Expert Systems With Applications 200 (2022) 117074

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Statistical emotion control: Comparing intensity and duration of emotional reactions based on facial expressions

F. Javier Otamendi *,1

Universidad Rey Juan Carlos, Brain Research Lab, Spain

A R T I C L E I N F O

Keywords:
Emotions
Biosensors
Statistical process control
iMotions
Emotion characterization

A B S T R A C T

The aim is to develop an intelligent automatic facial expression recognition and emotion analysis (AFEREA) algorithm that, first, characterizes the time-based raw signals of biosensors in quantitative indicators of the emotional state of the individuals participating in an experiment and, second, compares the emotional reactions across them in terms of intensity and duration. The proposed Statistical Emotion Control (SEC) intelligent algorithm is based on statistical process control (SPC) theory. After representing the individuals' baseline behaviour in a non-normal I-chart and describing the output per subject in emotional peaks with their corresponding duration in terms of relative cutoffs, SEC uses Poisson c-charts to compare across subjects in terms of the quantity of peaks and binomial p-charts in terms of length of the emotional reactions. To validate the data-driven algorithm, the state-of-the-art iMotions software and its AFFDEX face recognition and emotion analysis algorithm is used to record the individuals while receiving the results of their economic decisions when playing an experimental business game. The SEC intelligent algorithm is proven to be useful to take the raw output of the biosensors, to characterize the intensity and duration of the emotional reactions as well as to compare across subjects by emotion. SEC recognizes "out of control" negative emotions more often (7.25% vs. 2.00%) and positive emotions as often (15.63%) by setting relative cutoffs instead of traditional absolute thresholds. The results show significant pairwise discrepancies among both tested settings in 7.86% of the recorded 560 combinations of emotions and individuals, with a high 43.59% among those timeseries with the maximum recorded value above the traditional threshold of 50.

1. Introduction

1.1. Emotions and facial expressions

Emotional reactions measurement and analysis is one of the topics that has received a lot of attention in the last decade using one or several of the following techniques: (1) Eye Tracking (ET) (Wedel & Pieters, 2008; Ramsøy et al., 2012; De Oliveira et al., 2015), (2) analysis of facial micro-expression (Teixeira et al., 2012; Lewinski, Fransen, & Tan, 2014; Wedel & Pieters, 2013; Taggart et al., 2016), (3) functional magnetic resonance imaging (fMRI) (Bakalash & Riemer, 2013; Venkatraman et al., 2015; Couwenberg et al., 2017), (4) virtual reality (Bigné et al., 2016), and (5) electromyography (EMG) (Kulke et al., 2020; Xi et al., 2020).

Its applications in the last few years in diverse fields are numerous, for example:

- geriatric care (Taggart et al., 2016),
- forensics (Kielt et al., 2018; Lei et al., 2017),
- pain studies (Xu & de Sa, 2020),
- sport and physical exercise (Timme & Brand, 2020),
- the influence of negative emotions on driving (Braun et al., 2019), and
- consumer satisfaction from tourism (González-Rodríguez et al., 2020).

In this research, we focus on automatic facial expression recognition and emotion analysis (AFEREA) as a means to understand emotional reactions, with the aim of not only measuring and characterizing each individual's reactions but also comparing across individuals in any field of study.

The pillar of any AFEREA is FACS (Facial Action Coding System), a system that codes the micro-movements of the face in action units

* Address: Universidad Rey Juan Carlos, Facultad de Ciencias Jurídicas y Sociales, Paseo Artilleros s/n, 28032 Madrid, Spain.
E-mail address: franciscojavier.otamendi@urjc.es.
[1] ORCID: 0000-0001-6064-2724.

https://doi.org/10.1016/j.eswa.2022.117074
Received 13 September 2021; Received in revised form 16 March 2022; Accepted 28 March 2022
Available online 31 March 2022
0957-4174/

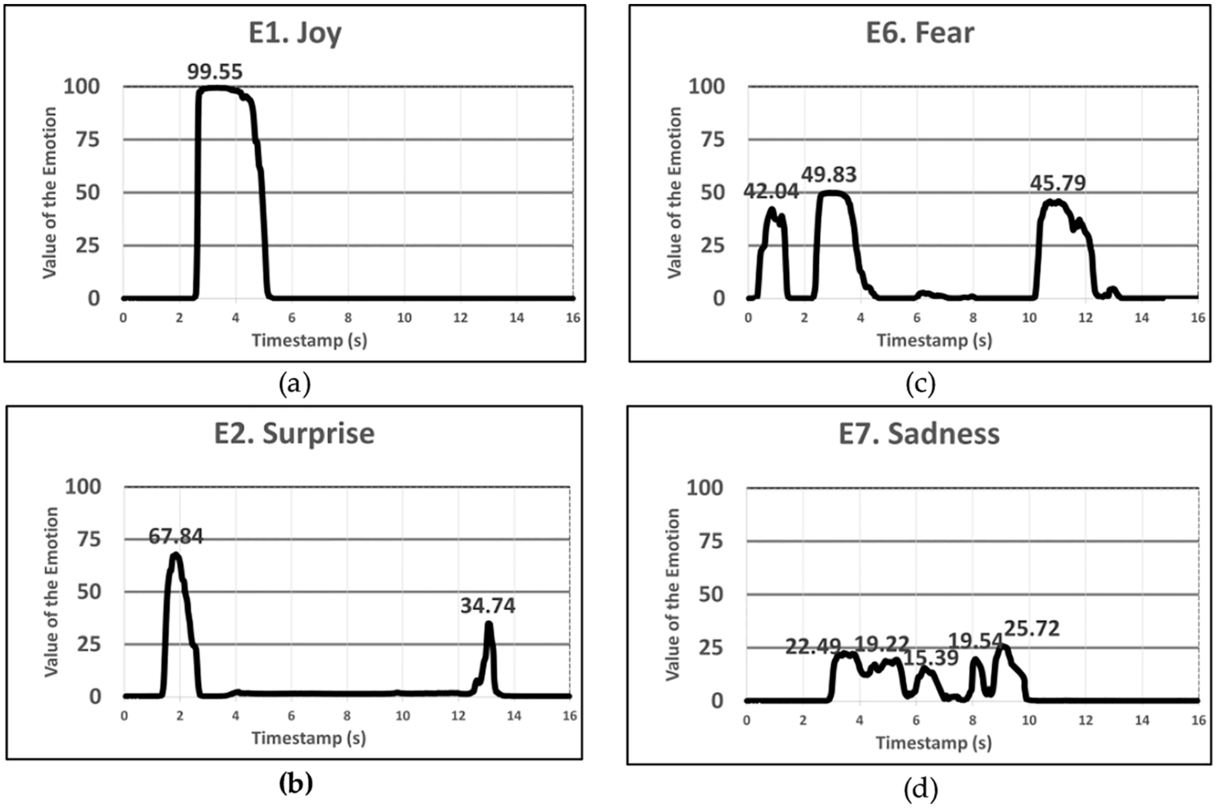


Fig. 1. Sample emotion timeseries: (a) Joy; (b) Surprise; (c) Fear; (d) Sadness.

(Ekman & Friesen, 1976, Ekman et al., 2002). These action units are readily converted into the so-called basic emotions (originally 6 (Ekman, 1972) and currently 7 (Ekman Group, 2021)): 2 positive ("E1. Joy" and "E2. Surprise") and 5 negative ("E3. Anger", "E4. Contempt", "E5. Disgust", "E6. Fear" and "E7. Sadness"). In other words, after recording along time the face of the participants while paying attention to a stimulus (face-to-face, video, audio…), the micro-movements of the muscles of the face are coded into emotions.

To make the process automatic, different software is available in the market, for example, FaceReader-FEBE system (Lewinski, Fransen, & Tan, 2014), GfK-EMO Scan software (Hamelin et al., 2017) and FACET and AFFDEX (McDuff et al., 2016; Stöckli et al., 2018) to carry the recording and to code the FACS and classify facial expressions into the 7 universal basic emotions. The result for each of the individuals is a set of timeseries that represents the reaction by emotion to the stimulus.

However, after the raw timeseries are available from the biosensor, the post-processing and the statistical analysis is left to the researcher. This analysis stage usually involves the comparison among individuals or their clustering to give answer to the questions that have been stated in the research using traditional statistics and related models, for example, ANOVA (Samant & Seo, 2020), regression analysis (Brand & Ulrich, 2019) or structural equation modelling (SEM) (González-Rodríguez et al., 2020).

In other words, the output per individual in the form of timeseries of action units and emotions is included in the biosensor that records the facial expressions and monitors each individual reactions, whereas the post-processing and statistical analysis to determine emotional reactions must be carried out in a case-by-case basis to answer the research questions and objectives.

1.2. Post-processing

Therefore, the individual reactions that are measured by the AFEREA biosensor(s) in the form of raw timeseries must be intelligently characterized with indicators that are robust and liable to be used to perform the statistical analysis. Intelligent algorithms and models must therefore be designed to be used in different contexts. The objective of this research is to design and test a novel algorithm that might be consistently used by researchers in many situations for characterizing one individual and for comparing among the group of participants in any emotional reactions experimental research.

To understand the problems associated to the time-based statistical analysis, and the necessity to develop reaction indicators and algorithms that post-process the raw timeseries, Fig. 1 includes 4 examples of the timeseries with emotional signals that have been recorded per individual during the experiment that has been used in this research. The seven basic emotions were recorded over time for 16 s for each subject. The graphs show the theoretical range of 0 to 100 (in intervals of 25 for displaying purposes) in which the AFEREA software codes the emotions from action units of facial movements. The local maxima are highlighted.

The shape of the timeseries is different in each case as a function of the number and duration of peaks of the emotions. The start, the length and the value of each peak are also different. And so are the maximum recorded values in each case. As such:

- Panel 2(a) represents "E1. Joy". The timeseries starts at 0 whereas the values range over the whole feasible range, from 0 to 100. The timeseries shows 1 peak above 50[2] with a duration of about 2.5 s.

[2] Absolute thresholds to determine presence of emotions are usually set. 50 is an arbitrary value for illustrative purposes and to understand the difficulty of characterizing the emotions within the algorithm. 50 is used since it is a value proposed by iMotions as default to determine if an emotion is present (iMotions, 2021). iMotions is the state-of-the-art AFEREA software that is used in this research.

- Panel 2(b) represents “E2. Surprise”. The timeseries starts at 0, and peaks twice, the first time over 50 for less than 1 s and the second time under 50 for about 0.5 s. The maximum is only 67.84, never reaching the theoretical maximum of 100.
- Panel 2(c) represents “E6. Fear”. The timeseries starts once again at 0, never reaches 50 (maximum of 49.83) and shows three significant peaks or waves, with the last one taking over 2 s.
- Panel 2(d) represents “E7. Sadness”. The timeseries starts at 0, and never reaches 50 (maximum of 25.72). It shows 5 peaks, 4 of them lasting around 1 s and the other, the one corresponding to 19.54, less than 0.5 s.

Intentionally, Fig. 1 includes four reactions that differ in terms of number and duration of peaks, as well as the value of the maxima. “E1. Joy” seems to be the strongest emotion while “E6. Fear” and “E7. Sadness”, although peaky, do not reach 50 in absolute terms.

Several technical and applied questions arise by comparing the panels of Fig. 1, questions that are the basis for this research on post-processing biosensor’s timeseries into emotional reactions:

- Is an emotional reaction higher than other just by being higher in absolute terms?
- Is it good to always set an absolute threshold? Should it be set at 50?
- Do the values cover the whole range between 0 and 100 for each and every subject and emotion? Or is there an experimental set-up influence (glasses, beard, gender, ethnicity…) (Magdin et al., 2019) that hinders the possibility of covering the whole range?
- Which of the emotions is more intense (peaky) or more durable?
- Does the initial emotional state of the individual, its baseline state at the beginning of the experiment, have an impact in the analysis of the timeseries?
- How can we compare among subjects and/or emotions?

These questions are partially answered in the literature in few articles with ad-hoc decisions, without demonstrating in any case the best strategy or values for the parameters of the analysis:

- Some use a baseline period (Kulke et al., 2020; Samant & Seo, 2020; Stöckli et al., 2018).
- Some others remove the first values from the analysis: 1 s (Stöckli et al., 2018) or 3 s (Mehta et al., 2021). Even others set a neutral image at the beginning of the experiment during 1 s (Ho et al., 2020) or 3 s (Brand & Ulrich, 2019; Stöckli et al., 2018).
- Some provide an indication of the reasonable length of exposure to the stimulus: 2 s (Ho et al., 2020), 4 s (Brand & Ulrich, 2019), 5 s (Taub, Sawyer, Lester et al., 2020), 10 s (Kulke et al., 2020) or 15 s (Samant & Seo, 2020).
- Some use a threshold, sometimes the threshold being absolute and some relative after normalizing the data. For example, and in terms of absolute thresholds, Triyanti et al. (2019) set two thresholds, 20 and 50, to discriminate between low (0–20), moderate (21–50) and high (51–100) intensity. Other subjective absolute settings are 10 (Brand & Ulrich, 2019; Timme & Brand, 2020) and 20 (Mele et al., 2019). For relative thresholds, Taub, Sawyer, Smith et al. (2020) propose a normalization step with 2 sigma limits.
- Some consider as indicator the maximum value over the whole timeseries (Stöckli et al., 2018), some the average (Kulke et al., 2020) and some others the percent of values above the threshold (Timme & Brand, 2020; Mele et al., 2019; Taub, Sawyer, Lester et al., 2020). Some subtract from the raw values the baseline average (Samant & Seo, 2020) or the baseline median (Stöckli et al., 2018).

There is no algorithm to our knowledge that generalizes the post-processing stage but, after summarizing the literature, some aspects are critical and need to be included in any algorithm for statistical analysis of the timeseries obtained from the biosensors:

- An estimation of the baseline state, so deviations from it mark the appearance of the emotional reaction
- A normalization of the timeseries so the ranges are comparable across individuals and emotions
- A selection of indicators to characterize each timeseries
- A threshold to determine significant values, with this threshold being relative or absolute

In fact, the main driver of this research is that data-driven relative measures that code the signals will bring further insights and understanding of the emotional states and reactions of each individual, facilitating the comparison across them in any field of study with face recording.

1.3. Objectives

This article includes a novel data-driven post-processing algorithm that provides additional knowledge to any AFEREA biosensor while characterizing the recorded timeseries of action units and the emotions in order to determine the intensity and duration of the extraordinary emotional reactions per individual, before comparing across subjects. The result is an expert and intelligent system to cope with the statistical analysis and comparison of emotional reactions that were captured with a face recognition system in different settings and applications.

The objective is therefore to develop an intelligent algorithm for the control and detection of emotional reactions through the analysis of facial expression software output, a post-processing algorithm that is adequate and robust to compare across individuals and emotions when different stimuli are present, a robust algorithm that might be used just by easily adjusting certain parameters in different research settings and for different biosensors.

The pillar of the intelligent post-processing algorithm is statistical process control (SPC), which is applied in industrial environments to control the process settings and detect deviations from in-control situations as soon as they appear (Montgomery, 2009). The SPC theory uses the so-called control charts based on confidence intervals to monitor the underlying baseline behaviour along time while counting the number and duration of “out-of-control” (or extraordinary or statistically different) situations.

If only baseline emotions could be controlled with this proven methodology, one that is used to count, record and normalize the data in such a way that relative indicators might be used to perform robust comparisons of “out-of-control” situations, facial expressions raw data could be post-processed so that comparison of emotional reactions across emotions and individuals is readily achieved.

The development of the Statistical Emotion Control (SEC) algorithm therefore uses SPC theory to calculate and graph indicators along time using control charts. This well-proven theory is used and applied for the first time to our knowledge to the control of baseline emotions and the detection of unexpected emotional reactions, through the application of 3 charts. The first control chart, called I-chart, is used to determine the baseline emotional state of the individual, that is, to determine normal “in-control” behaviour. The second chart, called c-chart, is used to calculate the number of waves (or count of peaks of emotions above the emotional baseline state) to provide an indication of intensity of the emotions (which are obviously “out-of-control” states). The third chart, called p-chart, is used to calculate the percent of “out-of-control“ recorded values (or percent of frames above the baseline state), providing an indication of duration of emotions. SEC might then be considered a novel algorithm that adds intelligence to AFEREA raw biosensor data.

This article is structured as follows. Section 2 explains the development of the Statistical Emotion Control (SEC) algorithm after introducing Statistical Process Control (SPC). Section 3 is devoted to the experimental setting that is used to validate SEC, which includes a business game and the state-of-the-art iMotions software as an example

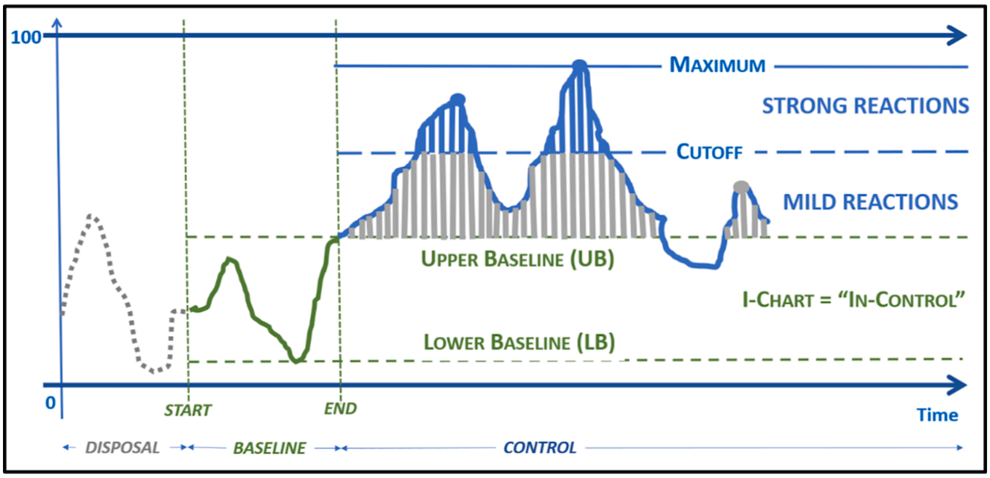


Fig. 2. Timeseries and its characterization.

of facial expression analysis software. Section 4 shows the results of the validation and Section 5 discusses about the research outcome.

2. The Intelligent System: Statistical Emotion Control (SEC)

Individuals and their emotional state are indeed non-industrial processes, but the analogy between monitoring and controlling industrial parts and individual emotional reactions is astonishing. If all the individuals are assumed to be in a neutral, identical emotional state of "in-control" when an experiment starts, the so-called BASELINE[3], the aim is to determine "out-of-control" individuals, in the sense that some emotional reactions might be recorded outside their basic emotional state when faced with different stimuli. Therefore, the proposed algorithm should first detect for each subject those extraordinary emotional reactions and characterize them in indicators of occurrence, intensity and duration. Second, the algorithm should compare across subjects in terms of overall intensity and duration.

More specifically, the proposal is a two-stage intelligent algorithm for the analysis of emotional reactions based on the combined use of 3 SPC control charts:

- Characterization of individuals and emotions
  - o Set the stable baseline emotional state of an individual or the "in-control" state, using the **I chart**.
  - o Set the indicators for "out-of-control" situations on the I-chart. In other words, values outside the limits of the control chart should be converted into indicators that characterize the emotional reaction by **counting the peaks and establishing the duration** of the "out-of-control" emotional reactions.
- Comparison across individuals
  - o Compare the number of peaks across individuals using a **c-chart** and determine those subjects with extraordinary "out-of-control" behaviour, that is, those with more peaks that deemed standard by the chart.
  - o Compare the duration of peaks across individuals using a **p-chart** and determine those with extraordinary "out-of-control" behaviour, that is, those with more duration that deemed standard by the chart.

2.1. *Characterizing individuals and emotions*

The whole strategy to characterize emotional reactions is based on the concept of BASELINE emotional state and its proper specification using SPC principles. Then, deviations from this baseline must be detected and accounted for in terms of number of peaks and their duration.

2.1.1. *Setting the baseline, periods and zones: an I-chart based on raw values*

Fig. 2 includes the proposed I-chart that sets the "in-control" baseline state that will be used to determine "out-of-control" emotional reactions. The characterization strategy includes three time PERIODS along the x-axis as well as three detection ZONES over the y-axis.

2.1.1.1. X_AXIS: Time periods. In order to set the BASELINE behaviour via I-charts, three different periods are proposed to be set in any experiment along the x-axis or time axis:

- BASELINE (I-chart): period in which the first set of valid measures is recorded in order to calculate the underlying variability of the emotional state of the individual. These valid measures should be enough to calculate the upper and lower baseline, or the limits of the control chart that set the "in-control" state. The length of the period is constant between START and END across individuals and emotions.
- CONTROL (I-chart continuation): period that covers the rest of the experiment after the END of the BASELINE period. The measures are all valid records and are used to calculate statistically significant "out-of-control" peaks and their duration.
- DISPOSAL: before the START of the BASELINE period, a transitory period that allows for the facial recognition algorithm to reach a steady state for the individual being recorded. All the recorded values are not used at all. It should have the necessary duration to let the individual reach the baseline emotional state. Its length is constant for each and every recorded emotion and it is biosensor dependent.

The two critical values to determine the three periods are therefore the **START and END of the BASELINE** period that specifies the

[3] Baseline is the term used in psychology by APA as "1. data or information obtained prior to or at the onset of a study…2. any stable level of performance used as a yardstick to assess the effects of particular manipulations or changes" (https://dictionary.apa.org/baseline).

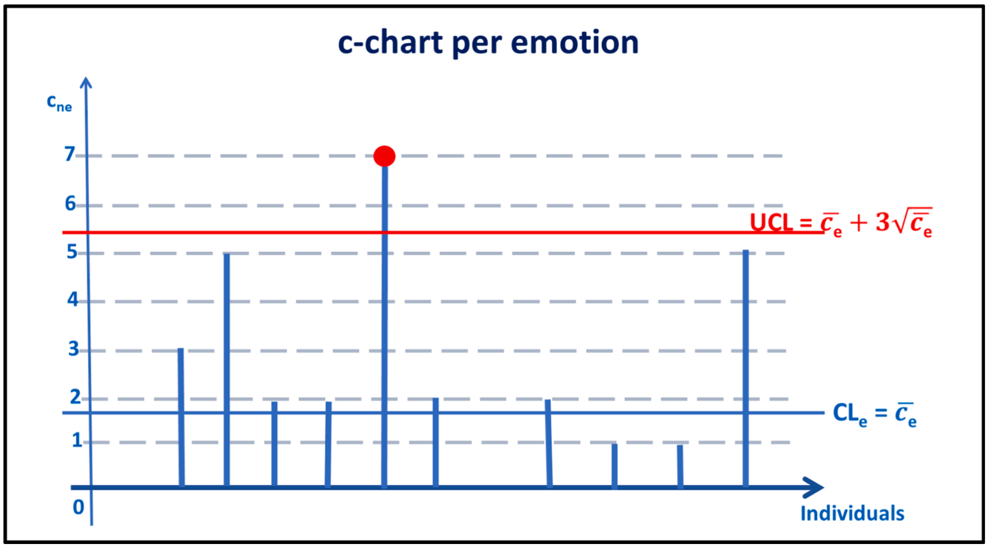


Fig. 3. A sample c-chart for a given emotion.

timeframe in which the I-chart control limits are set. Since each of the three periods must have enough measures, the recording rate of any AFEREA biosensor is critical to set minimum lengths for the periods. Frames are usually recorded at a rate between 10 and 20 FPS. We set the DISPOSAL period and the BASELINE period to at least 0.5 s each (Brand & Ulrich, 2019). Then, the length of the CONTROL period will be as long as the experimenter deems appropriate according to the objectives of the research.

The BASELINE period is obviously used to set upper and lower limits on the baseline since the baseline state of the subject is a random variable. Following statistical process control theory (SPC), the philosophy of the I-chart and its 3-sigma limits is used to calculate the boundaries to cover 99.73% of the normally distributed underlying distribution (Montgomery, 2009). In order to provide a robust algorithm, the normality assumption is relaxed, and the proposed SEC algorithm uses instead the 0.13th-percentile and the 99.87th-percentile of the recorded data in BASELINE to cover the same proportion of any underlying distribution. Therefore, the following definitions apply:

- Raw Upper Baseline: rUB = 99.87th-percentile of the raw BASELINE data
- Raw Lower Baseline: rLB = 0.13th-percentile of the raw BASELINE data

The values of the UPPER and LOWER BASELINES are further adjusted so that the upper baseline cannot be lower than a threshold (MINIMUM UPPER BASELINE, minUB, proposed to be at least 5), and the lower baseline cannot be higher than a threshold (MAXIMUM LOWER BASELINE, maxLB, proposed to be at most 95) to account for reliability of measurement and calibration.

- Upper Baseline: UB = max (rUB, minUB) = max (rUB, 5)
- Lower Baseline (LB) = min (rLB, maxLB) = min (uUB, 95)

After the BASELINE period, the CONTROL period is used to detect deviations from the baseline emotional state. It is in this period where extraordinary emotional reactions over the initial baseline state should be determined both in terms of the number of peaks (to measure intensity) or the percent of hits (to measure persistence or duration).

2.1.1.2. Y_AXIS: Value zones. The zone above the upper baseline along the CONTROL period goes up until the maximum value recorded. Since the height of this zone might be large compared to the baseline zone, the proposed option is to divide the range in two ZONES of the same height, by setting a CUTOFF (or relative threshold) that differentiates both zones[4]. This separation aims at addressing only strong emotional reactions. Therefore:

- MAXIMUM line: highest recorded value.
- CUTOFF: average between the upper baseline and the maximum = (UB + MAXIMUM)/2

Consequently, there are three resulting ZONES over the y-axis that represents emotion values:

- BASELINE zone: covers the baseline emotional state and ranges between the lower baseline and the upper baseline
- MILD REACTIONS zone: covers the range between the upper baseline and the cutoff.
- STRONG REACTIONS zone: covers the range between the cutoff and the maximum.

2.1.2. Counting peaks and hits

With all the periods and emotional zones relative to the "in-control" baseline I-chart in place, a set of indicators might be readily calculated to summarize the emotional reactions of an individual in the CONTROL period. The choice is to focus on the STRONG REACTIONS zone, although any researcher might as well focus on one or several zones. Three indicators are calculated per individual and emotion:

- StrongPeaks: count of local maxima that have been detected above the cutoff in the STRONG EMOTIONS zone (dark dots in Fig. 2).
- StrongHits: count of values that have been recorded above the cutoff in the STRONG EMOTIONS zone (dark bars in Fig. 2).
- StrongRate: out of the total records, percent of values (StrongHits) that have been recorded above the cutoff in the STRONG EMOTIONS zone (dark bars in Fig. 2).

Out of the proposed set of indicators, StrongHits (and therefore StrongRate) are easily accounted for by counting the number of recorded values that are above the cutoff. The determination of StrongPeaks

[4] Splitting into two zones of equal height is what iMotions proposes as well, by setting the threshold at 50, the midpoint between the expected minimum of 0 and the expected maximum of 100.

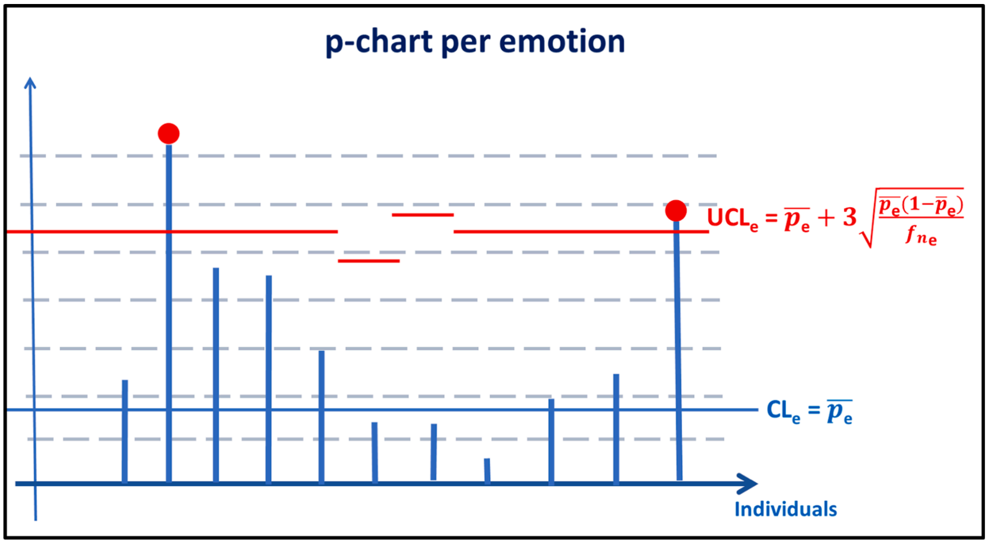


Fig. 4. A sample p-chart for a given emotion.

is performed by implementing an algorithm that compares neighbouring values, in this case 13 (6 before, the current value and 6 after). The choice of 13 is not random but designed to cover at least 0.5 s of the experiment (although it should be adjusted ad-hoc since its value depends on the frames per second that are recorded by the biosensor).

2.2. Comparing across individuals

These indicators of intensity (StrongPeaks or just PEAKS) and duration (StrongHits and StrongRate or just HITS and RATE) are the pillars to compare across individuals in terms of their emotional reactions after recording their facial micro-expressions over time. Once again, SPC theory and the control charts are used to determine those behaviours that are statistically different than the rest.

Out of the available SPC chart options, the choice is to jointly use c-charts and p-charts to characterize emotional reactions. If the number of "out-of-control" peaks of emotions is recorded, the c-chart is suited to represent the intensity of the emotional reactions. If the rate or percent of "out-of-control" hits is tracked, the p-chart is suited to represent the duration of the emotional reactions.

2.2.1. Intensity: A c-chart based on number of peaks

c-charts (Fig. 3) are used to monitor "number of defects", in this case, "number of peaks" or the number of local maxima per individual and emotion pinpointed by the peak detection algorithm. For a single emotion, the chart represents the number of peaks for each individual, in this case, with just 1 individual with more peaks that deemed appropriate (7 peaks, more than the calculated UCL≈5.5).

The number of peaks is well represented by a Poisson distribution, so the calculation of the control limits for the c-chart for a single emotion is as follows (Montgomery, 2009):

- Obtain the number of peaks for a given emotion e = {1, …,7} for each of the N individuals in the experiment, n = {1, …, N} and call them $c_{ne} = \{0, 1, 2, \dots\}$
- Calculate the average for all the individuals for a single emotion, and call it $AVG_e = \overline{c_e} = \frac{\sum_{n=1}^{N} c_{ne}}{N}$
- Calculate the standard deviation, and call it $STDEV_e = \sqrt{\overline{c_e}}$
- Calculate the statistics for 3-sigma limits as follows:
  o $CL_e = \overline{c_e}$
  o $LCL_e = \max\{0, \overline{c_e} - 3\sqrt{\overline{c_e}}\}$
  o $UCL_e = \overline{c_e} + 3\sqrt{\overline{c_e}}$

Then, those values c_{ne} that are outside the limits for that particular emotion are considered to be significantly different than the rest. Individual extraordinary intense emotional reactions are highlighted this way.

2.2.2. Duration: a p-chart based on percent of hits

p-charts (Fig. 4) are used to monitor "percent defective", in this case, "percent of hits" or the number of strong hits (values outside the cutoffs) divided by the number of recorded frames. In Fig. 4, for a single emotion, the chart represents the rate or percent of values above the cutoff for each individual, in this case with just 2 individuals with an "out-of-control" value (0.8 and 0.6, more than the calculated UCL ≈ 0.55). Notice that the control limits may vary per individual, since the number of valid recorded frames per individual might be different.

The percent of hits is well represented by a Binomial distribution, so the calculation of the control limits is as follows (Montgomery, 2009):

- Obtain the number of hits for each of the N individuals in the experiment for a given emotion e = {1, …,7}, h_{ne}, and divide it by the number of recorded frames, f_{ne}, and call them $p_{ne} = \{0,\dots,1\}$
- Calculate the average for all the individuals, and call it $AVG_e = \overline{p_e} = \frac{\sum_{n=1}^{N} h_{ne}}{\sum_{n=1}^{N} f_{ne}}$
- Calculate the standard deviation, and call it $STDEV_{ne} = \sqrt{\frac{\overline{p_e}*(1-\overline{p_e})}{f_{ne}}}$
- Calculate the statistics for 3-sigma limits as follows:
  o $CL_e = \overline{p_e}$
  o $LCL_{ne} = \max\{0, \overline{p_e} - 3\sqrt{\frac{\overline{p_e}*(1-\overline{p_e})}{f_{ne}}}\}$
  o $UCL_{ne} = \overline{p_e} + 3\sqrt{\frac{\overline{p_e}*(1-\overline{p_e})}{f_{ne}}}$

Then, those values p_{ne} that are outside the limits for a particular emotion are considered to be significantly different than the rest. Individual extraordinary durable emotional reactions are highlighted this way.

2.3. Flowchart

As a summary, the flowchart of the proposed intelligent SEC (Statistical Emotion Control) algorithm is the following:

- Obtain the results of the face recognition algorithm in a timeseries format, with the values ranging between 0 and 100.

- For each subject and emotion, calculate the number of peaks and rate of hits as follows:

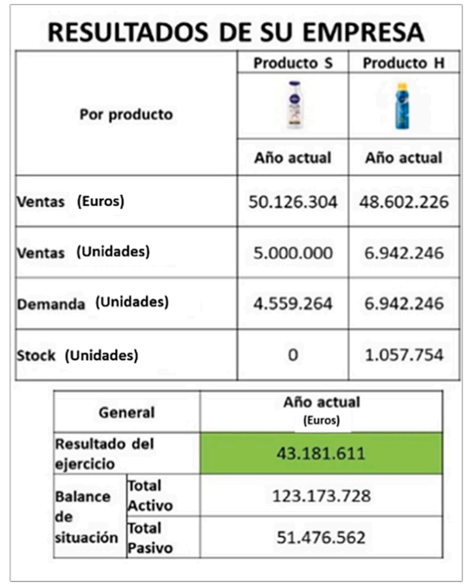
  X_AXIS.

- o Set the values of START and END to define the baseline period.
  Y_AXIS.
- o Eliminate the first milliseconds, those corresponding to the disposal period, that is, before the START of the baseline period.
- o Calculate the bounds of the baseline by using **non-normal I-charts**, that is, calculate the 0.13th and 99.87th percentiles of the data recorded between START and END of the baseline period.
- o The upper baseline should be at least the MINIMUM UPPER BASELINE, in this case, a value of 5 is proposed.
- o The lower baseline should be at most the MAXIMUM LOWER BASELINE, in this case, a value of 95 is proposed.
- o Calculate the MAXIMUM after the END of the baseline period.
- o Calculate the CUTOFF as the midpoint between the upper bound for the baseline and the maximum.
  INDICATORS.
- o Calculate the number of local maxima above the cutoff and in the control zone, and call it PEAKS.
- o Calculate the number of values above the cutoff and in the control zone, and call it HITS.
- o Calculate the number of records after the END of the baseline period.
- o Calculate the percentage of hits above the cutoff and in the control zone by dividing the number of hits over the number of records, and call it RATE.
- To compare among subjects within emotions
  - o Use a c-chart per emotion to detect extraordinary peaky, intense emotional reactions.
    - ▪ Calculate the upper control limit using PEAKS.
    - ▪ Count individuals with more peaks than the control limit.
  - o Use a p-chart per emotion to detect extraordinary durable emotional reactions.
    - ▪ Calculate the upper control limit using RATE or percent of hits.
    - ▪ Count individuals with more rate than the control limit.

Therefore, the parameters to be set are the START and END of the BASELINE period, as well as the CUTOFF (which is proposed to be the midpoint between the upper baseline and the maximum). Additionally, the MINIMUM UPPER BASELINE is proposed to be 5 and the MAXIMUM LOWER BASELINE to 95.

3. The experimental setting for validation

An experiment is set so it is used to validate the SEC algorithm. The experiment is based on a business game played on a computer over a 1-month period. The simulator called PRAXIS MMT (https://www.praxismmt.com/) is employed to perform a business game and iMotions (https://imotions.com/) is used to record the facial expressions of the participants while viewing the impact of their decisions on the screen after each game. The study aims at understanding the emotional reactions of the participating individuals while visualizing the effect of their economic decisions on the performance of the company they are virtually managing[5,6].



RESULTADOS DE SU EMPRESA

Por producto	Producto S	Producto H
	Año actual	Año actual
Ventas (Euros)	50.126.304	48.602.226
Ventas (Unidades)	5.000.000	6.942.246
Demanda (Unidades)	4.559.264	6.942.246
Stock (Unidades)	0	1.057.754

General		Año actual (Euros)
Resultado del ejercicio		43.181.611
Balance de situación	Total Activo	123.173.728
	Total Pasivo	51.476.562

Fig. 5. Screenshots with profits.

3.1. The business game

Business simulators are used primarily for training. Real environments are emulated so participants in business games can take economic decisions related to the supply chain and receive a feedback on how they are performing in terms of the P&L (Profit and Loss) statement or the Balance Sheet of the company they are managing.

Marketing students at a Spanish university were requested to participate via email. The business game was independently played twice, so two sets of measurements were available for each subject. The experiment was composed of 5 sessions:

- Session 1 (Feb 13th, 2019): The experiment was explained to the potential participants while showing a power-point presentation. Out of those that finally agreed to participate 40 were randomly selected. A confidential agreement form was signed prior to the start of the experiment. A sample game was carried to make the students familiar with the simulation environment. A password was then assigned to each of the participants and a link to the tool was provided so they could open the business game and prepare the decisions.
- Session 2 (Feb 20th, 2019): The students input their decision at one of the main laboratories of the campus. They all did it simultaneously within the maximum allotted time of 1 h.
- Session 3 (Feb 26th, 2019): Individually, the students looked at their results on the screen of the same computer so their behaviour could be recorded. The timeseries results were called "**Year1**".
- Session 4 (Mar 5th, 2019): The students input their second-round decisions, once again, at one of the main laboratories of the campus. They all did it simultaneously within the maximum allotted time of 1 h.
- Session 5 (Mar 10th, 2019): Individually, the students look at their results on the screen of the same computer so their behaviour could be recorded. The timeseries results were called "**Year2**".

The output image that is shown to the students after each game (Sessions 3 & 5) was depicted in tabular form (Fig. 5). For each product in the market, total sales (Ventas) both in euros and units were shown, as well as the demand figures (Demanda) and the final stock level (Stock). As for the company as a whole, the profits were highlighted and shown (Resultado del ejercicio) as well as the balance sheet totals (Balance de situación): Assets (Activo) and Liabilities and Equity (Pasivo).

[5] In fact, the aim is to understand their expectations of obtaining an economic profit. These expectations, defined as "a state of tense, emotional anticipation" (https://dictionary.apa.org/expectations), are one key individual characteristic that is measured using neuroscience (Delgado et al., 2008).

[6] The relationship between emotions and simulated business games is a current line of research (Bakker et al., 2011; Hühn & Rausch, 2022; Robson et al., 2015).

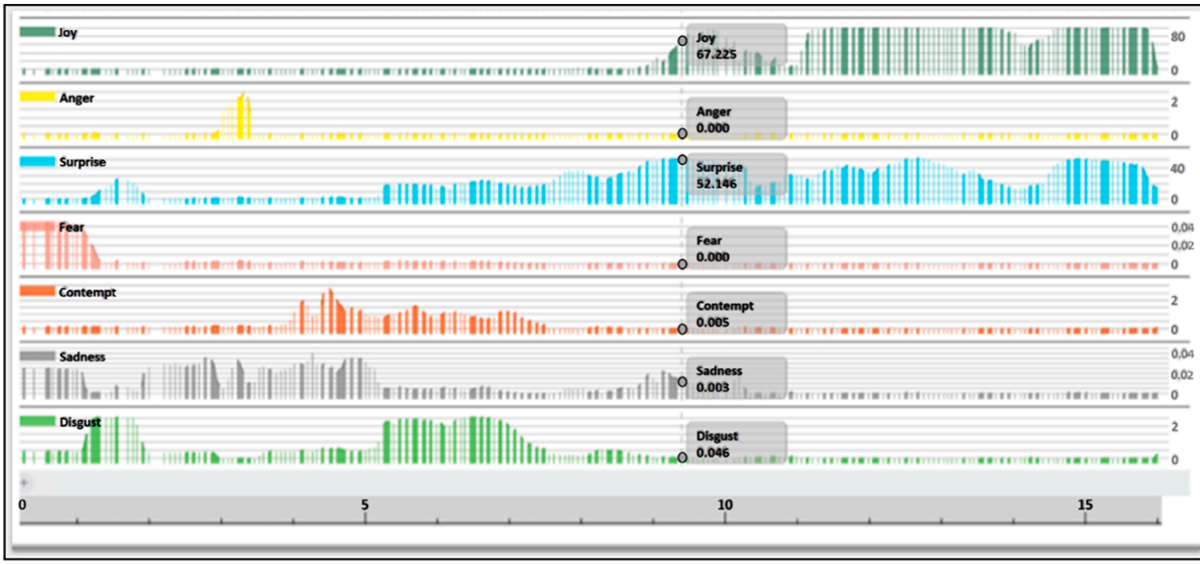


Fig. 6. iMotions screenshot.

For Session 3 and Session 5, the output screen in Fig. 5 was shown to the participants with their specific results for 16,000 ms or 16 s. Videos of subjects' faces were recorded at a frame rate of 12 frames per second.

The length of the stimulus is enough to set the DISPOSAL and BASELINE periods at 0.5 s each (Brand & Ulrich, 2019), and a CONTROL period with the remaining time of 15 s (Samant & Seo, 2020).

Two games ("Year1" and "Year2") were played by 40 subjects for a total of 80 sets of measurements. Besides, information about gender and economic results was also recorded.

3.2. The AFEREA Biosensor: iMotions

To carry out the emotional measurements in this study, and out of the available options in the market, a software platform for biometric measurements research called iMotions was used (iMotions, 2021). Version 7.0 was used in this research.

The software records biometric measurements or action units per frame while an experimental subject is watching the stimulus on the computer screen: 34 core facial landmarks (jaw, brows, nose…), inter-ocular distance and head position (yaw, pitch and roll).

The recorded values for the action units for each frame are then transformed into Ekman's 7 basic emotions (Ekman et al., 2002) by iMotions using AFFDEX (McDuff et al., 2016). An indicator for each emotion per frame is provided by the software based on the probability of appearance of the emotion, so the range of values for each of them is from 0 to 100.

Fig. 6 shows an example of an output screen for 1 individual. Fig. 6 is a shot of a video that can be played on the screen, shortly after the subject has finished the experiment.

The software allows also for saving the results per frame in a log file with a text format, that is readily opened in MsExcel or accessed by other statistical software.

3.3. Post-processing interface per individual and emotion

All the calculations to automatically convert the raw output or log file into baseline emotions as well as to develop comparison c-charts and p-charts is carried out in MsExcel, although it could easily be performed in python or other programming languages.

The SEC algorithm takes the log file of iMotions and generates an output graph for each combination of individual and emotion (for example, Fig. 7 shows four outcomes). Indicators of emotional reactions for two different settings of the parameters to determine strong emotions are shown in each graph. TH50 corresponds to an absolute strategy where a THRESHOLD of 50 is set; values below the threshold are not considered as emotional reaction. SEC corresponds to a relative strategy where a CUTOFF is calculated as the average of the maximum and the upper baseline; values below the cutoff are not considered as emotional reaction. Peaks, hits and percent of hits are shown for each strategy, as well as the decision on intensity and duration after comparing across individuals. "NO" means that the emotion for this particular subject was not extraordinary when compared with the rest of individuals whereas "YES" indicates that the reaction of the individual was extraordinarily strong.

Intentionally, the selected four examples are the same as the ones that were used to set the research questions as well as the aim to develop a robust post-processing intelligent algorithm (Fig. 1). After the execution of the proposed algorithm, both with absolute TH50 and relative SEC settings:

- The subject 23 in year 1 for "E1. Joy" (Fig. 7(a)) shows 1 peak at 99.55 and 40 hits (percent hits = 0.15) both with the threshold at 50 and the cutoff at 52.28. In terms of extraordinary behaviour when comparing with the rest of individuals, the emotional reaction is neither intense nor durable (NO/NO) under absolute TH50 rules whereas is considered not intense but extraordinarily durable under relative SEC rules (NO/YES).
- The subject 15 in year 1 for "E2. Surprise" (Fig. 7(b)) shows 1 peak at 67.84 with 13 hits over 50 and 16 hits with the cutoff at 36.42. The emotional reaction is not intense but extraordinarily durable (NO/YES) under absolute TH50 rules. However, the emotional reaction is not extraordinary neither in intensity nor duration under SEC rules (NO/NO).
- The subject 33 in year 2 for "E6. Fear" (Fig. 7(c)) shows no values above 50 (maximum at 49.83) and 1 peak and 21 hits with the cutoff at 45.93. It is worth mentioning the high upper baseline of 42.03 since the timeseries shows a wave during the baseline period that already started during the disposal period. The emotional reaction is neither intense nor durable (NO/NO) under absolute TH50 rules whereas is considered not intense but extraordinarily durable under relative SEC rules (NO/YES).
- The subject 7 in year 2 for "E7. Sadness" (Fig. 7(d)) shows no values above 50 (maximum at 25.72) and 5 peaks and 51 hits with the cutoff at 15.36. The emotional reaction is neither intense nor durable (NO/NO) under absolute TH50 rules whereas is considered both

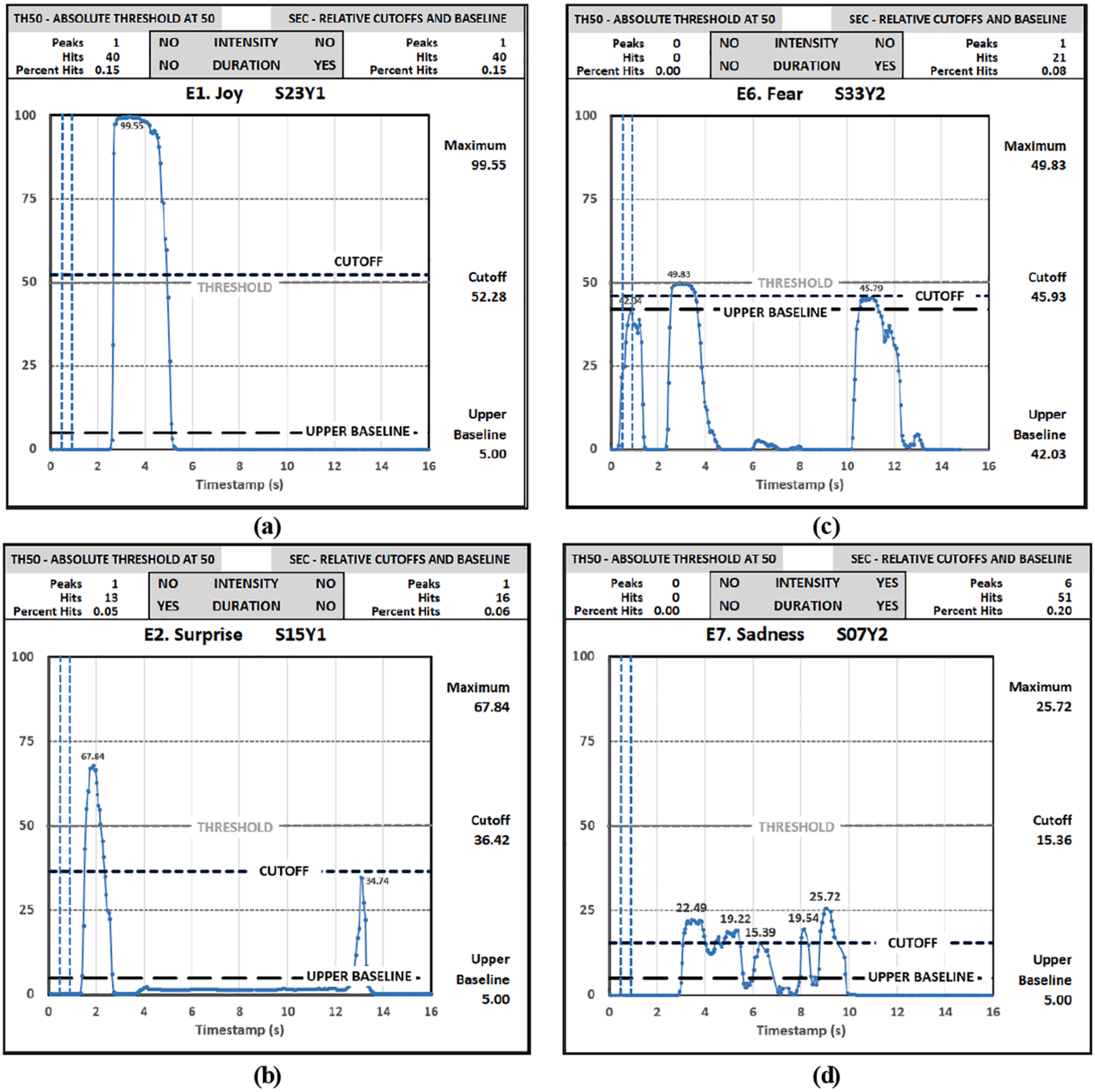


Fig. 7. Sample emotions in the MsExcel interface: (a) Joy; (b) Surprise; (c) Fear; (d) Sadness.

extraordinarily intense and durable under relative SEC rules (YES/YES).

These examples obviously show that the two settings (absolute TH50 vs relative SEC) indeed provide different results. Some individuals show "out-of-control" emotional reactions in either or both settings, in intensity and/or duration. Therefore, the examples show the importance of properly setting the disposal and the baseline periods, as well as the relative cutoff to account for variability in the measurements across individuals. In fact, if the threshold had been 50, the results would have been different, especially because both the negative emotions barely reach 50 and "E1. Joy" sometimes starts and remains very high due to the major influence of smiles on the results of the positive emotions, especially joy (Triyanti et al., 2019).

4. Results

This section shows the results that were obtained during the business game for all the 560 timeseries (40 subjects * 2 years * 7 emotions), using the novel SEC's relative strategy based on statistical process control theory, detection of changes to the baseline behaviour and implementation of relative cutoffs. The results are compared with the TH50 strategy in which absolute thresholds at 50 are maintained.

4.1. Range of the emotional responses

Concerning the biosensor's raw output, Fig. 8 shows the maximum and minimum values of the 560 timeseries. The representation is a boxplot with its quartiles, both for the maximum and minimum values recorded by subject and emotion.

The boxplots show anticipated positive asymmetry, meaning that only a small subset of individuals provides high values, especially above the iMotions threshold of 50 (only 21 of the 80 timeseries for "E1. Joy", 10 for "E2. Surprise", 2 for "E3. Anger", 4 for "E4. Contempt", 2 for "E5. Disgust", 0 for "E6. Fear" and 0 for "E7. Sadness"). The total number of timeseries with values above the absolute threshold TH50 is 39 out of the 560 recorded timeseries, 31 positive and only 8 negative.

Moreover, the feasible theoretical range is not fully covered by each

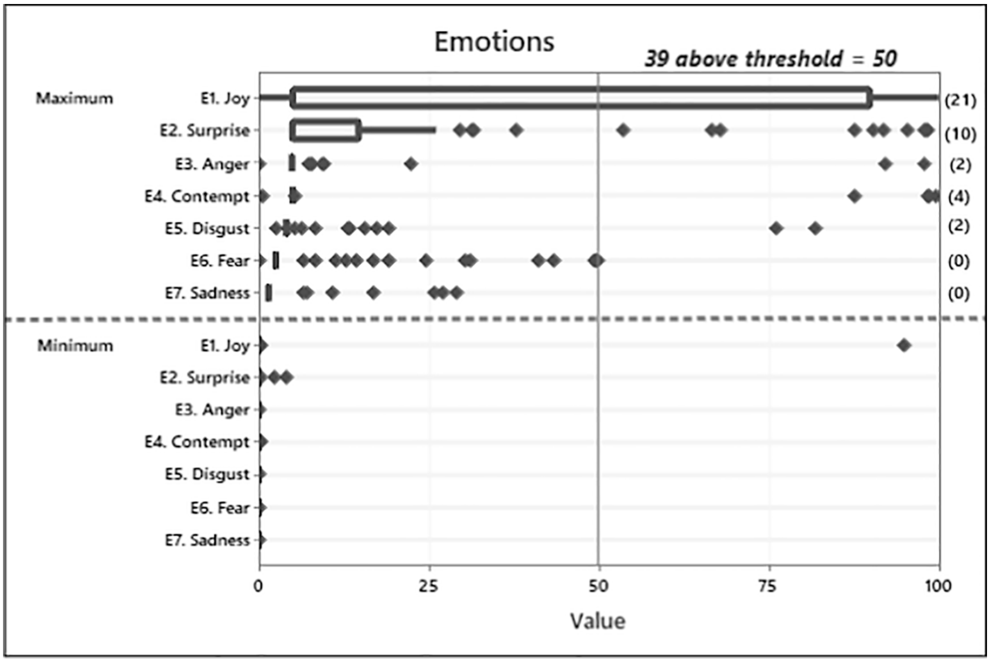


Fig. 8. Boxplots with the raw results per emotion.

and every emotion[7]. "E1. Joy" almost reaches the maximum 100 (in fact 99.93), with 25% of the subjects above 90, but "E2. Surprise", "E3. Anger" and "E4. Contempt" barely reach the upper bound of the range. "E5. Disgust" reaches 80 (but only 2 subjects above 25), "E6. Fear" almost 50 and "E7. Sadness" does not even reach 30. Either these last emotions are not present, or they have not been properly characterized during the experiment (game rules, subject positioning and/or personal characteristics, recording…).

Concerning minima, it is worth mentioning that one subject shows "E1. Joy" almost all the time (minimum of 93), which indicates that probably there was something not emotional that drove the measurements concerning this positive emotion.

4.2. Extraordinary emotional reactions

After characterizing the individual reactions with the baseline stage using non-normal I-charts, the extraordinary emotional reactions in intensity and duration among individuals are found using SPC theory and its c-charts and p-charts. Fig. 9 shows the c-charts of peaks and the p-charts of percent of hits, both for the absolute threshold setting of TH50 and the relative cutoff setting of SEC.

As expected, the results are different whenever absolute thresholds and relative cutoffs are used to pinpoint extraordinary reactions. For TH50, 25 individuals show intense and 32 durable emotional reactions (mainly positive); for SEC, 28 show intense and 49 durable reactions (both positive and negative).

Table 1 is used to summarize the total number and the proportion of extraordinary reactions by emotion (intense, durable or both) as well as to statistically test the overall difference in proportions across settings.

Following TH50, 33 emotional reactions (5.89%) are classified as extraordinary out of the 560 total timeseries. It is worth remembering that only 39 showed maxima above 50, so almost all of them showed intensity and/or duration. Following SEC, the total raises to 54 (9.64%) since there was no absolute threshold in this case but a relative cutoff. Not surprisingly, only "E1. Joy" shows less reactions under SEC (13) in comparison to TH50 (16).

Statistically, the difference in proportions of emotional reactions across settings of 3.75% is significant at the 0.05 level (p-value = 0.0188). No differences are found regarding positive emotions (0%) whereas the differences are outstanding in negative emotions (5.25%, p-value = 0.0000). By emotion, a higher proportion of individuals shows "E5. Disgust", "E6. Fear" and "E7. Sadness" (p-values of 0.0473, 0.0209 and 0.0402, respectively) under the relative SEC settings.

These significant differences call for a more detailed comparison among the two settings, using indicators of concordances (C) and discrepancies (D) between TH50 and SEC for each individual and emotion. A full concordance for a timeseries results whenever the emotional reaction is the same under both settings: both establish that there is no reaction, or both depict only intensity, or both pinpoint only duration, or both settings determine that the reactions are intense and durable. A discrepancy therefore results whenever there is no full concordance. Table 2 summarizes the number and the proportion of discrepancies (%D), separating the analysis for those timeseries with maxima below or above 50.

The percentage of discrepancies across settings for the whole set of timeseries is %D = 7.86% (D = 44 discrepancies out of 560, p-value = 0), with about the same number of discrepancies for positive (D = 19; %D = 11.88%, p-value = 0) and negative emotions (D = 25; %D = 6.25%, p-value = 0), with all the emotions showing discrepancies across settings except for "E4. Contempt".

Concerning "Emotion less than 50", out of the 521 timeseries with no values above 50, 27 (%D = 5.18; p-value = 0) are considered only by SEC to show extraordinary reactions[8]. It is striking to see that just 5 of the 27 correspond to positive emotions and 22 to negative emotions. Statistically, significant discrepancies are found for"E2. Surprise", "E5. Disgust", "E6. Fear" and "E7. Sadness".

A closer look to the 39 timeseries that show values above 50 ("Emotion $\geq$ 50") indicates that there are 17 discrepancies (%D = 43.59%, p-value = 0): 14 positive (%D = 45.16%, p-value = 0, both

[7] It is documented that AFFDEX uses a machine learning algorithm that potentiates positive emotions and sometimes underscores negative emotions (Stockli et al., 2018).

[8] The detailed analysis of the timeseries that result in a different decision on extraordinary reactions is provided as Supplementary Material.

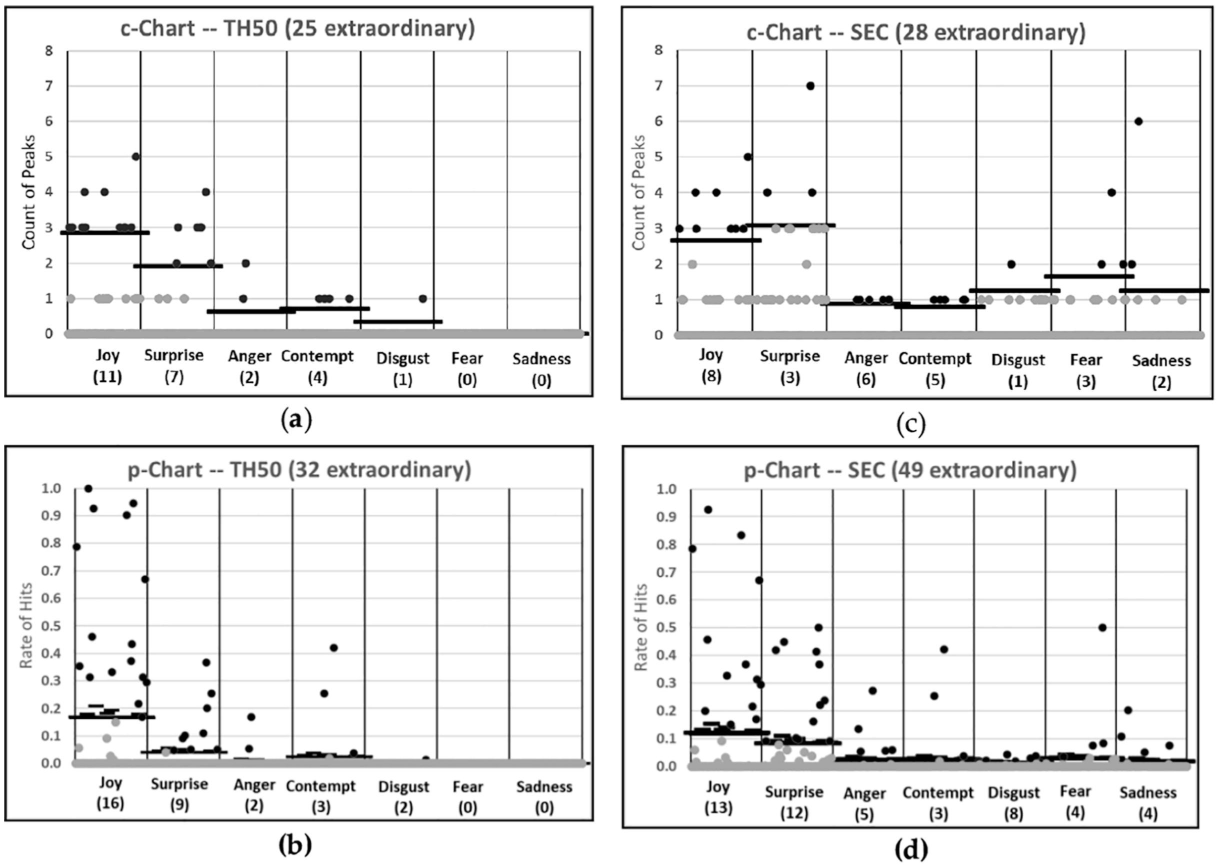


Fig. 9. SEC results per emotion: (a) Peaks detected using TH50 absolute settings; (b) Percent of hits above TH50 cutoff; (c) Positive peaks using SEC relative settings; (d) Percent of hits above SEC relative upper cutoff.

Table 1
Summary of extraordinary emotional reactions by setting. Column titles: E1. Joy, E2. Surprise, E3. Anger, E4. Contempt, E5. Disgust, E6. Fear, E7. Sadness. (*, **, means significant at 5% and 1%, respectively, after performing hypothesis tests based on proportions, Ho: %TH50-%SEC = 0).

Emotional Reactions	E1	E2	E3	E4	E5	E6	E7	Total
TH50	16	9	2	4	2	0	0	33
Proportion (% TH50)	20% 15.63%	11%	3% 2.00%	5%	3%	0%	0%	5.89%
SEC	13	12	7	5	8	5	4	54
Proportion (% SEC)	16% 15.63%	15%	9% 7.25%	6%	10%	6%	5%	9.64%
Differences in proportions (% TH50-%SEC)	**−4%** **0%**	**4%**	**6%** **5.25%****	**1%**	**8%***	**6%***	**5%***	**3.75%***

emotions significant on their own) and 3 negative emotions (%D = 35.75%, p-value = 0, although no significant emotion on its own).

The results therefore show that the discrepancies related to positive emotions mainly come from timeseries with values above 50 while those related to negative emotions come from timeseries with maxima below 50. Indeed, SEC with its relative cutoffs determines that some timeseries related to positive emotions should not be considered as extraordinary whereas some related to negative emotions should be considered as showing extraordinary reactions, even if TH50 with its traditional absolute thresholds dictates otherwise.

4.3. Conclusions

Summarizing the causes of discrepancies between absolute and relative thresholds, two come out clearly:

Table 2
Summary of discrepancies by setting. Column titles: E1. Joy, E2. Surprise, E3. Anger, E4. Contempt, E5. Disgust, E6. Fear, E7. Sadness. (*, **, means significant at 5% and 1%, respectively, after performing hypothesis tests based on proportions, Ho: %Discrepancies = 0).

Discrepancies *TH50* vs *SEC*	E1	E2	E3	E4	E5	E6	E7	Total
Total	6	13	6	1	9	5	4	44 *(19 + 25)*
% D	8%** *11.88%***	16%*	8%** *6.25%***	1%	11%*	6%**	5%**	*7.86%***
Emotion < 50	0	5	5	1	7	5	4	27 (5 + 22))
% D	0% 3.74%**	7%*	6%* 5.75%**	1%	9%**	6%*	5%*	5.18%**
Emotion > 50	6	8	1	0	2	0	0	17 (14 + 3)
% D	29%** 45.16%**	80%**	50% 35.75%**	0%	100%	0%	0%	43.59%**

- There is a need to set relative CUTOFFS to reduce the variability among subjects and account for the experimental error, both in terms of the biosensor and its software as well as individual face characterization and conversion into positive and negative emotions. A proper minimum value for the cutoff must be set, most probably at 20 (Mele et al., 2019; Triyanti et al., 2019).
- There is a need to properly set the BASELINE both to easily pick intense or durable strong reactions and to avoid the effect of the initialization of the experiment. The use of a disposal period is necessary.

Therefore, the use of the proposed intelligent SEC algorithm and its relative baselines and cutoffs based on the well-proven SPC theory has been validated, although minor adjustments should be made both in the experimental environment, setting an adequate baseline period, and in the experimental analysis, setting the relative cutoff around 20–25.

5. Discussion

There is a "need to improve the application of automated facial expression analysis in real life settings" (Stöckli et al., 2018), especially due to the increase in the number of experiments being performed with real subjects in different areas. Whereas most of the studies focus on the algorithm to detect emotions out of facial expressions, there is also the necessity to create new knowledge in the post-processing phase of the output of the automated facial expression recognition and emotion analysis (AFEREA) software.

Obviously, the better the recognition of faces is, and the better the algorithms to establish the emotional state of the subjects during the recording are, the greater potential for application in real life situations is. Faces however show many differences across subjects, and therefore the recognition and the results might be affected by many external factors like (1) age or race, (2) position relative to the camera, (3) scars, (4) glasses or piercings… It is also known that some parts of the face or some grins may favour positive emotions. In fact, the state-of-the-art iMotions AFEREA software usually captures "E1. Joy" more often than negative emotions.

So, while AFEREA software developers and researchers in facial analysis keep on focusing on improving the quality of their recognition tasks, frame by frame, applied researchers should strive for developing algorithms and tools to facilitate statistical analysis taking the output of the AFEREA software as input.

On that regard, the fulfilled objective of the current article was to develop a novel post-processing algorithm that could be used to relatively characterize and compare individuals with different biometric traits and grins in many situations of different fields. Based on proven statistical process control (SPC) theory, the proposed intelligent algorithm, called by similarity statistical emotion control (SEC), allows for the detection of extraordinary emotional reactions not only for a single individual and emotion but also across individuals and emotions using relative cutoffs instead of absolute thresholds.

As such, the robust algorithm SEC uses the same statistical tools for emotion detection as the background SPC uses for "out-of-control" detection. A combination of 3 SPC charts is used to characterize individual behaviour and allow for the comparison across individuals based on non-normal distributions: (1) a percentile-based I-chart to determine baseline or "in-control" state based on raw measurements, (2) a Poisson c-chart to quantify intensity or amplitude of the emotion based on the number of peaks, and (3) a Binomial p-chart to compute the duration of the emotions based on the rate of hits above cutoffs or thresholds.

SEC also incorporates ideas that might be found in different applications or uses of AFEREA in real life: (1) a disposal period is used to allow for the AFEREA software to properly initialize and precisely recognize the faces, (2) a baseline period to determine "in-control" initial state, (3) thresholds to account for incorrect settings of the experiment, and (4) relative cutoffs to mitigate the within-individual and within-emotion differences.

Of course, and a first line for future research, the results call for a further methodological analysis to precisely determine (or at least give a range) on absolute parameters like the start and end of the baseline or the minimum value of the upper baseline, or even if the cutoff should be the midpoint between the upper baseline and the maximum. On that regard, a calibration exercise should be performed following validation articles (Stöckli et al., 2018; Lewinski, den Uyl, & Butler, 2014) and via additional applications.

The second line for future research relates to SEC's applicability when using other AFEREA software. If the output of the software provides, as it is usually the case, a log file with timestamps, no problems are foreseen. Only minor adjustments according to the recording rate may be needed.

The validation stage in this research was based on an experimental business game monitored using the state-of-the-art iMotions. The proposed SEC relative algorithm performs as expected while properly characterizing individuals in terms of peaks and their duration, especially for negative emotions that were not assigned values over the available whole range 0–100. The comparison across individuals has also proven robust, highlighting the "out-of-control" individuals.

Therefore, the development of an intelligent system has been successful. The system uses facial recognition platforms and emotional characterization algorithms as the input to an intelligent post-processing algorithm that focuses on the measurement and characterization of emotional reactions with the aim of comparing across individuals and emotions and that is why we have resorted to the well-proven SPC theory. Choosing to characterize baseline emotional states and determine emotional reactions in terms of peaks and hits has allowed for the use of non-normal I-charts to set "stability", and Poisson c-charts and Binomial p-charts to detect "out-of-statistical control" emotional individuals.

As a conclusion, the SEC algorithm has proven to be an adequate robust intelligent methodology to take raw biosensor data on emotions after recording facial expressions, clean it and use it to compare emotional reactions among individuals over time.

CRediT authorship contribution statement

F. Javier Otamendi: Conceptualization, Methodology, Data curation, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eswa.2022.117074.

References

Bakalash, T., & Riemer, H. (2013). Exploring ad-elicited emotional arousal and memory for the ad using fMRI. *Journal of Advertising, 42*(4), 275–291. https://doi.org/10.1080/00913367.2013.768065

Bakker, A. R., Pacheco, B. D. M., D'Ipolitto, C., Xexéo, G., & de Souza, J. (2011). Emotions in Business Game: Case Study in Desafio Sebrae. In 2011 Brazilian Symposium on Games and Digital Entertainment. Salvador, Bahia, November 7-9 (pp. 141-147). IEEE.

Bigné, E., Llinares, C., & Torrecilla, C. (2016). Elapsed time on first buying triggers brand choices within a category: A virtual reality-based study. *Journal of Business Research, 69*(4), 1423–1427. https://doi.org/10.1016/j.jbusres.2015.10.119

Brand, R., & Ulrich, L. (2019). I can see it in your face. Affective valuation of exercise in more or less physically active individuals [Original Research]. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.02901

Braun, M., Schubert, J., Pfleging, B., & Alt, F. (2019). Improving driver emotions with affective strategies. *Multimodal Technologies and Interaction, 3*(1), 21. https://www.mdpi.com/2414-4088/3/1/21.

Couwenberg, L. E., Boksem, M. A. S., Dietvorst, R. C., Worm, L., Verbeke, W. J. M. I., & Smidts, A. (2017). Neural responses to functional and experiential ad appeals: Explaining ad effectiveness. *International Journal of Research in Marketing, 34*(2), 355–366. https://doi.org/10.1016/j.ijresmar.2016.10.005

De Oliveira, R., Dos Santos, R., Rocha, J., & Giraldi, J. (2015). Eye tracking in neuromarketing: A research agenda for marketing studies. *International Journal of Psychological Studies, 7*(1), 32–42. https://doi.org/10.5539/ijps.v7n1p32

Delgado, M. R., Gillis, M. M., & Phelps, E. A. (2008). Regulating the expectation of reward via cognitive strategies. *Nature neuroscience, 11*(8), 880–881. https://doi.org/10.1038/nn.2141

Ekman Group. (2021). Universal Emotions. Retrieved from https://www.paulekman.com/universal-emotions/. Accessed September 9, 2021.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska Symposium on Motivation 19* (pp. 207–282). University of Nebraska Press.

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental psychology and nonverbal behavior, 1*(1), 56–75. https://doi.org/10.1007/BF01115465

Ekman, P., Friesen, W. V. & Hager, J. C. (2002). Facial Action Coding System: The Manual on CD-ROM. Salt Lake City, UT: A Human Face.González-Rodríguez, M. R., Díaz-Fernández, M. C., & Pacheco Gómez, C. (2020). Facial-expression recognition: An emergent approach to the measurement of tourist satisfaction through emotions. Telematics and Informatics, 51, 101404. doi: 10.1016/j.tele.2020.101404.

Hamelin, N., Moujahid, O. E., & Thaichon, P. (2017). Emotion and advertising effectiveness: A novel facial expression analysis approach. *Journal of Retailing and Consumer Services, 36*, 103–111. https://doi.org/10.1016/j.jretconser.2017.01.001

Ho, M. W.-R., Chien, S. H.-L., Lu, M.-K., Chen, J.-C., Aoh, Y., Chen, C.-M., Lane, H.-Y., & Tsai, C.-H. (2020). Impairments in face discrimination and emotion recognition are related to aging and cognitive dysfunctions in Parkinson's disease with dementia. Scientific Reports, 10(1), 4367. https://doi.org/10.1038/s41598-020-61310-w iMotions. (2021). https://imotions.com/. Accessed September 9, 2021.

Hühn, C., & Rausch, A. (2022). Collaboration and emotions during simulation-based learning in general management courses. *Studies in Educational Evaluation, 73*, Article 101130. https://doi.org/10.1016/j.stueduc.2022.101130

Kielt, L., McGregor, M., & Jasra, S. (2018). The forensic biometric analysis of changes in facial response provoked by emotional arousal during initial and subsequent exposure to stimuli. *Journal of Emerging Forensic Sciences Research, 3*(1), 91–136. https://jefsr.uwindsor.ca/index.php/jefsr/article/view/5658.

Kulke, L., Feyerabend, D., & Schacht, A. (2020). A comparison of the affectiva imotions facial expression analysis software with EMG for identifying facial expressions of emotion [Original Research]. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.00329

Lei, J., Sala, J., & Jasra, S. (2017). Identifying correlation between facial expression and heart rate and skin conductance with iMotions biometric platform. *Journal of Emerging Forensic Science Research, 2*(2), 53–83.

Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics, 7*(4), 227–236. https://doi.org/10.1037/npe0000028

Lewinski, P., Fransen, M. L., & Tan, E. S. H. (2014). Predicting advertising effectiveness by facial expressions in response to amusing persuasive stimuli. *Journal of Neuroscience, Psychology, and Economics, 7*(1), 1–14. https://doi.org/10.1037/npe0000012

Magdin, M., Benko, Ľ., & Koprda, Š. (2019). A case study of facial emotion classification using AFFDEX. *Sensors, 19*(9), 2140. https://www.mdpi.com/1424-8220/19/9/2140.

McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J., & Kaliouby, R. E. (2016, May). AFFDEX SDK: a cross-platform real-time multi-face expression recognition toolkit. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, San Jose California USA May 7 - 12(pp. 3723-3726).

Mehta, A., Sharma, C., Kanala, M., Thakur, M., Harrison, R., & Torrico, D. D. (2021). Self-reported emotions and facial expressions on consumer acceptability: A study using energy drinks. *Foods, 10*(2), 330. https://www.mdpi.com/2304-8158/10/2/330.

Mele, M., Millar, D. & Rijnders, C. (2019). Explicit and implicit measures in video quality assessment. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)* (pp. 38-49). doi: 10.5220/0007396100380049.

Montgomery, D. C. (2009). *Introduction to statistical quality control* (7th ed.). John Wiley & Sons.

Ramsøy, T. Z., Friis-Olivarius, M., Jacobsen, C., Jensen, S. B., & Skov, M. (2012). Effects of perceptual uncertainty on arousal and preference across different visual domains. *Journal of Neuroscience, Psychology, and Economics, 5*(4), 212–226. https://doi.org/10.1037/a0030198

Robson, K., Plangger, K., Kietzmann, J. H., McCarthy, I., & Pitt, L. (2015). Is it all a game? Understanding the principles of gamification. *Business horizons, 58*(4), 411–420. https://doi.org/10.1016/j.bushor.2015.03.006

Samant, S. S., & Seo, H.-S. (2020). Influences of sensory attribute intensity, emotional responses, and non-sensory factors on purchase intent toward mixed-vegetable juice products under informed tasting condition. *Food Research International, 132*, Article 109095. https://doi.org/10.1016/j.foodres.2020.109095

Stöckli, S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. C. (2018). Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior Research Methods, 50*(4), 1446–1460. https://doi.org/10.3758/s13428-017-0996-1

Taggart, R. W., Dressler, M., Kumar, P., Khan, S., & Coppola, J. F. (2016). Determining emotions via facial expression analysis software. In Proceedings of Student-Faculty Research Day, CSIS, Pace University, May 6th. http://csis.pace.edu/~ctappert/srd2016/2016PDF/c2.pdf. Accessed September 09, 2021.

Taub, M., Sawyer, R., Lester, J., & Azevedo, R. (2020). The impact of contextualized emotions on self-regulated learning and scientific reasoning during learning with a game-based learning environment. *International Journal of Artificial Intelligence in Education, 30*(1), 97–120. https://doi.org/10.1007/s40593-019-00191-1

Taub, M., Sawyer, R., Smith, A., Rowe, J., Azevedo, R., & Lester, J. (2020). The agency effect: The impact of student agency on learning, emotions, and problem-solving behaviors in a game-based learning environment. *Computers & Education, 147*, Article 103781. https://doi.org/10.1016/j.compedu.2019.103781

Teixeira, T., Wedel, M., & Pieters, R. (2012). Emotion-induced engagement in internet video advertisements. *Journal of Marketing Research, 49*(2), 144–159. https://doi.org/10.1509/jmr.10.0207

Timme, S., & Brand, R. (2020). Affect and exertion during incremental physical exercise: Examining changes using automated facial action analysis and experiential self-report. *PLOS ONE, 15*(2), Article e0228739. https://doi.org/10.1371/journal.pone.0228739

Triyanti, V., Yassierli, Y., & Iridiastadi, H. (2019). Basic emotion recogniton using automatic facial expression analysis software. *Jurnal Optimasi Sistem Industri, 18*(1), 55. https://doi.org/10.25077/josi.v18.n1.p55-64.2019

Venkatraman, V., Dimoka, A., Pavlou, P. A., Vo, K., Hampton, W., Bollinger, B., … Winer, R. S. (2015). Predicting advertising success beyond traditional measures: New insights from neurophysiological methods and market response modeling. *Journal of Marketing Research, 52*(4), 436–452. https://doi.org/10.1509/jmr.13.0593

Wedel, M., & Pieters, R. (2008). Eye tracking for visual marketing. *Foundations and Trends® in Marketing, 1*(4), 231–320. https://doi.org/10.1561/1700000011

Wedel, M., & Pieters, R. (2013). Looking at vision: Eye/face/head tracking of consumers for improved marketing decisions. In L. Moutinho, E. Bigné, & A. K. Manrai (Eds.), *The Routledge companion to the future of marketing* (pp. 177–189). Routledge.

Xi, X., Zhang, Y., Hua, X., Miran, S. M., Zhao, Y.-B., & Luo, Z. (2020). Facial expression distribution prediction based on surface electromyography. *Expert Systems with Applications, 161*, Article 113683. https://doi.org/10.1016/j.eswa.2020.113683

Xu, X., & R de Sa, V. (2020). Exploring multidimensional measurements for pain evaluation using facial action units. http://cogsci.ucsd.edu/~desa/AAP_workshop.pdf. Accessed September 9, 2021.