Emotion Detection: A Technology review

Jose Maria Garcia-Garcia
Computer Science Research Institute
University of Castilla-La Mancha
Albacete, Spain
josemaria.garcia14@alu.uclm.es

Victor M. R. Penichet
University of Castilla-La Mancha
Albacete, Spain
Victor.Penichet@uclm.es

Maria D. Lozano
University of Castilla-La Mancha
Albacete, Spain
María.Lozano@uclm.es

ABSTRACT

Emotion detection has become one of the most important aspects to consider in any project related to Affective Computing. Due to the almost endless applications of this new discipline, the development of emotion detection technologies has brought up as a quite profitable opportunity in the corporate sector. Many start-up enterprises have emerged in the last years, dedicated almost exclusively to a specific type of emotion detection technology. In this paper, we present a thorough review of current technologies to detect human emotions. To this end, we explore the different sources from which emotions can be read, along with existing technologies developed to recognize them. We also explore some application domains in which this technology has been applied. This survey has let us identify the strengths and shortcomings of current technology for emotion detection. We conclude the survey highlighting the aspects that requires further research and development.

CCS CONCEPTS

• **Human-Centered Computing → Interaction Design.**

KEYWORDS

Affective Computing, emotion recognition, technologies

1 INTRODUCTION

Affective Computing, as it was defined in 1995 in [23], is the "computing that relates to, arises from, or influences emotions", or in other words, any form of computing that has something to do with emotions. Due to the strong

relation with emotions, their correct detection is the cornerstone of Affective Computing and will be the focus of this paper. Even though each type of technology works in a specific way, all of them share a common core in the way they work, since an emotion detector is, fundamentally, an automatic classifier. The creation of an automatic classifier involves collecting information, extracting the features which are important for our purpose, and finally training the model, so it can recognize and classify certain patterns [24]. Later, the model will be asked for classifications of new data. For example, if we want to build a model to extract emotions of happiness and sadness from facial expressions, we have to feed the model with pictures of people smiling, tagged with "happiness", and pictures of people frowning, tagged with "sadness". After that, when it receives a picture of a person smiling, it identifies the shown emotion as "happiness", while pictures of people frowning will return "sadness" as a result.

In real life, the creation of a model is not that simple. Not only there is a lot of information to consider, but an effort of interpretation is also needed, as we will expose later. Humans express their feelings through several channels: facial expressions, voices, body gestures and movements, etc. Even our bodies experiment visible physical reactions to emotions (breath and heart rate, pupils size, etc.).

Because of the high potential of knowing how the user is feeling, this kind of technology (emotion detection) has experienced an outburst in the business sector. Many technology companies have recently emerged, focused exclusively on developing technologies capable of detecting emotions from specific input. In the following sections, we present a review of each kind of affective information channel, along with some existing technologies capable of detecting that kind of information, if any.

The rest of the paper is organized in six sections. Each one of the first five sections is focused on one of the channels from which we can get affective information: emotion from speech, emotion from text, emotion from facial expressions, emotion from body gestures and movements and emotion from physiological states. The sixth section discusses the findings presented in the previous ones, highlighting the strengths and shortcomings identified in the review. The last section presents the conclusions and future works.

2 EMOTION FROM SPEECH

One of the exploitable channels for gathering emotional information from the user of a system is their voice. When a person starts talking, they generate information in two different channels: primary and secondary [11]. The primary channel is linked to the syntactic-semantic part of the talking (what the person is literally saying), while the secondary channel is linked to paralinguistic information of the speaker (tone, emotional state, and gestures). E.g., someone says "That's so funny" (primary channel) with a serious tone (second channel). By looking at the information of the primary channel, we get the message that the speaker thinks that something is funny, and by looking at the information received by the second channel, we get to know that the real meaning of the message is that the speaker is lying or being sarcastic.

In the next sub-sections, four technologies of this category are presented: *Beyond Verbal*, *Vokaturi*, *EmoVoice* and *Good Vibrations*.

Each of them has been tested (when possible), and conclusions got from these experiments are gathered in the last sub-section.

2.1 Beyond Verbal

Beyond Verbal, an Israel-based company active since 2013, works on software which extracts, decodes and measures human moods, attitudes and decision-making profiles based on their voices. It offers its technology as a API-style licensed service, with which users can make their voice-based applications emotion-aware [6]. *Beyond Verbal* technology is available by payment of certain month fees that go from 58$ to 298$, including a custom pricing options for volumes of audio bigger than 20,000 minutes monthly [5]. However, they include a free option that allows users to try *Beyond Verbal* technology for a month with no cost, but with a limit of 100 minutes.

Since *Beyond Verbal* services are cloud-based, using them on a project is as simple as making a request to their API. The only aspect that needs to be consider is that this technology needs, at least, 13 seconds of spoken voice to give a result. This conversation voice should also be clear enough (although their algorithms tolerate certain noise). After analysing some audio tracks, the application returns the information in a JSON object, containing the following data [6]:

- *Temper*. Reflects the speaker's temperament and it is represented with a value in the range between 0 (lowest temper) and 100 (highest temper). Higher tempers are associated with aggressive emotions (anger, hatred, hostility, etc.) while lower tempers are more linked to depressive emotions (sadness, pain, regret, fear, anxiety, etc.). On the other hand, medium tempers express neutral or positive emotions. There are also three temper groups:

Low, Medium and High, each one corresponding to a segment of the previous range. Limits between these groups are blurry, so the value of the temper can be seen better with a value in a range. E.g., a value of 30 on a 0-100 scale is between low and medium temper, while a value of 70 is between medium and high.

- *Valence*. Output that measures speaker's level of negativity/positivity. This parameter, along with arousal, are the most used parameters to express emotion according to a dimensional approach [27]. Valence can be neutral, positive or negative and it is also represented with a category (negative, neutral, positive) and with a value in a range. However, this is a new addition to Beyond Verbal API and it is still on beta phase.

- *Arousal*. Output that measures a speaker's degree of energy ranging from bored, to neutral to energetic. Expressed with a category (low, mid and high) and a value in a range.

- *Mood groups*. Indicator that expresses the emotion detected on the audio by using a sentence in natural language. *Beyond Verbal* API distinguishes four groups, each one bigger and more complex than the previous one. In Table 1 we can see an example of the kind of elements considered in each group. Groups 7, 11 and 21 -numbers indicate the number of emotions included in each group-express the emotion detected in an uncomplicated way (one word), while Composite (group with 432 elements) expresses compound, more complex emotions. Responses usually contain just emotions from groups 11 and Composite and for each one of these, they distinguish two emotions: the primary emotion, which corresponds to the strongest emotion detected within that group, and the secondary emotion, which corresponds to the second strongest emotion detected within that group.

The preliminary assessment performed on *Beyond Verbal* API has shown satisfactory results in a quiet environment. The response also contains information about the quality of the audio sent and the gender of the person detected on it.

2.2 Vokaturi

Vokaturi, founded in 2016 and based in Amsterdam, develops software which reflects the state of the art in emotion recognition from the human voice. They have developed several libraries, in C and Python, so developers can integrate emotion detection from speech in their applications. *Vokaturi* offers three kinds of licenses to use their technology [34]:

- *OpenVokaturi*. Open-source version of the *Vokaturi* library, distributed under GPL license. It has an accuracy on classification of 66.5 %.

- *VokaturiPlus*. Closed-version of *Vokaturi* library. This version is used for applications to be commercialized. It has an accuracy on classification of 76.1 %.

- *VokaturiPro*. This version allows the client to train the *Vokaturi* to recognize customize emotions. It has an accuracy on classification of 76.1 %.

Table 1. Beyond Verbal mood groups

Group	Emotions
Group 7	Angry; Cool; Enthusiastic; Frustrated; Happy; Sad; Worried.
Group 11	Creative, Passionate; Criticism, Cynicism; Defensiveness, Anxiety; Friendly, Warm; Hostility, Anger; Leadership, Charisma; Loneliness, Unfulfillment; Love, Happiness; Sadness, Sorrow; Self-Control, Practicality; Supremacy, Arrogance.
Group 21	Admiration; Anger; Anxiety; Belief; Creativity; Disliking; Dominance, etc.
Composite	Emotional coping. Internal struggle or interpersonal conflict; Assertiveness and ambition to achieve goals; Goal-oriented anger. Forcefulness to achieve goals, etc.

The *OpenVokaturi* version was the one checked for comparative purposes. One the one hand, it operates on the computer where it is being used, so it doesn't need access to Internet. On the other hand, it's not as powerful as other services that return us the results calculated by a powerful net of computers.

Vokaturi results are expressed as a combination of Paul Ekman's six basics emotions [14] (except for surprise and disgust) and also considers neutrality. The presence and/or intensity of each emotion is denoted with a value between 0 and 1, which represents the weight of that emotion over the whole emotions. The sum of every value is equal to 1.

In testing conditions, similar to the ones used for testing *Beyond Verbal*, the free version of *Vokaturi* has given poor results: most of the audio samples used were misclassified with anger as the strongest emotion.

2.3 EmoVoice

EmoVoice is a comprehensive framework for real-time recognition of emotions from acoustic properties of speech (without using word information) [33]. It has been developed by the Human-Centered Multimedia department from the Universität Augsburg University, and it has been recently integrated in the Social Signal Interpretation (*SSI*) framework [35] [31]. It offers tools to record, analyse and recognize human behaviour in real-time. While the previous technologies were services already implemented and ready to be used, this one works by creating a model and training it to recognize emotions, which can be later connected to the Internet to be requested analysis and classifications.

One remarkable aspect of this technology is that it allows users to create their own speech databases, along with the fact that the SSI framework is a patch-based technology, which gives users more flexibility: using XML

to specify the different elements in each step, developers can create a completely customized emotion detector.

Although it has been used in several research projects, *EmoVoice*, as well as *SSI*, is a complex low-level tool that can take too much time to set and configure for a small project, or for a project in which there is not enough time to program the emotion part.

2.4 Good Vibrations

Other option in the market, regarding emotion recognition from speech, is the Good Vibrations Company. They also offer a service in the form of a SDK, although they do not offer any trial version. The most we can check is the manual of their technology [13].

2.5 Comparative study and discussion

Table 2 shows the results of the comparative study performed on the four analysed technologies. Only two of them could be tested, *Beyond Verbal* and *Vokaturi* (Open version). *Good Vibrations* does not offer any kind of demo or trial version, although the manual of their SDK is available on [13].

Table 2. Comparison of emotion detection technologies from speech

Name	API/ SDK	Requires Internet	Information returned	Difficulty of use	Free Software
Beyond Verbal	API	Yes	Temper Arousal Valence Mood (Up to 432 emotions)	Low	No
Votakuri	SDK	No	Happiness, neutrality, sadness, anger and fear	Medium	Yes
EmoVoice	SDK	No	Determined by developer	High	Yes
Good Vibrations	SDK	-	Happy level, relaxed level, angry level, scared level and bored level	Medium	No

3 EMOTION FROM FACIAL EXPRESSIONS

As in the case of speech, facial expressions reflect the emotions that a person can be feeling. Eyebrows, lips, nose, mouth, muscles of the face: they all reveal the emotions we are feeling. Even when a person tries to fake some emotion, still their own face is telling the truth. The technologies used in this field of emotion detection work in an analogous way to the ones used with speech: detecting a face, identifying the crucial points in the face which reveal the emotion expressed and processing their positions to decide what emotion is being detected.

In the next sub-sections, we describe four technologies used to detect emotions from facial expressions: *Emotion API* (Microsoft Cognitive Services), *Affectiva*, *nViso*, and *Kairos*. Each of them has been tested (when possible), and conclusions got from these experiments are gathered in the last sub-section.

3.1 Emotion API

One of the services offered by the Microsoft Cognitive Service pack is the *Emotion API* [21]. This API receives a picture or a video and identifies the faces on it and the emotions expressed on them. The presence of each emotion is denoted with the confidence with which that emotion is detected. E.g., in a picture of a smiling person, happiness will be detected with a high level of confidence, while sadness and fear will be detected with a low value of confidence.

The emotions detected are the six basic ones established by Paul Ekman [14] plus contempt and neutral (absence of emotion). In [21], there is a demo available for testing purposes, where users can submit a picture and check the results produced. Although the service is cloud-based, Microsoft offers a SDK for Android, Python and .NET, which allows an easy integration into projects with these technologies. This service has several licensing options, including a free version, which offers, monthly, 30.000 images analysis requests, 300 video uploads and 3.000 video status queries [21].

3.2 Affectiva

Founded in 2009 by Rana el Kaliouby and Rosalind Picard, this company was born in the MIT Media Lab and has experienced one of the biggest growths in the emotions detection from facial expressions sector [1]. Their technology identifies seven emotions and twenty points in the face, including also gender, age and ethnicity. By analysing the pixels from these twenty zones, *Affectiva* classifies the facial expression detected, taking into account Paul Ekman's Facial Action Coding System [3]. *Affectiva* offers a SDK for Java, Objective-C, C++, Unity, Javascript and Windows, being one of the easiest technologies to integrate in a project (at least the Javascript SDK). *Affectiva*'s technology is available completely for free (except for companies generating more than $ 1,000,000 yearly with non-academic purposes) [2].

3.3 nViso

nViso, a Switzerland-based company founded in 2009, "provides the most scalable, robust, and accurate artificial intelligence solutions to measure instantaneous emotional reactions of consumers in online and retail environments"

according to [25]. By using a system like the one used by *Affectiva*, *nViso* indicates the emotion detected with Ekman's basic emotions plus the neutral state (absence of emotion). *nViso* does not offer any demo or trial version, and pricing info has to be requested in order to use it.

3.4 Kairos

Another company working on emotions detection from facial expressions is *Kairos* [18], although emotions detection is not their only asset. *Kairos* services include also face detection, identification and verification, emotion, age, gender, sentiment, ethnicity and multi-face detection, attention measurements and face grouping. These services can be accessed via API, although there is also an SDK which works offline.

In their corporate website, *Kairos* offers several demos of their services, along with a sandbox to start trying their API directly on the browser. The services offered need pictures or videos to produce a result. In the case of videos, the service slices the video into 0.033 seconds-length segments, analysing each segment to find the faces on it and the emotions expressed on each one of them. *Kairos* offers three different licensing options to use their services, although there is a free version for personal purposes that provides access to all the features. As said before, *Kairos* has also an off-line version (SDK), but this is not free.

While results can be wrong when analysing a single picture, one benefit of using video as input is the fact that some wrong results in some frames do not spoil one thousand right ones in the rest of the video.

One drawback about this technology is that it requires uploading the content we want to analyse to some online platform; besides, the results are not immediately provided and you have to manually and periodically check if the results are available. This can add some serious latency to our request, blocking us from simulating a real-time behaviour.

3.5 Comparative study and discussion

Table 3 shows a comparative study of the previously described technologies. *nViso* has not been tested due to their lack of trial versions or demos.

As far as the results are concerned, every tested technology showed a considerable accuracy. However, several conditions (reflection on glasses, bad lightning) mask important landmarks on the face, generating wrong results. E.g., an expression of pain, in a situation in which eyes and/or brows cannot be seen, can be detected as a smile by these technologies (because of the stretching, open mouth).

As far as time is concerned, *Emotion API* and *Affectiva* show similar times to scan an image, while *Kairos* takes much longer to produce a result. Besides, the amount of values returned by *Affectiva* provides much more

information to the developers, and it is easier to interpret the emotion that the user is showing than when we just have the weight of six emotions, for example. It is also remarkable the availability of *Affectiva*, which provides free services to those dedicated to research and education or producing less than $1,000,000 yearly.

Table 3. Comparison of emotion detection technologies from facial expressions

Name	API/ SDK	Requires Internet	Information returned	Difficulty of use	Free Software
Emotion API	API/ SDK	Yes	Happiness, sadness, fear, anger, surprise, neutral, disgust, contempt	Low	Yes (Limited)
Affectiva	API/ SDK	Yes	Joy, sadness, disgust, contempt, anger, fear, surprise [1]	Low	Yes, with some restriction
nViso	API/ SDK	No	Happiness, sadness, fear, anger, surprise, disgust and neutral	-	No
Kairos	API/ SDK	Yes	Anger, disgust, fear, joy, sadness, surprise [2]	Low	Yes, only for personal use

[1] besides, it also detects different facial expressions, gender, age, ethnicity, valence and engagement.

[2] besides, it also detects user head position, gender, age, glasses, facial expressions and eye tracking.

4 EMOTION FROM TEXT

There are certain situations in which the communication between two people, or between one person and one machine, has not the visual component given by the face-to-face communication. In a world dominated by telecommunications, words are powerful allies to discover how a person may be feeling. Although emotion detection from text (also referred as sentiment analysis) must face more obstacles than the previous technologies (spelling errors, languages, slang), it is another source of affective information to consider. Since emotion detection from texts analyses the words contained on a message, the process to analyse a text take some more steps than the analysis of a face or a voice. There is still a model that needs to be trained, but now text must be processed in order to use it to train a model [7]. This processing involves tasks of tokenization, parsing and part-of-speech tagging, lemmatization, stemming, among others. In the next sub-sections, four technologies of this category are presented: *Tone Analizer*, *Receptiviti*, *BiText*, and *Synesketch*. Each of them has been tested (when possible), and conclusions got from these experiments are gathered in the last sub-section.

4.1 Tone Analyzer

Tone Analyzer is a service developed on Bluemix which uses the power of IBM Watson to detect emotions on a piece of text [16]. *Tone Analyzer* service is based on the theory of psycholinguistics, a research field that explores the relationship between linguistic behaviours and psychological theories [15]. This technology, which works via API, analyses the relationship between text tones and the linguistic characteristics of the text. The results returned by *Tone Analyzer* are organized in three categories [17]:

- *Emotional tone*. This is the tone which expresses emotions. *Tone Analyzer* uses the six basic emotions previously commented, indicating the presence or absence of each emotion with a value between 0 and 1.
- *Social tone*. Measures social tendencies on the text using the Big Five personality model categories.
- *Language tone*. These values describe how the writing is perceived.

At this moment, only English (and French in Bluemix Premium) is supported by this API. *Tone Analyzer* offers a demo on [16] that allows users to test the service. For further use, a membership on Bluemix in needed.

4.2 Receptiviti

Receptiviti [28] is a Canadian start-up company which has been working on text analysis for more than 20 years. The core of *Receptiviti* is their text analysis application, LIWC (Linguistic Inquiry and Word Count). LIWC, whose development started on 1993, is a dictionary composed of almost 6.400 words, each one tagged in various categories according to their affective meaning. E.g., "smile" would be tagged as "verb", "happiness", "positive emotions", etc. *Receptiviti*'s API offers three services, each of them with a minimum number of words to work with properly (though the services still work with less words). *Receptiviti* analysis (emotion detection) for example, recommends a minimum of 300 words. Besides, before requesting an analysis, developers must specify the kind of text they are going to send: social media text, product reviews, etc. Although this service supports both English and Spanish, emotion scores are only returned for English input. The information returned by the API, along with the range of each score, includes not only emotions detected, but hints about the writer's personality. *Receptiviti* can be used through API calls or their web interface, which makes experimenting an easy task. *Receptiviti* offers a free powerful demo for a month, but prices for longer licenses are not available.

4.3 BiText

Another option in this field is *BiText* [10]. Founded on 2007 and settled in San Francisco and Spain, *Bitext* offers a service to detect emotions on text, and it is very focused on

marketing and customer satisfaction. Combining machine learning techniques with deep linguistic analysis [9], *Bitext* obtains high values of accuracy. They also offer several case studies to try their technology. *Bitext* API distinguishes English, Catalan, German, French, Italian, Dutch, Portuguese and Spanish and offers four services: sentiment analysis, concepts extraction, entities extraction and categorization [8]. For detecting the emotion in a bunch of text, *Bitext* API (sentiment analysis) chops the text received into pieces (sentences) and analyses the topic developed in that sentence and the valence of the emotion expressed. One sentence can contain several topics and emotions. *Bitext* offers a free license which allows up to 100 daily requests, but also three paid options for bigger request volumes.

4.4 Synesketch

Synesketch [20] is a free open-source software for textual emotion recognition, standing out because of their artistic way to express the detected emotions. *Synesketch* expresses the detected emotions using Ekman's basic emotions, each of them with its associated weight, and valence (positive or negative). *Synesketch* detection is based on keyword finding, sharing the same spirit that the other technologies.

Synesketch is a Java SDK which runs on the device where it is downloaded, thereby permitting developers to build a server with *Synesketch* so it can return results worldwide. The SDK is ready to generate results on text files, but this can be changed.

4.5 Comparative study and discussion

Due to the big presence of social media and writing communication in the current society, this field is, along with emotion detection from facial expressions, one of the most attractive to companies: posts from social media, messages sent to "Complaints" section, etc. Companies which can know how their customers are feeling have an advantage over companies which cannot. Table 4 shows a comparative study of some of the key aspects of each technology. The only application that has not been fully tested is *Synesketch*, since it required a complex coding to be correctly tested. The rest of APIs have been tested using the trial versions provided by each company. It is remarkable that as far as text is concerned, most of the companies offers a demo o trial version on their websites, while companies working on face or voice recognition are less transparent in this aspect. Regarding their accuracy, the four technologies have shown good values. On the one hand, *BiText* has proved to be the simplest one, as it only informs if the emotion detected is good or bad. This way, the error threshold is wider and provides less wrong results. On the other hand, *Tone Analyzer* has proved to be less

clear on its conclusions when the text does not contain some specific key words.

Table 4. Comparison of emotion detection technologies from text

Name	API/ SDK	Requires Internet	Information returned	Difficulty of use	Free Software
Tone Analyzer	API	Yes	Emotional, social and language tone	Low	No
Receptiviti	API	Yes	See [28]	Low	No
BiText	API	Yes	Valence (Positive/Neg ative)	Low	No
Synesketch	SDK	No	Six basic emotions	Medium	Yes

As far as the completeness of results is concerned, *Receptiviti* has been the one giving more information, revealing not only affective information but also personality-related information. The main drawback is that all these technologies (except *Synesketch*) are paid services, and may be not accessible to everyone. Since *Synesketch* is not as powerful as the rest, it will require an extra effort to be used.

5 EMOTION FROM BODY GESTURES AND MOVEMENT

Even though people do not use it to communicate information in an active way, their body is constantly broadcasting affective information. Tapping with the foot, crossing the arms, tilting the head, changing our position a lot of times while being sat, etc. Body language reveals what a person is feeling the same way our voice does [22].

However, this field is pretty new and there is not a clear understanding about how to create systems which read emotions in a body gesture. Most researchers have focused on facial expressions (over 95 per cent of the studies carried out on emotions detection have used faces as stimuli), almost ignoring the rest of channels through which people reveal affective information [19].

Despite the newness of this field, there are several proposals focused on recognizing emotional states from body gestures and use these results for other purposes. Experimental psychology has already demonstrated how certain types of movements are related to specific emotions [26]. For example, people experimenting fear will turn their bodies away from the point which is causing that feeling; people experimenting happiness, surprise or anger will turn their bodies towards the point causing that feeling.

Since there are no technologies available for emotion detection from body gestures, there is not consensus about what data is needed to detect emotions. Usually, experiments on this kind of emotion detection use frameworks (as for instance, *SSI*) or technologies to detect

the body of the user (as for instance, Kinect), so the researches are responsible for elaborating their own models and schemes for the emotion detection. These models are usually built around the joints of the body (hands, knees, neck, head, elbows…) and the angle between the body parts that they interconnect [30], but in the end, it is up to the researchers.

6   EMOTION FROM PHYSIOLOGICAL STATES

Physiologically speaking, emotions originate on the limbic system. Within this system, the amygdala generates emotional impulses which create the physiological reactions associated with emotions [4]. Electric activity on face muscles, electrodermal activity (also called galvanic skin response), pupil dilatation, breath and heart rate, blood pressure, brain electric activity, etc. Emotions leave a trace on the body, and this can be measured with the right tools.

Nevertheless, information coming directly from the body is harder to classify, at least with the category system used in other emotion detection technologies. When working with physiological signals, the best option is to adopt a classification system based on a dimensional approach [27]. An emotion is not just "happiness" or "sadness" anymore, but a state determined by various dimensions, like valence and arousal. It is because of this that the use of physiological signals is usually reserved for research and studies, like, for example, related to autism. There are not emotion detection services available for this kind of detection based on physiological states, although there are plenty of economical sensors to read these signals.

7   CONCLUSIONS

Emotion detection, together with Affective Computing, is a thriving research field. Few years ago, this discipline did not even exist, and now there are hundreds of companies working exclusively on it, and researchers investing time and resources on building affective applications. However, emotion detection has still many aspects to improve in the future years.

Applications which extracts information from the voice needs to be able to work in noisy environments, to detect subtle changes, maybe even to recognize words and more complex aspects of the human speech, like sarcasm.

The same applies for applications that extracts information from the face. Most people use glasses nowadays, and thus, it can complicate the face detection greatly.

Applications reading body gestures do not even exist right now, even though it is a source of affective information as valid as the face. There are already applications which can detect the body (Kinect) but there is not any technology like *Affectiva* or *Beyond Verbal* for the body yet. Physiological signals are even less developed,

because of the imposition of sensors that this kind of detection needs. However, researchers [12] are working on this issue so physiological signals can be used as the face or the voice. In a not too distant future, reading the heartbeat of a person with just a mobile with Bluetooth may not be as crazy as it may sound.

There are other ways to extract affective information we have not considered yet [29]. Previous technologies analyse the impact of an emotion in our bodies, but, what about our behaviour? A stressed person has a tendency towards making more mistakes. In the case of a person interacting with a system, this will be translated as faster movements in the interface, more mistakes when selecting elements or typing, etc. This can be logged and used as another indicator of the affective state of a person.

All these technologies are not perfect. Humans can see each other and estimate how other people are feeling within milliseconds, and with a small threshold error, but these technologies only can try to figure out how a person is feeling according to some input data. To get more accurate results, more than one input is required, so multimodal systems are the best way to warranty the highest precision of results.

"Union means strength" is a saying that also fits in emotion detection field. Human interaction is, by definition, multimodal [30]. Unless the communication is through phone or text, people can see the face of the people they are talking to, listen their voices, see their body, etc. Humans are, at this point, the best emotion detectors as we combine information from several channels to estimate a result. That is how multimodal systems work. It is important to remark that a multimodal system is not just a system which takes, for example, affective information from the face and from the voice and calculates the average of each value. The hard part of implementing one of these systems is to combine the affective information correctly. E.g., a multimodal system combining text and facial expressions that detects a serious face and the message "it is very funny" will return "sarcasm/lack of interest", while the result of combining these results in an incorrect way will return "happy/neutral". It is proven that by combining information from several channels, the accuracy of the classification improves significantly.

Although the accuracy of multimodal systems is better than the accuracy of systems using affective information from one source, there are no services of this kind at the moment (beyond the framework *SSI* [32]), but we have the individual services at our disposal to be combined.

The interest of companies about the possibility of collecting affective information from their clients has produced a boost to this field. However, this growth has a strong economic interest behind, as these services are rarely available for free. Even though trial versions and demos can be enough for a test, they may not be enough for

researches trying to create affective applications. For this reason, a stronger presence of researchers in this field is needed.

REFERENCES

[1] Affectiva. 2017. About us – Affectiva. Retrieves from: http://www.affectiva.com/who/about-us/ [Accessed May 2017]

[2] Affectiva. 2017. Affectiva Developer Portal - Pricing. Retrieves from: http://developer.affectiva.com/#pricing [Accessed 11 May 2017]

[3] Affectiva. 2016. The Emotion Behind Facial Expressions. Retrieves from: http://blog.affectiva.com/the-emotion-behind-facial-expressions [Accessed May 2017]

[4] alive Editorial. 2015. Emotions and Physiology. Retrieved from: http://www.alive.com/health/emotions-and-physiology/ [Accessed May 2017]

[5] Beyond Verbal. 2017. Beyond Verbal Developers site/ api. Retrieved from: http://developers.beyondverbal.com/Home/api [Accessed May 2017]

[6] Beyond Verbal. 2017. Beyond Verbal – the emotions analytics company. Retrieved from: http://www.beyondverbal.com/ [Accessed May 2017]

[7] H. Binali and V. Potdar. 2012. Emotion detection state of the art. In *Proceedings of the CUBE International Information Technology Conference* (CUBE '12). ACM, New York, NY, USA, 501-507. DOI: http://dx.doi.org/10.1145/2381716.2381812

[8] Bitext. 2017. Bitext API. Retrieved from: https://api.bitext.com [Accessed May 2017]

[9] Bitext. 2017. Machine Learning & Deep Linguistic Analysis in Text Analytics. Retrieved from: https://blog.bitext.com/machine-learning-deep-linguistic-analysis-in-text-analytics [Accessed May 2017]

[10] Bitext. 2017. Sentiment Analysis. Retrieved from: https://www.bitext.com/sentiment-analysis/ [Accessed May 2017]

[11] S. Casale, A. Russo, G. Scebba and S. Serrano. 2008. Speech Emotion Classification Using Machine Learning Algorithms. In *IEEE International Conference on Semantic Computing*. Santa Clara, CA, 158-165. DOI: 10.1109/ICSC.2008.43

[12] A. Conner-Simons and R. Gordon. (2016). Detecting emotions with wireless signals. Retrieved from: http://news.mit.edu/2016/detecting-emotions-with-wireless-signals-0920 [Accessed May 2017]

[13] Good Vibrations. 2017. Good Vibrations Company B.V. – Recognize emotions directly from the voice. Retrieved from: http://good-vibrations.nl/ [Accessed May 2017]

[14] Steven Handel. 2014. Classification of Emotions. Retrieved from: http://www.theemotionmachine.com/classification-of-emotions/ [Accessed May 2017]

[15] IBM. 2017. Science behind the service – Tone Analyzer. Retrieved from: https://www.ibm.com/watson/developercloud/doc/tone-analyzer/science.html [Accessed May 2017]

[16] IBM. 2017. Tone Analyzer. Retrieved from: https://tone-analyzer-demo.mybluemix.net/ [Accessed May 2017]

[17] IBM. 2017. Understand your tone score – Tone Analyzer. Retrieved from: https://www.ibm.com/watson/developercloud/doc/tone-analyzer/understand-tone.html [Accessed May 2017]

[18] Kairos. 2017. Face Recognition, Emotion Analysis & Demographics. Retrieves from: https://www.kairos.com/ [Accessed May 2017]

[19] A. Kleinsmith and N. Bianchi-Berthouze. Affective Body Expression Perception and Recognition: A Survey. In IEEE Transactions on Affective Computing, vol. 4, no. 1. Jan.-March 2013. 15-33. DOI: 10.1109/T-AFFC.2012.16

[20] U. Krčadinac. 2016. Synesketch. Retrieved from: http://krcadinac.com/synesketch/ [Accessed May 2017]

[21] Microsoft. 2017. Microsoft Cognitive Services – Emotion API. Retrieved from: https://www.microsoft.com/cognitive-services/en-us/emotion-api [Accessed May 2017]

[22] Mind Tools. 2017. Body Language Understanding Non-Verbal Communication. Retrieves from: https://www.mindtools.com/pages/article/Body_Language.htm [Accessed May 2017]

[23] MIT Media Lab. 2017. Affective Computing. Retrieved from: http://affect.media.mit.edu/index.php [Accessed May 2017]

[24] Ivan Morgun. 2015. Types of machine learning algorithms. Retrieved from: http://en.proft.me/2015/12/24/types-machine-learning-algorithms/ [Accessed May 2017]

[25] nViso. 2016. Artificial Intelligence Emotion Recognition Software – nViso. Retrieves from: http://nviso.ch/ [Accessed May 2017]

[26] S. Piana, A. Staglianò, F. Odone, A. Verri and A. Camurri. 2014. Real-time Automatic Emotion Recognition from Body Gestures.

[27] Rosalind W. Picard. 2009. Future affective technology for autism and emotion communication. In *Philosophical Transactions of The Royal Society B: Biological Sciences, 364* (1535). 3575–3584. DOI: 10.1098/rstb.2009.0143

[28] Receptiviti. 2016. Receptiviti. Retrieved from: http://www.receptiviti.ai/ [Accessed May 2017]

[29] Olga C. Santos. 2016. Emotions and personality in adaptive e-learning systems: an affective computing perspective. In *Emotions and Personality in Personalized Services* (pp. 263-285). Springer International Publishing. DOI: 10.1007/978-3-319-31413-6_13

[30] J. Tao and T. Tan. 2005. Affective computing: a review. In Proceedings of the First international conference on Affective Computing and Intelligent Interaction (ACII'05), Jianhua Tao, Tieniu Tan, and Rosalind W. Picard (Eds.). Springer-Verlag, Berlin, Heidelberg, 981-995. DOI=http://dx.doi.org/10.1007/11573548_125

[31] Universität Augsburg University. 2017. EmoVoice - Real-time emotion recognition from speech. Retrieved from: https://www.informatik.uni-augsburg.de/en/chairs/hcm/projects/tools/emovoice/ [Accessed May 2017]

[32] Universität Augsburg University. 2017. OpenSSI. Retrieved from: https://hcm-lab.de/projects/ssi/ [Accessed May 2017]

[33] Vogt T., André E., Bee N. 2008. EmoVoice — A Framework for Online Recognition of Emotions from Voice. In: André E., Dybkjær L., Minker W., Neumann H., Pieraccini R., Weber M. (eds) Perception in Multimodal Dialogue Systems. PIT 2008. Lecture Notes in Computer Science, vol 5078. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-540-69369-7_21

[34] Vokaturi. 2016. Vokaturi. Retrieved from: https://vokaturi.com/ [Accessed May 2017]

[35] J. Wagner, F. Lingenfelser, and E. Andre. 2011. The Social Signal Interpretation Framework (SSI) for Real Time Signal Processing and Recognitions. In *Proceedings of INTERSPEECH 2011*. Florence, Italy, 2011.