intervals will be simply denoted as the measure for HRV.

From the speech signal, all three features, as described in the previous section, were processed. See Eqs (1)–(4) and [4] for their definitions.

All data was analyzed through two RM ANOVA, with four measures: HRV determined from the ECG signal and the SD F0, intensity, and energy of the speech signal, as is also denoted in Fig. 1. In line with

the adopted model of emotion (see Section 2), the first set of analyses adopted the two dimensions of emotion that were defined: valence (positive, negative, neutral) and arousal (high/low). These two dimensions served as within subject factors in this set of analyses. In a second set of analyses, the six emotion classes, as denoted in Table 1, were analyzed separately.

Four between subject factors were included in the analyses: the environment (office/living room), gender (male/female), and the two personality traits extroversion and neuroticism. From preliminary analyses appeared that age was of no influence on any of the measures. Therefore, age was omitted from further analyses. See also Fig. 1.

For both sets of analyses, the multivariate test will be reported first, including all four measures. Next, for each measure the univariate tests will be reported. With all analyses, the interaction effects will be reported.

4.6. Results on experienced valence and arousal

4.6.1. Multivariate analysis

No main effect of arousal on the physiological parameters/measures was found. However, in interaction with gender, arousal showed an effect on the measures, $F(4,15) = 4.999$, $p = .009$. Also in interaction with the environment, arousal showed an effect on the measures, $F(4,15) = 3.509$, $p = .033$.

A main effect of valence on the physiological parameters/measures was found, $F(8,66) = 4.490$, $p < .001$. Moreover, in interaction with both gender ($F(8,66) = 2.850$, $p = .009$) and the environment ($F(8,66) = 2.622$, $p = .015$) valence showed an effect on the measures.

A main interaction effect of arousal and valence on the physiological parameters/measures was determined, $F(8,66) = 6.192$, $p < .001$. Also, the interaction of arousal and valence with both gender ($F(8,66) = 2.081$, $p = .050$) and the environment ($F(8,66) = 2.524$, $p = .018$) showed an effect on the measures. In addition, the four-way interaction between arousal, valence, gender, and the environment showed an effect on the measures, $F(8,66) = 3.365$, $p = .003$. No interaction effects with the personality traits were shown.

4.6.2. Univariate analyses

No main effect of arousal on any of the physiological parameters/measures was found. However, in interaction with gender, arousal did show an effect on both HRV ($F(1,18) = 7.813$, $p = .012$) and

SD F0 ($F(1,18) = 12.863$, $p = .002$). Also, in interaction with the environment, arousal showed an effect on HRV, $F(1,18) = 16.318$, $p = .001$. Moreover, a three-way effect of arousal, gender, and the personality trait extroversion was determined on HRV, $F(1,18) = 8.700$, $p = .009$. No interaction effects with the personality trait neuroticism were found.

A main effect of valence on HRV was identified, $F(2,36) = 24.937$, $p < .001$. Moreover, in interaction with gender, an effect of valence on both HRV ($F(2,36) = 4.828$, $p = .014$) and SD F0 ($F(2,36) = 8.186$, $p = .001$) was detected. Also, in interaction with the environment, valence showed an effect on HRV, $F(2,36) = 10.307$, $p < .001$. Moreover, three three-way interaction effects were found. The interaction between valence, gender, and the environment showed an effect on the intensity of speech, $F(2,36) = 4.831$, $p = .014$. The interaction between valence, gender, and the personality trait extroversion showed an effect on SD F0, $F(2,36) = 7.435$, $p = .002$. The interaction between valence, location, and the personality trait neuroticism showed an effect on the intensity of speech, $F(2,36) = 5.036$, $p = .012$.

A main interaction effect of arousal and valence on HRV was found, $F(2,36) = 29.089$, $p < .001$. Moreover, three three-way interaction effects were determined. The interaction between arousal, valence, and gender showed an effect on the intensity of speech, $F(2,36) = 4.265$, $p = .022$. The interaction between arousal, valence, and the environment showed an effect on HRV, $F(2,36) = 10.135$, $p < .001$. The interaction between arousal, valence, and the personality trait neuroticism showed an effect on HRV, $F(2,36) = 3.694$, $p = .035$. Moreover, the four-way interaction effect of arousal, valence, gender, and the environment on HRV was detected, $F(2,36) = 15.041$, $p < .001$.

In none of the analyses, effects of either arousal or valence on the energy of speech were found.

4.7. Results on the six emotion classes

4.7.1. Multivariate analysis

The multivariate analysis showed a strong effect for the emotion classes on the set of physiological parameters/measures, $F(20,342) = 6.111$, $p < .001$. In addition, in interaction with both gender ($F(20,342) = 2.872$, $p < .001$) and environment ($F(20,342) = 2.898$, $p < .001$), an effect of the emotion classes on the measures was found. In line with these interaction effects, a three-way interaction effect between the emotion classes, gender, and the environment was

found on the measures, $F(20,342) = 2.514$, $p < .001$. No interaction effects with the personality traits were found.

4.7.2. Univariate analyses

A strong main effect was found for the emotion classes on HRV, $F(5,90) = 23.772$, $p < .001$. An interaction effect of the emotion classes and both gender ($F(5,90) = 4.128$, $p = .002$) and environment ($F(5,90) = 10.966$, $p < .001$) on HRV was found. In line with the two-way interaction effects on HRV, a three-way interaction effect on HRV between the emotion classes, gender, and environment was found, $F(5,90) = 7.456$, $p < .001$.

A strong interaction effect between the emotion classes and gender on SD F0 was determined, $F(5,90) = 5.501$, $p < .001$. In addition, a three-way interaction effect on SD F0 between the emotion classes, gender, and the personality trait extroversion was identified, $F(5,90) = 3.918$, $p = .003$.

No effects of the emotion classes were found on either the intensity of speech or the energy of speech. Moreover, no interaction effects with the personality trait neuroticism were detected.

4.8. Discussion

In line with the results section, we discuss both the analyses of experienced arousal and valence and the analyses of the six emotion classes separately. Subsequently, we will describe the relations between emotions and the measures. Finally, we relate both sets of analyses to each other and draw conclusions from them.

4.8.1. Experienced valence and arousal

When gender is taken into account, the experienced arousal is clearly reflected in both HRV and SD F0. The effect on HRV is also influenced by both the environment and the personality trait extroversion. No effect of arousal is found on the two other speech parameters: intensity and energy of speech.

For both the F0 of speech and HRV it is known that male and females have different characteristics. Hence, an influence of gender was expected and will always be of importance. Moreover, the environment has to be taken into account. The difference between the environments assessed in this research was limited; hence, in practice this effect could be more substantial. Further, it is noteworthy that personality traits have shown to be of limited or no influence.

The experienced valence influences both HRV and speech parameters. When gender is taken into account, the experienced valence is clearly reflected in both HRV and SD F0. Moreover, indications were found for the influence of valence on the intensity of speech. However, this needs to be further investigated, before firm conclusions can be drawn. The speech parameter energy was not sensitive for experienced valence. Further, it has to be noted that personality traits showed to have a limited influence on these effects.

4.8.2. The six emotion classes

Through HRV, the six emotion classes could be reliably distinguished. However, both gender and environment influence this effect. Personality traits did not influence this effect.

In interaction with gender, SD F0 also showed to be a good discriminator among the six emotion classes. The personality trait extroversion influenced this effect. The personality trait neuroticism was not of any influence. The speech parameters intensity and energy of speech did not discriminate among the emotion classes.

4.8.3. Relations between measures and emotions

The factors valence and arousal heavily influenced each other. Moreover, various other factors had their influence as well. Consequently, it was hard to relate the behavior of the recorded measures to both dimensions of emotion. Nevertheless, three general relations between valence and arousal and the physiological parameters were observed. These observations also partly explain the results found with the analyses.

A positive valence was accompanied with a HRV that is higher than it was with a neutral or negative valence. With a neutral or negative valence, high arousal is reflected through both a low HRV and a low SD F0 of speech, as was also reported in [44]. Compared to low arousal, high arousal was accompanied with a higher intensity of the speech signal.

4.8.4. Conclusions

The film fragments were classified using three dimensions: arousal, positive valence, and negative valence. From this two characteristics were derived: arousal and valence. These were also applied in the first analysis. In a second analysis, these two dimensions were ignored and the emotions were treated as separate classes. Regrettably, emotion theory lacks true standards. This made it hard to determine what was the best approach. However, both analyses had

enough results in common to provide some general guidelines for follow-up research.

Both HRV and SD F0 of speech showed to be good discriminators between emotions, when the gender of the participants was taken into account. Hence, both measures can be validated through each other. However, it should be noted that the variety among emotions is rich and only six were assessed in the current research. Moreover, it is unknown how sensitive both measures are for emotion discrimination. Hence, further research is needed on this issue.

It should be noted that on both HRV and SD F0 of speech, the environment is of influence. This effect will probably be of more influence in real world settings, which are not as controlled as the current research was. Luckily, such ambient awareness is already among the challenges true AmI faces.

5. General discussion

Both the F0 of speech and the HRV can be considered as physiological parameters that can be determined indirectly or at least unobtrusively. This makes them par excellence suitable for AmI purposes. This study was the first that reports the use of both signals simultaneously to unravel user's emotional state. See Fig. 1 for an overview of USE's processing scheme.

The results of this study show that the combination of these measures provides a reliable, robust, and unobtrusive method to penetrate user's emotional state. Moreover, the signals validate each other. Both HRV and SD F0 seem to indicate influences of experienced valence and arousal in parallel.

That emotion is a crucial factor in making AmI a success is illustrated by the AI community. AI explains its lack of success in seeking true intelligence by their ignorance of the topic of emotions (e.g., [33,43]), where traditionally logic-based reasoning used to be dominant. In HCI and affiliated communities, emotion was accepted as crucial in understanding human behavior and, hence, in efficient interaction between man and machine already [24,43]. This is no different for ubiquitous computing, especially for AmI, which should be sensitive and responsive to people's presence [1,50].

The successful introduction of the paradigm USE can be considered as a first step to true AmI. Future developments should extent and strengthen the USE concept. We propose to take the three key elements of

AmI, concerning "the adjustment of electronic systems in response to users" [1], as a starting point:

1. Personalization: which refers to adjustments on a short time scale.
2. Adaptation: adjustments to changes in user behavior over longer periods of time.
3. Anticipation: system adjustments that differ over a very long period of time.

For the system's users, a similar distinction exists, also based on different time scales: emotions, moods, and personality. So far, we neglected this distinction and used a general definition. However, evidence for the need of a more hierarchical theory of emotions slowly begins to get shape; e.g., [16,18,40,41]. The three user characteristics are:

1. Emotion: a short reaction (i.e., a matter of seconds) to the perception of a specific (external or internal) event accompanied by mental, as well as behavioral and physiological changes [15,45].
2. Moods: are long lasting and change gradually (over the course of minutes or hours, or even longer), are experienced without concurrent awareness of their origin, and are not object-related. Moods do not directly affect actions, but do influence our behavior indirectly [15,18,45].
3. Personality: A person's set of distinctive traits and behavioral and emotional characteristics. A thorough overview on this topic is provided by Cooper and Pervin [9] that comprises reprints of various seminal articles, including various articles on Norman's five-factor model; e.g., Barrick and Mount (1991) and Goldberg (1993).

This triplet of user characteristics perfectly maps on the three key elements of AmI [1]. Hence, Aarts' key elements can be defined in terms of emotions or should at least take them into account.

How emotion should be described and modeled remains a topic of debate. In this paper, we have adopted the definition of Kleinginna and Kleinginna [22]. However, even in the same decade, various seminal works on emotion have been published; e.g., Frijda (1986) and Orotony, Clore, and Collins (1988). Both of these works included their own definition of emotion; e.g., Orotony, Clore, and Collins [35] defined emotions as: *valenced reactions to events, agents, or objects, with their particular nature being determined by the way in which the eliciting situation is construed* (Chapter 1, p. 13 and Chapter 5, p. 191). Since the 80s, a vast number of books, opinions, and research papers

have been published, illustrating the lack of a generally accepted, multi disciplinary theory on emotions. For a concise, more recent overview of the various theories on emotions, we refer to [37].

The debate on what emotions are is intriguing. However, more practical considerations should also be noted. For example, the use of wearable computing facilitates the communication between user and AmI. Since the early work of Steve Mann [30], wearable computing devices evolved rapidly. In the last years, various prototypes have been developed, which enable the recording of physiological signals; e.g., [17,26,27]. Adopting this paradigm, in addition to speech recordings and the ECG as measures, other signals can be applied to achieve an even higher probability of correct interpretation [3,15,45].

The F0 represents the frequency with which the vocal folds open and close or vibrate. As is shown, this information can be derived from the speech signal. However, a more direct and, thus, more robust method is to use electrodes attached to (or near) the throat at the level of the glottis. Through impedance variations, they can record vocal fold vibrations. This method was already used half a century ago (e.g., [29]); however, due to technical limitations it was not reliable. Nowadays, the technical problems are solved and sensors can be worn unobtrusively but, regrettably, this method seems to have been forgotten. Obtaining noise-free F0 of speech could be considered as an alternative for the more indirect speech processing.

Taking it all together, AmI, following AI, has to embrace emotion as an essential element in pursuing its intelligence. It is surprising that the combination of speech and ECG had not been used to unravel user's emotions before. Par excellence, these signals could be exploited in parallel for AmI purposes, as is illustrated through USE. Both SD F0 of speech and HRV parameters unravel users' emotion space. Moreover, various manners of implementation of the required sensors secure an unobtrusive recording of both signals. This having said, the current study provides a significant leap forward in making AmI a success.

Acknowledgments

The authors thank the two anonymous reviewers and the special issue editors, who provided valuable comments on this article. Frans van der Sluis (University of Twente, The Netherlands), Joris H. Janssen (Eindhoven University of Technology, The Netherlands /

Philips Research, The Netherlands), Leon van den Broek (Radboud University Nijmegen, The Netherlands / University of Utrecht, The Netherlands), and Winnie Teunissen are gratefully acknowledged for reviewing earlier drafts of this article. Moreover, we thank Frans van der Sluis for plotting Fig. 2.

References

- [1] E. Aarts, Ambient Intelligence: Vision of our future, *IEEE multimedia* **11** (2004), no. 1, 12–19.
- [2] B.M. Appelhans and L.J. Luecken, Heart Rate Variability as an Index of Regulated Emotional Responding, *Review of general psychology* **10** (2006), no. 3, 229–240.
- [3] J.N. Bailenson, E.D. Pontikakis, I.B. Mauss, J.J. Gross, M.E. Jabon, C.A. Hutcherson, C. Nass, and O. John, Real-Time Classification of Evoked Emotions using Facial Feature Tracking and Physiological Responses, *International journal of human-computer studies* **66** (2008), no. 5, 303–317.
- [4] P. Boersma, Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound, *Proceedings of the Institute of Phonetic Sciences*, vol. 17, University of Amsterdam, 1993, pp. 97–110.
- [5] P.P.G. Boersma and D.J.M. Weenink, Praat 4.0.4, 2006, URL: <http://www.praat.org> [Last accessed on April 12, 2009].
- [6] J.T. Cacioppo and G.G. Berntson, Relationship Between Attitudes and Evaluative Space: A Critical Review, With Emphasis on the Separability of Positive and Negative Substrates, *Psychological bulletin* **115** (1994), no. 3, 401–423.
- [7] J.T. Cacioppo, L.G. Tassinary, and G. Berntson, *Handbook of Psychophysiology*, 3rd ed., Cambridge University Press, USA, 2007.
- [8] I.C. Christie and B.H. Friedman, Autonomic specificity of discrete emotion and dimensions of affective space: A multivariate approach, *International journal of psychophysiology* **51** (2004), no. 2, 143–153.
- [9] C.L. Cooper and L.A. Pervin, *Personality: Critical concepts in psychology*, 1st ed., *Critical concepts in psychology*, New York, NY, USA: Routledge, 1998.
- [10] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J.G. Taylor, Emotion Recognition in Human–computer Interaction, *IEEE signal processing magazine* **18** (2001), no. 1, 32–80.
- [11] A. Daly, Movement analysis: Piecing together the puzzle, *TDR – the drama review: A journal of performance studies* **32** (1988), no. 4, 40–52.
- [12] K. Ducatel, M. Bogdanowicz, F. Scapolo, J. Leijten, and J.-C. Burgelman, Scenarios for Ambient Intelligence in 2010, Tech. report, European Commission Information Society Technologies Advisory Group (ISTAG), February 2001.
- [13] A. Duke and M.M. Novicevic, Historical foundations of social effectiveness? Dale Carnegie's principles, *Social influences* **3** (2008), no. 2, 132–142.
- [14] H.J. Eysenck and S.B.G. Eysenck, *Manual of the Eysenck personality scales (EPS Adult)*, Hodder and Stoughton: London, 1991.
- [15] S.H. Fairclough, Fundamentals of physiological computing, *Interacting with computers* **21** (2009), no. 1–2, 133–145.

- [16] N.H. Frijda, *The emotions*, Cambridge, New York, USA: Cambridge University Press, 1986.
- [17] H. Gamboa, F. Silva, H. Silva, and R. Falcão, *PLUX – Biosignals Acquisition and Processing*, 2009, URL: <http://www.plux.info> [Last accessed on April 12, 2009].
- [18] G.H.E. Gendolla, On the impact of mood on behavior: An integrative theory and a review, *Review of general psychology* **4** (2000), no. 4, 378–408.
- [19] H. Gunes and M. Piccardi, Automatic Temporal Segment Detection and Affect Recognition From Face and Body Display, *IEEE transactions on systems, man, and cybernetics – part b: Cybernetics* **39** (2009), no. 1, 64–84.
- [20] A. Haag, S. Goronzy, P. Schaich, and J. Williams, Emotion recognition using bio-sensors: First steps towards an automatic system, *Lecture notes in computer science (affective dialogue systems)* **3068** (2004), 36–48.
- [21] D. Jiang, M. He, Y. Qiu, Y. Zhu, and S. Tong, Long-range correlations in heart rate variability during computer-mouse work under time pressure, *Physica a: Statistical mechanics and its applications* **388** (2009), no. 8, 1527–1534.
- [22] P.R. Kleinginna and A.M. Kleinginna, A categorized list of emotion definitions, with a suggestion for a consensual definition, *Motivation and emotion* **5** (1981), no. 4, 345–379.
- [23] P.J. Lang, The emotion probe: Studies of motivation and attention, *American psychologist* **50** (1995), no. 5, 372–385.
- [24] E. Leon, G. Clarke, V. Callaghan, and F. Sepulveda, A user-independent real-time emotion recognition system for software agents in domestic environments, *Engineering applications of artificial intelligence* **20** (2007), no. 3, 337–345.
- [25] R. Likert, A Technique for the Measurement of Attitudes, *Archives of psychology* (1932), no. 140, 5–53.
- [26] C.L. Lisetti and F. Nasoz, Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals, *EURASIP journal on applied signal processing* **2004** (2004), no. 11, 1672–1687.
- [27] P. Lukowicz, Wearable computing and artificial intelligence for healthcare applications, *Artificial intelligence in medicine* **42** (2008), no. 2, 95–98.
- [28] M. Malik and A.J. Camm, *Heart Rate Variability*, Armonk, NY, USA: Futura Publishing Company, Inc., 1995.
- [29] B. Malmberg, *Manual of Phonetics*, 1st ed., Amsterdam, The Netherlands: North-Holland Publishing Company, 1968.
- [30] S. Mann, Wearable computing: A first step toward personal imaging, *IEEE computer* **30** (1997), no. 2, 25–32.
- [31] I.B. Mauss and M.D. Robinson, Measures of emotion: A review, *Cognition and emotion* **23** (2009), no. 2, 209–237.
- [32] M. Minsky, *Emotion*, New York, NY, USA: Simon & Schuster Paperbacks, 1985, ch. 16, pp. 162–172.
- [33] M. Minsky, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, New York, NY, USA: Simon & Schuster, 2006.
- [34] S.A. Neumann and S.R. Waldsein, Similar patterns of cardiovascular response during emotional activation as a function of affective valence and arousal and gender, *Journal of psychosomatic research* **50** (2001), no. 5, 245–253.
- [35] A. Ortony, G.L. Clore, and A. Collins, *The cognitive structure of emotions*, New York, NY, USA: Cambridge University Press, 1988.
- [36] R.W. Picard, *Affective Computing*, Boston MA, USA: MIT Press, 1997.
- [37] J.J. Prinz, *Gut Reactions: A Perceptual Theory of Emotion*, Philosophy of Mind, New York, NY, USA: Oxford University Press, Inc., 2004.
- [38] P. Rainville, A. Bechara, N. Naqvi, and A.R. Damasio, Basic emotions are associated with distinct patterns of cardiorespiratory activity, *International journal of psychophysiology* **61** (2006), no. 6, 5–18.
- [39] E. Ruiz-Padial, J.J. Sollers, III, J. Vila, and J.F. Thayer, The rhythm of the heart in the blink of an eye: Emotion-modulated startle magnitude covaries with heart rate variability, *Psychophysiology* **40** (2003), no. 2, 306–313.
- [40] J.A. Russell, Core affect and the psychological construction of emotion, *Psychological review* **110** (2003), no. 1, 145–172.
- [41] K.R. Scherer, Vocal communication of emotion: A review of research paradigms, *Speech communication* **40** (2003), no. 1–2, 227–256.
- [42] H.A. Simon, Motivational and emotional controls of cognition, *Psychological review* **74** (1967), no. 1, 29–39.
- [43] K. van Deemter, B. Krenn, P. Piwek, M. Klesen, M. Schröder, and S. Baumann, Fully generated scripted dialogue for embodied agents, *Artificial intelligence* **172** (2008), no. 10, 1219–1244.
- [44] E.L. van den Broek, Emotional Prosody Measurement (EPM): A voice-based evaluation method for psychological therapy effectiveness, *Studies in health technology and informatics (medical and care computatics)* **103** (2004), 118–125.
- [45] E.L. van den Broek, J.H. Janssen, J.H.D.M. Westerink, and J.A. Healey, Prerequisites for Affective Signal Processing (ASP), *Biosignals 2009: Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing (Porto – Portugal)* (P. Encarnação and A. Veloso, eds.), 2009, pp. 426–433.
- [46] R. Ventura and C. Pinto-Ferreira, Responding efficiently to relevant stimuli using an emotion-based agent architecture, *Neurocomputing [in press]* (2009).
- [47] D. Ververidis and C. Kotropoulos, Emotional speech recognition: Resources, features, and methods, *Speech communication* **48** (2006), no. 9, 1162–1181.
- [48] J.H.D.M. Westerink, E.L. van den Broek, M.H. Schut, J. van Herk, and K. Tuinenbreijer, Computing emotion awareness through galvanic skin response and facial electromyography, *Philips Research Book Series*, vol. 8, Springer: Dordrecht, The Netherlands, 2008, ch. 14, pp. 137–150.
- [49] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan, Towards Practical Smile Detection, *IEEE transactions on pattern analysis and machine intelligence [in press]* (2009).
- [50] Z. Zeng, M. Pantic, G.I. Roisman, and T.S. Huang, A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions, *IEEE transactions on pattern analysis machine intelligence* **31** (2009), no. 1, 39–58.