This is a pre-print of the following publication:
Brodny G., Landowska A.: Integration in Multichannel Emotion Recognition, accepted for publication in: Proceedings of Human Systems Interaction Conference, Gdansk, Poland (HSI, 2018)

Integration in Multichannel Emotion Recognition

1st Grzegorz Brodny
Faculty of Electronics, Telefommunications and informatics,
Gdansk University of Technology
80-233 Gdansk, Poland
grzegorz.brodny@pg.edu.pl

2nd Agnieszka Landowska
Faculty of Electronics, Telefommunications and informatics,
Gdansk University of Technology
80-233 Gdansk, Poland
nailie@eti.pg.gda.pl

***Abstract*—The paper concerns integration of results provided by automatic emotion recognition algorithms. It presents both the challenges and the approaches to solve them. Paper shows experimental results of integration. The paper might be of interest to researchers and practitioners who deal with automatic emotion recognition and use more than one solution or multichannel observation.**

***Keywords*—affective computing, late fusion, early fusion, emotion recognition, emotion representation models, emotion mapping, system integration, multi-sensor fusion, human-machine interfaces**

I. Introduction

The paper concerns challenges in integration of automatic multimodal affect recognition. Multichannel observation of human emotions is applied in multiple domains: usability and user experience evaluation [1] [2] [3], educational software and resources designed for e-learning [4][5][6], affect-aware games and other intelligent personalized systems [7][8][9] as well as in optimization of processes [10][11][12]. Although it might seem, that the domain of affective computing is well established and there are even numerous off-the-shelf solutions, the reliability, accuracy and granularity of emotion recognition remains a challenge [13].

Nowadays, there are numerous emotion recognition algorithms that differ on input information channels, an emotion representation model and recognition method. A recognition algorithm might use one or a combination of the following channels:

- visual information from cameras,
- body movements mattes,
- textual input of a user,
- voice signals,
- standard input devices usage,
- physiological measurements.

As all emotion recognition channels are susceptible to noise and differ in accuracy, the common approach is to combine the channels (multimodal emotion recognition) [14]. This approach requires either integration of source data (early fusion approach) or results from different algorithms (late fusion approach). Both have some disadvantages and the challenge of multimodal integration constitutes the research problem addressed in this study.

In 2013 a project was started at Gdansk University of Technology (GUT) to build an emotion monitor stand, that uses existing technologies in order to extend human-systems interaction with emotion recognition and affective intervention. The concept of the stand assumed combining multiple modalities used in emotion recognition in order to improve the accuracy of affect classification [15]. The emotion monitor stand objective is to conduct experiments on computer users affective states retrieval and analysis. The stand is equipped with computers, cameras and a set of biosensors, which allow to monitor user activities and record multiple user observation channels at the same time [14].

Integration of the existing technologies, input channels and solutions turned out to be very challenging. The fusion struggle on this particular stand, led to more general observations, that might be applied to other integration solutions.

We decomposed the challenge of integration in emotion recognition into the following subproblems:

(1) the solutions (both off-the-shelf and home-made) differ in terms of reliability of the recognized emotional state, especially while some of the input channels they use are temporarily unavailable. This is particularly the challenge while using early fusion approach. Moreover, availability of the channels is context-, task- and/or user-dependent [16] [17].

(2) missing 'standard' emotion representation model; There are multiple models for representing emotions and there is no standard one. As a result, emotion recognition algorithms and solutions use diverse (and sometimes unique) emotion representation models [18]. Moreover, the reported emotional state might be provided using diverse scales and precision. This challenge emerges when using late fusion approach and requires additional mapping algorithms [19].

(3) in late fusion discrepancies between results (recognized emotional states) from different input channels and algorithms are observed; Assuming the same investigation, time and a person, in some experiments we observed huge discrepancies between recognized emotional states among different algorithms. Not only solutions based on diverse input channels exhibit the discrepancy, but also the same channel recorded twice (e.g. two cameras recording the same face with different angle) results in different results [16] [20].

(4) temporal uncertainty of emotion recognition results, that differs among algorithms and contexts; Some solutions return a prediction of an emotional state even if conditions for prediction were suboptimal (large camera angle, insufficient lighting or other noise). At the same time, most of the emotion

This work was supported in part by Polish-Norwegian Financial Mechanism Small Grant Scheme under the contract no Pol-Nor/209260/108/2015 as well as by DS Funds of ETI Faculty, Gdansk University of Technology.

recognition tools do not report the reliability of the predicted state.

(5) Time synchronization in late fusion; If results are combined from different channels and/or algorithms, they might be obtained with delays depending on data size, computational complexity and communication delay.

The study presented in this paper aims at addressing selected of the above listed challenges. The paper proposes a hybrid-approach integration solution, enabling combination of both early and late paradigms. A case study of complex emotion monitor stand is presented and the solution is evaluated in terms of integration ability and robustness. Implementations and results in this paper focus on late fusion in sentiment analysis. The thesis of the paper might be formulated as follows: *"The proposed hybrid approach for fusion in emotion recognition allows to provide integration ability and robustness."* The paper describes the concept, design and evaluation of emotion monitoring integration tool and is organized as follows: related works section reports the previous research that was relevant for this study; integration methods section describes the proposed approach; evaluation method section describes the criteria and experimental design and is followed by results and discussion sections.

II. Related works

Works that are mostly related to this research fall into three categories: research on late and early fusion, studies on emotion recognition based on multiple input channels, and papers on emotion representation models and approaches to mapping between them.

The methods of detecting emotional states could be categorized into four classes: (1) single algorithm (no integration required), (2) early fusion, (3) late fusion, (4) hybrid fusion.

(1) **Single algorithm**. Nowadays, many affective solution use only one input channel and one emotion recognition algorithm based on that [21]. These solutions are very specific, dedicated for one problem and often reveal only two classes of emotional states e.g. a positive versus negative state, stress versus lack-of-stress [5] [9]. As literature on emotion recognition is broad and has already been summarized several times, here we focus on fusion and for an extensive bibliography on single modality algorithms one may refer to Gunes, et al. [18], Zeng et al. [22] or Poria et al. [14].

(2) **Early fusion (called also feature-level fusion).** The early fusion method uses data from multiple input channels that are combined during the data collection step into one input vector (before classification) [19]. All data types are processed at the same time. This method usually provides high accuracy [21], but as the number of input channels increases, the processing complexity becomes more challenging. Another challenges include: time synchronisation for data streams coming from input channels; temporal unavailability of selected input channel resulting in missing values in feature vectors, large feature vectors when fusing numerous channels, (feature selection techniques are used to maximize the performance of the classifier) [23]. Moreover, adding a new channel or module often requires retraining and/or rebuilding solution (low scalability).

(3) **Late fusion (also called decision-level fusion)**. In late fusion, in contrast to early fusion, integration of data is performed during decision step. This method is based on an independent processing of data from each input channel and training multiple classifiers. Each of the classifiers provides one hypothesis on emotional state. The integration function provides a final estimate of emotional state based on partial results. This method provides more scalability than the early fusion because a new module is just one more result to integrate and in that, this solution allows to integrate algorithms using a black-box approach. The main challenge in the method remains time synchronization for data from diverse modules – integrate a subset of results or wait for all modules to provide a hypothesis. Another challenge is mapping the output to unified emotion representation model [21].

(4) **Hybrid fusion.** The hybrid methods are a combination of late and early fusion. Each module has a separate classifier as in the late fusion but also has access to input data from all input channels. The main advantage of this method is preservation of algorithms independence, while still using combined information from multiple channels. However, the synchronisation issue remains unsolved.

Hupont et al. report, that current solutions mostly use one input channel only and possible integration methods frequently are ad-hoc designed [21]. Some studies inform, that the best recognition results are obtained when fusing information from diverse input channels. Hupont et al. claim, that multimodal fusion improves robustness and accuracy of human emotion analysis.

Gunes and Piccardi provide a comparison of single, early and late fusion solutions. In their experiments they used facial expression and body gestures channel. The best recognition accuracy was achieved using early fusion, but authors consider too small training set as a threat to validity of their study. Authors propose to consider also a combination of the two approaches – a hybrid fusion [23].

In late and hybrid fusion one of the challenges is integration of results provided in different emotion representation models [24]. Models fall into three categories: (1) categorical, (2) dimensional and (3) componential.

(1) Categorical models are the most intuitive for human, but not for the computers [18]. They present each emotion as a combination of labelled emotional states. An example of those is a popular Ekman's model, that combines basic emotions: joy, fear, anger, surprise, sadness and disgust to represent complex emotional states [25].

(2) Dimensional models (usually two- or three-dimensional) represent emotions as compound of bipolar entity for example: valence (pleasant vs unpleasant), arousal (relaxed vs arousal) and dominance/power/control (submissiveness vs dominance) [18]. Emotions in these models are represented as a point in 2D, 3D or more dimensional space [26]. These models are less intuitive for humans but are more efficiently computed by applications. To

be understandable for people, the points require some mapping to emotion labels.

(3) Componential models of emotions are based on appraisal theory. The models are more complex and concentrate on how emotions are generated [27] [26] [28]. Some authors claim that categorical models could be mapped to dimensional ones and vice versa. However, most mappings are not lossless [18]. The researchers proposed a few mappings which are mainly derived from correlation coefficients. For example mapping between a big five personality model (categorical) and PAD (3D) was proposed as a function [29][30]. The mapping was created by calculating a correlation between the factors from both models and the correlation coefficients were used as weights in the mapping functions.

The next case was mapping between PAD and the Ekman model (categorical) [30], which was created in an analogical way. Mehrabian and Russel calculated correlation coefficients between PAD and models of personality [29]. A more accurate mapping between PAD and the Ekman's six basic emotions was recently proposed [19].

The next example is a mapping of the emotion labels to dimensional space proposed by [21]. The mapping provides weights that are derived from a database of coordinates from dimensional space to each label. The model can be used directly (e.g. in sentiment analysis). Another method of mapping was used by [31] in OCC (componential model) to PAD (3D). In this mapping, the points from PAD space were created for each OCC parameters (24) as a label (e.g. anger, fear, distress). The PAD's coordinates were used as weights to mapping functions.

III. Proposed Integration method

The proposed integration solution model

The main purpose of the integration solution is to perform the integration of emotional activation information from multiple channels and algorithms. The integration solution conceptual model is provided in Figure 1 with integration functionality marked with dotted line.

The proposed integration method follows several design decisions:

(1) Multiple input channels are recorded by a number of applications, dedicated for specific channel. Input channels recorded by a single software are synchronized. Some channels and off-the-shelf systems require additional synchronisation effort. Architecture supports late fusion of emotion recognition results provided by algorithms from diverse vendors and from different programming languages.

(2) Selected algorithms share the data using common memory space. The data might be shared on any level: source, intermediate (pre-processed), final estimates on emotional state.

(3) The approach used in this research is a hybrid fusion method, however it is primarily focused on late fusion. We focused on late fusion approach because early fusion one was not possible with the use of existing off-the-shelf solutions, including commercial software, we also intended to use. Late or the hybrid fusion method supports integration, exchange and modifiability of modules for emotion recognition ("black box" approach).

(4) We have chosen dimensional PAD model as a common integration scheme, mapping other results towards it. Based on the research review, dimensional models, and PAD in particular, were more universal and allowed for representation of all emotions. All results represented with other models are mapped, with a single mapping algorithm if possible.

(5) Currently, data labeling and training classifiers is performed post-hoc, in a batch mode, after the experiment is finished. Although the concept was to train classifiers off-line and then use them in real time, this would require more research, than was assumed. One of the problems we encountered is synchronization issue, as data streams from input channels have diverse size and frequencies. Moreover, processing of some data is more complex than others, and as a

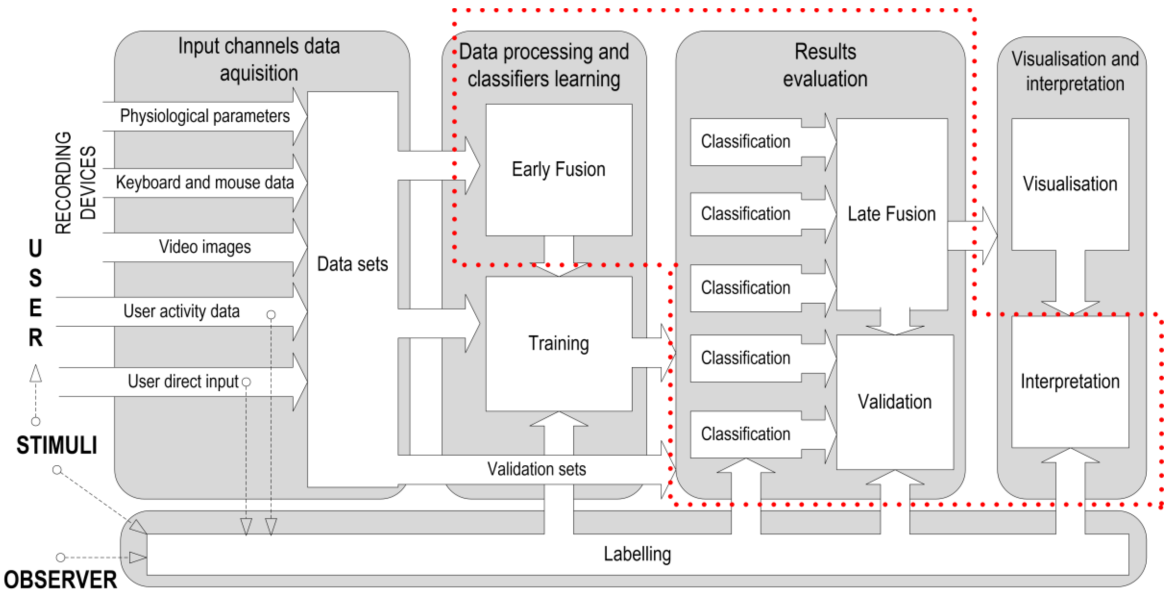

Fig. 1. Conceptual model of integration solution for emotion recognition based on hybrid approach

result emotion estimate is obtained with varying delays.

(6) Algorithms providing the emotion estimates are evaluated based on the final integrated result. The evaluations are shared in a common memory space and are used in further decision processes.

Implementation of integration solution model

The software layer of the implemented solution consists of a set of off-the-shelf and home-made applications for data recording, processing and visualization, accompanied by the Emotion Monitor home-made software to integrate them. Four layers were implemented for the solution: emotion recognition solution layer (1), integration and evaluation layer (2), presentation layer (3) and communication layer (4), that correspond to the conceptual model. The implementation architecture (component diagram) of the solution is provided in Figure 2.

Emotion recognition solution layer (1) consists of: an application to store and track biometric data, tools for observation and recording of video images, keyboard and mouse usage tracker and user activity logger. Apart from the applications recording input channels, the main emotion monitor's application is the one, that combines input channels and multiple classifiers in order to provide an affective state estimate.

Emotion Recognition Solution Layer encapsulates emotion recognition algorithms, called experts, which are treated with black-box approach. Implementation supports algorithms written in Python, Java, C++ or C# programming language and additionally any external program, communicating by command line might be attached. We have already

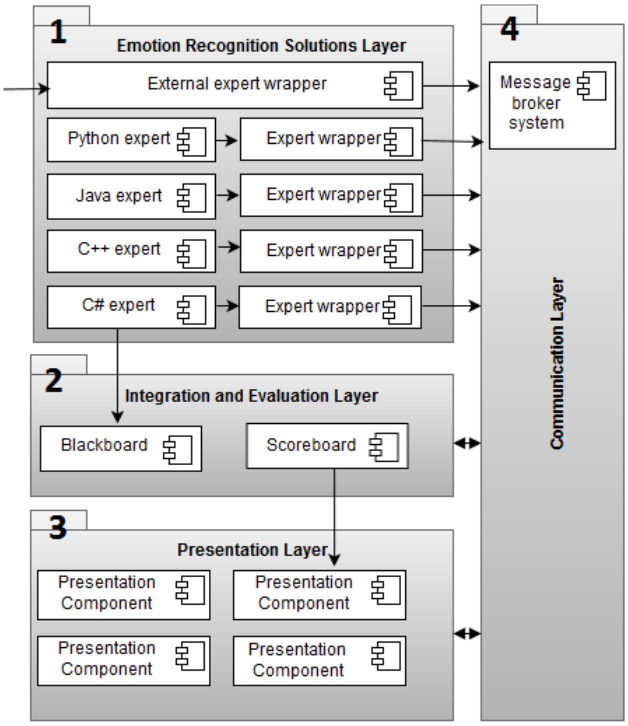

Fig. 2 Component Diagram of Emotion Monitor

successfully integrated Java, C++ and C# classifiers as well as external batch program for emotion recognition. As the integration tool is written in C#, the C# algorithms (experts) don't require a wrapper class. For other technologies wrapper classes are prepared. Only wrapper implementation for Python language was not verified, because we lack algorithms in this technology.

Wrappers communicate with a *Message broker* component in the Communication Layer (4). The *Message broker* uses a *RabbitMQ* component and potentially each solution, that is able to communicate with the queue system might be integrated.

Algorithms are expected to work on input channels data and send their hypothesis on a human emotional state along with a time stamp and evaluation of certainty. The data provided by the experts are shared in integration layer (2). Experts might share a final outcome as well as some intermediate results using *Blackboard* component (common memory space based on Blackboard design pattern).

The late fusion is provided in the second layer of the solution, which uses a number of techniques and design approaches to provide an agreed, reliable result. One of the aspects of the integration is the evaluation of experts' results consistency per case in a certain context. The early versions of the integration algorithm were based on simple voting mechanism, however that approach was not favorable for continuous inputs. The current version of integration algorithm:

- maps all emotion hypothesis to PAD model using the matrix provided as Equation 1 [30].
- enables consistency measures per dimension;
- takes into account certainty factor as reported by the recognition expert;
- evaluates experts based on inconsistency with others;
- uses expert evaluation in choosing experts to launch (if only selected ones are launched due to performance reasons).

$$PAD = [Anger, Disgust, Fear, Happiness, Sadness] = \\ = \begin{bmatrix} -0{,}51 & -0{,}4 & -0{,}64 & 0{,}4 & -0{,}4 \\ 0{,}59 & 0{,}2 & 0{,}6 & 0{,}2 & -0{,}2 \\ 0{,}25 & 0{,}1 & -0{,}43 & 0{,}15 & -0{,}5 \end{bmatrix} \quad \textbf{(1)}$$

The final emotional states and experts scores are sent to presentation layer (4), which enables multiple presentation forms.

IV. Evaluation Criteria

Thesis of the paper was formulated as follows: *"The proposed hybrid approach for fusion in emotion recognition allows to provide integration ability and robustness."* The two factors of *integration* and *robustness* are defined as follows.

Integration is a feature of the architectural design representing ability to combine results provided by diverse algorithms. The criteria should be understood as ability to integrate and to provide the results, that are not worse than individual experts (algorithms).

Robustness is a feature, that results from proper integration. If an invalid expert provides the results for the integration, the final result should be compromised as little as possible.

In order to evaluate the integration and robustness factors, an one experiment and one simulation were held, that showed how the integration solution performs (in terms of emotion recognition accuracy), with a number of algorithms of diverse individual accuracies. The experiment used 7 versions of opinion mining algorithms and compared the integrated result with the individual algorithms' accuracy. The textual inputs dataset used for evaluation was prepared (6 sets of 40 sentences, described further as S1,S2…S6 sets).

Several scenarios of the test were performed: (1) all algorithms scenario; (2) three strong and one weak algorithms scenario; (3) one strong and three weak algorithms scenario. The first scenario allows to evaluate integration factor, while the second and third scenario aimed at evaluation of solution robustness. We name the algorithms as being "strong" or "weak" based on the relative difference in accuracy among them.

Hypothesis of the experiment and the simulation might be formulated as follows:

H0 – there is no difference between integrated result and individual algorithms accuracy

H1 – the integrated result is at least as good as the individual algorithms accuracy.

V. Experimental evaluation of integration and robustness

Integration feature was tested on multiple versions of home-made algorithm for sentiment analysis in text. The algorithm is an naive version of weighted word-based sentiment retrieval, with accuracy varying according to version of affect-annotated lexicon used [32]. The algorithm and lexicons were not explored in this study, as they are not novel - in forthcoming tables they are simply indicated with numbers. The focus of evaluation process was integration mechanism. The reference algorithm that was based on the referential ANEW dictionary [33] was assumed as "ground truth" for evaluation of the integration accuracy. Table I presents accuracy of individual algorithms and the integrated solution for scenario (1), when all algorithms are valid. Solution marked as "blended" represents integrated results (based on late fusion approach).

Integrated results were more accurate than individual algorithms' results for all sentence sets. This scenario allows to report, that late fusion approach used in this model leads to better accuracy. The result is consistent with results obtained by previous studies.

Robustness was evaluated with scenario (2) and (3), with a combination of more (strong) and less (weak) accurate algorithms, while attribution to the strong and weak groups was based on results of scenario (1). As the differences in accuracies between the algorithms were not large, the division was relative.

In scenario (2) three strong and one weak algorithms were applied. The resulting accuracies are provided in Table II. The integrated result was better than individual algorithms' results for sentence set 6, and for the rest of sets it was worse than the best algorithm (with difference beings less than 1 percentage point). The integrated result was better than for the three weaker individual solutions.

In scenario (3) tested integration function while using one strong and three weak algorithms. The reference algorithm was used as a 'strong' solution. The integrated results was a little better than the best of the weak algorithm, which is presented in Table III. However, the three weak algorithms tend to skew the integrated result and the one is worse than the

Table I.
Accuracy of emotion recognition from text as an estimate of integration characteristics

Algorithm	Accuracy per set of sentences					
	S1 [%]	S2 [%]	S3 [%]	S4 [%]	S5 [%]	S6 [%]
Algorithm v. 2.0	72,16	71,64	65,70	72,24	71,49	70,98
Algorithm v. 2.0.1	71,63	73,36	68,42	72,65	72,04	70,10
Algorithm v. 2.1	71,87	69,03	68,09	74,00	67,66	70,82
Algorithm v. 2.2	71,87	74,06	68,42	72,64	72,04	70,10
Algorithm v. 2.2.1	73,42	73,39	68,42	74,28	69,12	70,10
Algorithm v. 2.2.2	70,23	n.a.	75,00	64,27	84,58	70,04
Algorithm v. 2.3 AD	67,61	63,95	74,33	68,64	70,00	67,63
Blended	84,39	82,91	79,10	80,92	89,04	74,09

Table II.
Accuracy of emotion recognition from text as an estimate of robustness characteristics in Scenario 2

Algorithm	Accuracy per set of sentences					
	S1 [%]	S2 [%]	S3 [%]	S4 [%]	S5 [%]	S6 [%]
Algorithm v 2.2	71,87	74,06	68,42	72,64	72,04	70,10
Algorithm v 2.2.1	73,42	73,39	68,42	74,28	69,12	70,10
Algorithm v 2.2.2	70,23	n.a.	75,00	64,27	84,58	70,04
Algorithm v. 2.3 AD	67,61	63,95	74,33	68,64	70,00	67,63
Blended (selected)	73,27	73,22	74,99	73,67	77,80	70,42

Table III.
Accuracy of emotion recognition from text as an estimate of robustness characteristics in Scenario 3

Algorithm	Accuracy per set of sentences					
	S1 [%]	S2 [%]	S3 [%]	S4 [%]	S5 [%]	S6 [%]
Algorithm v 2.0	72,16	71,64	65,70	72,24	71,49	70,98
Algorithm v 2.1	71,87	69,03	68,09	74,00	67,66	70,82
Algorithm v 2.2	71,87	74,06	68,42	72,64	72,04	70,10
Reference solution	100,00	100,00	100,00	100,00	100,00	100,00
Blended (selected)	73,09	74,22	72,17	73,10	72,20	70,89

best algorithm result. This is mainly due to the consistency of results provided by the three 'weak' algorithms (they were three versions of the same algorithm using different dictionaries).

As the observation based on sentiment analysis was inconclusive in terms of robustness criteria, another simulation was performed, assuming large differences between the algorithms taking part in obtaining the final solution. In this case we used home-made simulations of emotion recognition algorithms named Algorithm 1-5. 'Strong' algorithms generated random output ranging <0,2; 0,8>, providing acceptable consistency, while 'weak' algorithm provided a value from range <-1; 0>. Execution of the algorithms was performed 100 times to get significant results (score range is <0,1>). Table IV provides execution count and score per algorithm after 100th run. The 'weak' expert was quickly spot and after 100^{th} run got it score of 0,46. As a result, it was chosen for execution only 42 times, while the others were executed more intensively.

As a result of the experiment and the simulation, the observations concerning integration and robustness might be formulated as follows:

- integrated results are slightly more accurate than the ones of individual algorithms, assuming, the proportion of strong and weak ones is balanced;
- it is possible to automatically spot the low-accuracy algorithm on the run and to adapt the frequency of its execution;

The integration function used in the solution prefers consistency among algorithms. Therefore in unbalanced combination, one strong algorithm vote weights less than three votes of weak algorithms. One might consider an alternative evaluation and integration functions. The adjustment of integration function would require changing one class only. In the proposed solution multiple integration functions might be tested in parallel.

VI. SUMMARY OF RESULTS AND DISCUSSION

In order to verify the integration model used an experiment and a simulation were performed. The results might be summarized as follows:

- the results provided by the experts are integrated based on their consistency, which provides slightly better accuracy than individual algorithms;
- run-time evaluation of algorithms allows for reducing influence of solutions with low accuracies in the certain context; however robustness factor is dependent on the balance of high- and low-performing algorithms.

The results blend into the observations made in the related research so far: the late fusion provides better results than individual algorithms. However, none of the studies before used a robustness factor in evaluating unbalanced combination of algorithms.

Practical implications of this study on the proposed integration model in emotion recognition include:

(1) It is possible to use (almost) any algorithm or of-the-shelf solution for emotion recognition, assuming existence of API in late/hybrid fusion.

(2) It is possible to integrate the results using late fusion approach, assuming, there is some mapping to the emotion representation model used in the integration layer.

(3) As algorithms have diverse accuracies and execution time, the integration is made off-line. The on-line integration requires synchronization and was not addressed in this study.

The results obtained at this particular implementation might have some implications for the research on emotion recognition solutions and their integration. The following list of challenges have been identified during this study:

(1) Temporal unavailability of the input channels might be bypassed if the algorithms provides some estimate of the quality of provided result (although currently no algorithm does).

(2) The integration requires a common affect representation model or some mapping between the models. It's quite hard to integrate and compare results based on labels - any discrete or continuous model might be considered instead.

(3) On-line continuous integration requires solving the issue of diverse latency in processing input channels by different algorithms. Streams of emotional states estimates must be somehow synchronized.

The authors are aware of the fact, that this study is not free of some limitations. First of all, in this study we have treated both the sentiment analysis algorithms and the integration solution as a black-box and no details were provided on the used integration nor evaluation algorithms. This approach was chosen intentionally, as the algorithms might be changed and adjusted. The proposed solution allows to possibly use a number of consistency measures and evaluation techniques and compare them, which is part of our future research.

Another limitation is the use of sentiment analysis algorithms only, while usually multimodal integration is performed.

Although some limitations exits, we are convinced, that the research thesis *"The proposed hybrid approach for fusion in emotion recognition allows to provide integration ability and robustness"* was addressed. Using multiple channels and integration mechanism based on late fusion data, it is possible to improve solution accuracy.

TABLE IV.
NO OF TIMES AND EVALUATION SCORES FOR THE EMOTION RECOGNITION ALGORITHMS ILLUSTRATING ROBUSTNESS PROPERTY

Name	Condition	Execution count (out of 100)	Score achieved [0-1]
Algorithm 1	Strong	98	0,989
Algorithm 2	Strong	99	0,993
Algorithm 3	Strong	98	0,993
Algorithm 4	Reference	100	0,990
Algorithm 5	Weak	42	0,465

Based on this study, if one wants use out-off-the-shelf algorithms one should consider late or hybrid integration method. Future research would extend the solution with synchronization and then on-line fusion on emotion recognition might be performed. Moreover, we plan the experiments with multimodal fusion, using facial expression analysis, physiological data and behavioral observations.

VII. CONCLUSION

The reliability and the accuracy of the provided estimate of an emotional state of a human being depends on many conditions: availability and quality of the input channels, environmental variables (noise) and the expressivity of an individual. No matter how many sophisticated algorithms we use, we should not forget, that emotion is an internal phenomena and we have some insight only into the external symptoms of it. So the objective of the emotion recognition is to have clues convincing enough to assume we are close to the ground truth, which remains unknown.

ACKNOWLEDGMENT

This work was supported in part by Polish-Norwegian Financial Mechanism Small Grant Scheme under the contract no Pol-Nor/209260/108/2015 as well as by DS Funds of ETI Faculty, Gdansk University of Technology.

REFERENCES

[1] T. Partala and A. Kallinen, “Understanding the most satisfying and unsatisfying user experiences: Emotions, psychological needs, and context,” *Interact. Comput.*, vol. 24, no. 1, pp. 25–34, Jan. 2012.

[2] R. L. Hazlett and J. Benedek, “Measuring emotional valence to understand the user’s experience of software,” vol. 65, no. 4, pp. 306–314, Apr. 2007.

[3] P. G. Zimmermann, P. Gomez, B. Danuser, and S. G. Schär, “Extending usability: putting affect into the user-experience,” *Proc. Nord.*, pp. 27–32, 2006.

[4] H. H. H. H. Binali, C. Wu, V. Potdar, Chen Wu, and V. Potdar, *A new significant area: Emotion detection in E-learning using opinion mining techniques*. IEEE, 2009, pp. 259–264.

[5] A. L.-H. S. I. (HSI), 2013 The 6th, undefined 2013, and A. Landowska, *Affect-awareness framework for intelligent tutoring systems*. 2013.

[6] A. Landowska, M. S.-I. with Computers, and undefined 2016, “Methodology of Affective Intervention Design for Intelligent Systems,” *academic.oup.com*.

[7] K. Hone, “Empathic Agents to Reduce User Frustration Empathic Agents to Reduce User Frustration: The Effects of Varying Agent Empathic Agents to Reduce User Frustration 2 Empathic Agents to Reduce User Frustration: The Effects of Varying Agent Characteristics,” *Interact. Comput.*, 2006.

[8] A. Kolakowska *et al.*, *Emotion Recognition and Its Applications*. 2014, pp. 51–62.

[9] L. Chittaro and R. Sioni, “Affective Computing vs. Affective Placebo: Study of a Biofeedback-Controlled Game for Relaxation Training,” *Int. J. Hum. Comput. Stud.*, 2014.

[10] T. Partala and V. Surakka, “The effects of affective interventions in human-computer interaction,” *Interact. Comput.*, 2004.

[11] A. Kołakowska *et al.*, *Emotion recognition and its application in software engineering*. 2013.

[12] M. R. Wrobel, “Emotions in the software development process,” *2013 6th Int. Conf. Hum. Syst. Interact.*, pp. 518–523, 2013.

[13] A. Landowska, “Emotion monitoring - Verification of physiological characteristics measurement procedures,” *Metrol. Meas. Syst.*, 2014.

[14] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, “A review of affective computing: From unimodal analysis to multimodal fusion,” *Inf. Fusion*, 2017.

[15] A. Landowska, “Emotion monitor - Concept, construction and lessons learned,” in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015*, 2015, no. October 2015, pp. 75–80.

[16] A. Landowska and J. Miler, “Limitations of Emotion Recognition in Software User Experience Evaluation Context,” in *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems*, 2016, pp. 1631–1640.

[17] A. Landowska, W. Elektroniki, T. I. Informatyki, P. Gdańska, and G. Brodny, “Postrzeganie inwazyjności automatycznego rozpoznawania emocji w kontekście edukacyjnym,” pp. 26–41, 2017.

[18] H. Gunes and B. B. Schuller, “Categorical and dimensional affect analysis in continuous input: Current trends and future directions,” *Image Vis. Comput. Affect Anal. Contin. Input*, vol. 31, no. 2, pp. 120–136, Feb. 2013.

[19] A. Landowska and Agnieszka, “Towards New Mappings between Emotion Representation Models,” *Appl. Sci.*, vol. 8, no. 2, p. 274, Feb. 2018.

[20] A. Landowska, G. Brodny, and M. R. Wrobel, “Limitations of emotion recognition from facial expressions in e-learning context,” *CSEDU 2017 - Proc. 9th Int. Conf. Comput. Support. Educ.*, vol. 2, 2017.

[21] I. Hupont, S. Ballano, S. Baldassarri, and E. Cerezo, “Scalable multimodal fusion for continuous affect sensing,” in *IEEE SSCI 2011 - Symposium Series on Computational Intelligence - WACI 2011: 2011 Workshop on Affective Computational Intelligence*, 2011.

[22] Zhihong Zeng *et al.*, “A survey of affect recognition methods: Audio, visual, and spontaneous expressions,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.

[23] H. Gunes and M. Piccardi, “Affect recognition from face and body: early fusion vs. late fusion,” *Syst. Man Cybern. 2005 IEEE Int. Conf.*, vol. 4, pp. 3437–3443, 2005.

[24] G. Valenza, A. Lanata, and E. P. Scilingo, “The role of nonlinear dynamics in affective valence and arousal recognition,” *IEEE Trans. Affect. Comput.*, 2012.

[25] K. Scherer and P. Ekman, “Approaches to emotion,” 2014.

[26] K. S. D Grandjean, D Sander, “Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization,” *Conscious. Cogn.*, 2008.

[27] P. C. E. Johnny R.J. Fontaine, Klaus R. Scherer, Etienne B. Roesch, “The World of Emotions is not Two-Dimensional,” *Psychol. Sci.*, vol. 18, no. 12, pp. 1050–1057, Dec. 2007.

[28] A. Ortony, G. Clore, and A. Collins, “The cognitive structure of emotions,” 1990.

[29] A. Mehrabian, “Analysis of the Big-five Personality Factors in Terms of the PAD Temperament Model,” *Aust. J. Psychol.*, vol. 48, no. 2, pp. 86–92, Aug. 1996.

[30] Z. Shi, J. Wei, Z. Wang, J. Tu, and Q. Zhang, “Affective transfer computing model based on attenuation emotion mechanism,” *J. Multimodal User Interfaces*, vol. 5, no. 1–2, pp. 3–18, Mar. 2012.

[31] P. Gebhard and Patrick, “ALMA,” in *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems - AAMAS ’05*, 2005, p. 29.

[32] A. Kolakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wrobel, “Evaluation criteria for affect-annotated databases,” in *Communications in Computer and Information Science*, 2015.

[33] M. M. Bradley and P. P. J. Lang, “Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings,” *Psychology*, 1999.