THE RUSSIAN EMOTIONAL CORPUS: COMMUNICATION IN NATURAL EMOTIONAL SITUATIONS

Kotov A. (kotov@harpia.ru)
National Research Centre "Kurchatov Institute", Moscow, Russia

Budyanskaya E. (davemonty2@gmail.com)
Research Computing Center, M. V. Lomonosov Moscow State University, Moscow, Russia

The Russian Emotional Corpus (REC) includes annotated video recordings of natural communication in tense emotional situations: oral university exams and talks between clients and officers at a municipal office regarding utility bills. Annotation of the corpus describes speech (text, syntagmatic structure, speech acts, face threatening acts and the cases of irony), facial expression, gaze direction and hand gestures. In the recorded emotional situations informants show diverse emotional cues and numerous rational and emotional communicative strategies. The annotation allows us: (a) to search for a specific cue (like smiles or squints) and describe its usage and functions in natural communication, or (b) to search for specific patterns — like behavior markers of hesitation or facial cues at the end of utterances with face threatening acts. These observations help us to animate emotional computer agents, which receive semantic trees at their input, simulate rich emotional dynamics and produce semantic trees, ready-made phrases and gestures for the output.

Key words: multimodal corpus, emotional expression, emotional communication, emotional agents

1. Introduction

Linguistic corpora recently became an important base for theoretical studies as well as for training and verification of applied linguistic tools. Widespread text corpora are combined with audio corpora and multimodal corpora, allowing the study of speech, phonetics, facial expression, gestures and other aspects of real face to face communication. We collect and annotate the Russian Emotional Corpus (REC) with video records of natural emotional communication. The corpus is mainly aimed at the creation of emotional computer agents, which recognize emotional cues in communication and simulate emotional behavior in a dialogue.

1.1. Usage and types of emotional corpora

The development of entertainment technologies, computer agents (like game characters, interface assistants) and mobile robots as well as attention to emotional aspects in all types of communication stimulate the development of emotional corpora. Emotional multimodal corpora are represented by (a) the observed public data: records from mass media [Abrilian, Devillers et al., 2005], (b) performed data: where actors have to perform given emotions [Bänziger, Scherer, 2007], and (c) experimental data: where the emotional arousal is induced in experimental situations [Zara, Maffiolo et al., 2007].

At the same time, this data is still different from real face to face interaction. The behavior of informants is restricted (induced) by the situation (environment) of publicity, acting or experiment. As a result this data may miss a number of natural emotional phenomena: conflicts between internal motivation, goals and suppressed emotions — on one side, and strategies of politeness, necessity to simulate "polite" or "desirable" emotions (like etiquette smiles) — on the other side.

This requires the collection of real emotional data, as for example, made in [Scherer, Ceschi, 1997]. This is a difficult task for several reasons: (a) during communication subjects should remain in front of a fixed camera (unlike in spoken corpora), (b) camera interferes with communication, (c) we cannot manipulate the situation, for example asking a salesman to frustrate the buyer (unlike in experimental cases), (d) severe legal limitations apply to all public cases of video records, and (e) ethical and legal considerations force us to limit access to the corpus.

At the same time, we may get ecologically valid data (subject to possible influence of (b)) containing diverse communicative strategies and complicated emotional patterns, suitable for studies and modeling.

1.2. Emotional computer agents

Natural emotional data is important for dialogue support systems, which should "understand" clients in natural situations, and for the development of software agents with rich emotional architectures [Rehm, André, 2008].

For example, Max computer agent uses different incoming phrases to change its emotional state and to balance between "good" and "bad" mood, producing different expressive cues and phrases for each local state [Becker, Kopp, Wachsmuth, 2004]. Greta computer agent uses two different (conflicting) emotional arousals (internal emotions and masking superficial emotion) to construct compound patterns of facial expression — "blended emotion" [Ochs, Niewiadomski, et al., 2005], as observed in mass media corpus [Abrilian, Devillers, et al., 2005; Maglogiannis, Karpouzis, et al. 2006].

In our studies we develop emotional computer agents, which receive semantic trees at their input, simulate emotional dynamics and produce semantic trees, ready-made phrases and gestures in BML format for the output. The agents are designed to simulate complicated emotional dynamics in communication, in particular:

(i) emotional oscillation: an incoming event or phrase may simultaneously activate numerous responses (like anger, confusion, desire for reconciliation, etiquette regulations, rational intention to solve the situation etc.), which compete for the output, forcing the agent to oscillate between several short communicative positions or emotions (*microstates*) in its behavior [Kotov, 2007];

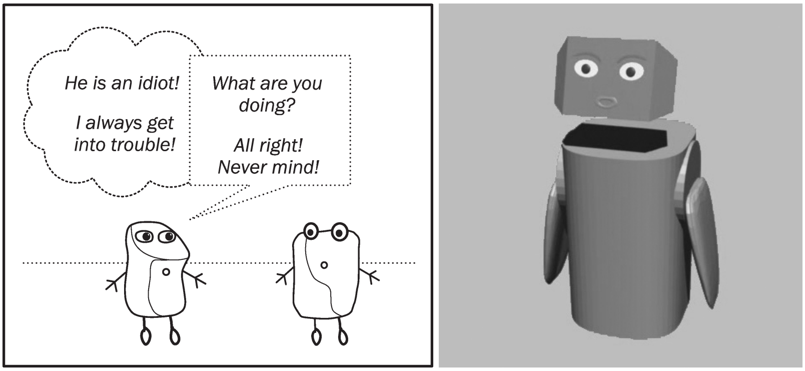

Fig. 1. 2D and 3D emotional computer agents

(ii) top-down emotional effects (as discussed in [Clore, Ortony, 2000]): in a negative state the agent treats incoming stimuli as negative, changing the semantic representation of an incoming phrase or event — if we 'touch' the agent, it believes, we 'beat him' (if in a negative state), or that we cherish him (if in a positive state) [Kotov, 2007];

(iii) hypocrisy and sarcasm: if we 'beat' the agent, forcing its negative arousal, he may suppress the immediate expression and automatically select the best positive reaction (like *It's good you have paid attention to me!*) to use it in a sarcastic way [Kotov, 2009].

(iv) switches in communication: the agent may address different parties around — the main counteragent (the person, who has just committed an action or was affected by the agent's action) and observers with different levels of the agent's confidence (friends or strangers); the agent may also convey its "thoughts" (too aggressive to too sincere responses) to a special callout.

As the architecture is based on a number of conflicting tendencies, it should be well balanced in order to produce believable compound patterns. To observe and study naturally appearing complicated patterns we require a video corpus of real behavior in tense emotional situations.

2. Principles of collection

We collect records of emotional situations, where the investigated party (student, client) has very strong motivation to communicate and achieve his goal. Obstacles, that appear, may force internal conflicts, numerous expressive cues and usage of different influence strategies.

During the communication one of the parties (examiner, client) gives the instruction, and the other party (student, officer) should perform in order to satisfy the opponent. From this point of view we investigate two symmetrical situations, similar to future interactions with mobile robots and dialogue systems, which should fulfill human requests or otherwise satisfy the opponent through the emotional interaction.

The legislation obliged us to inform the recorded party and respect his privacy. We have turned off the camera in the cases of refusal (3 cases during the collection of the student corpus). We also gave an obligation not to publish or disclose the records, limiting their usage only to scientific purposes.

2.1. Oral university exams

Oral university exams have been widely replaced by written tests to eliminate the contribution of personal influence and impressions to the final score. This "undesirable" emotional interaction was exactly the object of our interest in the collected records. This corpus section contains 295 records of university exams and tests (length from 26 sec to 60 min, average 6 min, total length 29,5 hours; 5 humanitarian university courses, 2 universities). Each record includes full communication with one student: video of the student and audio record of the conversation. 194 records are annotated to the present moment.

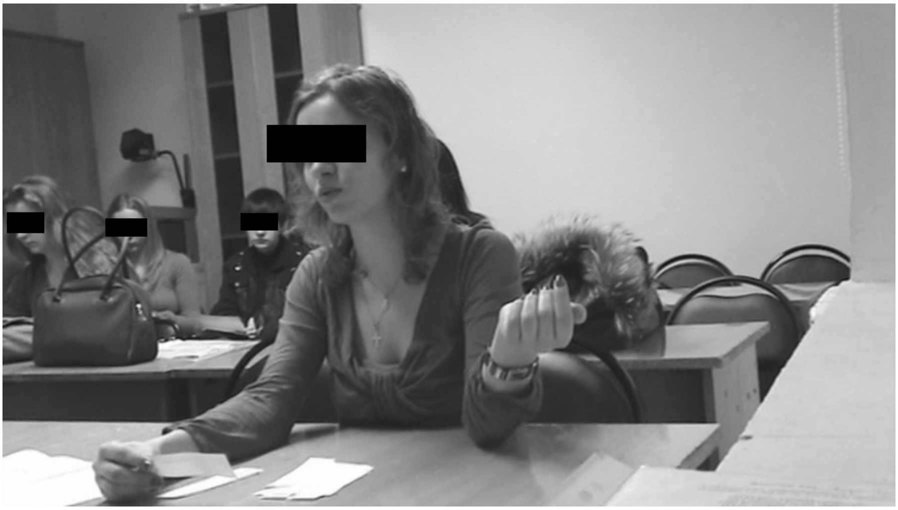

Fig. 2. Exam situation sample frame

2.2. "One window" municipal service

"One window" municipal service interacts with clients on all the questions regarding provided community facilities (water, central heating), issued bills, residential rental, discounts etc. REC contains test series, recorded in June 2009 (4 days) and the main series, recorded in the middle of October 2009 (15 days, 2,5 weeks at 6 working days). Corpus covers the full working time of the window (average 8 hours per day for the main series). The total recorded time (146 hours, including idle time) was divided into 510 records, where each record includes full communication with one subject (length from 5 sec to 30 min, average 3,8 min, total length 32 hours). The records include video of a client and full audio of the conversation. For this corpus section only microstates are annotated to the present moment.

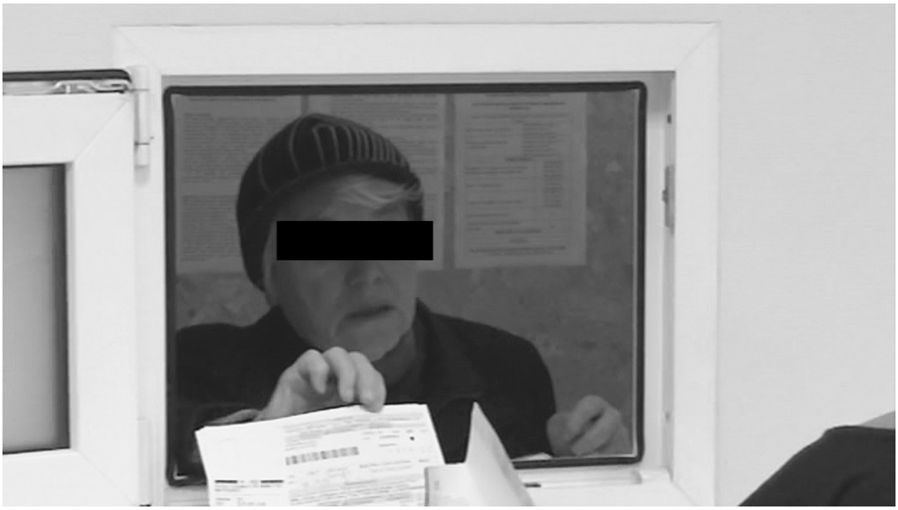
[redacted]

Fig. 3. "One window" sample frame

Many factors show, that the influence of camera was greatly reduced: in the exam corpus some students tried to cheat when an examiner left the room, in "one window" corpus only a few subjects noted the camera (not only the announcement about video recording).

3. Principles of annotation

The main goal for the annotation was to annotate expressive cues, specific for emotional interaction and distinguishing current emotional examples from imaginary steady and non-emotional behavior. We had to provide (a) reliable and simple superficial annotation, suitable for different search and pattern extraction tasks, and (b) specific functional annotation, oriented to our research purposes.

The corpus is annotated with the help of ELAN annotation software[1]. The annotation is divided into "tiers" and describes speech, facial expression and hand gestures, as shown in Table 1. We also annotate rapid changes in expressive or communicative strategies (changes in microstates) to rely on these examples during modeling of the agent's behavior. Head movements and body postures are not yet annotated.

[1] ELAN annotation software is developed at Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. Software URL: http://www.lat-mpi.eu/tools/elan/

Annotation scheme is always a compromise between accuracy (e.g. annotation of action units or allophones) and simplicity (e.g. annotation of cues and words/ phrases). We follow the second strategy: we try to provide a simple index to a wide number of cases and annotate cues, which may be significant for emotional interaction, taking into account, that minor actions, constituents of each cue, can be studied later. To reduce the redundancy we join tiers for (a) eyes, eyebrows, nose and gaze and (b) lips, jaw and tongue — we consider, that mainly one cue (from our list) may be expressed in each of these areas and other units usually follow the expression (e.g. when subjects widely open their eyes, they usually move eyebrows up); although in this annotation we lose some cues, where eyes and eyebrows act separately (e.g. closing eyes and moving eyebrows up).

The records include diverse, indefinite and incomplete movements and gestures, so annotation of hand gestures has four independent parameters: active_organ, passive_organ, mode and trajectory (as in Table 1). This allows the description of non-standard, incomplete and partial gestures.

Annotation is made by one annotator, checked by a supervisor and stored on a version control server.

Table 1. Annotation scheme of the Russian Emotional Corpus

Group	Tier	Comment	Annotation type
Text	text1	Text of the main participant	Free text annotation, divided by speech segments
- " -	text_stru	Syntagmatic structure	Describes syntagmatic position of the current speech segment
- " -	text_comgoal	Illocutionary force	Indicates communication goal or illocutionary force
- " -	text_fta	Face threatening acts	Indicates possible face threatening acts and politeness strategies
- " -	text_irony	Irony	Indicates the cases of irony or explicit role-playing
- " -	text2	Text of the main addressee	(the above annotation tiers are duplicated for this line)
- " -	text3	Text of all other participants	- " -
Eyes	eyes	Eye, eyebrows and nose movements, specific gazes	Fixed list: *looks aside, looks up* (gazes are considered as expressive cues, continuous gaze direction is not annotated)*, opens* [eyes] *wide, squints, closes, winks, rises eyebrows, frowns, wrinkles nose, blinks rapidly, other*

Group	Tier	Comment	Annotation type
Mouth	mouth	Mouth movements	Fixed list: *licks lips, smiles, laughs, purses lips, chews* (moves lower jaw)*, compresses lips, sucks in lips, bites lip, grins, moves lips* (as if silently speaking)*, coughs, clicks, spits, inhales, exhales, opens mouth, other*
Hands	active_organ	Active organ for gestures	Fixed list: *single finger, fingers, palm, fist* (or back of the hand), *arm, adapter, other*
- " -	passive_organ	Passive organ or object for gestures	Fixed list: *none, hair, elbow, forehead, eyes-eyebrows, nose, cheek, mouth-lips, ear, chin, fingers, palm-hand, fist* (or back of the hand), *arm, own clothes or body, an object, human, environment* (immobile object, like table or wall), *other*
- " -	mode	Mode of operation	Fixed list: *manipulates, points, counts, iconic* (iconic gestures, manipulation with imagined objects), *touches, knocks, rubs, scratches, waves, demonstrates* (keeping an object with hands), *closes* (e. g. mouth with a hand), *supports, crosses* (fingers, palms, arms), *other*
- " -	trajectory	Trajectory of operation	Fixed list: *fixed* (cue has no trajectory), *to an object, to the addressee, away* (e. g. waving with hand away from the body), *towards the body, up* (e. g. scratches the nose upwards with a finger), *down, in discourse space* (manipulations in an imaginable space*), in the environment* (e. g. arranges objects on the table), *round, excursion* (preparation), *recursion* (release), *reverse* (movement to and from, e. g. when hesitating)*, standby, fragment* (incomplete gesture), *other* (e. g. fiddles with clothes)
Micro-states	microstates	Microstates and sequences of microstates	Keywords for microstates — units and markers for the developed computer agent architecture
- " -	m_phases	Phases of microstates	Each expressive phase is marked as a separate annotation

4. Patterns of emotional interaction

Emotional communication in REC has certain distinctions: (a) informants show numerous and compound reactions on incoming stimuli, (b) informants try different strategies (both rational arguments and emotional influence) in order to get to their goal in communication, and (c) in both cases informants may show substitutive expressive means — substituted gestures and simulated emotions.

4.1. Multiple reactions

Subjects may balance between different speech strategies, mixing cues of their own arousal with required performance strategies. In the following example (Table 2) a student (female) gives a definition for "zeugma" and hesitates during the answer phrase, showing different speech and facial hesitation cues (record 20081226-zhurn-a1, 04:20).

The following dialogue in "one window" service shows switches in communication, when the client announces his "thoughts" in the second phrase, and changes in communication strategies, when the client after an emotional appraisal moves on to discuss a rational solution:

> < The client (female) wanted to get a statement of her residential account, but forgot her passport >
> **Officer:** *Only with passport.*
> **Client:** *So only with passport, yeah?* < looks above the window > *Well, I'm not a lucky one.* < looks in the window > *But you're open every day, right?*
> (clip 20091022-a03)

These examples constitute a valuable material on how to construct and balance different reactions (calling expressive cues or speech strategies) for an emotional computer agent.

4.2. Multiple strategies

In addition to numerous strategies of politeness, described in [Brown, Levinson, 1987], people show diverse emotional strategies to influence the addressee. In the exam section students (female) may start to smile and flirt with the examiner (male) after a wrong answer in order to smooth over the bad impression and achieve benevolence through the emotional interaction. Students and clients may simulate and exaggerate cues for pain, tiredness, frustration, nervousness in order to show, that the situation is negative and should be solved by the opponent: examiner or municipal officer [Kotov, 2011]. This may be combined with other strategies: attempts to command the opponent and to make him follow the applicable rules. In particular, in 20081225-zhurn-b3 a student (female) during a 18 sec fragment consecutively (a) tries to provoke sympathy and come to take the test another time, (b) tries to provoke indulgence by showing, she does not feel/understand very well, (c) rationally asks to specify the task, (d) accuses the opponent, indicates, that the task is not precise, (e) laughs in order to reduce a negative impression from her past words. In the intense exam situation here the influence strategies are changed without correct coordination every 3,6 sec (on average), which makes this example a valuable base to design computer agents, able to enumerate possible emotional strategies in communication.

Table 2. Speech protocol with facial expression

Speech	**Mouth**	**Eyes**
Zeugma is…		
	licks lips	looks aside < right > rapidly blinks
when…		looks aside < left >
	opens mouth	
	compresses lips; clicks	looks aside < down >
ep… ekh…		looks aside < right >
some word…		<looks at the examiner> rapidly blinks
mostly the verb		looks aside < right > looks aside < down >
can be omitted.		<looks at the examiner>

5. Discussion

The corpus gives us numerous examples of emotional expressive cues and strategies, which are difficult to observe in other types of data.

Search function, provided by ELAN annotation software, allows to look for a specific cue or a combination of cues (possible expressive pattern): e. g. 'raises eyebrows and compresses lips' (n = 47) or 'compresses lips and immediately after licks the lips' (n = 94). By browsing the search results we can (a) classify the superficial expressive types for a cue, e. g. types of smiles (mouth open, mouth closed etc.) or clicks (bilabial, labiodental, coronal etc.), and (b) classify contexts and functions for a specific cue or expressive pattern, e. g. position of smiles in speech, functions of smiles. We also can check, if the combination of cues constitutes a steady expressive pattern with some specific function, or if the two cues just co-occur on the timeline. This allows the design of an expressive library for a computer agent with "grammar" for each expressive cue, which further may allow the establishment of functional annotation for the superficial cues in corpus.

Observation of compound examples suggests the architecture of the computer agent, which should be sufficient to construct dynamically these examples during its operation. In particular, examples with numerous expressive cues and numerous influence strategies require the architecture, where different reactions are activated for a given stimulus (or goal) and compete for the output (speech and animation of the agents figure).

6. Conclusions

The corpus provides ecologically valid data with annotation and can be used for a variety of scientific and applied purposes:

- analysis of functions and "grammar" of facial expression and gestures;
- analysis of communication strategies: humor, politeness strategies, exploitation of emotions ("pull" emotions);
- extension for the theories of conversational analysis and design of automatic dialogue systems;
- automatic construction of "speaker's intent" or "the theory of mind", when a computer system interprets observed expressive cues in order to detect the internal state and intent of the speaker;
- automatic pattern extraction;
- training of automatic classifiers (e. g. computer neural networks) in the area of dialogue simulation;
- database of complicated expressive patterns for cartoon animators and actors;
- training base for state and corporate officers, interacting with clients; optimization of corporate client services.

In our case, the observations made on the basis of the corpus, allow the extraction of expressive patterns for emotional computer agents, and the design of the internal architecture of the agents to simulate the observed compound reactions, involving internal emotional arousal as well as surface etiquette and simulated emotional expressions.

References

1. *Abrilian S., Devillers L., Buisine S., Martin J.-C.* EmoTV1: Annotation of Real-life Emotions for the Specification of Multimodal Affective Interfaces. HCI International 2005. Las Vegas, USA, 2005.
2. *Bänziger T., Scherer K. R.* (2007) Using Actor Portrayals to Systematically Study Multimodal Emotion Expression: The GEMEP Corpus, in Affective Computing and Intelligent Interaction, Lecture Notes in Computer Science 4738. Springer-Verlag, Berlin, Heidelberg, pp. 476–487.
3. *Becker C., Kopp S., Wachsmuth I.* (2004) Simulating the Emotion Dynamics of a Multimodal Conversational Agent, in Affective Dialogue Systems, Lecture Notes in Computer Science 3068, Springer-Verlag, Berlin, Heidelberg, pp. 154–165.

4. *Brown P., Levinson S. C.* (1987) Politeness : Some Universals in Language Usage, Cambridge.
5. *Clore G. L., Ortony A.* (2000) Cognition in Emotion: Always, Sometimes, or Never? in Cognitive Neuroscience of Emotion, Oxford Univ. Press, pp. 24–61.
6. *Kotov A.* (2007) Simulating Dynamic Speech Behaviour for Virtual Agents in Emotional Situations, in Affective Computing and Intelligent Interaction, Lecture Notes in Computer Science 4738. Springer-Verlag, Berlin, Heidelberg, pp. 714–715.
7. *Kotov A.* Accounting for irony and emotional oscillation in computer architectures. International Conference on Affective Computing and Intelligent Interaction ACII 2009. Amsterdam, IEEE, 2009, pp. 506–511.
8. *Kotov A.* Types of Simulated Emotional Expressive States in the Russian Emotional Corpus Komp'iuternaia Lingvistika i Intellektual'nye Tekhnologii: Po Materialam Ezhegodnoi Mezhdunarodnoi Konferentsii "Dialog" [Computational Linguistics and Intellectual Technologies: Materials of the International Conference "Dialog"]. Bekasovo, 2011, pp. 315–324.
9. *Maglogiannis I., Karpouzis K., Bramer M., Martin J.-C., Caridakis G., Devillers L., Karpouzis K., Abrilian S.* (2006) Manual Annotation and Automatic Image Processing of Multimodal Emotional Behaviors in TV Interviews. Artificial Intelligence Applications and Innovations, Vol. 204, pp. 369–377.
10. *Ochs M., Niewiadomski R., Pelachaud C., Sadek D.* (2005) Intelligent Expressions of Emotions, in Affective Computing and Intelligent Interaction, Lecture Notes in Computer Science 3784, Springer-Verlag, Berlin, Heidelberg, pp. 707–714.
11. *Rehm M., André E.* (2008) From Annotated Multimodal Corpora to Simulated Human-Like Behaviors, in Modeling Communication with Robots and Virtual Humans, Lecture Notes in Computer Science 4930, Springer-Verlag, Berlin, Heidelberg, pp. 1–17.
12. *Scherer K. R., Ceschi G.* (1997) Lost luggage emotion: A field study of emotion-antecedent appraisal, Motivation and Emotion, Vol. 21, pp. 211–235.
13. *Zara A., Maffiolo V., Martin J.-C., Devillers L.* (2007) Collection and Annotation of a Corpus of Human-Human Multimodal Interactions: Emotion and Others Anthropomorphic Characteristics, in Affective Computing and Intelligent Interaction, Lecture Notes in Computer Science 4738, Springer-Verlag, Berlin, Heidelberg, pp. 464–475.