Organization of Information Support for a Bioengineering System of Emotional Response Research

© N.N. Filatova © N.I. Bodrina © K.V. Sidorov © P.D. Shemaev
Tver State Technical University,
Tver, Russia
nfilatova99@mail.ru vavilovani@mail.ru bmisidorov@mail.ru pshemaev@rambler.ru

Abstract. Nowadays studying a mechanism of human emotional responses attracts much attention. Information about a human personality and condition, which is expressed in a manner of speech, is just as important as his statements. However, computer synthesis and speech recognition systems do not currently use this information. It is possible to numerically assess certain physiological characteristics related to emotions (cardiogram, muscle curves, EEG, speech). In order to assess an emotion objectively, it is necessary to use a complex approach including testee's self-evaluation and recording characteristics of certain body functional systems. There are widely distributed databases containing examples of such characteristics. Previous bases contain recordings of scenic speech with imitated emotions. Modern researchers prefer working with natural emotions caused by irritants – incentives. The paper specifies a multi-channel bioengineering system for studying emotions "EEG-Speech+", which is created in TSTU, and how to work with it. It also describes two series of experiments. The first one includes searching for signs of emotion valence by a speech signal attractor. The second one includes investigating the emotion dynamics by an EEG signal. The authors describe the structure of an extended multimodal emotion base, which stores the results of all experiments. They also consider its open online version.

Keywords: emotional response, speech signal, electroencephalogram, emotion, incentives, stimulated emotion, imitated emotion, database of emotional response examples.

1 Introduction

Studying human emotional response mechanism refers to an interdisciplinary field of knowledge, which attracts more and more attention nowadays. Research on the works of Anokhin P.K., Simonov P.V., Leontyev, Ilyin E.P., Danilova N.N., Izard K., Rusalova M.N., Ukhtomsky A.A., Fress P., Chomskaya E.D., Everly G., Rosenfeld R., Hebb D., etc., allows identifying several basic conclusions, which are not disputed by the scientific community at this stage:

- emotions are inherent not only in a human, but in all intelligent representatives of mammals;
- the emotional response mechanism is innate, some emotions are shown at the earliest stages of life;
- emotions are most often a reaction in response to an external or internal irritants – an incentive;
- the system of human emotional reactions is developing; it is formed in the process of accumulation of his personal experience and formation of cognitive functions.

The mechanism of emotional responses is a further development of reflex systems of the mammalian organism. It solves two sets of tasks: improves the means of adapting an organism to changes in external conditions and creates an apparatus for implementation of communicative processes and maintenance of socially significant contacts. Human emotional responses are related to brain activity and are revealed in functioning peculiarities of certain body functional systems.

In a colloquial human interaction, extralinguistic information about speaker's personality and state, which is expressed in his manner of speech, is as important as the text of a statement. However, computer synthesis and speech recognition systems do not currently use information about emotions, which is a very important factor in communication.

Systems that are capable of generating emotionally colored speech and recognizing human emotional state will be in demand in virtual learning, for studying brain dysfunction, identifying network content and interactive entertainment. In addition, they will be useful for people who have different speech deviations. Modern speech synthesizers do not model emotional speech. The algorithms for recognizing human emotional state are only being developed.

Nowadays there are no objective means of measuring quantitative characteristics of emotions. However, there are opportunities for quantitative assessment of certain physiological characteristics related to them (cardiogram, galvanic skin response, muscle curves, electroencephalograms, and speech patterns).

Considering testee's subjective assessments in an emotion, objective emotion evaluation requires an integrated complex approach including both testee's self-assessment and recording characteristics of certain functional systems.

Successful development of emotion recognition modules by various signals recorded in a person, who is

Proceedings of the XX International Conference "Data Analytics and Management in Data Intensive Domains" (DAMDID/RCDL'2018), Moscow, Russia, October 9-12, 2018

experiencing an emotion, is possible when there is a big volume of such signals. Geographical and ethnic studies show that an emotional expression is formed and changes with the course of the history of linguistics. Consequently, the sources of emotional responses should be carriers of an appropriate language.

Initially, the bases with the records of emotionally colored speech have become widespread. They are gradually expanded. Other biomedical signals (cardiogram, galvanic skin response, heart rate, muscle curves, electroencephalograms, etc.) taken at the moment when a testee demonstrates an emotional response are added to speech samples.

2 Modern bases of emotional response examples

Early studies of emotional responses are based on the records of scenic speech with imitated emotions [1, 3, 6, 12, 18, and 19]. Usually, exterior listeners recognize such emotions correctly. The analysis of acoustic characteristics is based on the records of identic texts. Nevertheless, it is not known how well an actor is able to represent all speech characteristics that ordinary people show when they experience similar emotions. Imitated emotions are reproduced on assignment and do not need incentives.

In studies, the difference between experienced and expressed emotions is minimal. In everyday social interactions, it is often appropriate to suppress emotions. Moreover, it is preferable to express emotions that people do not really experience at the moment. A computer synthesizer of an emotional speech, which is created based on studying only simulated emotions, might deform user intentions.

Therefore, the majority of modern researchers work with stimulated emotions (Table 1) instead of using emotion imitations. Such emotions are natural and are triggered by specially prepared emotiogenic incentives. Information support of the bases includes these incentives or their descriptions. There are some papers that pay attention to classification, evaluation or marking of incentives [7, 14].

The need to confirm the desired emotion in a testee leads to expanding a list of types of biomedical signals stored in databases [7, 13, and 17]. In experiments, testees are usually instructed not to restrain their emotions, but in real social interactions personal feelings are not expressed so openly. For this reason, some researchers use other people as sources of emotionogenic incentives [13, 17]. In an experiment, a testee together with an assistant must solve some problem. Interaction, communication with an assistant is an incentive.

3 Bioengineering system “EEG-Speech+”

A specialized bioengineering system “EEG-Speech+” has been created and developing at the Department of Automation of Technological Processes of the Tver State Technical University [16]. The system has a multichannel scheme for recording testee's responses to external emotiogenic incentives. Simultaneous recording of several types of biomedical signals allows confirming changes in testee’s emotions according to the scenario of the experiment.

Table 1. The bases of examples of stimulated emotional responses

Name, year, language	Incentives	Data
DEAP data, 2005, Eng. [7]	1-minute video with sound (more than 120)	• EEG; • physiologic measuring; • face video; • assessments
Film Stim, 2010, Eng., French. [14]	1–7-minute video with sound (more than 70)	• assessments
Cognitive Human Computer Interaction Lab, 2011, Eng. [8]	recordings of classical music	• EEG
MAHNOB-HCI, 2012, Eng. [17]	video with sound (more than 30) and images (more than 20)	• EEG; • physiologic measuring; • face and body video; • speech; • position of the pupil; • assessments
Recola Database, 2013, French. [13]	interaction	• EEG; • ECG; • speech; • face video; • assessments

Fig. 1 shows the composition and interaction scheme of the components of the bioengineering system “EEG-Speech+”. By now, the system has been expanded to five channels for recording emotional response (Ch1 - Ch5 in fig. 1): video, sound, electroencephalogram (EEG), muscle curve (EMG) and information (testee's report).

A personal computer B serves to present visual or acoustic incentives and contains a base of incentives, as well as all software necessary for their reproduction. A special device [4, p. 78] delivers olfactory incentives to a testee. The main workstation A controls the process of presenting olfactory incentives.

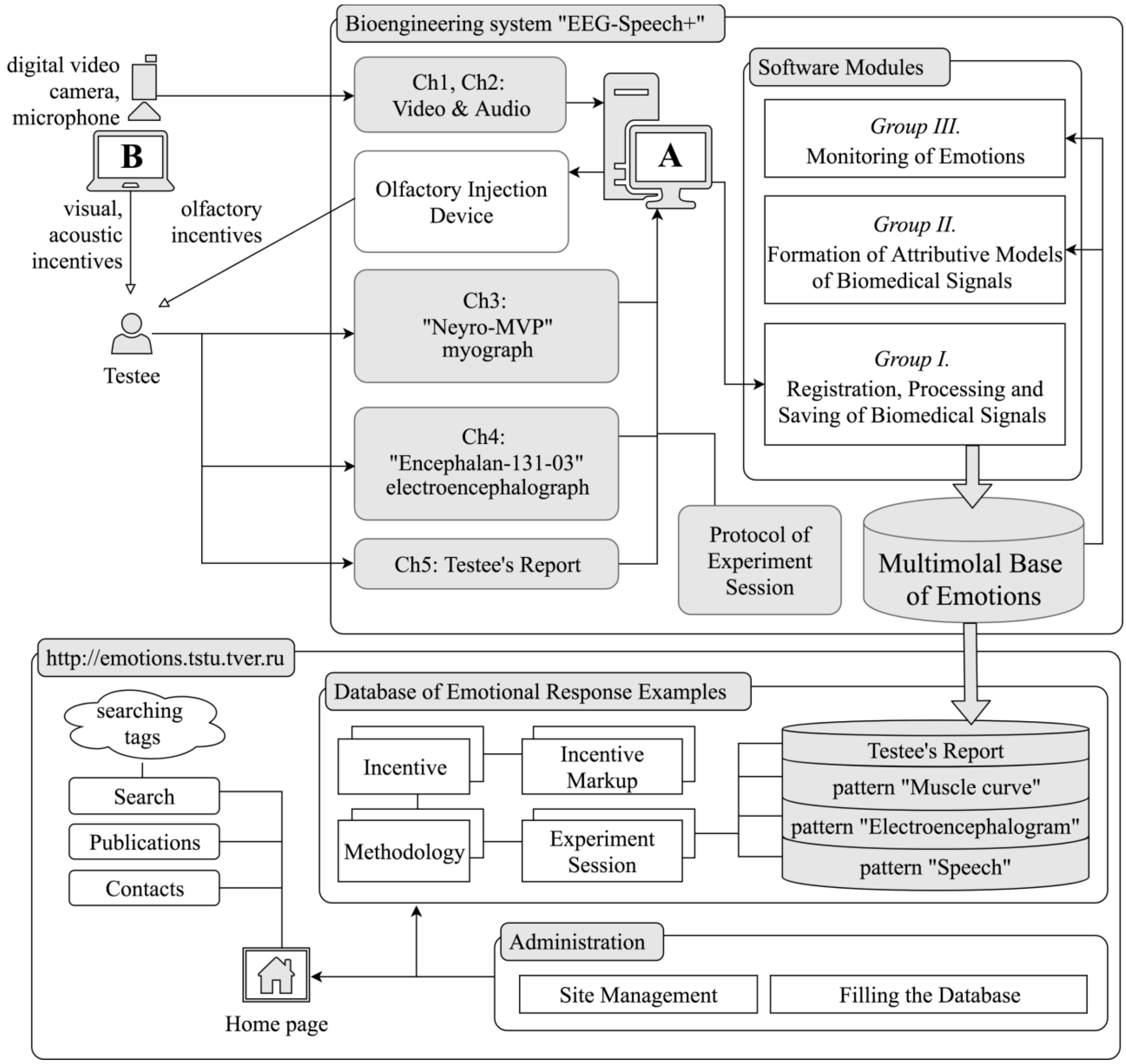

Fig. 1. The bioengineering system “EEG-Speech+” and a database of emotional response examples

Each experiment session has a specially prepared scenario (Table 2). The workstation A receives biomedical signals from all channels used in the current experiment. The signals are stored in the appropriate database of testees. The received signals are processed and cleared of interference and artifacts. The bioengineering system software includes three groups of modules (Modules of groups I, II and III in Fig. 1) [4]:

- registration, processing and saving biomedical signals;
- formation of attribute models of biomedical signals;
- monitoring of emotions.

The software modules are implemented in MATLAB in C# language. The bioengineering system software is installed on the main workstation A, but can be used on any personal computer, so that processing of experimental results can be remote and in a distributed mode.

4 Experiments and results

There are a lot of experiments with the bioengineering system “EEG-Speech+”. The studies include several directions:

- search for signs of biomedical signals related to an emotional response;
- determining the direction of emotion development (growth, fading).

4.1 Signs of emotion valence in a speech signal

Most biomedical signals are not stationary and irregular, i.e. a probability distribution of signal parameters is random. Therefore, the methods of nonlinear dynamics become relevant for their processing.

In particular, in order to identify individual characteristics of emotions based on the initial biomedical signal, there is a reconstruction of an attractor, which becomes an object of research later.

Table 2. An example of experiment scenario

Time	Scenario activity	Expected emotional response
• placing electrodes for long-term recording of biomedical signals; • tuning the channels selected for the experiment; • start recording biomedical signals.		
3 min.	Background demonstration	neutral
10 min.	Incentive demonstration "+"	positive
6 min.	Background demonstration	fading of positive, transition to neutral
1 min.	A short survey of a testee: confirmation of the expected emotional response	
3 min.	Background demonstration	neutral
10 min.	Incentive demonstration "–"	negative
6 min.	Background demonstration	fading of negative, transition to neutral
1 min.	A short survey of a testee	
• stop recording biomedical signals; • cutting off the channels; • detaching electrodes.		
10-20 min.	Detailed survey of a testee: playback of incentives and their marking	

So, a number of authors use an index of a restored attractor correlation dimension [9, 11] to recognize a sign of emotions.

The paper [11] used this feature when comparing EEG of the signal recorded for five testee's states: grief, joy, time counting, a background with closed eyes and a background with open eyes. The author notes a significant increase in the correlation dimension under conditions of emotional experience comparing with a neutral state.

One of other signs of emotion recognition is the Lyapunov exponent. The paper [9] uses the Lyapunov exponent to assess testee's emotional state by certain phonemes in a speech signal. The author notes a significant difference between the state of "calmness" and when there are negative emotions (anger, disgust).

Studying of attractors reconstructed from Russian speech patterns showed that when a testee experiences positive emotions, the attractor form expands, in the case of negative one it gets narrow. Consequently, the number of points in the center changes. Thus, we can assume that a correlate of a sign of emotions can be the point density indicator of attractor trajectories.

The hypothesis was checked through the research that involved students and postgraduates of the Tver State Technical University at the age of 18–25. The testees were offered to watch videos of up to 3 minutes, which can be conditionally divided into three groups:
1. a positive incentive (k+);
2. a negative incentive (k-);
3. a neutral incentive (N).

After each video the participants had to say a challenge phrase.

As a measure of the attractor density in the center, we used the indicator [5]:

$$\rho_j = k_j / S_j, \quad k_j = h_j + r_j/2, \tag{1}$$

Which is the ratio of the number of attractor points related to one of the cells of an orthogonal grid covering an attractor projection, (k_j) to the cell area (S_j). h_j is the number of points inside each j-th cell. The number of points (r_j) on the boundary of the j-th and j+1-th cells is divided equally between boundary cells.

The autocorrelation function determines the optimal value of the time delay τ, which varies depending on a testee.

Attractor properties were analyzed using the first projection of the attractor, or rather the area of the greatest cluster of points localized near the origin of coordinates (Fig. 2).

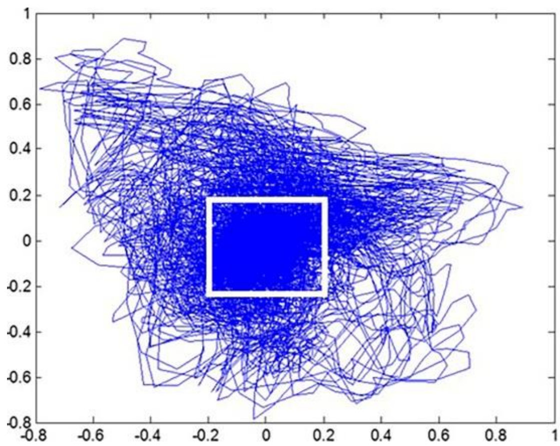

Fig. 2. A projection of an attractor, which was reconstructed from a speech signal, with a selected area of the greatest cluster of points

The duration of each received speech record for analysis was 20,000 readings (≈1 seconds). The records went through auto-normalization with the removal of artifacts.

It has been experimentally established that the presence of a noise component does not affect the classifying ability of the parameter ρ_j [15].

In total, we analyzed 74 speech signal fragments from 8 testees (3 incentives for each sign of emotions).

Figure 3a shows a diagram of changes in the averaged values of ρ_j attractor density in the center.

It should be noted that a negative incentive causes an increase of the ρ_j index in relation to a neutral state (from 2 to 55%) almost in all testees. On the contrary, with a positive video incentive, this parameter tends to decrease (from 5 to 38%). The obtained result confirms the hypothesis about the interrelation between an emotional impact sign and an attractor density.

Similar experiments were performed with samples of voice recordings from the international database Emo-DB [1], which contains audio recordings of emotionally colored speech in German from 10 different speakers. We analyzed signals with a negative (disgust), positive (happiness) and neutral incentive. Figure 3b shows the results of ρ_j averaged values for several testees on the same phrase.

Unlike the samples of Russian speech, German speech is characterized by an increase (on average by 20%) of the number of points in the attractor center affected by positive incentives in relation to a neutral state. Negative incentives also cause an increase in ρ_j density (on average by 10%).

Conclusion. It is established that the sign of emotions significantly affects the number of points of the reconstructed attractor in the center. This is true both for Russian speech samples and for studying phrases in German. The density parameter ρ_j available from experiments can be used to construct a classifier.

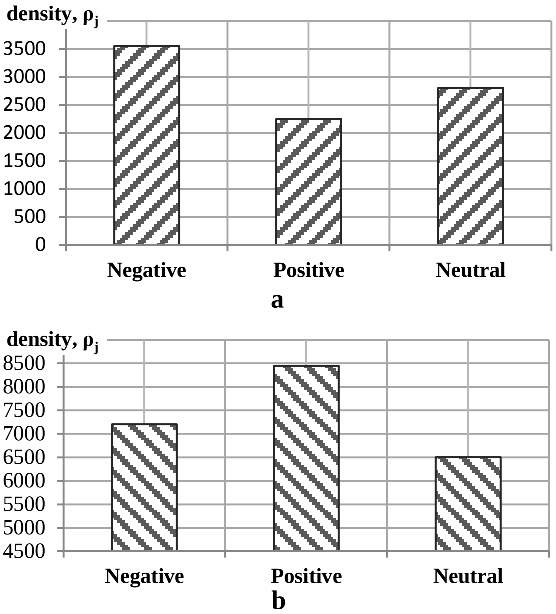

Fig. 3. Dependence graphs of ρ_j attractor density at the center on the sign of emotions for the samples of Russian (a) and German speech (b)

4.2 Research on an emotion dynamics based on the analysis of EEG signals

A series of experiments included using 2–4-minute video clips with sound as emotiogenic incentives. The testees were TSTU students and postgraduates aged 18–25.

Each video incentive was pre-marked by a testee according to a sign of an emotional response.

In a series of experiments, a testee was consistently presented with several negative incentives (-E), and then several positive ones (+E). Before changing an incentive sign, a testee was presented with neutral frames with a green background. Each experiment lasted no less than 20 and not more than 25 minutes.

While watching incentives, testee's EEG was continuously recorded. His speech was recorded after each incentive. The processing of the experimental results had two stages.

The first stage included creating fragments of biomedical signals free from noise (for speech signals) and artifacts (for EEG signals).

Perception of incentives of the same sign (-E or +E) resulted in sequences of EEG fragments. Their characteristics contain information on changes in testee's emotional responses.

The second stage of processing the experimental results included identification and quantitative evaluation of these latent characteristics. The bioengineering system "EEG-Speech+" provides calculation of signal power spectral analysis (EEG or speech signals), as well as attractor reconstruction based on them.

Figure 4 shows a projection of an attractor constructed from an EEG fragment (lead C4-A2), which is correlated with the terminal part of the first negative incentive.

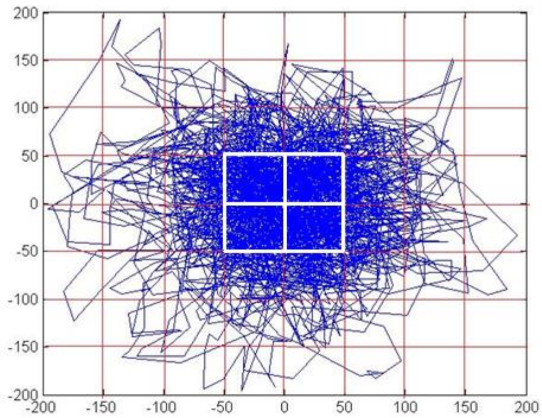

Fig. 4. A projection of an attractor constructed from an EEG signal (lead C4-A2)

The experiments showed that leads F7-A1 and F8-A2 had the strongest changes in power spectra when a testee was watching positive and negative incentives.

However, the reproducibility of this result was not high. Therefore, each lead had reconstructed attractors with their properties depending on the sign of testee's emotional response, as shown in previous studies [10].

To characterize attractors, we used the features proposed in [5]: an attractor trajectory density near its center ρ_j (1) and a number of empty cells in a grid covering the attractor projection k_0 (Fig. 4). Grid dimensions are fixed: 196 cells, a step is 50 readings.

Observation of changes in the signs of ρ_j and k_0 showed that in most experiments there is their correlation with a sign of an emotional response.

When a testee experiences positive emotions, k_0 decreases. It increases during experiencing negative emotions.

Conclusion. Preliminary results show the possibility of using an attractor density as a sign of EEG signals, illustrating the development of an emotional state at a certain time interval. The observation interval does not have imposed limitations.

5 A multimodal emotion database and a public emotion database

The experimental results are a basis for a multimodal emotion database, which contains examples of signals with a bright and slightly expressed emotional color. At the first stage, the database has speech patterns and associated EEG patterns [4]. The "entity-relation" model of the extended multimodal emotion database is supplemented by descriptions of incentives and new channels (Fig. 5).

The examples of emotional responses in the database are not labeled with the names of emotions ("anger", "fear", "joy", etc.). We use only natural emotional responses, so we determine the valence of an emotion (positive, negative or neutral) and its level (strong, weak, etc.).

The multimodal emotion database includes:

- 266 patterns of a challenge phrase lasting 2–6 seconds, pronounced by different speakers who are not actors in response to a presentation of a video incentive;
- 2660 vowel phonemes lasting 0.025–0.25 seconds segmented from challenge phrases;
- 240 EEG patterns cleared from artifacts lasting for 12 seconds.

Since 2016, there is a public database containing examples of emotional responses [2]. The database is developed in PHP language with MySQL DBMS. There is a website (http://emotions.tstu.tver.ru) to access the database using cms joomla. For now, several series of experiments are available to the public:

1. Recordings of speech signals (in .wav) of 17 testees. There are up to 10 samples for certain testees. Emotiogenic incentives are specially prepared videos with sound, which cause positive, negative and neutral emotional states.
2. Recordings of speech signals (in .wav) and EEG signals (in .txt) of 9 testees. There are several recording sessions for certain testees. Registration of speech and EEG was parallel. Emotiogenic incentives were also videos with sound. Parallel recording of speech signals and EEG allowed objectively fixing the presence of positive, negative and neutral responses of testees to incentives.

Acknowledgments. The reported study has been funded by RFBR according to the research projects: № 17-01-00742, № 18-37-00225.

References

[1] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W.F., Weiss, B.A.: Database of German Emotional Speech. In: 9th European Conference on Speech Communication and Technology (Interspeech) Proceedings, pp. 1517-1520. ISCA. Lisbon, Portugal (2005).

[2] Database of Emotional Response Examples, http://emotions.tstu.tver.ru, last accessed 2018/05/15.

[3] Engberg, I.S., Hansen, A.V.: Documentation of the Danish Emotional Speech Database (DES). Aalborg University, Denmark (1996).

[4] Filatova, N.N., Sidorov, K.V.: Computer Models of Emotions: Construction and Methods of Research. RITs TSTU, Tver' (2017) (in Russ., Komp'yuternye modeli emotsiy: postroenie i metody issledovaniya: monografiya)

[5] Filatova, N.N., Sidorov, K.V., Terekhin, S.A.: A Software Package for Interpretation of Nonverbal Information by Analyzing Speech Patterns or Electroencephalogram. Software & Systems 111(3), 22–27 (2015). doi: 10.15827/0236-235X.111.022-027 (in Russ., Programmnye Produkty i Sistemy)

[6] Haq, S., Jackson, P.J.B., Edge, J.D.: Audio-Visual Feature Selection and Reduction for Emotion Classification. In: International Conference on Auditory-Visual Speech Processing (AVSP) Proceedings, pp. 185-190. ISCA. Australia (2008).

[7] Koelstra, S., Muehl, C., Soleymani, M., Lee J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I.: DEAP: a Database for Emotion Analysis Using Physiological Signals. IEEE Transaction on Affective Computing 3(1), 18-31 (2012). doi: 10.1109/T-AFFC.2011.15

[8] Liu, Y., Sourina, O., Nguyen, M.K.: Real-Time EEG-Based Human Emotion Recognition and Visualization. In: Proceedings of the 2010 International Conference on Cyberworlds, pp. 262-269. IEEE Computer Society. Singapore (2010). doi: 10.1109/CW.2010.37

[9] Mekler, A.A.: The program complex for the analysis of electroencephalograms by methods of the dynamic chaos theory: Ph.D. Thesis. IHB RAS, St. Petersburg.(2006) (in Russ., Programmnyy kompleks dlya analiza elektroentsefalogramm metodami teorii dinamicheskogo khaosa)

[10] Mekler. A.A., Gorbunov, I.A.: Relation between the pattern of experienced emotions and characteristics of the EEG complexity. In: The Fifth International Conference on Cognitive Science Proceedings, pp. 528–529. Kaliningrad, Russia. (2012) (in Russ., Pyataya mezhdunarodnaya konferentsiya po kognitivnoy nauke)

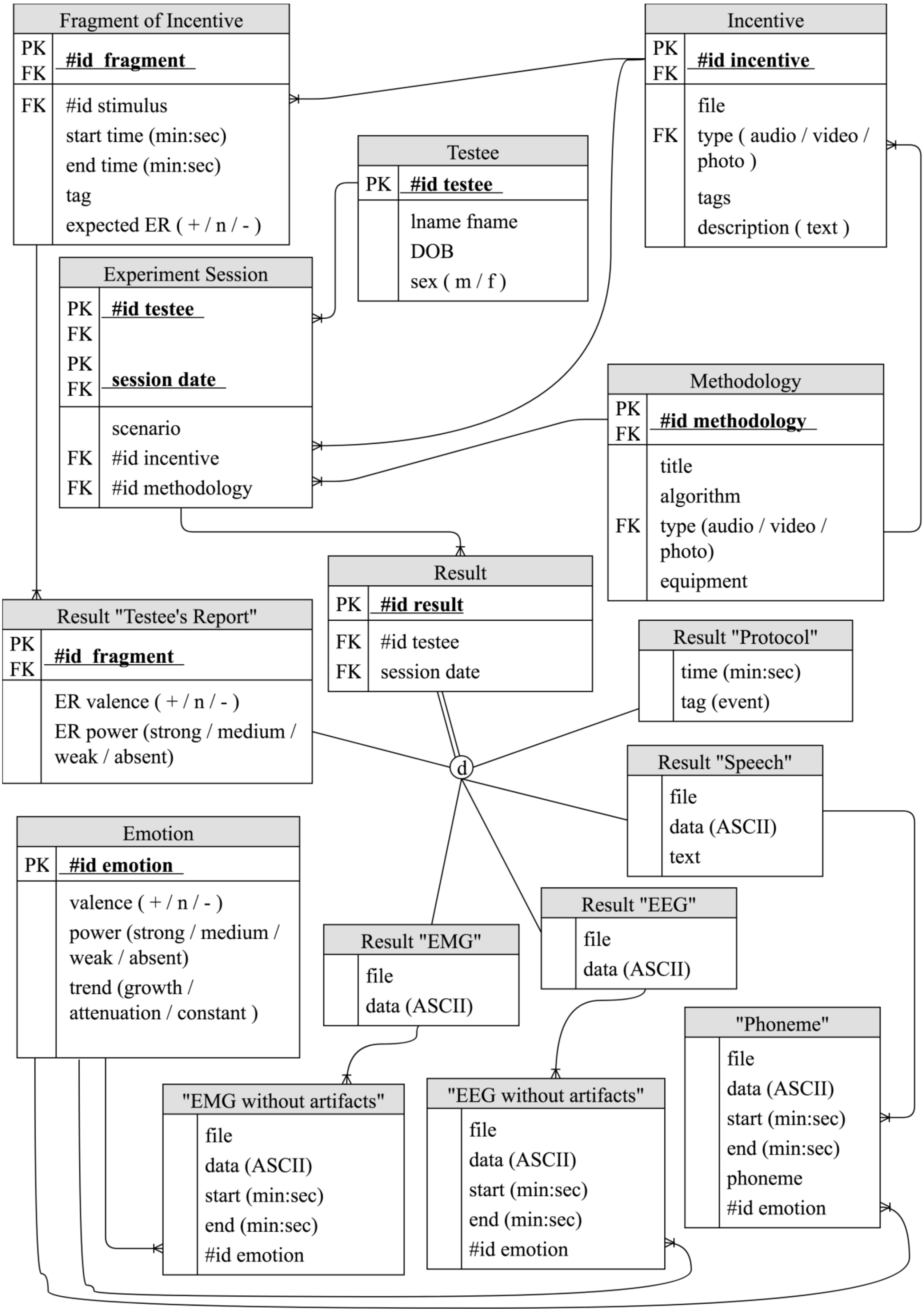

Fig. 5. An ER-model of an expanded multimodal emotion database

[11] Perervenko, U.C.: Investigation of invariants of the nonlinear dynamics of speech and principles of building an audio analysis system of the psychophysiological state: Ph.D. Thesis. TTI UFU, Taganrog (2009) (in Russ., Issledovaniye invariantov nelineynoy dinamiki rechi i printsipy postroyeniya sistemy audioanaliza psikhofiziologicheskogo sostoyaniya)

[12] RAVDESS Speech/Song Database, https://smartlaboratory.org/ravdess/, last accessed 2018/05/08.

[13] Ringeval, F., Sonderegger, A., Sauer, J., Lalanne, D.: Introducing the RECOLA Multimodal Corpus of Remote Collaborative and Affective Interactions. In: Proceedings of 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, pp. 1-8. IEEE. Shanghai (2013). doi: 10.1109/FG.2013.6553805

[14] Shaefer, A., Nils, F., Sanchez, X., Philippot, P.: Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researches. Cognition and Emotion 24(7), 1153-1172 (2010). doi: 10.1080/02699930903274322

[15] Shemaev, P.D., Filatova, N.N.: Investigation of the influence of noise in the voice signal on the recognition of the characteristics of the emotion’s valence. In: Proceedings of conference «BIOMEDSYSTEMS-2015», pp. 90–93. RSREU. Ryazan, Russia (2015) (in Russ., Vserossiyskaya konferentsiya "BIOMEDSISTEMY-2015")

[16] Sidorov, K.V.: Biotechnical System of Human Emotions Monitoring by means of Speech Signals and Electroencephalogram: Ph.D. Thesis. TSTU, Tver' (2015) (in Russ., Biotekhnicheskaya sistema monitoring emotsiy cheloveka po rechevym signalam i elektroentsefalogrammam)

[17] Soleymani, M., Lichtenauer, J., Pun, T., Pantic, M.: A multimodal database for affect recognition and implicit tagging. IEEE Transactions on Affective Computing 3(1), 42-55 (2012). doi: 10.1109/T-AFFG.2011.25

[18] Tillmann, H.G., Draxler, Chr., Kotten, K., Schiel, F.: The Phonetic Goals of the new Bavarian Archive for Speech Signals. In: Elenius, K., Peter Branderud, P. (eds.) 13th International Congress of Phonetic Sciences Proceedings, vol.4, pp. 550-553. Congress organizers at KTH and Stockholm University. Stockholm, Sweden (1995).

[19] Wang, Y., Guan, L.: Recognizing human emotional state from audiovisual signals. IEEE Transactions on Multimedia 10(5), 936–946 (2008). doi: 10.1109/TMM.2008.927665