



## Effects of emotional prosody on skin conductance responses in French

Caterina Petrone<sup>1</sup>, Francesca Carbone<sup>2</sup> & Maud Champagne-Lavau<sup>1</sup>

<sup>1</sup> Aix-Marseille Université, CNRS, LPL UMR 7039, 13100, Aix en Provence, France

<sup>2</sup> Università degli Studi di Napoli «Federico II», Napoli, Italy

caterina.petrone@lpl-aix.fr, carbonefrancesca@hotmail.it

maud.champagne-lavau@univ-amu.fr.

### Abstract

This pilot study investigates the effects of emotional prosody in French, as resulting by listeners' electrodermal activity. Differently from responses to standard perception tasks, skin conductance responses (SCRs) are automatic, thus allowing evaluating spontaneous reactions of subjects to external stimuli in a non-invasive way. Based on an identification task on 20 listeners, a set of 4 sentences was selected for the skin conductance study. The sentences were composed of words whose meanings were not emotionally laden. They were uttered with four prosodic patterns each conveying four basic emotions (neutral, joy, anger, sadness). The corpus included 36 stimuli, i.e., 4 natural stimuli and 32 stimuli in which the tempo and pitch range of the whole utterances were independently manipulated. In the skin conductance study, ten listeners rated the arousal and valence of each stimulus a 5-points Likert scale. At the same time, SCRs were collected. Results collected so far indicate that: (1) emotional prosody has an effect on the peripheral nervous system activity, even in absence of visual cues (e.g., images, facial expressions); (2) amplitude of SCRs is triggered by auditory stimuli varying both in valence and intensity and (3) SCRs are modulated by manipulation of the prosodic cues.

**Keywords:** emotional prosody, tempo, pitch range, Skin Conductance Response, French.

### 1. Introduction

When listening to our interlocutors, we use prosodic information not only for, e.g., parsing the continuous speech stream or recover the most prominent words. Prosodic characteristics of an utterance may also convey speakers' emotional state and listeners can experience an emotional response to them. We refer to this function as "emotional prosody". Emotional prosody might involve changes in listeners' behavior, even in absence of face-to-face interactions. For instance, angry prosody might signal a potential danger for the listener, who might rapidly adapt his behavior to the threatening situation [1]. A sad or a happy prosody can lead listeners to experience immediate emotional reactions such as tears.

A great deal of work has been carried out so far on prosodic correlates of emotions, such as pitch, intensity, voice quality and speech rate [2, 3]. Research has also discussed such prosodic characteristics along two main dimensions organizing emotional responses, i.e., valence (the intrinsic attractiveness or aversiveness of an emotion) and arousal (a state of physiological activation, i.e., how intensely an emotion is felt). For instance, it has been found that prosodic characteristics

can be used by listeners to predict whether the speaker is about to give a happy or sad news, even in absence of lexical information [4, 5]. Based on the level of arousal, researchers have distinguished two main prosodic patterns [6]. The first pattern is characterized by high average of pitch, wide range and variability of pitch, high intensity and high speech rate, is typical of emotions with high arousal, like joy and anger. The second is characterized by low fundamental frequency, range and variability of intonation narrow, weak intensity and reduced speed, is typical of emotions with low arousal, like sadness [7, 8, 2, 4]. It is possible that emotions which have the same level of arousal, share acoustic characteristics like pitch range and speech rate [9, 7].

In psychophysiological literature, it has been found that emotional arousal of the listeners is typically accompanied by changes at the level of the autonomic nervous system (ANS) such as heart rate, breathing and skin conductance responses (SCRs). In particular, SCRs are abrupt changes in skin conductance as an immediate response to, e.g., external stimuli or events ("phasic SCR"). SCR depends on the activity of sweat glands innervated by the ortosympathetic branch of the ANS. The most frequently used measure of phasic SCRs is amplitude [10], which is measured from the baseline (level of skin conductance in absence of any external stimulus) to the peak (maximum value of SCR which takes place after presentation of an external stimulus and during a specified time). Another standard measure of SCRs is latency, which is the period between the stimulus onset and the SCR initiation. The latency window is usually between 1 and 3 seconds. SCRs are a reliable measure of emotional arousal in different domains than prosody, such as the visual domain and music [11]. These studies show that SCR amplitude increases with stimulus' arousal and that SCR amplitudes are higher in pleasant/unpleasant stimuli than in neutral ones. However, research on the effects of speakers' emotional prosody is rather scarce [1, 12].

In this study, we explore for the first time the effects of emotional prosody and their prosodic cues on SCRs in French. We focus on the four basic emotions of Power & Dalglish's model [13], i.e., fear, anger, happiness and sadness (see section 2) as they engage emotional responses varying in both valence (unpleasant vs. pleasant) and arousal (high vs. low). Two experiments were carried out. In an identification task, listeners were asked to identify the emotion conveyed by natural stimuli varying only in their prosodic pattern. The best identified stimuli were used for the second experiment, an EDA study. We hypothesized that SCR amplitude will be higher when listening to angry and happy prosody (as they are associated to emotions with high arousal level) than when listening to sad prosody (as it is associated to an emotion with

low arousal level). A secondary aim of this experiment is to explore the contribution of tempo (in terms of speech rate) and pitch range on SCRs. The relevance of these two acoustic parameters is verified by their independent manipulation. We expect higher SCRs amplitude as a reaction to enhancement of both cues.

## 2. Methods

The stimuli used in the physiological experiment were selected from a small database of natural utterances produced by two French native speakers. The selection was based on the results of an identification task which was conducted prior to the skin conductance study.

### 2.1. Identification task

This study helped in the selection of the material. Listeners had to identify sentences produced with different emotional prosodic patterns in a forced choice task.

#### 2.1.1. Corpus

The material was recorded in the anechoic room of the *Laboratoire Parole et Langage* (LPL). Two native female speakers of Southern French, i.e., a phonetician with a long training on prosody (Speaker 1) and a professional actress (Speaker 2) produced 3 sets of utterances composed of words whose meanings were not emotionally laden (e.g., set 1: *Amélie a lu la lettre*, “Amélie read the letter”, set 2: *Denise va parler avec lui*, “Denise is going to speak with him”; set 3: *Vincent va venir ici* “Vincent is coming here”). Speakers produced them in a way that they intended to convey 6 emotions by means of prosodic variation (fear, joy, sadness, anger, surprise and neutral state). Sentences uttered with a neutral prosody served as a reference production for each speaker. Utterances conveying surprise were inserted as fillers to balance the number of positive and negative emotions, but they were not considered for the skin conductance study because produced with a different sentence type (incredulity questions). Each target utterance was preceded by an appropriate context aimed at eliciting the 6 emotions on the target sentences. Speakers read silently the contexts and then uttered the target sentences. In total, we collected 36 sentences (3 sets X 2 speakers X 6 emotions).

#### 2.1.2. Procedure

Twenty native speakers of Southern French (10 F and 10 M), aged 20-50 y.o., were recruited for the experiment. They had to decide on the emotion conveyed by the speaker by pressing one of six keyboard buttons corresponding to the list of the six emotions appearing on the computer screen. The auditory stimuli were played from a laptop through the software PERCEVAL [14]. Each stimulus was heard one times in 3 randomized blocks. The statistical analysis was performed on 720 obs. (20 listeners \* 36 sentences). It included a series of logit models with mixed effects, in which the factors EMOTION (6 levels: neutral, happy, angry, sad, afraid) and SET (3 levels: set1 “Amélie”, set2 “Denise”, set3 “Vincent”) and SPEAKER (2 levels: speaker1, speaker2) were the fixed factors and LISTENER was the random intercept. Likelihood-ratio tests were run comparing full models (e.g., which contained a fixed factor) with simpler ones (e.g., without that factor). This allowed estimating the main effects of the factors. The cutoff for significance is  $p < .05$ .

### 2.1.3. Results

The histograms show the perception of correct identification (y-axis) by the six emotions (x-axis), split by sentence set and speaker. Globally, emotions were relatively well identified both when uttered by Speaker 1 (mean across emotions and sentence sets: 64.4%) and by Speaker 2 (mean: 60.5%). However, there is strong variation in listeners’ identification scores for each speaker across emotions and sentence sets. A series of likelihood ratio tests showed that dropping out the factor EMOTION ( $\chi^2 = 212.12$ ,  $p < .001$ ), SPEAKER ( $\chi^2 = 53.16$ ,  $p < .001$ ), SET ( $\chi^2 = 172.41$ ,  $p < .001$ ) as well as their interactions ( $\chi^2 = 149.59$ ,  $p < .001$ ) led to a significant decrease in model fitting, confirming significant main effects and interactions among the fixed factors. Sentences with neutral prosody were better identified for Speaker 1 (76.6%) than for Speaker 2 (58.3%) across sentence sets. The correct identification of neutral prosody was a crucial criterion to choose utterances by Speaker 1 for the skin conductance study. Indeed, the maximal amplitude of the SCR elicited by the other emotions is expected to be higher in comparison of that elicited by neutral prosody. For Speaker 1, SET 2 (based on the utterance *Denise va parler avec lui*) was the best identified sentence set. Hence, it was chosen as base for the corpus employed for the skin conductance study experiment. From this set, we discarded the stimulus uttered with fear, since listeners were not able to safely identify the intended emotion. To sum up, we selected 4 utterances from the SET2 and uttered by Speaker 1. These utterances were produced with angry, happy, sad and neutral prosody. The pitch range and duration of these stimuli is illustrated in Table 1.

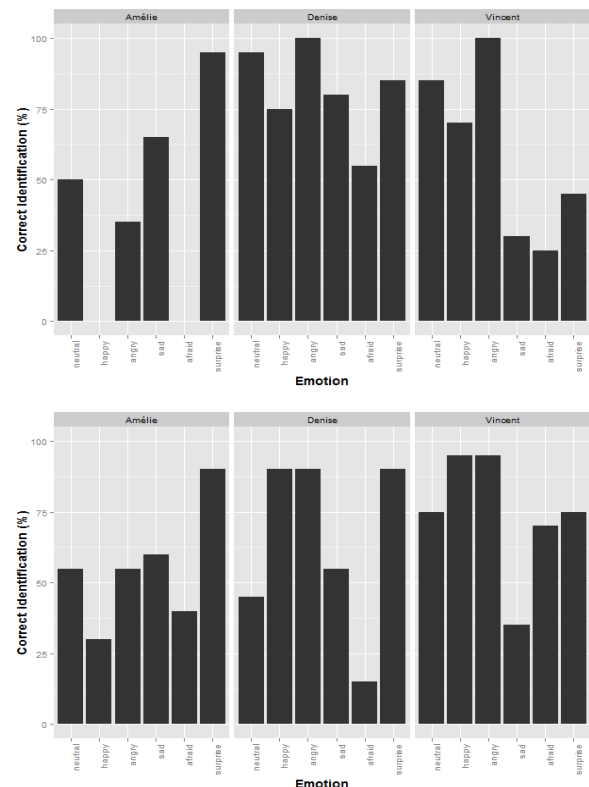


Figure 1. Participant’s responses for utterances produced by speaker1 (top) and speaker2 (bottom), split by sentence set and emotion.

Table 1: *Prosodic correlates of sentence “Denise va parler avec lui” uttered by phonetician.*

| Emotion   | F0 range (st) | Sentence duration (ms) |
|-----------|---------------|------------------------|
| tristesse | 5.2           | 155                    |
| neutre    | 10.7          | 143                    |
| Joie      | 10.6          | 145                    |
| colère    | 11.07         | 118                    |

## 2.2. Skin Conductance Response (SCR)

The goal of this experiment is to determine whether different emotional prosodies trigger SCRs in absence of visual cues. A preliminary analysis of the effects of speech rate and pitch range on peripheral autonomic responses is also offered.

### 2.2.1. Corpus

The corpus included 36 stimuli, i.e., 4 natural utterances and 32 stimuli in which the tempo and pitch range of the whole utterances were independently manipulated. This low number of stimuli minimized possible habituation effects, which are known to reduce SCRs [15]. All manipulations were done by means of Praat [16] In particular, we modified: 1) only tempo (fast, normal, slow); 2) only pitch range (expanded, normal, reduced) 3) and both tempo and pitch range (resulting in four combinations). Stimuli with normal tempo and normal pitch range corresponded to the natural utterances. For tempo manipulations, we first calculated the difference between the natural utterances with the longest (which corresponded to the stimulus with sad prosody) and the shortest duration (which corresponded to the stimulus with angry prosody). This difference is 37 ms. Hence, we got for each utterance a factor to increase tempo by adding 40 ms from the duration of the natural utterances divided by the duration of natural utterances. We then performed the same manipulations to reduce tempo by subtracting 40 ms from the duration of natural utterances divided by the duration of natural utterances.

For pitch, we either increased or decreased the pitch range of each utterance by 3 st. This criterion was chosen for two reasons. First, by trying to applying higher values than 3st, we changed the phonological make up of the sentences, thus resulting in unnatural sounding stimuli. Also, this value is midway between the highest (angry: 11.07 st) and the lowest pitch range (sad: 5.2 st) in the natural utterances.

### 2.2.2. Procedure

Twenty native speakers of Southern French (16 F and 4 M), aged 20-50 y.o., who did not participate in the identification task, were recruited. They were asked to listen to the stimuli and to rate the arousal and valence of each stimulus on a 5-points Likert scale. Values ranged from 1 to 5 (with 1 meaning “very weak” and 5 “very strong” on the arousal scale; and with 1 meaning “very unpleasant” and 5 “very pleasant” on the valence scale). Sound presentation and rating task were accomplished through *Eprime 2.0*. [17]. At the same time, SCRs were recorded by means of the *Biopac MP36R* system and *Acqknowledge software 4.4* [18]. Evoked SCR were recorded with electrodes on the volar surface of the medial and distal phalanges of the fingers on the non-dominant hand. The experiment was run in a silent room and we asked the listeners to sit in a comfortable chair and to put on professional headphones. A short training session preceded the experiment,

in which the listeners had to rate 3 unrelated utterances. Listeners signed an informed consent form before the experiment.

### 2.2.3. Results

We report here preliminary results on ten listeners. The analysis of the remaining ten listeners is still in progress. Fig. 2 shows mean values of arousal (x-axis) and valence (y-axis) for the natural utterances, as rated by our listeners. Utterances with angry and happy prosody were rated more intense than utterances with neutral and sad prosody in the scale of arousal. In contrast, utterances with neutral and sad prosody were judged more pleasant than the one with angry prosody, while happy prosody had the highest score on the valence scale.

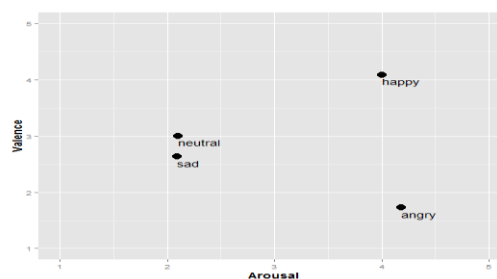
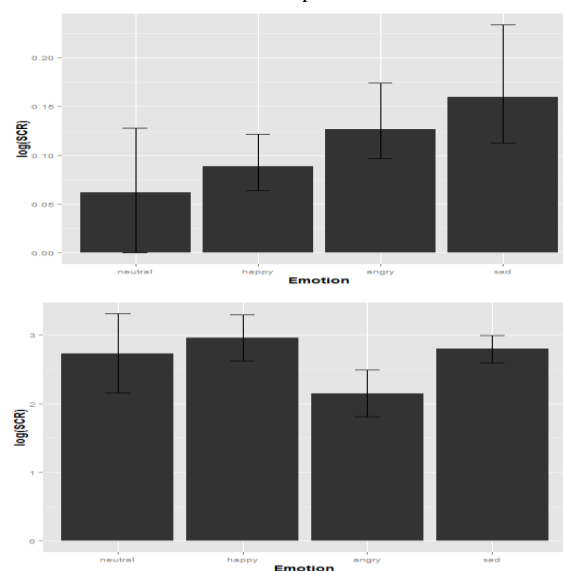


Figure 2. *Mean values of the rating task for the natural stimuli.*

Concerning SCRs, Fig. 3 and Fig. 4 show the amplitude and latency values for the natural stimuli split by emotion. Utterances with angry, happy and sad prosody triggered higher amplitude than the utterance with neutral prosody. The highest amplitude is associated with sad prosody, suggesting that valence plays a major role compared to arousal. Moreover, latency values are lower for angry prosody, meaning that this emotion entails a faster SCR response than the other emotions.



Figures 3-4. *Means and standard error for SCR amplitude (top) and latency (bottom) for the four natural utterances.*

The results described so far indicate that listeners have different peripheral responses on emotional prosody, but are these responses modulated by changes in the acoustic parameters of emotional prosody? Although we do not have

yet quantitative robust analyses, graphical exploration suggests that SCRs are sensitive to both tempo and pitch range manipulations. These results are illustrated in Fig. 5.

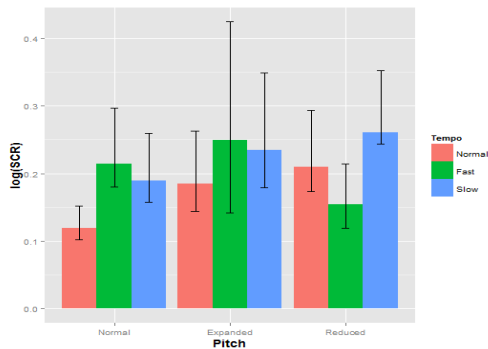


Figure 5. Means and standard error for SCR amplitude (top) and latency (down) split by tempo (top) and pitch (down) manipulations. Data are collapsed across emotions.

At a normal pitch range, we found higher SCR amplitude in stimuli with either fast or slow tempo than in stimuli with normal tempo. The SCR amplitude progressively increases when normal or slow tempo is combined with expanded or reduced pitch, with stimuli characterized by fast tempo and reduced pitch triggering the highest SCR amplitude value. An interaction between tempo and pitch range emerges for utterances with fast tempo, since reduced pitch triggers lower amplitude compared with normal pitch.

### 3. Discussion

Concerning the rating task, we found that that angry and happy prosody are rated more intense and less pleasant, while neutral and sad prosody were considered more pleasant but less intense. This is in line with e.g., literature on music perception [19]. As for the EDA study, we don't have yet quantitative robust analyses (the results so far are based only on ten listeners) which allow to draw conclusions. However, graphical exploration already suggests that skin conductance responses are sensitive measures of emotions induced by prosody. SCRs were in fact triggered in absence of any other verbal (e.g., emotionally laden words) and non-verbal (e.g., facial expressions) cues. Prosody could induce skin conductance responses depending on the underlying dimensions of emotions, i.e., valence and arousal. Specifically, sad, angry and happy prosody induced higher SCR amplitudes than neutral prosody, which was our baseline condition. The fact that sad and angry prosodies activated higher SCRs than happy prosody is quite surprising, since literature on music perception [e.g., 19], reported that emotions with high arousal level (happiness and anger) have a major impact on SCRs than emotions with low arousal level (sadness). Our preliminary result might be interpreted with respect to the study by Balconi on perception of affective pictures [21], who focused on anger and fear vs. happiness and connected this to "significance they have for the subjects' safeguard". He found higher SCR amplitude for visual stimuli with low valence (unpleasant pictures) than for stimuli with high valence (pleasant pictures) whatever the intensity of the stimuli. Moreover, we found an effect of the manipulation of the acoustic parameters on listeners' electrodermal activity. Speeding/slowing speech rate and expanding/reducing pitch range generally increased the

SCR peak. The effect was visible when speech rate and pitch range were presented both as single or combined manipulated cues. Interactions between these two cues are also possible, in that stimuli with fast speech and reduced pitch range showed lower SCR amplitude than the unmanipulated counterparts. Moreover, SCR latency values are lower for angry prosody, meaning that this emotion entails a faster SCR response than the other emotions. Though latency is a standard measure of SCRs, few studies have investigated the changes of SCR latency in relation to emotional dimensions. [19] found that there are no differences in SCR latencies when listeners are exposed to visual stimuli varying in valence, such as pleasant vs. unpleasant images. As far as we know, there are no studies on SCR latency in connection with arousal. There are at least two possible explanations to our results. First, anger is the negative emotion with the highest arousal. From both a phylogenetic and ontogenetic standpoint, angry prosody is of primary importance because it might signal a potential danger for the listeners. Listeners' rapid physiological adaptation is required to support rapid action preparation for the purpose of their own safeguard. Hence, it is possible that stimuli with angry prosody captured the highest degree of attention by the listeners. This might be independent of the involvement of voluntary processes, since SCRs are automatic. A second hypothesis concerns the phonetic and phonological aspects of the stimuli. It might be possible that stimuli uttered with angry prosody contained salient acoustic changes in pitch/rhythm (such as high pitch excursions or modulations in the durational patterns) or specific intonation contours (in terms, e.g., of number and type of pitch accents). Thus intrinsic prosodic properties might have drawn listeners' attention to stimuli with angry prosody compared to the other stimuli. Future experiments might help in deciding among these two hypotheses. For instance, the first hypothesis could be addressed by comparing different auditory stimuli signaling emotions with same arousal and valence as anger, such as fear. We should expect in fact similar SCRs latencies in both emotions. The second hypothesis might be addressed through a study in which the prosodic contour is more strictly controlled and acoustic variations in the signal are statistically correlated with SCRs. It is supposed that salient acoustic variations should affect SCR latency independent of the type of emotion.

Finally, we would like to point out that this study has shown for the first time the effects of different emotional prosodies and prosodic cues on peripheral autonomic system. Hence, we suggest that skin conductance responses are a valid methodology to investigate psychophysiological effects of emotional prosody. Differently from verbal activity, electrodermal activity is automatic, which allows evaluating listeners' spontaneous reactions to external stimuli in absence of voluntary processes. Moreover, this technique is non-invasive which is another possible advantage compared to methods which directly evaluate brain activity. Hence, it might be easily applied to larger samples of subjects and across different populations, e.g., to study the effects of emotional prosody on patients with disorders of emotional experience.

### 4. Acknowledgements

We thank Daniel Hirst for technical help with Praat. This study is also supported by a grant of the Agence Nationale de la Recherche (ANR 107923) to the first author and by an Erasmus Placement grant for mobility offered by Pr. F.M. Dovetto (University of Naples) to the second author.

## 5. References

- [1] T. Aue, C. Cuny, D. Sander, D. Grandjean, "Peripheral responses to attended and unattended angry prosody: A dichotic listening paradigm", *Psychophysiology*, vol. 48, 385-392, 2011.
- [2] R. Banse, K. R. Scherer., "Acoustic profiles in vocal emotion expression", *Journal of personality and social psychology*, vol. 70, no. 3, pp. 614-636, 1996.
- [3] K. Hammerschmid, U. Jürgens, "Acoustical correlates of affective prosody", *Journal of voice*, vol. 21, no. 5, pp. 531-540, 2007.
- [4] M. Swerts, J. Hirschberg, "Prosodic predictors of upcoming positive or negative content in spoken messages", *Journal of the Acoustical Society of America*, vol. 128, no 3, pp.1337-1344, 2008.
- [5] E. Rodero, "Intonation and emotion: influence of pitch levels and contour type on creating emotions", *Journal of voice*, vol. 25, no. 1, pp. e24-e34, 2011.
- [6] J. Pittam, K. R. Scherer., *Vocal expression and communication of emotion*, in M. Lewis, J. M. Haviland. (Eds.), *Handbook of emotions*. New York: Guilford Press, pp. 185-197, 1993.
- [7] K. R. Scherer, "Vocal affect expression: A review and model for future research", *Psychological Bulletin*, vol. 99, no. 2, pp. 143-165, 1986.
- [8] K. R. Scherer, "Vocal communication of emotion: A review of research paradigms", *Speech Communication*, vol. 40, 227-256, 2003
- [9] J. Bachorowski, "Vocal expression and perception of emotion", *Current directions in psychological science*, vol.8, no. 2, pp. 53-57, 1999.
- [10] W. Boucsein, *Electrodermal activity*, New York: Plenum Press, 1992.
- [11] S. Khalfa et al., "Role of tempo entrainment in psychophysiological differentiation of happy and sad music?.", *International Journal of Psychophysiology*, vol. 68, no 1, pp.17-26, 2008.
- [12] E. Bernat, C. J. Patrick, S. D. Benning, A. Tellengen, "Effect of picture content and intensity on affective physiological response", *Psychophysiology*, vol. 43, no.1, pp. 93-103, 2006.
- [13] M. Power, T. Dalgleish , *Cognition and Emotion. From Order to Disorder*. Erlbaum: Psychology Press, 1997.
- [14] C. André, A. Ghio, C. Cavé, B. Teston, "Perceval: a Computer-Driven System for Experimentation on Auditory and Visual Perception", *Proceedings of XVth ICPhS, Barcelone*, p. 1421-1424, 2003.
- [15] J. Braithwaite et al., "A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments", *Psychophysiology*, vol. 49, 1017-1034, 2013.
- [16] P. Boersma, "Praat, a system for doing phonetics by computer", *Glott International*, vol.5, no 9/10, pp. 341-345, 2001.
- [17] W. Schneider, A. Eschman, A. Zuccolotto, *E-Prime User's Guide*, Pittsburgh: Psychology Software Tools, Inc., 2002.
- [18] M. Balconi, E. Brambilla, L. Falco., " Appetitive vs. defensive responses to emotional cues. Autonomic measures and brain oscillation modulation.", *Brain Research*, vol. 1296, pp. 72-84, 2009.
- [19] C. Amrhein, et al., "Modulation of event-related brain potentials during affective picture processing: a complement to startle reflex and skin conductance response?.", *International Journal of Psychophysiology* , vol. 54, no.3, pp. 231-240, 2004.