COMPREHENSIVE BRAINWAVE AUTHENTICATION USING EMOTIONAL STIMULI

Violeta Tulceanu

Faculty of Computer Science, ”Alexandru Ioan Cuza” University of Iasi
violeta.tulceanu@infoiasi.ro

ABSTRACT

In this paper we propose a new concept of authentication. Beyond deciding whether a user is entitled to the resources he claims, the authentication process provides information about the state of the user and his capacity to responsibly handle the resources he has the right to access. We chose emotions as credentials because aside from being inextricably linked to the user, their variation over time provides valuable meta-information about the user's state and evolution. On a practical level, the first target of this type of authentication are disabled persons who could thus gain higher life quality by having access to online banking, education etc.

Index Terms— brainwave authentication, BCI, brain computer interfacing, emotions, emotion classification

1. INTRODUCTION

For approximately forty years research is being carried on in the direction of brain computer interfacing (BCI). Since the early work of J.J.Vidal [1] the domain has evolved significantly. There has been much work [2] done in the area of spelling applications over the last years, making it possible for disabled patients to communicate via a BCI. Furthermore, research is being carried out in thought recognition and intention expression. These, together with attempts made to use BCI to enable web searches, have made us consider the possibility of allowing disabled persons to use web applications that require authentication. Such examples would be online banking, using e-mail, sitting online examinations etc., which would provide such patients with the possibility of leading a less confined life, by giving them access to education, communication and the ability to handle their own finances. Thus, one of the purposes of this paper is to introduce an emotion-based brainwave authentication scheme that will allow disabled patients to gain more control over their lives. Previous work on brainwave authentication such as [3, 4, 5, 6, 7, 8, 9] proposes schemes that share the fact that they all focus exclusively on establishing identity, without further concern about information that may be obtained about the subject by analyzing his brainwaves, and without taking into account how the subjects' response will vary over long periods of time.

In this paper we propose a method of identifying a user via brainwaves and uses their full potential of providing meta-information about the user during the authentication process. We use auditive and visual emotional stimulation. This enables us to observe the emotional changes of the user over time, and beyond being able to state whether the user is entitled to a resource, we are able to state whether the user is capable of handling the given resource responsibly. Perturbed emotional states associated to manic episodes, minor epileptic seizures, etc. would be visible when using this type of authentication, allowing the system to react accordingly and limit access to resources. Slow changes in emotional responses over time will allow the system to detect and predict potential crises and long-term depressive episodes.

Thus, in this authentication scheme, we are interested in the following: a) establishing the user's identity and deciding whether the credentials come from him or an impersonator; b) establishing the user's current state of mind and if he can responsibly handle the resources he is entitled to; c) detecting if the user is acting of his free accord; d) predicting the user's future behavior; e) providing time-adapting user model.

2. EMOTIONS AS CREDENTIALS

Emotions are triggered by imagination and anticipation [10, 11, 12]. However, [10] mentions a bias towards forming unconscious emotional memories in periods of intense stress. Thus memory could substitute anticipation in eliciting an emotional response. Rememorating a strong emotion and the context where it occurred should trigger an emotional response (as also stated by [13, 14, 15, 16, 17, 18]). The amplitude of this response will depend on the user's current state of mind. Plain rememoration of an unpleasant event is expected to trigger a low emotional response if the subject is in a normal state of mind. However, if the mental state is not stable, then the emotional response should be augmented.

As we chose using the electroencephalogram (EEG), we measure brain activity at scalp level. Subcortical activity cannot be measured by EEG, thus rendering direct assessment of the activity in the amygdala, hippocampus, thalamus etc. impossible. Instead, we can assess the activity in the cortical projections of the amygdala. Such areas would be the dorsolateral prefrontal cortex and the median prefrontal cortex [10]. We thus consider that we can deduce the emotional state of a subject, based on which cortical areas present increased activity, its nature, and which subcortical emotional systems have their projections in the given areas.

In this paper, the EEG signal recorded over 20 channels produces the set of moderator variables, and we use stimulation similar to that employed by Ekman [13].

3. PROPOSED AUTHENTICATION SCHEME

Authentication scheme In this authentication scheme, the credentials should be collected via a BCI with embedded functions for performing client-side processing, and a good bit transmission rate. Once the signal is processed as we describe in 5, an emotional pattern is extracted by selecting the most and least dominant centroids for each type of emotional stimulation, on each channel. The set of centroids is sent as an encrypted, timestamped message to the trusted authority(TA) that will decrypt and rewrite the string of PSDs(power spectral densities) as a string of words in the language induced by the user model. The TA will accept or not the credentials, after passing the resulting string through the automaton associated to the language determined by the user model.

1. EEG recording in a state of calm.
2. Auditive emotional stimulation. The subject listens to words related to circumstances in which he experienced contempt, disgust, fear, worry, pleasure, affection, love, pride, hope and sadness. He should rememorate the circumstances in which the emotions occurred and focus on that emotion for one or two seconds.
3. Client-side signal preprocessing: signal recovery, noise elimination and artifact removal, segmentation and k-means clustering. Also, the dominant and nondominant clusters are determined.
4. The client sends the trusted authority the following message(encrypted and timestamped):

   {em.no.}{ch.no.}{D}{dominant-centroid1}...
   {dominant-centroid10}{N}{nondominant-centroid1}...
   {nondominant-centroid10}

5. The trusted authority decrypts the message, verifies the timestamp and assigns each of the received PSDs to the user-model centroids to which it has the smallest log-spectral distance.
6. On doing this, the message will be rewritten as a sequence of words in the language induced by the user model, by assigning to each centroid in the message the label of the centroid in the user model, in the same category to which it has the least cross-spectral distance. For instance, the message could be of the form:

   $$\{em1\}\{ch.no.1\}\{1,2,3,4,\neg 8\}\ldots \{ch.no.20\}\{3,6,7,8,\neg 11\}$$
   $$\ldots$$
   $$\{em20\}\{ch.no.1\}\{3,5,7,\neg 10\}\ldots \{ch.no.20\}\{4,8,9,\neg 12\}$$

7. The trusted authority (TA) will pass the words through the dynamic pushdown automaton and then obtain whether the credentials can be accepted or not as part of the user language.
8. The TA stores a hash of message sent by the client.

It is important that the data acquisition device should be able to perform the processing very fast, since the amount of data to be processed is likely to be somewhat large.

Predicting future behavior Valid credentials should be accepted by the automaton associated to the language determined by the user model. If they are not accepted by the automaton, authentication fails. This means the input credentials clearly differed from the user model, suggesting they were either from an impersonator or the subject was in a highly disturbed mental state. However, this does not point up more subtle, long-term changes in the user's state of mind. Although these are not as obvious as short-term changes caused by distress or an epileptic seizure, if prolonged sufficiently, can have more dramatic effects.

To capture these we defined a centroid's sphere as the set of instances assigned to the centroid in training. We denote the sphere radius with ρ, defined as the greatest cross-spectral distance between the centroid and its assigned instances.

In the above automaton, let ***st*** be a state and PSD_{st} its corresponding PSD in the credential set.If the cross-spectral distance between the centroid corresponding to ***st*** and PSD_{st} is greater than ρ, then the next transition will be to a state ***st'*** that is equivalent to ***st***, allowing authentication, but marking that the equivalent state ***st'*** was used. When a given number of consecutive successful authentications have runs that include such equivalent states, the TA should react by updating the user model (recalculating the centroids).

When the user model is updated, the previous model is stored, along with the associations between the symbols and the centroids. In turn, these must be analyzed, in order to make predictions about long-term future behavior.

Possible attacks Potential weaknesses involve data interception, replay attacks and tampering. Data interception is impossible, as brain activity produces low voltage signal(10-100 microV) and there are no sensors capable of intercepting such a signal from a distance. Timestamps and watermarking prevent replay and tampering. Artificial brainwave generation uses previously captured credentials to deduce a model of the user's data and generate fresh, valid ones. However, we only send centroid labels over the network, not raw data. Exhaustive search by injecting messages with combinations of centroids would mean making $20 channels * (\sum_{k=1}^{12} C_k^{12}) * 10 emotions$ combinations. The TA can prevent this by limiting the number of consecutive authentication attempts.

4. DATA ACQUISITION

We performed full EEG using the 10/20 International electrode system in a clinic. The user profiles of 5 subjects were built, their average age being 50. Only the data from 3 subjects was valid. To acquire the data, we performed the experiment described below,consisting of three types of mental tasks that alternate with recording reference brain signal without stimulation. Prior to conducting the experiment ten emotions were selected to act as challenges: contempt, disgust, fear, worry, pleasure, affection, love, pride, hope and sadness.The subjects underwent two data acquisition sessions, with one or two days in between sessions.

Each subject was asked to set a list of three words he strongly associated with personal circumstances in which he had experienced each emotion, resulting in a 30 word-long list. Also, they were required to set three personalized pictures, and associate to them four words they related and four they did not relate to the images. When presented an emotionally charged word or picture, they were required to focus on the stimulus for five seconds and rememorate personal circumstances linked to it. There were 15 second breaks between tasks - EEG recording with eyes closed, without stimulation. The final task was setting a password, which they were supposed to later mentally recompose by seeing letters highlighted on screens with the alphabet. After seeing all the letters that composed their chosen password in the correct order, they could halt the experiment.

5. SIGNAL PROCESSING

We used a 300 Hz sampling frequency. To eliminate noise induced by the powerline, eye movements(EOG), muscular movements(EMG) and the cardiac rhythm(ECG), we used a B-form smoothing spline (tolerance 1). The resulting values were scaled by a 0.01 factor. To recover lost signal fragments, we used the signal before and after the gap to interpolate using a left-sided and right-sided autoregressive model and then performed cross-fading to obtain the missing values.

The signal is nonstationary because of recording duration(16 min), fatigue, switching between alpha and gamma brain activity, blinking, and alternation of visual and auditive stimuli. To eliminate nonstationarity, each signal sample was split into 1 second frames, and each frame was further split into 16 subframes each 62.5 mseconds long, with a 50 msecond overlap. For each of the 62.5 ms subframes we calculated the PSD using the Welch averaged modified periodogram approach. In order to distinguish between alpha(relaxation) and gamma(active processing) rhythms, we applied k-means on the PSDs extracted above, on each channel. We set 12 centroids, one for each of the 10 emotions, one for the indifferent states and one for alpha rhythm. Every subfragment was marked to allow easy rearrangement of the classified PSDs in their original order.

6. EMOTIONAL PATTERNS. BUILDING THE USER MODELS

As described in 5, we performed segmentation, splitting the filtered signal into 62.5 ms PSDs with 50% overlap. The PSDs obtained from each channel were classified with the k-means algorithm, grouping the data around 12 centroids. The result were 20 arrays of centroid identifiers (centroids from different channels with equal labels are not equal). We can represent a channel CH_i in two ways: in the first formula below, CH_i is a succession of words(one for each emotion Em_j) over the alphabet of labels assigned to the centroids around which the PSDs are grouped; in the second, CH_i is viewed as a union of compositional data points.

$$
\begin{aligned}
CH_i = \bigcup\{\cup CentroidLabel(PSD_j)| \\
PSD_j \in Em_j,\ 1 \leq j \leq 10,\ 1 \leq i \leq 20\} \\
CH_i = \bigcup\{\cup CentroidLabel(PSD_k)| \\
PSD_k \in Centroid_k,\ 1 \leq k \leq 12\},
\end{aligned}
$$

for all $1 \leq i \leq 20$.

We represent each of the 10 emotions Em_j as a compositional data point E_j with 12 parts, one standing for each centroid, and a probability distribution defined on the number of PSDs assigned to that emotion.

$$E_{[c_1,\ldots,c_{12}]} = \left[\frac{c_1}{\sum_{i=1}^{12} c_i}, \ldots, \frac{c_{12}}{\sum_{i=1}^{12} c_i}\right],$$

where $c_i = |\{PSD|PSD \in Centroid_c\}|$. The subjects were asked to focus on each word for 5 seconds, and listened to series of 12 words associated to a certain emotion. This produces 60 second sequences of emotional stimulation of a given type, meaning 960 segments 62.5 ms long. So, the total number of PSDs assigned to a certain emotion is 960 ($= \sum_{i=1}^{12} c_i$). Assuming a uniform distribution of emotions to centroids, we obtain $E_{c_i} = 80/960$, for all i.

However, since emotional stimulation does not induce a uniform distribution of PSDs to centroids, we considered any centroid c_i with $E_{c_i} > 90/960$ to be ”dominant”. The c_i centroid labels are symbols that form an alphabet ($A(CH_i)$) associated to each channel.

$$A(CH_i) = \{c|c \text{ is a centroid label}\},$$

for all $1 \leq i \leq 12$. For each emotion, on each channel, the symbols associated to the ”dominant” centroids form unique words that encode emotions in the formal language associated to the given channel. After analyzing the data from our subjects, we noticed the following:

1. The channel prefix is formed by permutations of a constant subword.
2. An emotion is expressed as a permutation of the channel prefix plus an emotion-specific suffix.
3. An emotion is uniquely represented on each channel.

4. Not all emotions are distinguishable over all channels. Example: on channel 12 we cannot distinguish between the first seven emotions, nor between pride and hope, but we can distinguish sadness and neutrality from among all others.

Observation 1 supports LeDoux's affirmation that the cortex is not preoccupied with the mechanisms underlying a certain emotion, but rather with the processing itself. An emotion is expressed over a channel as a permutation of the channel prefix to which a suffix is added or not, inducing a decision tree-like structure for each channel, with the channel prefix at its root and the emotions as leafs.

Observations 2, 3 and 4, as well as the fact that different emotions will activate different combinations of areas of the cortex, according to the cortical projections of the underlying subcortical emotional subsystems, lead us to assert that *the expression of an emotion is the concatenation of the words that encode it over the channels where it is distinguishable* .

Definition 1: A *word* is any string $E = c_1 \cdots c_n$, where each c_i is the label of a dominant centroid in the subsequence corresponding to a certain type of emotional stimulation.

Remark that each word has a unique representation in terms of PSD. As each symbol label marks a centroid, which in turn is a PSD, we define $f : \{1, \ldots, 12\} \rightarrow \cup PSD$ that translates a symbol to its corresponding PSD. The function f is applicable on words, concatenating the PSDs.

Let $channel_i \neq channel_j$, $E_i = c_1 ... c_n$ a word from $channel_i$, and $E_j = c'_1 ... c'_m$ a word from $channel_j$. Assuming $E_i = E_j$ at a symbolic level, this does not imply that they encode the same emotion. Each channel has its own language, and although E_i and E_j are equal, they belong to different languages and have different meanings: $f(E_i) \neq f(E_j)$.

Definition 2: Two emotions $E_1 = c_1 ... c_n$ and $E_2 = c'_1 ... c'_m$ are *distinguishable* over the same channel if $m \neq n$ or $c_i \neq c'_j$ for some i and j.

For instance, fear is distinguishable over channels 1, 3, 5, 6, 10, 13, 15, 18, 19, 20 for one of the subjects. This means that for the given subject we would be expressing fear as:

$$\begin{aligned} \text{Fear} \;\; = \;\; & \{\mathbf{ch1}\}\{7,2,4,11\}\{\mathbf{ch3}\}\{9,8,3,7\}\{\mathbf{ch5}\}\{8,9,1,3,2\} \\ & \{\mathbf{ch6}\}\{4,8,9,7\}\{\mathbf{ch10}\}\{10,8,6,5\}\{\mathbf{ch13}\}\{1,9,10\} \\ & \{\mathbf{ch15}\}\{8,12,7\}\{\mathbf{ch18}\}\{3,11,10,4,2\} \\ & \{\mathbf{ch19}\}\{1,2,5,8,9\}\{\mathbf{ch20}\}\{3,2,5,9\}. \end{aligned}$$

By applying the function $f(Fear)$ we obtain the representation of the words in terms of centroids.

More generally, an emotion $\mathbf{E} = (\{channelidentifier\}E)*$. The combination of patterns such as the one above for each emotion will produce the user model, which is a succession $E_1 ... E_{11}$, where $\mathbf{E} = (\{channelidentifier\}E)*$.

In other words, our user model captures the distinctive features of the brain signal when exposed to a certain type of stimulation, and the exact combination of brain areas that respond in a particular way to the given stimulation(applying the distinguishability property over channels, we select only channels with increased activity for a given stimulus, in contrast to those showing only the generic channel prefix).

7. CONCLUSIONS

In this paper we proposed a new brainwave authentication scheme based on emotional responses. We provided argumentation that emotions can serve as credentials, we proposed and performed an experiment to collect data we further processed in order to obtain the user models.

Each emotional response on a channel was encoded as a distinguishable word. When a given emotion was distinguishable over a channel in more than 50% of the recording sessions, we considered the channel stable and included it in the emotion's generic pattern. We thus obtained 11 cortical activation patterns(which we do not include for space considerations), corresponding to the types of stimulation used. On analyzing them we noticed they overlap cortical regions acknowledged to be associated with the given emotions. To assess the authentication scheme's performance, we restricted the user models to the stable generic activation patterns and compared the users two by two. The image below shows an overall authentication success rate of 29.16%, using only the stable generic activation patterns for three users.

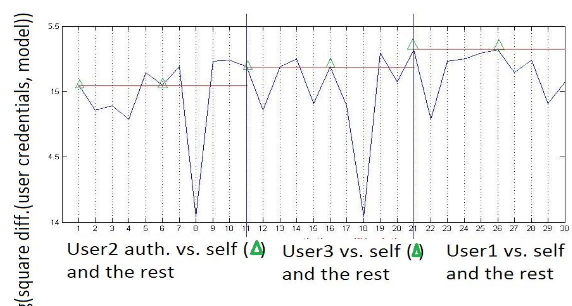

Fig.1: Authentication results using generic emotional patterns

When the restriction is relaxed to the distinguishable emotions on channels(Table 2), the user model includes the personal features that allow differentiation between users, and we obtain 100% authentication success rate.

Table 1: Personal emotional patterns - activated channels

Emotion	User 1	User 2	User 3
Contempt	1,2,6,8,13	1	14
Disgust	1,5,10,13	16	5
Fear	-	13,14	15,19
Worry	9	6,13	5,6,15
Pleasure	15	-	12
Affection	-	7,9,18	2
Love	5	10	4
Pride	-	15	1
Hope	9,13	-	-
Sadness	-	7,12	-
Indifference	-	7	-

Emotional authentication thus becomes an issue of deciding if a set of words(the current credentials) pertain to the language associated to the user. A potential model would be a tree automaton. By composing the trees for each channel the user model is obtained, and the linearization of the resulting tree represents the user's credential set. As future work, we consider modeling the emotions using a dynamic automaton obtained by composition of signature input-output nondeterministic pattern tree pushdown automata(SIOA).

We can extrapolate our model to that of [13], which views emotional patterns as decision trees, the nodes being various autonomic measurements(skin conductance, heart rate etc.), and emotions are their unique combinations. In our model, the tree is relabeled, and Ekman's [13] measurements are replaced with particular activity on specific combinations of channels. Thus, for a given channel, the root will be the channel prefix, and the suffixes will produce the given emotions.

In conclusion, our results highlight the potential of using emotions as credentials, although verification on a larger test group is needed.

8. REFERENCES

[1] J.J.Vidal, "Toward direct brain-computer communication," *Annu. Rev. Biophys. Bioeng.,2(1)*, pp. 157–180, 1973.

[2] Christoph Guger et al., "State of the art in bci research: Bci award 2010," in *Recent Advances in Brain-Computer Interface Systems*, pp. 193–222. InTech, 2011.

[3] M.Poulos, M.Rangoussi, V.Chrissikopoulos, and A.Evangelou, "Person identification based on parametric processing of the eeg," *Proceedings of the 6th IEEE International Conference on Electronics, Circuits and Systems, Vol.1*, pp. 283 – 286, 1999.

[4] J.Thorpe, P.C. van Oorschot, and A.Somayaji, "Passthoughts: Authenticating with our minds," *Proceedings of the New Security Paradigms Workshop (NSPW)*, pp. 45–56, 2005.

[5] S.Marcel and J. del R.Millán, "Person authentication using brainwaves(eeg) and maximum a posteriori model adaptation," *IEEE Trans. on Pattern Analysis and Machine Intelligence, Special issue on biometrics, 29(4)*, pp. 743 – 752, 2007.

[6] R.Palaniappan, "Two-stage biometric authentication method using thought activity brain waves," *International Journal of Neural Systems,18(1)*, pp. 59–66, 2008.

[7] I.Nakanishi and C.Miyamoto, "On-demand biometric authentication of computer users using brain waves," in *Networked Digital Tech., Communications in Computer and Information Science, Vol. 87, Part 5*, pp. 504–514. Springer-Verlag, Berlin Heidelberg, 2010.

[8] I.Nakanishi, S.Baba, and L.Shigang, "Evaluation of brain waves as biometrics for driver authentication using simplified driving simulator," *International Conference on Biometrics and Kansei Engineering (ICBAKE)*, pp. 71–76, 2011.

[9] A.Zúquete, B.Quintela, and J.P.Silva Cunha, "Biometric authentication with electroencephalograms: Evaluation of its suitability using visual evoked potentials," in *Biomedical Engineering Systems and Technologies, Communications in Computer and Information Science, Vol. 127, Part 4*, pp. 290–306. Springer-Verlag, Berlin Heidelberg, 2011.

[10] J.E.LeDoux and E.A.Phelps, "Emotional networks in the brain," in *Handbook of Emotions (M.Lewis, J.M.Haviland-Jones,L.Feldman Barrett)*, pp. 159–179. The Guilford Press, New York, 2008.

[11] A.Olsson and E.A.Phelps, "Social learning of fear," *Nature Neuroscience, 10(9)*, pp. 1095–1102, 2007.

[12] J.T.Larsen, G.G.Berntson, K.M.Poehlmann, T.A.Ito, and J.T.Cacioppo, "The psychophysiology of emotion," in *Handbook of Emotions (M.Lewis, J.M.Haviland-Jones,L.Feldman Barrett)*, pp. 180–195. The Guilford Press, New York, 2008.

[13] P.Ekman, R.W.Levenson, and W.V.Friesen, "Autonomic nervous system activity distinguishes among emotions," *Science, New Series, 221(4616)*, pp. 1208–1210, 1983.

[14] G.E.Schwartz, P.L.Fair, P.Salt, M.R.Mandel, and G.R.Klerman, "Facial muscle patterning to affective imagery in depressed and nondepressed subjects," *Science, 192(4238)*, pp. 489–491, 1976.

[15] J.T.Cacioppo, R.E.Petty, M.E.Losch, and H.S.Kim, "Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions," *Journal of Personality and Social Psychology, 50(2)*, pp. 260–268, 1986.

[16] P.J.Lang, M.K.Greenwald, M.M.Bradley, and A.O.Hamm, "Looking at pictures: Affective, facial, visceral and behavioral reactions," *Psychophisiology, 30(3)*, pp. 261–273, 1993.

[17] J.T.Larsen, C.J.Norris, and J.T.Cacioppo, "Effects of positive and negative affect on electromyographic activity over *zygomaticus major* and *corrugator supercilii*," *Psychophisiology, 40(5)*, pp. 776–785, 2003.

[18] M.M.Bradley and P.J.Lang, "Affective reactions to acoustic stimuli," *Psychophisiology, 37(2)*, pp. 204–215, 2000.