



Author(s) Rantanen, Ville; Kumpulainen, Pekka; Venesvirta, Hanna; Verho, Jarmo; Spakov, Oleg; Lylykangas, Jani; Vetek, Akos; Surakka, Veikko; Lekkala, Jukka

Title Capacitive facial activity measurement

Citation Rantanen, Ville; Kumpulainen, Pekka; Venesvirta, Hanna; Verho, Jarmo; Spakov, Oleg; Lylykangas, Jani; Vetek, Akos; Surakka, Veikko; Lekkala, Jukka 2013. Capacitive facial activity measurement. Acta IMEKO vol. 2, num. 2, 78-85.

Year 2013

Version Publisher's PDF

URN <http://URN.fi/URN:NBN:fi:tty-201409171432>

Copyright © 2013 IMEKO. This is an open-access article licensed under a Creative Commons Attribution 3.0 License, <http://creativecommons.org/licenses/by/3.0/>.

Capacitive facial activity measurement

Ville Rantanen¹, Pekka Kumpulainen¹, Hanna Venesvirta², Jarmo Verho¹, Oleg Špakov², Jani Lylykangas², Akos Vetek³, Veikko Surakka², Jukka Lekkala¹

¹ Sensor Technology and Biomeasurements, Department of Automation Science and Engineering, Tampere University of Technology, Korkeakoulunkatu 3, FI-33720 Tampere, Finland

² Research Group for Emotions, Sociality, and Computing, Tampere Unit for Computer-Human Interaction, School of Information Sciences, University of Tampere, Kanslerinrinne 1, FI-33014, Tampere, Finland

³ Media Technologies Laboratory, Nokia Research Center, Otaniementie 19, FI-02150 Espoo, Finland

ABSTRACT

A wide range of applications can benefit from the measurement of facial activity. The current study presents a method that can be used to detect and classify the movements of different parts of the face and the expressions the movements form. The method is based on capacitive measurement of facial movements. It uses principal component analysis on the measured data to identify active areas of the face in offline analysis, and hierarchical clustering as a basis for classifying the movements offline and in real-time. Experiments involving a set of voluntary facial movements were carried out with 10 participants. The results show that the principal component analysis of the measured data could be applied with almost perfect performance to offline mapping of the vertical location of the facial activity of movements such as raising and lowering eyebrows, opening mouth, raising mouth corners, and lowering mouth corners. The presented classification method also performed very well in classifying the same movements both with the offline and the real-time implementations.

Section: RESEARCH PAPER

Keywords: capacitive measurement, distance measurement, facial activity measurement, facial movement detection, hierarchical clustering, principal component analysis

Citation: Ville Rantanen, Pekka Kumpulainen, Hanna Venesvirta, Jarmo Verho, Oleg Špakov, Jani Lylykangas, Akos Vetek, Veikko Surakka, Jukka Lekkala, Capacitive facial activity measurement, Acta IMEKO, vol. 2, no. 2, article 14, December 2013, identifier: IMEKO-ACTA-02 (2013)-02-14

Editor: Paolo Carbone, University of Perugia

Received May 31st, 2013; **In final form** November 20th, 2013; **Published** December 2013

Copyright: © 2013 IMEKO. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Funding: This work was funded by Nokia Research Center, the Finnish Funding Agency for Technology and Innovation, and the Finnish Cultural Foundation.

Corresponding author: Ville Rantanen, e-mail: ville.rantanen@tut.fi

1. INTRODUCTION

Measuring facial movements has many possible applications. Human-computer and human-technology interaction (HCI and HTI) can use information of voluntary facial movements for the interaction [1]-[7]. Other applications, for example in behavioural science and medicine, can also benefit from the automated analysis of human facial movements and expressions [8]-[15].

In the context of HCI, the use of facial movements has been studied already for a decade. The first implementations were based on measuring electromyographic (EMG) signals that reflect the electrical activity of the muscles [1]. The measurement system by Barreto et al. [1] utilised only bioelectric signals for pointing and selecting objects, but later on the EMG measurement was adopted as a method to indicate selections in HCI when gaze is used for pointing [2], [3], [4], [6].

Recently, a capacitive detection method has been introduced as an alternative for the facial EMG measurement [5]. It provides a contactless alternative that measures facial movement instead of the electrical activity of the muscles that the EMG measures. Studies about pointing and selecting with the capacitive method in combination with head-mounted, video-based gaze tracking have been also published [7], [16], [17], [18].

Facial Action Coding System (FACS) is a vision-based method that characterises facial actions based on the activity of different facial muscles [19], [20]. Each facial expression has certain activated muscles that can have different levels of contraction. FACS and the detection of active muscles have been used as a basis for automatically analysing facial expressions, for example, for the use of behavioural science and medicine [9], [10], [11], [13], [14], [15]. These studies describe automated implementations of FACS by using vision-based

methods in the analysis. However, facial EMG can also register facial actions and provide information that is highly similar to the one provided by FACS [11]. EMG has also been shown to be suitable for measuring emotional reactions from the face [8]. This has been done long before EMG was first applied in the HCI context to detect voluntary facial movements.

The presented method applies capacitive measurement principles to measure the activity of the face. It has several advantages over the other methods that can be used for the task. Compared to EMG measurements, the presented method allows the measurement of more channels simultaneously. It is contactless and it does not require the attachment of electrodes to the face. Attached electrodes significantly limit the maximum number of measurable channels and they may also affect facial movements that are being targeted with the measurement [11], [21]. When compared to vision-based detection of facial activity, the capacitive method allows easier integration of the measurement to mobile, head-worn equipment, is unaffected by environmental lighting conditions, and can be carried out with computationally less intensive signal processing.

For the current study, a wireless, wearable prototype device was constructed, and analysis of data from controlled experiments was done to identify the location of facial activity and to classify it during different voluntary movements. Voluntary facial movements have previously been detected by identifying transient peaks from the signals [5]. The presented method provides a more robust way to analyse the facial activity based on multichannel measurements.

2. METHODS

2.1. Capacitive Facial Movement Detection

The measurement method for measuring facial activity is based on the capacitive detection of facial movements that was introduced in [5]. It applies the same measurement principle as capacitive push buttons and touchpads, and a single measurement channel requires only a single electrode that produces an electric field. The produced field can be used to detect conducting objects in its proximity by measuring the capacitance because the capacitive coupling between the electrode and the object changes as the object moves. In principle, the distance between the target and the electrode is measured.

2.2. Prototype Device

The wearable measurement prototype device is shown in Figure 1. The device was constructed as a headset that should fit most adults. The earmuffs of the headset house the necessary electronics, and the extensions seen in front of the face include the electrodes for the capacitive measurement channels. The device contains 22 electrodes in total, 11 for both sides of the face. The top extensions have 4 electrodes each, the middle ones have 3 each, and the lowest ones have 4. The electrodes are printed circuit board pieces with a size of 12 x 20 mm. They are connected to the measurement electronics with thin coaxial cables that shield the signals. The capacitive measurements are carried out with a programmable controller for capacitance touch sensors (AD7147 by Analog Devices). The sampling frequency was dictated by technical limitations and it was set to the maximum possible, 29 Hz. The device is battery-powered and a Bluetooth module (RN-41 by Roving

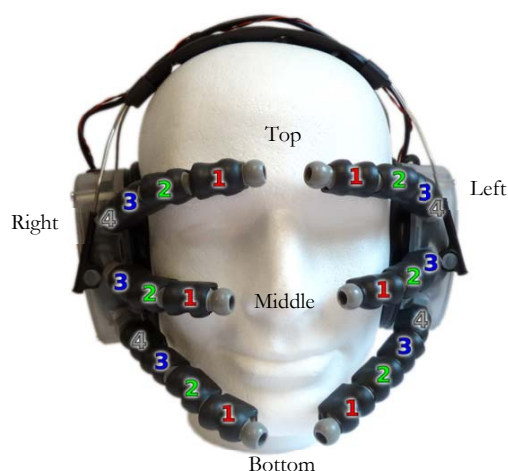


Figure 1. The wearable measurement device. The numbers represent the extension pieces that house the measurement electrodes. The actual electrode locations are on the pieces facing the face.

Networks) provides the connectivity to the computer. The device has the possibility for additional measurements such as inertial measurements via a 3D gyroscope and a 3D accelerometer. The operation of the device is controlled by Atmel's ATmega168P microcontroller.

2.3. Experiments

Ten participants (five male and five female, ages 22-33, mean age 27) were briefly trained to perform a set of voluntary facial movements. The participants were chosen to be inexperienced in carrying out the movements to avoid more easily measured overly expressive movements that experienced participants might perform. The movements were: lowering the eyebrows, raising the eyebrows, closing the eyes, opening the mouth, raising the mouth corners, lowering the mouth corners, and relaxation of the face. The relaxation was included to help the participant relax during the experiments while doing the other movements. The movements were instructed to be performed according to the guidelines of FACS [20]. The participants were instructed to activate only the needed muscles during each of the movements. After a brief practise period and verification that the participant made the movements correctly, the device was worn by the participant as shown in Figure 1: the top extensions targeted the eyebrows, the middle ones the cheek areas, and the bottom ones the jaw and mouth area. The distance of each of the measurement electrodes from the facial tissue was adjusted to be as close as possible without the electrodes touching the face during the movements. This way the distance was approximately 1 cm for all electrodes. In the experiments, synthesized speech was used to give instructions to the participants to perform each individual movement. After putting on the device, two repetitions of each of the movements were carried out in a controlled practise session to familiarise the participants with the experimental procedure. The actual procedure consisted of ten repetitions of each movement carried out in randomised order. Participants were given 10 seconds to complete each repetition. A mirror was used throughout the experiments to provide visual feedback of the facial movements to the participants.

2.4. Data Processing

2.4.1. Signal Processing Principle

Figure 2 shows a diagram of the pre-processing that was applied to the signals prior to further data processing.

First the capacitive signals were converted to signals proportional to the physical distance between the facial tissue and the measurement electrode. The conversion normalises the sensitivity of the measurement to the distance. The capacitance measurement was modelled with the equation of a parallel plate capacitor:

$$C = \frac{\epsilon A}{d}, \quad (1)$$

where ϵ is the permittivity of the substance between the plates, A the plate area, and d the distance between the plates. One plate is formed by the measurement electrode and the other by the targeted facial tissue. While the surface profile of the facial tissue is often not a plate, each unique profile can be considered to have such an equivalent plate that the equation 1 can be applied. Since the relationship between the capacitance and the distance is inversely proportional, the sensitivity of the capacitance to the distance is dependent on the distance itself. The absolute distance is not of interest, and a measure proportional to the distance can be calculated as

$$d_p = \frac{1}{C} = \frac{1}{C_s - C_b}, \quad (2)$$

where C_s is the capacitance value and C_b is the base level of the capacitance channel. Each channel has a unique base level that is affected by the length of the electrode cable and the surroundings of the electrode determined by its position on the extension. For the conversion, the base levels of all the capacitance channels were measured when the measurement electrodes were directed away from conducting objects.

Smoothing and baseline removal were applied to the distance signals computed with equation 2. These two steps were different when locating the facial activity and when classifying it. The differences are explained below in the corresponding sections. After processing the signals, only the first 4.5 seconds of the signals during each repetition of the movements were considered when calculating the results. The remaining 5.5 seconds of each 10-second repetition was

neglected from further analysis because all the participants had already finished the instructed movements by then, and they sometimes carried out other movements to relax during that remaining time.

2.4.2. Locating Facial Activity

The smoothing applied to the distance signals when locating the facial activity was done with a moving median filter with a length of 35 samples (approximately 1.2 seconds). This was done to remove noise. Further, the baselines of the signals were removed by subtracting the signal means during each repetition of the instructed movements. The baseline removal normalises the signal sequences so that they represent the relative changes in the physical distance.

Principal component analysis (PCA) was carried out to the processed signal sequences to find out the locations of the facial activity during the instructed facial movements. PCA is a linear method which transforms a data matrix with multiple variables, measurement channels in this case, to a set of uncorrelated components [22]. These principal components are linear combinations of the original variables and they describe the major variations in the data. PCA decomposes the data matrix \mathbf{X} , which has m measurements and n channels, as the sum of the outer product of vectors \mathbf{t}_i and \mathbf{p}_i and residual matrix \mathbf{E} :

$$\mathbf{X} = \mathbf{t}_1 \mathbf{p}_1^T + \mathbf{t}_2 \mathbf{p}_2^T + \dots + \mathbf{t}_k \mathbf{p}_k^T + \mathbf{E}, \quad (3)$$

where k is the number of principal components used. If all possible components are used, the residual reduces to zero. Vectors \mathbf{t}_i are called scores, and \mathbf{p}_i are eigenvectors of the covariance matrix of \mathbf{X} and are called loadings. The principal components in the equation 3 are ordered according to the corresponding eigenvalues.

To localise facial activity, the first principal component and its loadings were considered. The first principal component describes the major data variations, and, thus, the location of the most significant facial activity can be identified by analysing the loadings of the corresponding measurement channels. For the analysis, the loadings were normalised by dividing their absolute values with the sum of the absolute values of all channels. As a result, the sum of the normalised values is equal to 1.

To present the results, the vertical location of each repetition of the movements was mapped to the part of the face that introduced two of the three most significant relative loadings of the first principal component (M-out-of-N detector). For calculating percentages of successful mappings, the correct source of activity was considered to be the top extension channels for the lowering and raising eyebrows as well as the closing eyes movement, the bottom extension channels for the opening mouth and lowering mouth corners movements, and the middle or bottom extension channels for the raising mouth corners movement. Median loadings of the 10 repetitions of each movement were calculated for each participant and channel separately to verify the decisions about the correct sources of activity.

2.4.3. Classifying Facial Activity

Smoothing causes a delay. Therefore distance signals without smoothing were used when classifying the facial activity. Baseline removal was carried out to the distance signals

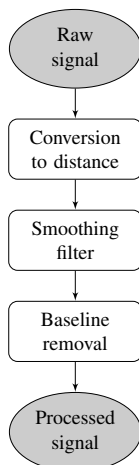


Figure 2. A block diagram of the signal processing.

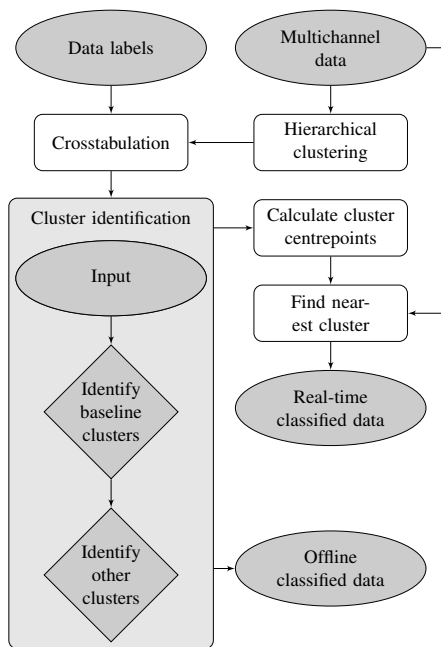


Figure 4. A block diagram of the classification of the facial movements based on the data.

directly. Figure 3 presents the algorithm used for solving the baseline for its removal.

The baseline calculation was based on a median filter. The median can perform well in the task because the signals were expected to have longer baseline sequences than the ones resulting from facial activity. The median filter applies a logic that only selects part of the samples as baseline points for the median calculation. The selection is based on a constant false alarm rate (CFAR) processor that calculates an adaptive threshold based on the noise characteristics of the processed signal [23], [24]. The distance signal was first pre-processed with a filter that implements a differentiator, a single-pole low-pass filter with a time constant of 20 ms, and a full-wave rectifier. This makes the input suitable for the CFAR processor. The current sample is used as a test sample for the processor. The implemented version of the processor uses samples before the test sample, referred to as reference samples, for calculating the threshold. The processor also leaves out samples closest to the

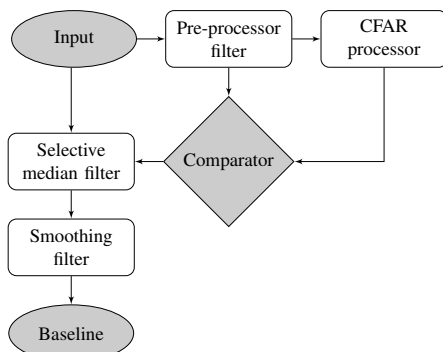


Figure 3. A block diagram of the baseline calculation for the baseline removal when classifying facial activity.

test sample as guard samples to reduce the information overlap between the test and reference samples. Samples closer than 1 second to the test sample were considered guard samples and the following 14 seconds were considered as the reference samples. The mean of the reference samples was then calculated and multiplied by a sensitivity parameter to obtain the adaptive threshold. The sensitivity parameter was chosen to be 0.5 in this case. The threshold was then fed to a comparator with the pre-processed test sample to find out if the test sample did not exceed it. The respective samples of the input signal were included in the median calculation by the selective median filter that had a length of 15 seconds. Finally, the baseline is calculated with a 2-second moving average filter from the median filtered signal to smooth step-wise transitions in the baseline level.

A method to classify facial movements based on the processed multichannel data was implemented. The classification method was based on hierarchical clustering. It used the Ward's linkage which forms clusters by minimising the increase of total within-cluster variance about the cluster centre [25], [26]. A fixed number of 14 clusters were chosen for the clustering based on the different events that the data represents (6 movements and the baseline). This selection allows 2 clusters for each event on average, which allows some deviation of the data when performing repetitions of the same movement and elongation of the data points during a movement, because the Ward's method is known not to be good at handling elongated clusters and outliers [26].

The work-flow of the classification is presented in Figure 4. The clustering first takes multichannel data with signal baselines removed and the labels of the data (information about the instructed movements) as an input. The data are first clustered and then cross tabulated against their 6 possible labels. Based on the tabulation, the clusters are identified so that first the clusters that represent the baseline data are identified. A cluster is identified as a baseline cluster if it contains data points with at least 5 different labels from the 6 possible (M-out-of-N detector). Other clusters are identified based on the label that has the largest number of samples in the cluster. In the offline classification, the data are then classified to represent the movement its cluster was identified with. A real-time classification can further be made based on previously identified clusters. First, the cluster centre points are calculated. Then each new data sample is classified to represent the movement that the cluster nearest to it was identified with. Thus, the real-time classification only requires the calculation of Euclidean distances to the cluster centres for each new data sample.

All the collected data were included in the offline classification. The real-time implementation of the classification was evaluated so that a randomly chosen repetition of each movement was included in the identification of the clusters, and the remaining 9 repetitions were used as test data to evaluate the performance of the method.

To present the results of the classification, the percentages of the data points that were classified as baseline were calculated. A high percentage would indicate problems in separating the movements from the baseline. From the data points that were not classified as the baseline, the percentages of correctly classified ones were calculated. Data points were considered to be correctly classified if they were classified as the movement that the participant was at that time instructed to perform.

3. RESULTS

Figures 5-8 show examples of the signals the measurement channels registered during the experiments and how the conversion from capacitance signals to ones that are proportional to the distance normalises the signals.

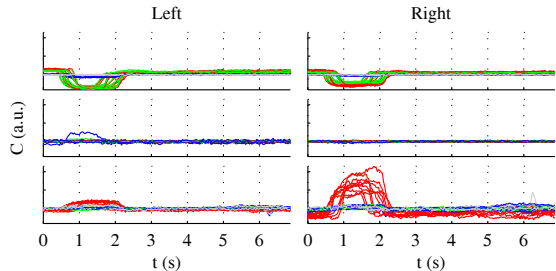


Figure 5. Raw capacitance signals from the 10 repetitions of the raising eyebrows movement with one participant. The different sides of the face are represented on the left and on the right. The top, middle, and bottom graphs represent the measurements from the corresponding extensions. The colours represent the different channels as shown in Figure 1: red, green, blue, and grey starting from the centre of the face. Signal baselines are aligned for the illustration.

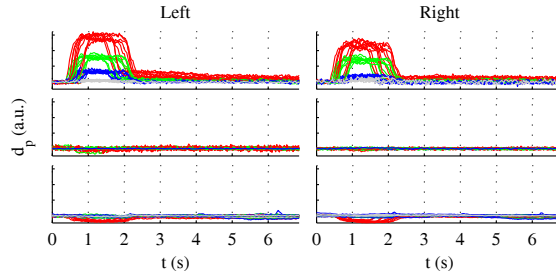


Figure 6. Signals after the conversion to distance signals from the 10 repetitions of the raising eyebrows movement with one participant. Signal baselines are aligned for the illustration.

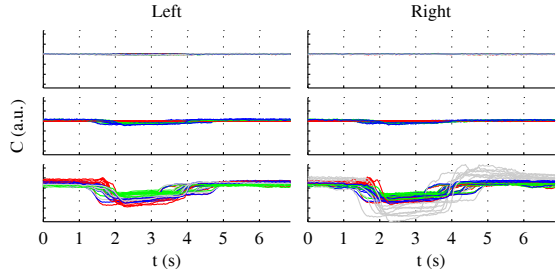


Figure 7. Raw capacitance signals from the 10 repetitions of the opening mouth movement with one participant.

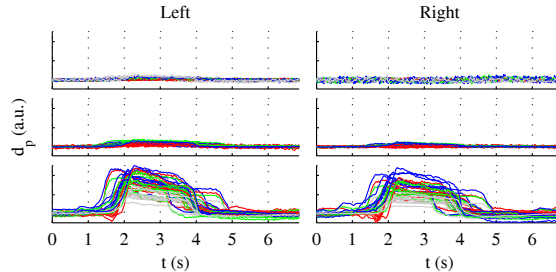


Figure 8. Signals after the conversion to distance signals from the 10 repetitions of the opening mouth movement with one participant.

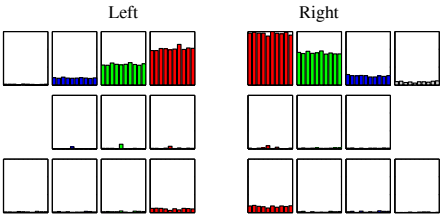


Figure 9. The facial activity as represented by the loadings of the first principal component during the raising eyebrows movement with one participant. Each graph represents the loadings of the 10 repetitions from the measurement channel of the corresponding physical location.

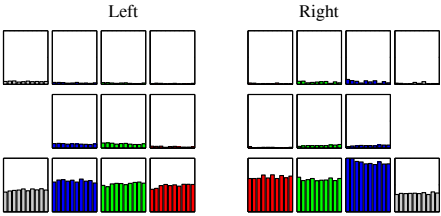


Figure 10. The facial activity as represented by the loadings of the first principal component during the opening mouth movement with one participant.

3.1. Locating Facial Activity

Examples of the detected facial activity presented as the loadings of the first principal component are shown in Figures 9 and 10.

The performance of locating the different movements based on principal component analysis is presented in Table 1. Out of the 6 included voluntary movements, opening mouth and raising mouth corners are located correctly in all the repetitions with all participants. Lowering eyebrows and raising eyebrows are located correctly in almost all the repetitions with all participants. Only a single repetition of each is incorrectly located. Lowering mouth corners is correctly located except for 4 repetitions with a single participant. Closing eyes has a limited success rate in the mapping with 7 out of 10 participants.

The locations of the three measurement channels that registered the most significant activity during the experiments

Table 1. The percentages of successful mapping of the vertical location of the movements. The last row shows the means and standard deviations for the movements.

Parti- cipant	Lowering eyebrows	Raising eyebrows	Closing eyes	Opening mouth	Raising mouth corners	Lowering mouth corners
1	100	90	50	100	100	100
2	100	100	60	100	100	100
3	100	100	90	100	100	100
4	100	100	100	100	100	100
5	100	100	30	100	100	100
6	100	100	100	100	100	100
7	100	100	40	100	100	100
8	90	100	80	100	100	100
9	100	100	100	100	100	60
10	100	100	80	100	100	100
Mean	99 ± 3	99 ± 3	73 ± 26	100 ± 0	100 ± 0	96 ± 13

according to the median loadings verified that the decisions regarding the correct sources of activity were justified according to the used order statistic, the median. The three most significant channels included incorrect locations only with one participant during the raising eyebrows movement and with 4 participants during the closing eyes.

3.2. Classifying Facial Activity

Examples of classified data are shown in Figures 11 and 12. Table 2 shows the percentages of samples that were classified as baseline. A paired t-test (significance level 0.05) did not reveal statistically significant differences between the percentages of the offline and the real-time classification. In the case of the closing eyes, the percentages show that the movement could not be classified as a movement but was classified as the baseline. The percentages for the other movements reflect the durations of the movements because the participants were not given any instructions about how long to hold them. The results of the offline and real-time classification methods are shown in Tables 3 and 4, and there are no statistically significant differences between the different methods according to a paired t-test (significance level 0.05).

4. DISCUSSION

The facial activity was mostly correctly located, but in a limited number of cases locating gave incorrect results. This can be a result of several factors. Firstly, the participants could not always carry out the movements exactly as instructed, but some unintentional activity of other muscles was included. Secondly, the measurement and the applied data processing both are slightly sensitive to the movement of the prototype device on the head. This may result in false detection of activity when the

Table 2. The average percentages and standard deviations of data points that were classified as baseline ones in the offline and real-time implementations of the classification. The number of samples is 1310 and 1179 for the two implementations, respectively.

	Lowering eyebrows	Raising eyebrows	Closing eyes	Opening mouth	Raising mouth corners	Lowering mouth corners
Offline	48 ± 18	50 ± 19	99 ± 2	34 ± 15	41 ± 17	34 ± 12
Real-time	53 ± 14	58 ± 14	99 ± 1	39 ± 15	49 ± 21	41 ± 12

Table 3. The percentages and standard deviations of correctly classified data points in the offline classification. The dashes mean that all the samples were classified as the baseline.

Partici- pant	Lowering eyebrows	Raising eyebrows	Closing eyes	Opening mouth	Raising mouth corners	Lowering mouth corners
1	98	100	-	79	99	95
2	72	100	0	70	100	98
3	85	100	-	97	36	100
4	100	100	-	91	100	90
5	100	100	-	100	100	100
6	96	100	0	81	95	100
7	93	100	0	88	96	100
8	100	75	-	100	100	100
9	100	100	-	100	90	100
10	100	100	-	100	100	100
Mean	94 ± 9	98 ± 8	0 ± 0	91 ± 11	91 ± 20	98 ± 3

Table 4. The percentages and standard deviations of correctly classified data points in the real-time implementation of the classification.

Partici- pant	Lowering eyebrows	Raising eyebrows	Closing eyes	Opening mouth	Raising mouth corners	Lowering mouth corners
1	95	100	-	84	94	98
2	100	100	-	100	100	100
3	87	100	-	60	83	100
4	100	100	-	99	100	87
5	100	100	0	100	98	100
6	94	100	0	92	78	100
7	94	100	-	78	98	100
8	96	83	-	90	100	100
9	98	100	0	98	89	100
10	100	100	-	100	100	100
Mean	96 ± 4	98 ± 5	0 ± 0	90 ± 13	94 ± 8	99 ± 4

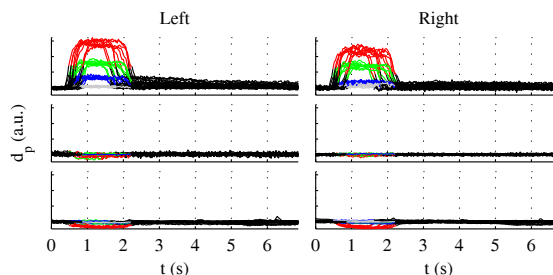


Figure 11. Classified data points after the baseline removal from the 10 repetitions of the raising eyebrows movement with one participant. The data points that were classified as the baseline are black, and the correctly classified data points are shown in colour.

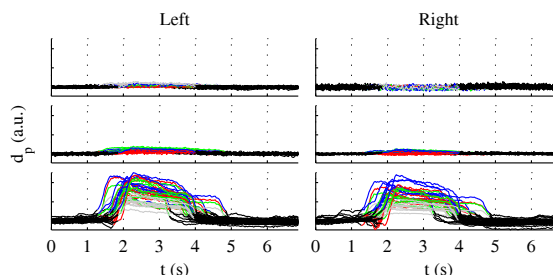


Figure 12. Classified data points after the baseline removal from the 10 repetitions of the opening mouth movement with one participant.

device moves instead of the facial tissue. Thirdly, including only one principal component may limit the performance when locating the activity. The amount of the variance explained by the first principal component was not analysed, but if it were it could be used to provide an estimate of the certainty in locating the activity. More principal components could be added to the analysis to reduce the uncertainty. Finally, the mentioned error sources are also affected by the noise in the measurement. The noise is dependent on the distance of the measurement electrodes from the target. The current implementation normalises the signal levels, but the normalisation also scales the noise so that measurements with the facial tissue further away from the measurement electrode include more noise than

when the tissue is closer. The smoothing could be more carefully considered to find the most suitable method for noise removal in this case. While the discussed factors may affect the performance, the reason for the limited performance with the closing eyes movement can be considered to be the small movement that it causes to the facial tissue at the measured locations.

It should be noted that the presented method for locating the activity only implements a rough mapping of the simple movements. Since the exact locations of the facial movements when certain muscle activation occurs varies between individuals, determining the precise location of the movements may not even provide additional value without first characterising the individual's facial behaviour. Thus, the classification was introduced to differentiate between the movements, and it could be applied also to more complex expressions. The classification was based on using hierarchical clustering to identify clusters formed from the measured data. Applying principal component analysis in real-time for the task was also considered. However, as a statistical method, it requires numerous samples to compute the principal components reliably. This causes delays dependent on the chosen window length. The processing of the implemented classification, however, does not impose additional delays since it only requires the calculation of distances between points after the clusters have been identified offline.

The percentages of data points that were classified as baseline show that the closing eyes movement is problematic also in the classification. The data points during the movement can be expected to be close to the baseline if at all visible in the data. The example graphs of the classified data points (Figures 11 and 12) show the changes from the baseline that are required for the classification to identify the data point as something else. The graphs also show that the delay for this is acceptable, even if the absolute delay cannot be calculated because no information about the true onset of the movements was extracted in this study.

The performances in classifying the data points correctly during the different movements show that the offline and the real-time versions both perform very well. This is a good result as the real-time version only used data from a single repetition of each movement for identifying the clusters compared to all the 10 repetitions in the offline one. Incorrectly performed movements, movement of the device on the head, and noise are possible sources for the errors also in the classification. In addition, the transition phases at the beginnings and the ends of the movements when the data points are close to the baseline can be expected to be more susceptible to incorrect classification.

The number of clusters chosen for the classification obviously affects how many movements and variations of the movements can be distinguished from one another. In this study, the number was chosen to be relatively small and the selection was based on the number of the included movements. The identification of the clusters used the information about the movement that the participant was instructed to perform to label each data point. Selecting a larger number of clusters would make it possible to identify variations of the movements, but it would also require more information for the labelling. One alternative would be to visually inspect video recordings to provide the labels. This could be done after the clustering to label each cluster rather than providing a label for each data point one by one.

This study only considered simple voluntary facial movements. Since complex facial expressions, even the spontaneous ones related to emotions, are formed by combinations of simple movements, they can be expected to be classified in the same way and as easily as the simple movements. They will just span a different volume in the multidimensional space of the measured data points. However, the movement ranges of facial tissue during spontaneous expressions are often more limited than in the simple movements of this study. This may introduce challenges in classifying some of the expressions.

5. CONCLUSIONS

A new method for mobile, head-worn facial activity measurement and classification was presented. The capacitive method and the prototype constructed for studying it were shown to perform well both in locating different voluntary facial movements to the correct areas on the face and in classifying the movements. Locating the movements with principal component analysis does not require a calibration of the measurement for the user, and the presented classification only required one repetition of each movement for identifying the movements before the classification could be carried out in real-time. The presented facial activity measurement method has clear benefits when compared to the computationally more intensive vision-based methods and the EMG that requires attachment of electrodes on the face.

Future research on the method should include verifying that the classification works with more complex expressions, i.e. with combinations of activity at different locations on the face. Furthermore, determining the intensity level of the activity of different facial areas could provide additional information. It could be studied how different activation levels can be distinguished from one another with the presented method, and whether even the smallest facial muscle activations can be distinguished.

ACKNOWLEDGEMENT

The authors would like to thank Nokia Research Center and the Finnish Funding Agency for Technology and Innovation (Tekes) for funding the research. Ville Rantanen would like to thank the Finnish Cultural Foundation for personal funding of his work, Nokia Foundation for the support, and IMEKO for the György Striker Junior Paper Reward that was received at the XX IMEKO World Congress.

REFERENCES

- [1] A. B. Barreto, S. D. Scargle, and M. Adjouadi, "A practical emg-based human-computer interface for users with motor disabilities", *Journal of Rehabilitation Research & Development* 37 (2000), pp.53–64.
- [2] T. Partala, A. Aula, and V. Surakka, "Combined voluntary gaze direction and facial muscle activity as a new pointing technique", *Proc. of IFIP INTERACT'01*, Tokyo, Japan, July 2001, pp.100-107.
- [3] V. Surakka, M. Illi, and P. Isokoski, "Gazing and frowning as a new human-computer interaction technique", *ACM Transactions on Applied Perception* 1 (2004), pp. 40–56.
- [4] C. A. Chin, A. Barreto, J. G. Cremades, and M. Adjouadi, "Integrated electromyogram and eye-gaze tracking cursor control system for computer users with motor disabilities", *Journal of Rehabilitation Research & Development* 45 (1) (2009), pp. 161-174.

- [5] V. Rantanen, P.-H. Niemenlehto, J. Verho, and J. Lekkala, "Capacitive facial movement detection for human-computer interaction to click by frowning and lifting eyebrows", *Medical and Biological Engineering and Computing* 48 (2010), pp. 39–47.
- [6] J. Navallas, M. Ariz, A. Villanueva, J. San Agustin, R. Cabeza, "Optimizing interoperability between video-oculographic and electromyographic systems", *Journal of Rehabilitation Research & Development* 48 (3) (2011), pp. 253–266.
- [7] O. Tuisku, V. Surakka, T. Vanhala, V. Rantanen, and J. Lekkala, "Wireless face interface: Using voluntary gaze direction and facial muscle activations for human-computer interaction", *Interacting with Computers* 24 (2012), pp. 1–9.
- [8] U. Dimberg, "Facial electromyography and emotional reactions", *Psychophysiology* 27 (5) (1990), pp. 481–494.
- [9] M. Pantic and L. J. Rothkrantz, "Automatic analysis of facial expressions: the state of the art", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (2000), pp. 1424–1445.
- [10] B. Fasel, J. Luetttin, "Automatic facial expression analysis: a survey", *Pattern Recognition* 36 (1) (2003), pp. 259–275.
- [11] J. F. Cohn, P. Ekman, "Measuring facial action", in: J. A. Harrigan, R. Rosenthal, K. R. Scherer (Eds.), *The New Handbook of Methods in Nonverbal Behavior Research*, Oxford University Press, Oxford, UK, 2005, Ch. 2, pp. 9–64.
- [12] E. L. Rosenberg, "Introduction", in: P. Ekman, E. L. Rosenberg (Eds.), *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action System (FACS)*, 2nd Edition, Oxford University Press, New York, NY, USA, 2005, pp. 3–18.
- [13] J. F. Cohn, A. J. Zlochower, J. Lien, T. Kanade, "Automated face analysis by feature point tracking has high concurrent validity with manual facs coding", in: P. Ekman, E. L. Rosenberg (Eds.), *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action System (FACS)*, 2nd Edition, Oxford University Press, New York, NY, USA, 2005, Ch. 17, pp. 371–392.
- [14] M. Pantic, M. S. Bartlett, "Machine analysis of facial expressions", in: K. Delac, M. Grgic (Eds.), *Face Recognition*, I-Tech Education and Publishing, Vienna, Austria, 2007, pp. 377–416.
- [15] J. Hamm, C. G. Kohler, R. C. Gur, and R. Verma, "Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders", *Journal of Neuroscience Methods* 200 (2011), pp. 237–256.
- [16] V. Rantanen, T. Vanhala, O. Tuisku, P.-H. Niemenlehto, J. Verho, V. Surakka, M. Juhola, and J. Lekkala, "A wearable, wireless gaze tracker with integrated selection command source for human-computer interaction", *IEEE Transactions on Information Technology in BioMedicine* 15 (2011), pp. 795–801.
- [17] O. Tuisku, V. Surakka, Y. Gizatdinova, T. Vanhala, V. Rantanen, J. Verho, and J. Lekkala, "Gazing and frowning to computers can be enjoyable", *Proc. of the Third International Conference on Knowledge and Systems Engineering (KSE)*, Hanoi, Vietnam, Oct. 2011, pp. 211–218.
- [18] V. Rantanen, O. Tuisku, J. Verho, T. Vanhala, V. Surakka, and J. Lekkala, "The effect of clicking by smiling on the accuracy of head-mounted gaze tracking", *Proc. of the Symposium on Eye-Tracking Research & Applications (ETRA '12)*, Santa Barbara, CA, USA, March 2012, pp. 345–348.
- [19] P. Ekman and W. V. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Consulting Psychologists Press, Palo Alto, CA, USA, 1978.
- [20] P. Ekman, W. V. Friesen, and J. C. Hager, *Facial Action Coding System: The Manual. A Human Face*, Salt Lake City, UT, USA, 2002.
- [21] A. J. Fridlund and J. T. Cacioppo, "Guidelines for human electromyographic research", *Psychophysiology* 23 (5) (1986), pp. 567–589.
- [22] J. E. Jackson, *A user's guide to principal components*, Wiley series in probability and mathematical statistics, John Wiley & Sons, New York, NY, USA, 1991.
- [23] M. I. Skolnik, *Introduction to Radar Systems*, 3rd Edition, McGraw-Hill, New York, NY, USA, 2001.
- [24] P.-H. Niemenlehto, "Constant false alarm rate detection of saccadic eye movements in electro-oculography", *Computer Methods and Programs in Biomedicine* 96 (2) (2009), pp. 158–171.
- [25] J. H. Ward, "Hierarchical grouping to optimize an objective function", *Journal of the American Statistical Association* 58 (301) (1963), pp. 236–244.
- [26] E. Rasmussen, "Clustering Algorithms", in: W. B. Frakes, R. Baeza-Yates (Eds.), *Information Retrieval: Data Structures and Algorithms*, 1st Edition, Prentice Hall, Upper Saddle River, NJ, USA, 1992.