Sentiment uncovering by employ of Biometrics techniques

Dr.Kashif Qureshi

Department Of Computer Science, tel: +966 (0)535598451,Jazan University, Jazan, Saudia
Srk1521@gmail.com

Abstract. The model Biometrics for Emotion Detection **(BED)** is introduced, which enables unobtrusive emotion recognition, taking into account varying environments. It uses the electrocardiogram (ECG) and speech, as a powerful but rarely used combination to untie people's emotions. BED was applied in two environments (i.e., office and home-like) in which 30 people watched 13 film scenes. It is shown that both heart rate variability (derived from the ECG) and, when people's gender is taken into account, the standard deviation of the fundamental frequency of speech indicate people's experienced emotions. As such, these measures validate each other. Moreover, it is found that people's environment can indeed of influence experienced emotions. These results indicate that BED might become an important paradigm for unobtrusive emotion detection.

*Section -*1

1. Introduction

A goal of human-centered computing is computer systems that can unobtrusively perceive and understand human behavior in unstructured environments and respond appropriately.

Three of the issues raised in this quote formed the starting point of the research described in this paper: 1) unobtrusively perceive, 2) understand, and 3) unstructured environments. Regrettably, in practice, the combination of these issues is rarely taken into account in research towards human-centered computing. This paper describes research that took into account all three aspects that Ire just denoted. To enable unobtrusive recordings of humans, I exploit the combination of the electrocardiogram (ECG) and

Speech. The combination of speech and biosignals (in general) to determine emotions is rare. However, some research has been conducted;. This research concluded that they did not achieve the same high gains that Ire achieved for audio-visual data which seems to indicate that speech and physiological data contain less complementary information.

This paper presents a study that challenges this conclusion. In combination with people's subjective experiences, these two signals should be able to untie generic principles underlying humans' emotions, at least partly.

It has been widely acknowledged that emotion elicitation is as crucial as is it complex for affective computing purposes. Moreover, context has been frequently mentioned as a factor of influence. This study explores whether or not context has an influence on emotion elicitation, as it comprises two identical studies in different settings. This is expected to provide a little grip on the influence of environmental factors / context. Since I want to explore generic principles behind human emotions, I ignore inter-personal differences. Please note that with this position, I do not challenge the idea that significant differences among humans exist; Ultimately, a personal- ized approach will need to be adopted; however, before that, generic principles have to identified to enable more efficient processing of cues.

First, I will briefly introduce the construct emotion (Section 2) and the model of emotion used. Section 4 describes the two studies conducted, with which I explore the use of the pattern BED. This includes a description of the **speech and ECG signals**. In Section 5, the analysis of the two studies

is presented. These comprise only a limited set of features, as an exhaustive search in feature space was not the aim of this study. Last, in Section 6, the implications of this research will be discussed and future directives will be provided.

*Section -*2

2. Sentiments

A thought, view, or attitude, especially one based mainly on emotion instead of reason: An anti-American sentiment swept through the country..

2.

• a. Emotion; feeling: Different forms of music convey different kinds of sentiment.

• b. Tender or romantic feeling.

• c. Maudlin emotion; sentimentality.

• 3. The emotional import of a passage as distinct from its form of expression.

• 4. The expression of delicate and sensitive feeling, especially in art and literature.

• Susceptibility to tender, delicate, or romantic emotion she has too much sentiment to be successful

• 2. (Often plural) a thought, opinion, or attitude

• 3. Exaggerated, overindulged, or mawkish feeling or emotion

• 4. an expression of response to deep feeling, esp. in art or literature

• 5. a feeling, emotion, or awareness a sentiment of pity

• 6. a mental attitude modified or determined by feeling there is a strong revolutionary sentiment in his country

• 7. a feeling conveyed, or intended to be conveyed, in words

Efforts at emotion recognition are inherently field unless one recognizes that emotion - intentions, action tendencies, appraisals and other cognitions, physiological and neuromuscular changes, and feelings is not readily ob-servable. In other words, emotions are a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems. Consequently, emotions can

• cause affective experiences such as feelings of arousal and (dis)pleasure;
• generate cognitive processes; e.g., emotionally relevant perceptual effects, appraisals, labeling processes;
• activate widespread physiological adjustments to arousing conditions; and
• lead to behavior that is often expressive, goal directed, and adaptive.

Much can be said both in favor and against this definition. However, with this work I do not aim to provide an exhaustive elaboration on the definition of emotions. Therefore, I adopt the previous definition of emotions as working definition.

In his seminal study, Russell (1980) introduced the circumflex model of emotion, which claims that all emotions can be characterized by two independent bipolar dimensions: judged valence (i.e., pleasure/positive or displeasure/negative) and arousal. The circumflex model allows a representation of emotions in a 2D space, which provides a visualization of emotions and their relation in a comprehensible way.

Parallel to Russell's model various other models of emotion have been introduced, using 1) categories, 2) unipolar or bipolar dimensions, 3) some other simple structure, and 4) a hierarchy. For an overview on models of emotion, I refer to .

For the current research, I adapted Russell's circumflex model . The bipolar valence dimension was replaced by two unipolar valence dimensions, untying positive and negative valence. The arousal dimension was not altered. A similar approach have been used earlier; e.g.,. The 3D model I propose tackles a major disadvantage of the circumflex model: the disability to handle mixed emotions; i.e., parallel experience of positive and negative valence.

3. Signals of Sentiment

People's emotional state can be accessed through processing various of their signals. When reviewing literature, it becomes apparent that these signals can be assigned to two groups: 1) A broad range of physiological measures signals Specialized areas of signal processing: speech processing, movement analysis, and computer vision techniques. These distinct measurement methods are seldom combined; where, on the one hand, several physiological measures are frequently combined and, on the other hand, speech processing, movement analysis, and computer vision are frequently combined

Physiological measures are often obtrusive and, hence, disregarded for user-centered applications, as Ami are. However, computing and wireless sensing technologies relief this problem .In contrast, speech and computer vision are unobtrusive but very noise sensitive. The audio recordings used for speech processing suffer 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 and with fixed light sources.

Speech and physiological measures, in particular the ECG, are rarely 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; .

From features of both the speech and the ECG signal, I 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, I chose for a controlled study to assess their feasibility for human-centered computing purposes. However, before the study is described, each of the signals used are introduced.

4. Philosophy and realization

Let us quote John F. Cohn , once more: Emotion can only be inferred from context, self-report, physiological indicators, and expressive behavior. The four factors addressed in this quote are incorporated in this research on BED. For each of them will be explained how they are embedded in BED and are implemented in the research.

4. Background

As was stated in the introduction, context is of the utmost importance. However, field research introduces a broad range of factors that cannot be controlled. These factors can be considered as severe forms of noise, when analyzing the data obtained. Most often, this results in qualitative data gathering and its analysis. In contrast, with the signals ECG and speech, I aim to gather quantitative data. Therefore, I chose to conduct a semi-controlled study.
40 people (average age: 27) voluntarily

participated in this research. Half of them was assigned to a living room environment and half of them was assigned to an office environment. Except for this difference, both groups participated in the same study, under the same conditions. All participants had (corrected to) normal vision and hearing and none of them had cardiovascular problems.

To trigger emotions, the participants watched six scenes (length: 3.18 min), adopted,. For a more exhaustive description of the film scenes, I refer to. The length of the film scenes enables a reliable extraction of HR variability from the ECG. The film scenes Ire presented in a random order. While watching the film scenes, their ECG signal was recorded using a modified Polar ECG measurement belt that was connected to a NIUSB-6008 data acquisition device. Its output was recorded in a Lab VIEW program (sample rate: 200 Hz). After the film scene was watched, the people Ire asked to talk about it. For this a standard microphone was used (sample rate: 44.100 Hz; sample size: 16 bit).

After instructions and a control of the equipment, the people read aloud a non emotional story. In this way it was checked if the people had understood the instructions, whether or not the equipment worked, and people's personal baseline for both the speech and the ECG signal was determined.

4. *identity* -report

After all film scenes Ire shown, the people rated the films, using 11 point Likert scales. Hence, their subjective attitudes Ire determined. Separate Likert scales Ire presented for positive and negative affect and for arousal; see also Section 2.

4. Physiological indicator: Electrocardiogram (ECG)

Electrocardiography (ECG or EKG from Gr eek: kardia, meaning heart) is a transthoracic (across the thorax or chest) interpretation of the electrical activity of the heart over a period of time, as detected by electrodes attached to the surface of the skin and recorded by a device external to the body.[1] The recording produced by this noninvasive procedure is termed an electrocardiogram (also ECG or EKG).

An ECG is used to measure the rate and regularity of heartbeats, as well as the size and position of the chambers, the presence of any damage to the heart, and the effects of drugs or devices used to regulate the heart, such as a pacemaker.

Most ECGs are performed for diagnostic or research purposes on human hearts, but may also be performed on animals, usually for diagnosis of heart abnormalities or research

The electrocardiogram (ECG) is an autonomic signal that cannot be controlled easily, just like the electro dermal activity. ECG can be measured directly from the chest. Alternatively, the periodic component of the blood flow in the finger or in an ear which can be translated into the ECG. From the ECG, the heart rate (HR) can be easily obtained; Research identified features of HR as indicators for both experienced valence and arousal [15].

In addition to the HR, a range of other features can be derived from the ECG. The most frequently used one is HR variability (HRV). HRV decreases with an increase in mental effort, stress, and frustration [9]. Moreover, some indications have been found that HRV is also influenced by the valence of an event, object, or action [15]. Taken together, HRV is one of the most powerful features to discriminate among emotions.

4. Expressive behavior: The speech signal

The Speech Signal Processing Toolkit (SPTK) is a suite of speech signal processing tools for UNIX environments, e.g., LPC analysis, PARCOR analysis, LSP analysis, PARCOR synthesis filter, LSP synthesis filter, vector quantization techniques, and other extended versions of them. This software is released under the Modified BSD license.

SPTK was developed and has been used in the research group of Prof. Satoshi Imai (he has retired) and Prof. Takao Kobayashi (currently he is with Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology) at P&I laboratory, Tokyo Institute of Technology. A sub-set of tools was chosen and arranged for distribution by Prof. Keiichi Tokuda (currently he is with Department of Computer Science and Engineering, Nagoya Institute of Technology) as a coordinator in cooperation and other collaborates (see "Acknowledgments" and "Who we are" in README).

The original source codes have been written by many people who took part in activities of the research group. The most original source codes of this distribution were written by Takao Kobayashi (graph, data processing, FFT, sampling rate conversion, etc.), Keiichi Tokuda (speech analysis, speech synthesis, etc.), and Kazuhito Koishida (LSP, vector quantization, etc.).

This version is accompanied by a Reference Manual. A small User's Manual "Examples for using SPTK" is also attached.

Speech processing, speech dialog, and speech synthesis can exhibit some form of intelligent, user-perceived behavior and, hence, are useful in designing human-centered computing environments [1]. However, speech comprises another feature: emotion elicitation.

The human speech signal can be characterized through 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 is present for the variability (e.g., standard deviation; SD) of the fundamental frequency (F0), energy of speech, and intensity of air pressure .Therefore, this feature is derived from speech as its emotion indicator.

5. Analysis

5.1 Data reduction

Before the actual data reduction could take place, noise reduction was applied for both the ECG and the speech signal. First, recording errors Ire removed. Of 11 participants either the ECG signal, the speech signal, or both was distorted (e.g., the microphone was positioned to close to the mouth, coughing, yawning) or not recorded at all. The data of these participants was omitted from further analyses.

The measurement belt for the ECG signal appeared to be sensitive to movements of the participant. For this type of noise was controlled and, if present, the ECG signal was corrected for it. Using the corrected ECG signal, people's HR was determined. Subsequently, HRV (i.e., SD of HR) was calculated.

The speech signal was segmented in such a way that for each film scene a separate speech signal was determined. Moreover, silences and utterances such as 'euh' were removed. This resulted noise-free speech signals. In order to cope with interpersonal differences in speech, the

signals Ire normalized by subtracting a baseline from the original signal. Subsequently, the speech processing environment Pratt was used to extract the SD F0 and the intensity of air pressure.

5.2 Subjective measurements

The people rated the film scenes on their experienced positive and negative valence and on arousal. For each film scene, peoples' average ratings on each of these scales are calculated. Together, these two

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

categorized dimensions of the valence-arousal model depict six emotion classes.

Each of the six emotion classes is represented in this research through one film fragment. The emotion classes with the values on the three dimensions, their categorization in valence and arousal, and the accompanying film fragment can be found in .

5.   Feature extraction

From both the speech signal and the ECG signal a large number of features can be derived [14,15]. This research did, hoIver, not aim to provide an extensive comparison of speech and ECG features. Instead, the combination of these two signals was explored. From the ECG signal, the intervals between the R-waves (R-R intervals) as Ill as their mean Ire determined. Subsequently, the HRV was defined as the SD of the R-R intervals.

Although no general consensus exists concerning the parameters of speech to be used for emotion detection, much evidence is present for the SD F0 and the Intensity of air pressure. They are useful for measuring experienced emotions. For a domain [0, T], the intensity of air pressure in the speech signal   is   defined   as:

where $x(t)$ is the amplitude or sound pressure of the signal in Pa (Pascal) and $P_0 = 2 \cdot 10^{-5}$ Pa is the auditory threshold [16]. It is expressed in dB (decibels) relative to P_0.

The F0 of pitch was extracted from the corrected speech signal through a fast Fourier transform over the signal. Subsequently, its SD is calculated. I refer to the documentation that accompanies [16], for a more detailed description of the extraction of F0 of pitch from the speech signal.

5.   Results

All data was analyzed through a RM ANOVA, with three measures: HRV determined from the ECG signal and the SD F0 and intensity of the speech signal. Two between subject factors Ire included in the analyses: the environment (office/living room) and gender (male/female). Age was omitted from the analysis since a preliminary analysis revealed that age was of no influence on any of the measures. First, the multivariate test will be

reported, including all four measures. Next, for each measure the univariate tests will be reported. With all analyses, the interaction effects will be reported.

The multivariate analyses should a strong effect for the emotion classes on the set of physiological parameters/measures, $F(15,11) = 29.688$, $p < .001$. In addition, in interaction with both gender ($F(15,11) = 7.266$, $p = .001$) and environment ($F(15,11) = 17.235$, $p = .000$), 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, $F(15,11)8.737$, $p < .001$.

A strong main effect was found for the emotion classes on HRV, $F(5,125) = 38.677$, $p < .001$. An interaction effect of the emotion classes and both gender ($F(5,125) = 7.678$, $p < .001$) and environment ($F(5,125) = 18.856$, $p < .001$) 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(5,125) = 10.967, p
< .001.

An interaction effect between the emotion classes and gender on SD F0 was found, F(5,125) = 2.553, p = .031. In addition, a three-way interaction effect on the intensity of speech between the emotion classes, gender, and environment was found, F(5,125) = 3.052, p= .013.

6. Conclusion

The emotions elicited by the film fragments are clearly discriminated by HRV. In interaction with gender, both HRV and SD F0 show to be a good discriminator among the six emotion classes. When both gender and environment are taken into account, both HRV and the intensity of speech reflect the emotions experienced.

Of the F0 of speech, it is known that male and females have different characteristics. Hence, the influence of gender was expected and will always have be taken into account. Moreover, environmental factors have to be incorporated in the processing scheme. Note that the difference between the environments assessed in this research was limited; hence, in practice this effect can be far more substantial. Further, research on the intensity of speech should reveal how robust the current findings for this parameter are.

Both HRV and SD F0 of speech should to be good generic unobtrusive discriminators between emotions. This makes them par excellence suitable as biometrics for unobtrusive emotion discrimination. This study is rare in that it reports the use of biosignals in combination with speech to unravel user's emotional state. However, it should be noted that the variety among emotions is rich and only six are assessed in the current research. Moreover, it is unknown how sensitive both measures are for emotion discrimination. Hence, further research is needed on this.

The results of this research should be seen in the perspective of the vast amount of research already done on emotions. This research has revealed a range of issues that can, and probably should, also be taken into account. First, the notion of time should be taken into consideration that helps to distinguish between emotions, moods, and personality . Second, BED distinguishes four factors: context, self-report, physiological indicators, and expressive behavior. The last three of them can be influenced by personality characteristics. For example, an introvert person will express his emotions in another way than an extrovert person. Moreover, probably also the self-reports and physiological indicators or biosignals will be influenced by this personality trait. Moreover, whether or not a person is neurotic will influence his behavior, in particular in relation to his environment (or context); e.g., see also .

Other more practical considerations should also be noted. For example, the advances made in computing and sensors facilitates the communication between humans and

systems; cf. [13]. This enables the use of more recordings of biosignals in parallel to speech recordings and ECG. In this way, an even higher probability of correct interpretation can be achieved .

It is surprising that the combination of speech and biosignals has not been used more often to unravel user's emotions; Par excellence, these signals could be exploited in parallel for generic human-centered computing purposes, as is illustrated through BED. Both speech and ECG parameters can unravel users' emotion space. Moreover, various manners of implementation of the required sensors secure an unobtrusive recording of both signals. This having said, with this article I hope to motivate
 a further exploration of the combination of speech and biosignals. Possibly, they enable

the significant step forward in making reliable unobtrusive emotion detection a success.

References

➢ Cohn, J.F.: Foundations of human computing: Facial expression and emotion. Lecture notes in Artificial Intelligence (Artifical Intelligence for Human Computing) **4451** (2007) 1-16

➢ Kim, J., Andre, E.: Emotion recognition using physiological and speech signal in short-term observation. Lecture Notes in Computer Science (Perception and Interactive Technologies) **4021**(2006)53-64

➢ Kim, J.: 15. In: Bimodal emotion recognition using speech and physiological changes. Vienna, Austria: I-Tech Education and Publishing (2007) 265-280

➢ Kim, J., Andre, E.: Emotion recognition based on physiological changes in music listening. IEEE Transactions on Pattern Analysis and Machine Intelligence **30** (2008) 2067-2083

➢ Kleinginna, P.R., Kleinginna, A.M.: A categorized list of emotion definitions, with a suggestion for a consensual definition. Motivation and Emotion **5** (1981) 345-379

➢ Russel, J.A., Barrett, L.F.: Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. Journal of Personality and Social Psychology 26 (1999) 805-819

➢ van den Broek, E.L., Isterink, J.H.D.M.: Considerations for emotion-aware consumer products. Applied Ergonomics **40** (2009) 1055-1064

➢ van den Broek, E.L., Lisy, V., Janssen, J.H., Isterink, J.H.D.M., Schut, M.H., Tuinenbrei- jer, K. In: Affective Man-Machine Interface: Unveiling human emotions through biosignals. Volume 52 of Communications in Computer and Information Science. Berlin/Heidelberg, Germany: Springer-Verlag (2010) 21-47

➢ Fairclough, S.H.: Fundamentals of physiological computing. Interacting with Computers **21** (2009) 133-145

➢ Ververidis, D., Kotropoulos, C.: Emotional speech

recognition: Resources, features, and methods. Speech Communication **48** (2006) 1162-1181

➢ Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence **31** (2009) 39-58

➢ Gunes, H., Piccardi, M.: Automatic temporal segment detection and affect recognition from face and body display. IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics **39** (2009) 64-84

➢ Pantelopoulos, A., Bourbakis, N.G.: A survey on Iarable sensor-based systems for health monitoring and prognosis. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews **40** (2010) 1-12

➢ Berntson, G.G., Bigger, J.T., Eckberg, D.L., Grossman, P., Kaufmann, P.G., Malik, M., Nagaraja, H.N., Porges, S.W., Saul, J.P., Stone, P.H., van der Molen, M.W.: Heart rate variability: Origins, methods, and interpretive caveats. Psychophysiology **34** (1997) 623-648

➢ Wikipedia.com