

# Cognitive Entropy in the Perceptual-Auditory Evaluation of Emotional Modal States of Foreign Language Communication Partner

Rodmonga Potapova<sup>1</sup>(✉) and Vsevolod Potapov<sup>2</sup>

<sup>1</sup> Institute of Applied and Mathematical Linguistics,  
Moscow State Linguistic University, Ostzhenka 38, Moscow 119034, Russia  
rkpotapova@yandex.ru

<sup>2</sup> Faculty of Philology, Lomonosov Moscow State University,  
GSP-1, Leninskie Gory, Moscow 119991, Russia  
volikpotapov@gmail.com

**Abstract.** The paper deals with the phenomenon of perceptual-auditory divergence in the evaluation of the foreign language communication partner's emotional-modal state. The problem of "human – human" interaction (vice versa "man – machine") is characterized by a very complex phenomenon associated with the communication "native language – foreign language" taking into account the idiosyncrasy of speech production and speech perception. The idiosyncrasy can be determined from the positions of the individual mixing of various types of below-specified information (e.g., biological, psychological, social, cognitive, etc.). All these factors affect the process of recognizing the communication partner's emotional-modal state. Therefore, one can assume that the idiosyncratic features of the perceiver (in this case, the listener) affect, in turn, the evaluation of the emotional-modal state, primarily that of a foreign language communication partner. In our pilot study special emphasis is laid on this problem considered on the basis of Russian-German and German-Russian matches. The obtained data suggest a new model of perceptual-auditory processing of verbal stimuli including such components as perceptual-auditory idiosyncrasy and auditory-perceptual cognitive entropy.

**Keywords:** Emotional-modal state · Divergence · Foreign language incentive · Perceptual-auditory evaluation · Spoken language communication · Idiosyncrasy · Cognitive entropy

## 1 Introduction

In the communication process it is important not only to understand the communication partner's verbal content, but (to a greater degree) also to recognize their emotional-modal state. An additional difficulty is found in searching a solution of a similar problem in the case of communication with foreign language and foreign-culture partner. The complexity of solving this problem is escalated in distant mediated communication (for example, mobile communications, social-network communication on the Internet, Voice over IP (VoIP), Skype, WhatsApp, Viber, Google Hangouts, etc.).

In the course of the study a hypothesis was formulated that in the communication act not only the principle of speech signal generation idiosyncrasy (stimulus) is implemented, but also the principle of signal auditory perception (reaction) idiosyncrasy.

## 2 Theoretical Background of the Research

It is known that a common unified theory of emotions does not exist. In particular, P.V. Simonov's need-information theory is widespread [22: 320–328]. According to this theory, the emergence of emotions is determined by certain needs and evaluation of the possibility to satisfy and meet these needs, which are formally expressed as follows:

$$\pm E = f[-N(I_r - I_a)],$$

where  $E$  is the emotion intensity and its sign;  $N$  is needs degree;  $I_r - I_a$  is an evaluation of the possibility to satisfy this need considering of the available experience;  $I_r$  is information of the means objectively required to satisfy the need;  $I_a$  is information of the means available to a person.

According to this theory, if there is an excess of information on the possibility to satisfy the need, then a positive emotion emerges; if there is a lack of information, then a negative emotion is produced. It is believed that the variety of emotions is determined by the variety of needs. This view coincides with the concept of deprivation [4], which is developing successfully nowadays and provides the possibility to “outreach” the solution of problems found in studying the nature of destructive actions, aggression and terror reflected on a significant scale in the information and communication media (e.g., in the social-network discourse) [21].

The classification is based on binary principle: every key set includes some subsets (e.g. situational subsets). Division may be, for example, into primary and situational subsets. Primary subsets include unsatisfied needs correlated with search of the target object. Situational emotions emerge as a result of evaluations made for steps of behavior and encourage either action in the same direction or modification of the behavior [24, 25]. Emotions are also divided by the nature of actions including overcoming, protection, and attack [23]. Emotions can be the result of two or more overlapping emotions.

To describe emotions W. Wundt [26] identified three features:

- Hedonic tone or emotion sign (positive–negative),
- Readiness for action (relaxation–tension),
- Activation level (tranquility–excitement).

H. Schlossberg developed Wundt's theory and introduced a feature including the opposition “acceptance-rejection” [5: 40].

In addition to the above approaches to the analysis of emotions worthy of mention is the concept that takes into account the ratio of an emotion and an event. In this case, anticipatory (before the event associated with the achievement/failure to reach the target) and summative (after the event associated with the achievement/failure to reach the target) emotions are distinguished. A significant factor is the orientation of

emotions: to or away from oneself. The primary function of emotions is the body mobilization for rapid reaction to the situation in most appropriate manner [3].

The above points of view relate primarily to interpretation of the concept of “emotion”. As for the emotional-modal state, the linguistic literature considers three aspects of modality: subjective, objective and secondary [8]. The subjective modality includes evaluative attitude of the speaker to the degree of cognition of these objective relations [6]: for example, doubt, certainty/uncertainty, presupposability. In other words, the subjective modality is understood as the modality of credibility degree.

Thus, the subjective modality expresses the speaker’s attitude to the verbal and non-verbal behavior of the percipient: emotional-expressive attitude, self-evaluation, evaluation derived from features of the content, etc. (e.g., confidence, diffidence, belief, presupposability, etc.).

As in the classification of emotions, all emotional-modal states (EMSs) are distributed in the following basic types: positive/negative; subjective/objective; primary/secondary; single-factor (uniform)/multifactor (mixed); strong/weak.

In our opinion, for any EMS classification, first of all, an evaluative criterion is common: evaluative; communicative-evaluative; situationally evaluative; socio-evaluative; ethnocultural-evaluative one.

### 3 Auditory Idiosyncratic Classification Features

Speech features that characterize the speaker and their psychophysiological features (idiosyncratic features) contain, as a rule, various types of information: verbal content (information-communicative content of the message); paraverbal content (pronunciation features of speech production “woven” into the prosodic and spectral-temporal substance of sound matter of any utterance); non-verbal content (facial expressions, gestures, proxemics, etc.); extraverbal content (gender, age, place of birth, places of long-term residence, upbringing, education, social status, situation, hairstyle, clothes, etc.).

The following classification of idiosyncratic information contained in the speech signal is proposed [1, 2, 7, 9, 10, 14, 15, 19]:

- Biological: anatomical, physiological, physical, psycho-emotional, psychomental, gender, age, sexual-genetic.
- Psychological: mental characteristic of the person as a single integrated functional system of behavior and activity regulation (consciousness, attention, memory).
- Sociobiological: sociobehavioral (signs indicating belonging to a certain group of people (for example, ethnic, cultural, regional and social one)).
- Socio-humanitarian: evolutionary-genetic, ethnic, interethnic, cultural-historical, general psychological, differential-psychological, psychogenetic.
- Multifactor: intellectual, verbal, cognitive, factor-analytical.

In describing the speaker’s “profile” by voice and speech (i.e., for example, individual attributes in forensics), three types of norms are distinguished: universal, group and idiosyncratic ones. The special role belongs to the voice information decoded at the level of auditory perception as follows [13, 14, 16, 17]:

- description that relates to the speaker's profile and their place in the perceiver's real existence;
- associative correlation with the speaker's name;
- speaker's psychophysical, psychophysiological and psychopathological forms (general constitution, face, gestures, manner of walking, etc.);
- speaker's voice, manner of phonation, articulation, coarticulation with regard to speech manner ("trophotropica" – "ergotropica") determined by distinct (acoustic, perceptual) features.

Speech features of the speaker are divided into controlled (external) and uncontrolled (internal) ones. Some experts mention potentially controlled features. The degree of control depends on two factors:

- speaker's ability to use auditory and proprioceptive feedback forms in the process of articulation program implementation;
- speaker's perceptual ability to use auditory forms of information to detect sound differences [2, 7].

Both of the above factors are part of the "control skills" concept. Factors beyond any control are conditioned by speaker's organic-genetic features: the structure of the speech apparatus including the length of the vocal tract, sizes of their tongue, soft palate, throat, jaw and mouth cavity; form (configuration) of the laryngeal tract and nasal cavity. This also may include so-called structural defects (e.g., presence of cleft in the hard palate ("cleft palate"), missing teeth, etc.) [7].

Controlled factors are not related ("derived") with organic-genetic constraints and include changes in the voice dynamics and all potentially controllable muscle articulatory gestures that characterize the manipulation components of the voice quality.

Uncontrolled features are considered as permanent and at the same time non-permanent organic basis for the speaker's features based on their anatomy and physiology and correlated with the invariant-norm of the physical state characteristic for the speaker's voice features.

Permanence and non-permanence of uncontrolled features is related to the differentiation of long-term and short-term features (the principle of identifying short-term uncontrolled features is speech production with sore throat, after running, fast ascent up the stairs, etc.). The short-term features cannot include, for example, features of voice break (puberty). Both controlled and uncontrolled features can be grouped on the basis of intraspeaker and interspeaker similarities and differences [9, 14]. Information on the speaker is hidden in the speech signal that relates to their anatomical features and muscular voice samples stored at the neuronal level that correlate, for example, with the speaker's constitution.

In addition to these special features, the study of emotions and emotional-modal states by voice and speech is especially complex, in particular, the problem of their perceptual-auditory identification [12, 13, 16–19].

## 4 Method, Experimental Results

In the preliminary research devoted to the study of emotions involving various groups of subjects (actors, subjects in a state of hypnosis and subjects with manic-depressive disorders ( $n = 540$ )), it was concluded that a fundamental distinction between “emotion” and “emotional-modal state” is needed, which allows classification of peoples’ emotive behavior considering the following [14]:

- basic, so-called primal unconscious emotions (e.g., anger, rage and fear/fear as unconscious or conscious reaction of neurons to stimuli of any kind, in particular, to danger (with active and passive (stupor) forms)/horror, joy, admiration);
- unconscious and conscious reactions to stimuli forming complex social emotional and cognitive systems correlated with the concept of “feeling” (e.g., love, hate, happiness, etc.);
- complementary emotional-modal states, reflecting a subjective self-estimate of a person, their communication partner, the current situation, reality, etc. (e.g., confidence, diffidence, doubt, indifference, contentment, compassion, credulity, depression, hopelessness, anxiety, dissatisfaction, satisfaction, disgust, contempt, shame, resentment, malice, etc.) [17–21].

This concept provides a comprehensive description and analysis of the multifaceted role that speech parameters play regarding basic (so-called primal) emotions, the latter, social emotions (feelings) and concrete situated emotional-modal states [11].

The results of the longitudinal experiment conducted in Russia and Germany for ten years with the different authentic materials (TV Talk-Shows of Russia and Germany; volume of 227 h) and for three years with authentic Skype dialogues (volume of 120 h) provided for evaluation of emotionally charged Russian and German speech by listeners native Russian and native German speakers without visual imagery (studies with visual imageries have also been carried out) revealed that auditory perception of the same emotional-modal stimuli is evaluated by representatives of the above-mentioned cultural and verbal communities in different ways [15–17].

The data used in this paper derive from two sources. The first source of data comes from the investigation on the basis of Russian Television Talk-Shows and the second source of data was Russian Skype dialogues. 120 native listeners of Russian (60 females, 60 males) were selected from the student population of Moscow State Linguistic University (Russia) and 120 native listeners of German (60 females, 60 males) from University Halle-Wittenberg (Germany) were used for this experiment.

The discursive presentation of focusing on different emotional-modal states regarding speech communication in Russian dialogues and polylogues on the bases of Talk-Shows and later Skype communication gave possibility to define peculiarities of perceptual-auditory divergence regarding evaluation of the foreign language emotional-modal state interpretation.

The task of Russian and German experiment participants was to listen every speech fragment without visual channel, time limitation and to answer all questions of a special questionnaire. The data base of emotional-modal states included positive and negative

lexemes ( $n = 55$ ) (e.g. neutral, natural, wistful worry, depressed, disappointed, aggressive, joyful etc.)

Listeners were requested to analyse the audio fragments and instructed to determine prosodic features of speech samples (pitch, speech rate, timbre, dynamics).

It was found out that the auditory-perceptual evaluation of emotional-modal state of foreign-language and foreign-culture communicants by their speech features is significantly different from innate and genetically structured neuro-emotional mechanisms of perceptual-auditory evaluation of the emotional-modal behavior of native speakers. During the experiments we also were able to establish statistically reliable specifics of the German-Russian and Russian-German speech tactics and speech strategies used in the decision-making process to evaluate the emotional-modal state of foreign language and foreign-culture partners in communication.

The data of the perceptual-auditory analysis of recognition of emotions and emotional-modal states were subjected to the subsequent statistical processing using signed rank tests with two-sided alternative hypothesis  $H_0$  with a 5% significance value, as well as to two-factor analysis of variance, which allowed to obtain reliable and significant differences.

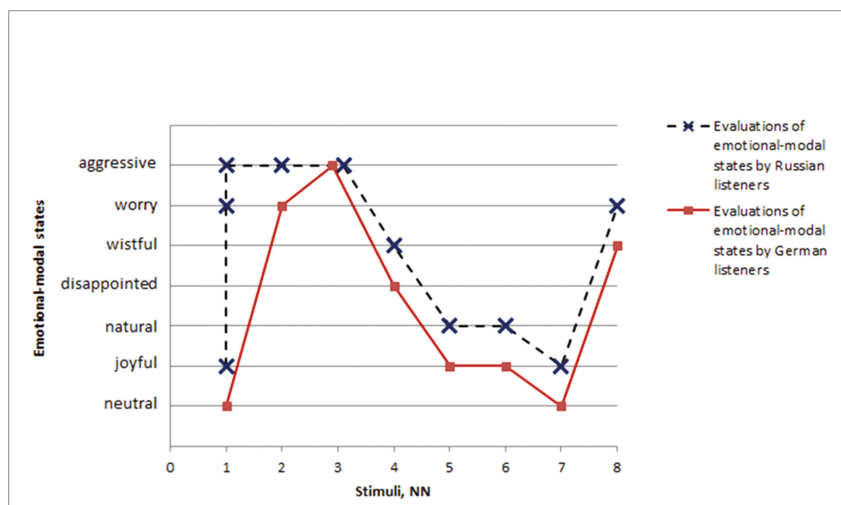
In the experiments on the perceptual auditory evaluation of the emotional-modal states of foreign-language and foreign-culture communicants special questionnaires were developed including such parameters as the melodic range with a certain gradation system; melodic register and its change stages; voice pitch; voice tone dynamics, tone with inclusion of a number of varied forms; dynamic and temporal features of speech production, etc.

Studying the same material with the inclusion of verbalics (parameters were separated as a special research provided knowledge of the languages) and non-verbalics (e.g. facial expressions, gestures, etc.) accompanying communication act.

Key findings related to the auditory evaluation of the verbal representations of foreign-language and foreign-culture communicants' EMSs can be reduced to the following.

The majority of German listeners perceive emotional-modal state of the Russian speech as "neutral" (66%), as "agitated" (27%) and as "aggressive" (7%); for speakers of Russian listeners it is typical to some extent to perceive emotional-modal state of German speaking communicants also as "neutral" (40%). Other emotional-modal states are evaluated with a high degree of variability. However, the same Russian speech stimuli-utterances evaluated by German listeners as "agitated" are, in most cases, perceived by Russian speaking subjects as "joyful" (27%). The evaluation "aggressive" falls into one connotative-emotional zone with the inclusion of other states, such as "agitated", "joyful", etc. (33%). The most significant is the presence of divergence in evaluating emotional-modal states "agitated" for German speaking subjects and "joyful" for Russian speakers [15].

All participants listeners (Russian and German native speakers) completed special questionnaires during the longitudinal experiment. Average results are presented in Fig. 1.



**Fig. 1.** Divergence in the perceptual auditory evaluation of emotional modal states of foreign language communication partners (data differ significantly ( $p$  – value  $<5\%$ ))

Comparing the prosodic means used by the listeners to evaluate emotional-modal states also revealed a number of divergent features:

- perceptual auditory evaluation by Russian listeners of emotional-modal states of German speaking subjects is based on evaluation of the perceived voice timbre (73%), melodic variation (20%) and temporal features (7%);
- perceptual auditory evaluation by German listeners of emotional-modal of Russian speaking subjects is based on the use of melodic features (60%), temporal features (20%), volume levels (13%), pausation features (7%).

## 5 Conclusion

The obtained results show primarily the functioning of the language categoricity mechanism with regard to foreign language stimulus auditory perception in the communication process that imposes specific restrictions on the interpretation of the results, in this case on the auditory analysis of foreign language spoken material.

Thus, idiosyncrasy on the example of spoken language communication can be of two types: not only at the level of speech production, as it is commonly believed, but also at the speech perception level. Regarding dialogic communication (in this case with regard to emotional-modal states reaction of the listener to spoken stimuli of the speaker due to the divergence process), reactions to stimuli can be of two main types: non-unisonant ( $S \neq R$ ) and approximately unisonant ( $S \approx R$ ). For foreign-language communicants the degree of non-unisonant perceptual auditory reactions increases.

It should be emphasized that perceptual-auditory idiosyncrasy, in our view, is compounded by the fact that it includes such components of the hearing, as physiological;

musical; emotional; subject-voice-identifying; associative; aesthetic; cognitive-individual; cognitive-social; cognitive-ethnic, etc.

During the speech communication with the participation of foreign-language and foreign-culture partners, idiosyncratic nuances of emotional-modal states, which are stimuli, may cause and usually cause the above-described reactions of non-unisonant or approximately unisonant type ( $S \neq R$ ,  $S \approx R$ ).

If we consider that the response is influenced by other individual factors (e.g., the physical state of the percipient, the general emotional background of communication, gender, age etc.), then the emotional-modal states reaction “acquires” a number of other connotations leading, in our opinion, to the emergence of the phenomenon which we propose to designate as “cognitive entropy”. At that, the measure of cognitive entropy uncertainty increases with the number of idiosyncratic individual-personal components of the person’s auditory perceptual-system, and their cognitive “equipment”.

The zone nature of cognitive entropy – in this case, in perceptual-auditory recognition of emotional-modal states – may relate, in our opinion, to other components of the human sensory system as well (e.g., non-verbalics), particularly in communication between foreign-language and foreign-culture participants of the communicative act.

The next stage of this research will include: (a) perceptual visual, (b) perceptual visual and perceptual auditory evaluation and analysis of emotional-modal state of non-verbal stimuli by native and non-native subjects without analysis of verbal speech signals.

Cross-language comparison of these data sets will give us the opportunity to define facts of divergences between reactions on the same emotional-modal stimuli and the peculiarities of the cognitive entropy regarding cognitive reactions on the same stimuli by aural and visual channels for native and non-native subjects. The data obtained and expected to be obtained in the future stages of the research may prove useful for fine-tuning voice and multimodal biometric analysis systems with regard to different languages possessing of different predominant distinctive parameters correlating to emotional-modal states and speaker-specific features in real communication, including both language- and culture-dependent phenomena.

**Acknowledgements.** This research is supported by Russian Science Foundation, Project №14-18-01059.

## References

1. Abercrombie, D.: *Elements of General Phonetics*. University Press, Edinburgh (1967)
2. Brown, R.: Auditory speaker recognition. In: *Forum Phoneticum*, Bd. 38. Helmut Buske Verlag, Hamburg (1987)
3. Ekman, P.: Basic emotions. In: Dalglish, T., Power, M. (eds.) *Handbook of Cognition and Emotion*. John Wiley and Sons, Ltd., Sussex (1999)
4. Gurr, T.R.: *Why Men Rebel*. Princeton University Press, Princeton (1970)
5. Isard, C.E.: *Human Emotions*. Plenum Press, New York (1977)
6. Krivonosov, A.T.: Modal particle as a signal of subjective modal significance in the German spoken language. In: *Structural Features of Spoken Language and Development of Conversational Skills*, Irkutsk, pp. 132–147 (1964). (in Russian)
7. Laver, J.: Voice quality and indexical information. *Br. J. Disord. Commun.* **3**, 43–54 (1968)



8. Lomtev, T.P.: Sentence and Its Grammatical Categories. MSU, Moscow (1972). (in Russian)
9. Potapova, R.K.: Subject-oriented perception of foreign language speech. *Issues Linguist.* **2**, 46–64 (2005). (in Russian)
10. Potapova, R.K.: Speech: Communication, Information, Cybernetics, 4th edn. Publ. House “Radio i svyaz”, Moscow (2010). (in Russian)
11. Potapova, R.: From deprivation to aggression: verbal and non-verbal social network communication. In: Materials of the VI International Conference “Global Science and Innovation”, vol. I, Chicago, pp. 129–137, 18–19 November 2015
12. Potapova, R., Komalova, L.: Auditory-perceptual recognition of the emotional state of aggression. In: Ronzhin, A., Potapova, R., Fakotakis, N. (eds.) *SPECOM 2015. LNCS (LNAI)*, vol. 9319, pp. 89–95. Springer, Cham (2015). doi:[10.1007/978-3-319-23132-7\\_11](https://doi.org/10.1007/978-3-319-23132-7_11)
13. Potapova, R.K., Potapov, V.V.: Auditory Perception of Speech by Non-Native Speakers. *The Phonetician*, CL-78, pp. 6–12 (1998)
14. Potapova, R.K., Potapov, V.V.: Language, Speech, Personality. *Yazyki slavyanskoj kul'tury*, Moscow (2006). (in Russian)
15. Potapova, R.K., Potapov, V.V.: Perception of emotional behavior of foreign-language and foreign-culture communicants. In: Phonetics and Non-phonetics. On Sandro Kodzasov's 70th anniversary, pp. 602–617. *Yazyki slavyanskikh kul'tur*, Moscow (2008). (in Russian)
16. Potapova, R.K., Potapov, V.V.: *Kommunikative Sprechtaetigkeit: Russland und Deutschland im Vergleich*. Boehlau Verlag, Koeln (2011)
17. Potapova, R., Potapov, V.: Auditory and visual recognition of emotional behaviour of foreign language subjects (by native and non-native speakers). In: Železný, M., Habernal, I., Ronzhin, A. (eds.) *SPECOM 2013. LNCS(LNAI)*, vol. 8113, pp. 62–69. Springer, Cham (2013). doi:[10.1007/978-3-319-01931-4\\_9](https://doi.org/10.1007/978-3-319-01931-4_9)
18. Potapova, R., Potapov, V.: Associative mechanism of Foreign spoken language perception (forensic phonetic aspect). In: Ronzhin, A., Potapova, R., Delic, V. (eds.) *SPECOM 2014. LNCS(LNAI)*, vol. 8773, pp. 113–122. Springer, Cham (2014). doi:[10.1007/978-3-319-11581-8\\_14](https://doi.org/10.1007/978-3-319-11581-8_14)
19. Potapova, R., Potapov, V.: Cognitive mechanism of semantic content decoding of spoken discourse in noise. In: Ronzhin, A., Potapova, R., Fakotakis, N. (eds.) *SPECOM 2015. LNCS(LNAI)*, vol. 9319, pp. 153–160. Springer, Cham (2015). doi:[10.1007/978-3-319-23132-7\\_19](https://doi.org/10.1007/978-3-319-23132-7_19)
20. Potapova, R., Potapov, V.: On individual polyinformativity of speech and voice regarding speakers auditive attribution (forensic phonetic aspect). In: Ronzhin, A., Potapova, R., Németh, G. (eds.) *SPECOM 2016. LNCS(LNAI)*, vol. 9811, pp. 507–514. Springer, Cham (2016). doi:[10.1007/978-3-319-43958-7\\_61](https://doi.org/10.1007/978-3-319-43958-7_61)
21. Potapova, R., Potapov, V.: Polybasic attribution of social network discourse. In: Ronzhin, A., Potapova, R., Németh, G. (eds.) *SPECOM 2016. LNCS(LNAI)*, vol. 9811, pp. 539–546. Springer, Cham (2016). doi:[10.1007/978-3-319-43958-7\\_65](https://doi.org/10.1007/978-3-319-43958-7_65)
22. Simonov, P.V.: Brain mechanisms of emotions. *J. High. Nerv. Activ.* **47**(2), 320–328 (1997). (in Russian)
23. Simonov, P.V.: *Emotional Brain*. Nauka, Moscow (1981). (in Russian)
24. Vilyunas, V.K.: *Psychology of Emotional Phenomena*. MSU Publ. House, Moscow (1976). (in Russian)
25. Vilyunas, V.K.: *Psychological Mechanisms of Person's Motivation*. MSU Publ. House, Moscow (1990). (in Russian)
26. Wundt, W.: *Grundriss der Psychologie*. W. Engelman, Leipzig (1896)