KEER2010, PARIS | MARCH 2-4 2010

INTERNATIONAL CONFERENCE ON KANSEI ENGINEERING AND EMOTION RESEARCH 2010

A WEARABLE INTERFACE FOR READING FACIAL EXPRESSIONS BASED ON BIOELECTRICAL SIGNALS

Anna GRUEBLER [*] [a], **Kenji SUZUKI**[a]

[a]*Artificial Intelligence Laboratory Graduate School of Systems and Information Engineering, University of Tsukuba, Japan*

ABSTRACT

This paper proposes a novel method to read expressions on the human face through an unobtrusive wearable device by applying computational methods to bioelectrical signals captured on the side of the face. The captured signals are considered to be a mixture of distal electromyographic signals and other biological signals and can be used to achieve a personal, pattern based identification of the facial expression. Over 90% accuracy of facial expression recognition was ascertained using this method even when presented with cross-talk from other muscles. We have been developing a wearable device, called *Emotion Reader*, which was not only able to identify emotional facial expressions in real-time but also to display the emotion in a continuous manner.

Keywords: Facial expression, Emotional behavior, Electromyography, Non-verbal communication, Wearable interface.

1. INTRODUCTION

Facial expressions play a significant role in interpersonal information exchange by providing additional information about the emotional state or intention of the person displaying them [1]. As stated by Ekman [2], the face is a multisignal system showing both voluntary and involuntary expressions and is therefore necessary for successful communication [3]. The smile can express affection, humor and put others at ease as well as induce a smile in others through facial mimicry [4]. Because of this, several approaches have been followed in order to read emotions automatically from the face.

The traditional approach to recognizing emotional facial expression uses video and photo-

[*]**Corresponding author:** Artificial Intelligence Laboratory, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8573, Japan. anna@ai.iit.tsukuba.ac.jp.

graphic cameras and subsequently computer vision algorithms to identify facial expressions [5, 6]. These methods however require a camera directed at the person's face and have little tolerance against occlusion or changes in lighting conditions or camera angles.

Another approach for facial expression recognition is the wearable approach. An example of this is the use of displacement sensors attached to the facial skin, as in MIT's Expression Glasses [7]. On the other hand, recognition of facial expressions can also be achieved through the analysis of the facial surface electromyographic (EMG) signals. Traditionally, facial EMG signals were acquired by placing electrodes on the epidermis directly over the facial muscles responsible for the expressions being analyzed, as recommended by Fridlund and Cacioppo [8], in order to achieve a stronger signal and to minimize cross-talk, which is the propagation of electrical signals to and from neighboring muscles [8, 9, 10]. The analyzed muscles are usually *corrugator supercilii* (activated while frowning, negative emotional valence) and *zygomaticus major* (activated while smiling, positive emotional valence) in order to try to objectively identify the valence of emotional experiences [10, 11, 12, 13, 14]. However, placing the electrodes directly on top of the muscles on the face has several drawbacks: First, the electrodes and tape or other apparatus used to attach them cover the face, hide the expression and look unnatural, in addition to having their own weight and inhibiting the facial movements, both of which are undesirable in an interface [15]. Finally, because of the movement of the face, electrodes are easily dislodged or moved.

Other research has shown that cross-talk, has been successfully used to predict finger movement by taking signals from the arm using distal EMG [16, 17]. To date, however, no reliable and unobtrusive interfaces to read facial expressions using bioelectrical signals and display them in real-time have been developed.

This work proposes the use of cross-talk to get information about facial expressions from bioelectrical potentials captured on areas on the side of the face making emotion reading unobtrusive. Because of the mixed nature of cross-talk it is necessary to transform the sampled signal. We propose a classification method by combining two techniques: Independent Component Analysis (ICA) to transform the signals into independent components and an Artificial Neural Network (ANN) to achieve accurate facial expression identification. Both just ANN [16], and ANN and ICA jointly [17], have been used to recognize hand and finger movements from distal EMG signals.

The goal of this research is to develop an emotional communication aid to improve human-human communication through an emotion reading system that can recognize the subject's emotions in real-time and is able to display the output in different formats. Additionally it must be unobtrusive to the user, not inhibit expressions and work in any environment regardless of changing lighting conditions and position of the subject. At this time, the research focuses on recognizing the smile from other expressions.

The *Emotion Reader* has applications in several areas, especially in therapy and assistive technology. It can be used to provide biofeedback to patients during rehabilitation and a quantitative smile evaluation during smile training. Additionally it can aid the visually impaired: the listener is able to perceive the speaker's facial expressions, through alternative forms of communication such as audio or vibro-tactile stimulation. It is also a tool in e-learning, distance communication

Figure 1: *Emotion Reader* Overview

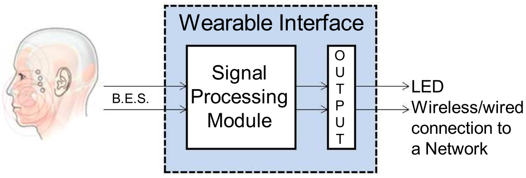

Figure 2: System Overview

and computer games because it can transmit the facial expression automatically without the need of high bandwidth. Because it is an unobtrusive wearable device it can be used outside the laboratory for continuous expression detection in environments where cameras are not supported or where subjects require high mobility. Another application lies in increasing the quality of life for patients suffering from facial paralysis, where the signals obtained from the healthy side of the face can be used to control a robot mask that produces an artificial smile on the paralyzed side [18]. In this paper, we first discuss the methodology used for the recognition of emotional facial expressions. Then we show several experiments on the subjects' bioelectrical signals. The obtained results show the significant rate of recognition in real-time and verify that the design of a wearable interface is feasible. Subsequently the wearable interface device will be introduced. Finally a discussion and conclusions will be presented.

2. METHODOLOGY

An overview of the *Emotion Reader* is illustrated in figure 1. The subject is required to wear the interface around the ear to obtain bioelectrical signals (B.E.S.) which are filtered and then run through the Signal Processing Module where ICA is used to separate independent components (I.C.). The signals are then filtered again and subsequently classified as patterns by an ANN. The output can be used to light a led on the wearable interface or be transmitted wired- or wirelessly through a network to be used by an application. The system overview is illustrated in figure 2.

2.1. Bioelectrical Signals measurement

As stated in previous studies [3], several muscles are activated for each facial expression. The main muscle activated during smiling for both voluntary and involuntary expressions is *zygomaticus major*, but *orbicularis oculi* (see figure 3a) is also contracted for the *felt* smile [3].

Because of cross-talk, when obtaining data from the side of the face, EMG from other muscles of the face not involved in smiling are also included in the sampled signal. A simplified model

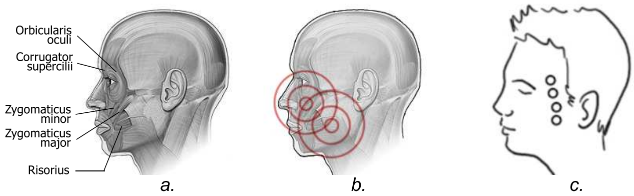

Figure 3: a. Typical muscles of the face [19] b. EMG cross-talk [19] c. an example of the proposed electrode placement.

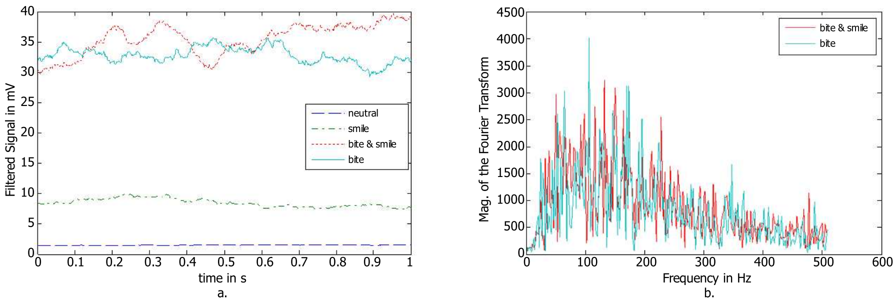

Figure 4: Bioelectrical signals. a. Averaged filtered signal for different facial expressions. b. Frequency domain of the raw signals for biting and simultaneous biting and smiling.

of cross-talk is shown in figure 3b. For the distal detection, four electrode pairs were placed on the subject's face, two pairs on each side of the face. The electrode locations that were used are shown in figure 3c. This is a new approach to bioelectrical signal capture on the face as opposed to the traditional electrode placement directly on top of the muscles, which makes it even more difficult to differentiate facial expressions from bioelectrical signals. Even when placing the electrodes directly over *zygomaticus major*, which should provide a clearer signal, it is not possible to differentiate between just "biting" and "simultaneous biting and smiling" because the signals for both these movements are similar in amplitude and their domains overlap, as described in figure 4a. As people sometimes bite down while smiling this poses a problem for the classification of the facial expression. The frequency domain does not offer significant information on the type of facial expression either because there are no distinctive frequency domains for different expressions as shown in figure 4b.

A more complex form of analysis is therefore necessary for the classification of facial expressions. In this case, ICA was implemented.

2.2. ICA and ANN based Classification

ICA separates statistically independent source signals iteratively from their sampled linear combinations (signal mixtures) and produces a demixing matrix that can be applied to the mixed signal and transforms it linearly into its independent components. It assumes that the source signals are independent and that there are usually at least as many different sampled mixtures of a set of source signals as there are source signals themselves [20].

It is appropriate to apply ICA to face EMG, because EMG signals originate from different

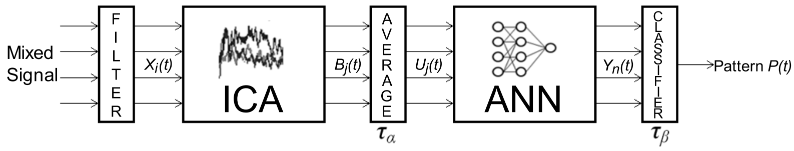

Figure 5: Signal Processing Module

muscle groups therefore the original signals are statistically independent [17]. In addition, because one can place electrodes on different positions on the face, it is possible to acquire different mixed signals for analysis where the original signals are present in different intensities.

In this implementation, the fastICA [21] and a Gaussian kernel, $g(\alpha) = \alpha e^{-\frac{2\alpha^2}{2}}$, are used in the signal processing module.

Four mixed signals were recorded and different number of I.C. were attempted in order to look for the best possible recognition rates.

Because of the iterative nature of ICA, it is impossible to determine the order of the independent components after demixing and their exact amplitude and sign also cannot be determined. Therefore an adaptive system that can be trained with the obtained independent components must be used. Multilayer ANNs can provide nonlinear classifications and can be trained with data from each subject and self-calibrate regardless of the order and amplitude of the obtained independent components from the ICA [22]. The ANN is a two-layer feed-forward network, with sigmoid hidden and output neurons. 4 neurons were used in the hidden layer. Input are the independent components and output is a binary vector categorizing the expression as “smile” or “not-smile”. For training of the ANN it was decided to use the patterns of neutral, smiling, biting and simultaneous smiling and biting face. The model was subsequently expanded to include the frown. An overview of the Signal Processing Module can be seen in figure 5.

The sampled and filtered data at time t, at the i-th electrode is defined as $x_i(t)$ ($i = \{1, 2, ..., N_e\}$ where N_e represents the number of electrode pairs). $X_i(t)$ is the collection of x_i used for ICA, where $X_i(t) = \{x_i(t), x_i(t+1), ...\}$. After separation by the ICA each independent component is defined as

$$b_j(t) = Wx_i(t)(j = \{1, 2, 3, ..., N_c\}) \tag{1}$$

where N_c represents the number of independent components and W is the demixing matrix. $B_j(t)$ is the collection of all $b_j(t)$, where $B_j(t) = \{b_j(t), b_j(t+1), ...\}$. Before the demixed I.C. are fed to the ANN, their absolute value is calculated and they are run through an averaging filter for smoothing (window τ_α=100 ms), $U_j(t)$ is the output of the averaging filter. Calculating the absolute value serves to avoid values of amplitude canceling each other out during the smoothing. The averaging filter discards outliers and gives a more accurate value for prediction of movement because each smoothed value represents a window of measurement.

$U_j(t)$ is used for training the ANN. Then, $Y(t)$ is the output of the ANN, where $Y(t) = \left\{y_1(t), y_2(t)..., y_{N_f}(t)\right\}$ and N_f is the number of expressions in which to classify the data. The pattern $P(t)$ is the classification result of the emotional facial expression, which is integrated

Figure 6: Electrode placement locations

within the constant time window τ_β where:

$$P(t) = max\left\{\int_{t-\tau_\beta}^{t}(y_1), \int_{t-\tau_\beta}^{t}(y_2), ..., \int_{t-\tau_\beta}^{t}(y_{N_f})\right\} \tag{2}$$

For calibration we need sampled data from each individual subject for T_Θ [s], which is used to create the demixing matrix and train the ANN. In this implementation $N_e = 4$ and $N_f = 2$ for classifying only the smile and $N_f = 3$ after including the frown.

3. EXPERIMENTS

In order to verify if distal sEMG signals can be used for accurate recognition of facial expressions, we first obtained recordings of bioelectrical signals from the side of the face (two bipolar electrode pairs on each side of the face) using Vitrode F Ag/AgCl disposable surface electrodes on 6 subjects. The electrode locations (N_e=4) are shown in figure 6. The subjects were 6 healthy individuals (3 female, 3 male), whose mean age was 31 years (ranging from 22 to 52 years). The subjects were asked to produce voluntary expressions or movements and maintain them for calibration. In this experiment, the calibration time T_Θ for each expression is 4 [s]. 4 recordings (sets), of 4 seconds each, were made of each expression and were then used as training or test sets (each set was used twice as training set and as test set for all other belonging to the same subject). The sampling rate was 1000 Hz. All signals were filtered with a pass-band (5-400 Hz) and a notch-filter (50 Hz) to avoid power noise. One group, 16 seconds in length (4 seconds for each expression {"neutral", "smile", "bite-smile", "bite"}) was used for the ICA and training the ANN. Recordings of another instance of the same expression were then separated into their I.C. by the produced matrix and fed through the trained ANN for classification as a test set.

3.1. Smiling face classification

Figure 7a shows a performance comparison between the training set and test set in terms of the classification rate. The time length of the test set is 16 seconds, N_f=2, {"smile", "non smile"}. Feeding the 4 raw signals directly to the ANN, offered in average only 71.33±10.21% accurate classifications through the ANN. On the other hand, after training with the signals separated in their I.C. (N_c=4), it was possible to achieve in average 93.62±7.07% accurate classification through the ANN in the test set.

In order to investigate the personal and user specific nature of the sEMG signal and the ICA, it was also attempted to classify the sEMG data from a subject by feeding it to an ANN trained with

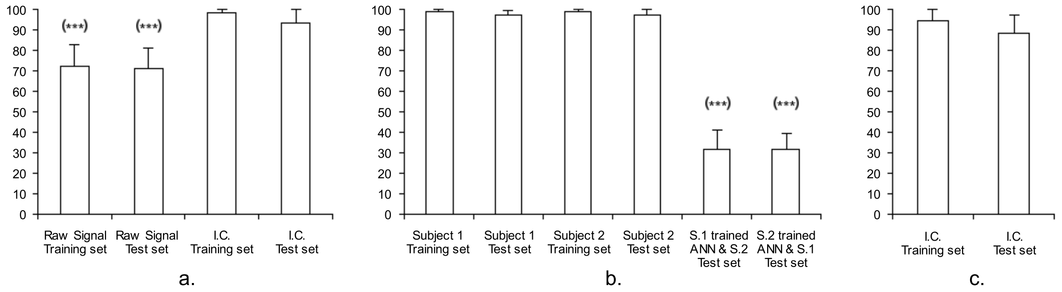

Figure 7: a. Raw Signal trained ANN vs. Independent Component trained ANN classification in %. b. Correct classification for subject 1, subject 2 and by testing with the trained ANN by the other subjects' data. c. Correct recognition of "smile", "frown" and "neither" in %

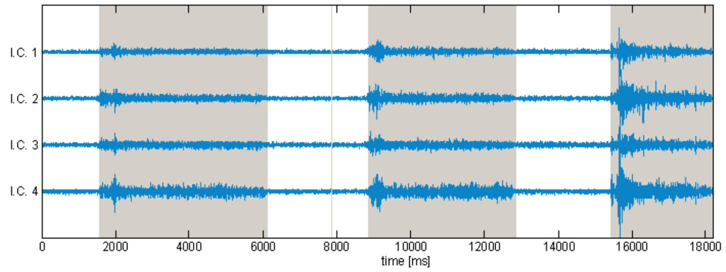

Figure 8: Smile detection (indicated by grey areas) from B.E.S. in a continuous data stream.

data from another. The results can be seen in figure 7b. It should be noted that the classification of the other subject's facial expressions is very poor.

3.2. 3-States Classification

We also expanded the proposed model to include the frown (N_f=3, {"smile", "frown", "neither"}), because facial EMG is usually used to assign positive or negative valence for emotional expressions. We carried out extended experiments for the classification of three facial expressions. The following results for correct classification (Fig. 7c) were obtained after training the network. The experimental conditions are the same as the previous experiments', with the addition of the "frown" to the training and test sets.

3.3. Realtime Smile Detection

By using a demixing matrix and an ANN trained with one of the previously recorded sets, it was possible to adequately classify the smile from a continuous data-stream. As can be seen in figure 8, smiles were predominantly correctly classified by the ANN. Real-time classification of the smile was possible by using the *Emotion Reader* (Fig. 9). The smile was displayed in the form of lit LEDs on the side of the face. The sampling window for classification τ_β was 0.25 s. The display is updated every 0,5 s allowing for signal processing. Furthermore, movements and expressions not part of the training set (blinking, head movements) were not misclassified by the device.

Figure 9: A prototype of the *Emotion Reader*

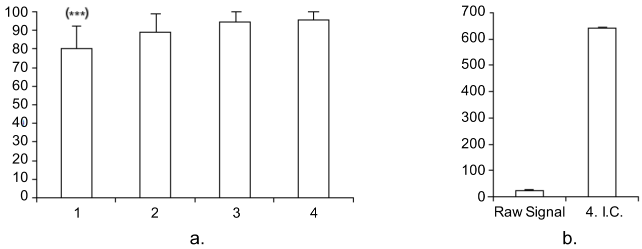

Figure 10: *a.* Number of I.C. (N_c={1, 2, 3, 4}) vs. average percentage of accurate classification. *b.* Total K-L Divergence for the raw signal and 4 I.C.

4. DISCUSSION

It has been shown that is possible to accurately identify smiling and non-smiling faces from bioelectrical signals on the side of the face. The ICA analysis works well for EMG signals because the different muscle groups generate statistically independent signals that propagate through the face and could be captured at different locations making ICA possible.

In order to investigate the characteristics of the independent components, we conducted several experiments with different conditions. We first tried to find a number of I.C. that provided good demixed signals for subsequent classification through the ANN. With regard to the different number of I.C.s that could be separated for classification, we obtained the average percentage of accurate classification for one subject. Figure 10a shows that the average correct classification increases as the number of I.C. used is increased.

By increasing the number of I.C., it was possible to increase the accuracy of the classification: One I.C. offered poor results, but 4 gave very accurate classification. Therefore a good classification is possible using 4 electrode pairs that can be attached comfortably to the side of the face making a wearable device feasible. Interference created by the jaw muscles during biting, a persistent problem in facial bioelectrical signal capture, is overcome with the ICA, providing good input vectors for the ANN. The use of an ANN for a pattern based identification of facial expressions instead of trying to identify signals belonging to a specific muscle is a new approach to facial expression analysis based on physiological signals. The use of an ANN for classification overcomes the weakness in the ICA regarding the uncertainty in the order in which the channels are decoded and their amplitudes. The Kullback-Leiber Divergence [22] of the probability distribution was calculated for each signal and for each expression in pairs. The results were added (Fig. 10b). As the K-L Divergence between components increases, so does the correct classification through the ANN.

It is the nature of bioelectrical signals to be a mixture of several source signals, but by increasing the independence of the components from each other through ICA, the classification of the ANN is improved. Even when using the ICA, ICA matrices that produce components with higher independence from each other, and therefore a higher K-L divergence, show higher classification rates. This can be used as a tentative measure to choose the best possible demixing matrix from a set before feeding the signal to the ANN.

5. CONCLUSIONS

In this paper we proposed a novel methodology to analyze facial expressions using distal sEMG and bioelectrical signals, independent component analysis and a trained ANN. It was shown that it is possible to detect the smile from signals captured on a location not directly on top of the muscles responsible for it in a continuous manner. A good identification can be achieved by using a small number of electrodes placed distally on the face. Further, it becomes apparent that a pattern based identification of facial expressions is necessary and that an adaptive mechanism like an ANN is required for classification because of the differences in the sEMGs from person to person.

The results of trying to identify one subject's facial expressions by using an ANN trained with data from another subject showed a very low recognition rate. This is to be expected because of the personal nature of EMG signals.

It was also shown that the K-L Divergence between independent components is greater than that of the raw signal, showing than an ICA approach for facial bioelectrical signal analysis is adequate as an increased K-L Divergence value leads to a better classification through the ANN.

It has finally been shown that it is feasible to design a wearable interface for emotion reading in real-time based on bioelectrical signals that, because of its wearable and unobtrusive design, can be used by subjects in an everyday environment outside the laboratory.

Future work should include increasing the number and degree of the recognized facial expressions.

REFERENCES

[1] Ekman, P, Friesen, W. V, and Ellsworth, P. What emotion categories or dimensions can observers judge from facial behavior? In Ekman, P, editor, *Emotion in the Human Face*. Cambridge University Press, 1982.

[2] Ekman, P. Strong evidence for universals in facial expressions: A reply to russells mistaken critique. *Psychological Bulletin*, 115(2):268–87, 1994.

[3] Surakka, V and Hietanen, J. Facial and emotional reactions to duchenne and non-duchenne smiles. *International Journal of Psychophysiology*, 29:23–33, 1998.

[4] Dimberg, U and Thunberg, M. Rapid facial reactions to emotional facial expressions. *Scandinavian Journal of Psychology*, 39:39–45, 1998.

[5] Cohn, J, Zlochower, A, Lien, J, and Kanade, T. Automated face analysis by feature point

tracking has high concurrent validity with manual facs coding. *Psychophysiology*, 36:35–43, 1999.

[6] Cohen, I, Sebe, N, Garg, A, Chen, L, and Huanga, T. Facial expression recognition from video sequences: temporal and static modeling. *Computer Vision and Image Understanding*, 91:160–187, 2003.

[7] Scheirer, J, Fernandez, R, and Picard, R. W. Expression glasses: A wearable device for facial expression recognition. In *CHI '99 Short Papers*, Pittsburgh, PA, 1999.

[8] Fridlund, A and Cacioppo, J. Guidelines for human electromyographic research. *Psychophysiology*, 23:567–589, 1986.

[9] Sato, W and Yoshikawa, S. Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition*, 104:1–18, 2007.

[10] Partala, T, Surakka, V, and Vanhala, T. Real-time estimation of emotional experiences from facial expressions. *Interacting with Computers*, 18:208–226, 2005.

[11] Achaibou, A, Pourtois, G, Schawarz, S, and Vuilleumier, P. Simultaneous recording of eeg and facial muscle reactions during spontaneous emotional mimicry. *Neuropsychologia*, 46:1104–1113, 2008.

[12] Jäncke, L. Facial emg in an anger-provoking situation: individual differences in directing anger outwards or inwards. *International Journal of Psychophysiology*, 23:207–214, 1996.

[13] M.Shut, Tuinenbreijer, K, van den Broek, E, van Herk, J, and Westerink, J. Computing emotion awareness through galvanic skin response and facial electromyography. *Probing Experience*, 8:149–162, 2008.

[14] Cacioppo, J, Petty, R, Losch, M, and Kim, H. S. Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology*, 50:260–268, 1989.

[15] Huang, C.-N, Chen, C.-H, and Chung, H.-Y. The review of applications and measurements in facial electromyography. *Journal of Medical and Biological Engineering*, 25(1):15–20, 2004.

[16] Tsenov, G, Zeghbib, A, Palis, F, Soylev, N, and Mladenov, V. Visualization of an on-line classification and recognition algorithm of emg signals. *Journal of the University of Chemical Technology and Metallurgy*, 43(1):154–158, 2008.

[17] Naik, G, Kumar, D, Singh, V, and Palaniswam, M. Hand gestures for hci using ica of emg. In *HCSNet Workshop on the Use of Vision in Human-Computer Interaction (VisHCI 2006)*, Canberra, Australia, 2006.

[18] Jayatilake, D and Suzuki, K. A soft actuator based expressive mask for facial paralyzed patients. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2008)*, Nice, France, 2008.

[19] Drake, R, Vogl, W, Mitchell, A, and Gray, H. *Grays Anatomy for Students*. Elesevier/ Churchill Livingstone, 2005.

[20] Haykin, S. *Unsupervised Adaptive Filtering*. John Wiley & Sons, Inc., USA, 2000.

[21] Gävert, H, Hurri, J, Särelä, J, and Hyvärinen, A. Fastica package for matlab. http://www.cis.hut.fi/projects/ica/fastica/, 2005. [Accessed 2009 January 20th].

[22] Duda, R, Hart, P, and Stork, D. *Pattern Classification*. John Wiley & Sons, Inc., USA, 2001.