Main Determinants of the Acmeologic Personality Profiling

Rodmonga Potapova[1(✉)] and Vsevolod Potapov[2]

[1] Institute of Applied and Mathematical Linguistics, Moscow State Linguistic University,
Ostozhenka 38, Moscow 119034, Russia
RKPotapova@yandex.ru

[2] Faculty of Philology, Lomonosov Moscow State University, GSP-1, Leninskie Gory,
Moscow 119991, Russia
volikpotapov@gmail.com

Abstract. This investigation aims at establishing a set of voice and speech personality identification features that predict some phonation and articulation gestures in regard to lexical-semantic, phonological, anthropometrical, acoustical, physiological, psychological, emotional and intellectual peculiarities of the "electronic personality" on the Internet and other automated digital communication means and devices. This problem is very significant for forensic investigations in the domain of speech communication. It is proposed to undertake special studies to find solutions to these problems in the field of forensic phonetics. In forensic application of the "electronic personality", it is necessary to be able to specify a temporal dynamics factor for a decision concerning acmeologic quantitative and qualitative changes of the personality in time.

Keywords: Social-Network Discourse · Relevant personality features
Forensic emotional personality profiling in dynamics
Acmeologic personality profiling · Perceptual-auditory analysis

1 Introduction

The development of the social-network discourse (SND) investigations on the Internet involves studying the mechanism of dependence between the acoustic prosodic-semantic interpretation of the speech utterance by the speaker and processing of the discourse construction by the listener considering such factors as: the cognitive-verbal base of the communicants' idiosyncratic peculiarities; the multimodal (verbal, para-verbal, non-verbal, extra-verbal) structure of coding (stimulus generation) and decoding (reaction to stimulus) of communication process items by the communicants [10, 11, 15–18]; the multi-level (phonological-phonetic, syntactic-semantic and pragmalin-guistic) structure of verbal coding (speech stimulus) and decoding (speech reaction to the stimulus) of the process by the communicants; paraverbal (emotional, emotional-modal and connotative) components of the speech stimulus and speech reaction to this stimulus (in the communication act); extraverbal (situational, individual – idiosyncratic, idiolectal [6], sociolectal, etc.) constituents of the speech stimulus and speech reaction

to the utterance taking into account the role of presupposition, as well as the recipient's previous experience in a particular subject area [15].

Experimental research and modeling of acmeologic variability of spoken social network discourse (SND) forms an important direction of modern communicative variantology [23], and forensic sciences require information about the multimodal individual dynamics of personalities or personality communities [9, 18]. Of particular importance is the above direction in connection with the SND functioning in the information and communication space of the Internet [9, 15, 16, 18]. Analysis of the deep mechanism of the prosodic-semantic variability of the verbal response to the stimulus between communicants within the spoken discourse in respect to the SND requires knowledge in various fields of speechology (spoken language sciences): in general, private and experimental phonetics, cognitive and communicative linguistics, speech acoustics, auditory perception, mathematical statistics, forensic linguistics, etc. [7]. The solution of the problem taking into account the multiversatile variability of the analyzed object includes the following: search for the pronouncing invariant and variants of the prosodic-semantic interpretation of the speech stimulus-utterance in the SND; determining the interaction of the various factors listed above in the process of SND construction; the degree of influence of the above factors on the final verbal product of the SND in the communication act; identification of the prosodic-semantic dominant within the SND; determining the acceptable range of variation of prosodic-semantic variforms – alloprosodosemants; determining speaker profiling, verification and identification; investigation of the variability of the speech and voice characteristics of the communicants, personality dynamics of the individual "portrait" in time with regard to the acmeologic method [15, 16, 18]. The research on the prosodic-semantic variability of multilevel verbal, para-, non- and extraverbal components of spoken discourse involves various modern methods of analysis, synthesis and modeling of sounding speech: acoustic, perceptual-auditory, associative, prosodic-semantic [7]. Acmeologic profiling of communicants on the Internet and in other automated means of communication includes first of all interdisciplinary researches: "detailed phonetic and linguistic description of the verbal behavior of an… individual, …careful analysis of dialectal and sociolectal features, speech defects, age, "voice quality," … a combination of traditional phonetic analysis, techniques, including analytical listening by a phonetician, and modern signal processing techniques…" [4: 80-99].

2   Conceptual Background of the Personality Profiling

Speech activities in the format of spoken social-network discourse (SND) – in particular, based on various modern IP-telephony facilities on the Internet, – can be presented taking into account the following level-by-level components: incentive level: external impact; motive; intent; communicative intention; formation level: the sense-forming phase; deep formation of the space-concept scheme; time (linear) development of the spatial-conceptual scheme of the utterance; formulation level: formulating phrase (choice of words); process of grammatical structuring; realization level: articulatory gestures (articulation); voice modulation (phonation); coarticulation transformations;

acoustic level: transformation of articulatory gestures at the output of the speech-formation system into a sound (acoustic) wave; auditory detection, auditory control and recognition of perceived acoustic stimuli; interpretation level: transformation of acoustic stimuli into verbal images, semantic content realization [9, 10, 12]:

In accordance with the expanded understanding of the object of research in speechology, the following techniques and methods can be mentioned: cognitive-communicative analysis of the text; indirect checking of models and hypotheses, for example, by studying speech errors, linguistic reactions, etc.; neurophysiological methods; bioelectric methods; registration and analysis of articulation, for example, using computed tomography, etc. [9]. The study of the realization and monitoring of motor programs should be related to the information processing system in the central and peripheral nervous systems. It is likely that in the central nervous system there is no functional center that would specialize in processing verbal information exclusively. Neural networks processing verbal information also include all functions.

3   SND Communication Analysis Considering Human Speech Functions

In investigations [8, 13–16, 18], the concept of SND was first substantiated based on its definition as a special electronic macro-polylogue with regard to a number of categories of form, content and functional weight. An example is one of the form categories on the basis of the "univector – polyvector" opposition. This opposition is correlated not only on the basis of location of communicative interaction vectors on the Internet, but also on the SND participants' interaction configuration, which is directly dependent on the number of communicants on the Internet.

When examining the speech behavior of the speaker in the SND format using, for example, IP-telephony, we proceed from the following postulate: human speech is both a symptom and a signal in relation to the real world: a symptom as a direct psychophysiological response to external stimuli and a signal as a sign language response of neuropsychological nature to stimuli of a more complex behavioral level in the communication act [11]. In this regard, speech is presented as a poly-informative and multifunctional phenomenon. The development of this issue assumes special importance in the study of spoken-speech communication with the help of IP-telephony: to solve a number of problems of forensic examination. "… The latter circumstance naturally led to the fact that phonograms of conversations by the channels of cellular communication and the Internet became objects of investigation for experts in forensic examination of sound recordings" [5: 129, 20].

The pronunciation of the speaker includes a set of specific properties of this individual manifested in formation of the sound flow in the speech apparatus and conditioned by the peculiarities of its structure, the features of the pronunciation-auditory skills, the specifics of thinking, and the formulation of thoughts with the help of linguistic means [10–12]. The speech "portrait" of the speaker includes verbal, paraverbal, non-verbal and extraverbal features. Verbal components refer to such aspects as the language used in the communication process (native, non-native, dialect, vernacular, sociolect, etc.)

For each speaker, an inventory of stable phonetic features is characteristic: pronouncing variants of phonemes, variants of intonemes, etc. Verbal speech features make it possible to determine such components of the speech portrait as nationality, places of the speaker's long residence, level of education, social status, economic status, upbringing, level of language proficiency, profession, level of intellectual skills, etc. It is thought that extraverbal features correspond with anthropometric (structure of the speech apparatus, body weight, height) [4], physiological (gender, age, norm/pathology), psychological (type of higher nervous activity (HNA) [22], emotional-volitional regulation), intellectual (specific thinking, cognitive level) aspects. Accordingly, it is possible to distinguish relatively stable speech extraverbal features in the speaker's speech portrait. Both verbal and extraverbal features have their own acoustic correlates that make it possible to recreate the "portrait" of the speaker. For example, gender and age can be characterized by some acoustic parameters [1]. There are various data for native Russian speakers. According to observations [12], the average value of the pitch frequency dynamics for males aged from 20 to 80 years increases ($\approx$ 100–130 Hz). For females aged 20 to 80 years, the reverse difference in value is observed ($\approx$ 220–180 Hz) [3, 10, 13–18, 21].

Proceeding from the basic premise, according to which the human speech is individually organized on the basis of phonation and articulatory gestures in direct connection with the socially-conditioned phonological representation of the utterance and its lexical and semantic features, it is proposed to conduct an express-analysis of the speaker's speech portrait taking into account the following stages: formation of the databases for correlates of anthropometric features; acoustic correlates of physiological features; acoustic correlates of psychological and emotional-psychological features; acoustic correlates of intellectual features [13–15, 17, 18]. Thus, the acoustic-linguistic algorithm of the speaker identification analysis is constructed taking into account the following stages: acoustic; anatomical-physiological aimed at decoding of the speech signal; socio-psychological aimed at decoding of the speech signal; intellectual-semantic decoded for the speech signal. In this regard, all the tasks can be conditionally characterized as tasks of compiling an individual portrait of the speaker, to which phonation (voice), articulatory segment (motor), prosodic (suprasegment) correlates of the speaker's speech should be attributed.

Speech characteristics of the speaker are divided into controlled (external) and uncontrolled (internal) ones. Some experts identify potentially controlled features. The degree of control depends on two factors [2, 11]: the speaker's ability to use auditory and proprioceptive forms of feedback in the implementation of the articulatory program; from his/her perceptual ability to use auditory forms of information to detect auditory differences. Therefore, information about the speaker is hidden in the speech signal, is correlated with his/her anatomical features and is stored at the neuronal level by the muscular speech patterns correlating with the speaker's physique [2, 3, 8].

4   Preliminary Results of the Investigation

When developing expert methods for speaker profiling by speech on the Internet, the following conditions for the speech signal realization are taken into account: speech

should be natural and be varied as much as possible relative to the speakers (interspeaker discrepancies), but rather homogeneous relative to each speaker (intraspeaker discrepancies); at the initial stage of development, the speech should not be influenced by noise, interference, etc., and should include special characteristics of transmission along the technical path; no distortion of the voice is allowed [20]. Particularly informative for speaker attribution by speech is the range of the pitch frequency (ΔF_0), which includes, first of all, such parameters as the pitch frequency range width (ΔF_0) and its register (very high, high, medium, lower medium, low, very low), which correlates with the following individual characteristics of the speaker: biological differentiation by gender, age, physique; and psychological differences in the speaker's behavior; idiosyncratic (individual) features at the biological, psychological and regional-social levels [6, 8, 10, 14, 17, 18, 21].

Individual features of the speaker are traditionally divided into two groups: acquired and non-acquired. Acquired features include such specific speech features that are formed under the influence of the external conditions of the speaker's life. Among the latter is primarily the process of language acquisition, and then its application in spoken and written communication. In this case, a special role is played by the dialect used by the immediate environment of the individual, especially when, during the phase of speech acquisition, which corresponds approximately to the time of schooling (age up to 18), the speaker lived in various dialectal societies. This includes the social conditions that define the so-called sociolect. The acquired features also include speech features resulting from various harmful factors, for example, smoking, alcohol and drug intoxication [1, 19]. Non-acquired features are correlated with organic-genetic data based on the anatomical and neurophysiological components of the speech apparatus. The latter include the size and spatial configuration (the so-called cavitary configuration) of the neck-laryngeal, nasal and pharyngeal tracts, the mobility and size of the tongue, and in particular the number of boundary conditions depending on the voice formation (the term of mathematics), as well as age and gender.

The pitch frequency can vary depending on such factors as loud speaking (for example, in a state of excitement, in noisy conditions (Lombard effect), etc.) In these cases, the pitch frequency changes upwards, and this should be taken into account when describing the speaker. At certain stages of mental illness, the voice can be not only lower, but also much more monotonous (for example, in a state of depression in manic-depressive patients). In speaker attribution by voice, along with the above characteristics, of great importance is information on the voice quality. In this case, features specific to the speaker are found. First of all, one should mention such a qualitative attribute as hoarseness. Here most informative is not this feature in itself, but rather its distribution in the speech flow: this phenomenon can occur where the voice for purely linguistic reasons is lowered, i.e. at the end of sentences and other syntactic or semantic units. In a number of speakers with low voices or voice pathology (for example, due to inflammation of the larynx, a tumor or nodes in the larynx, etc.), this symptom may appear in various other positions of the speech flow [8]. In the speaker attribution process, the rate of speech formation is also informative. The average speech rate for all languages is about 4.5–5 syllables per second. Extreme values are 3, 2–7, 5 syllables. Higher rate

leads to incomplete articulation or complete loss of sounds, syllables and even whole words.

As an example, the following requirements should be given that characterize the speaker's portrait by voice and speech: physical: gender, age, height, weight; civil status: parents, their mother tongue, origin, social status, etc.; linguistic: native/non-native, literary/non-literary, regional/dialectal language; educational: length of study (primary, secondary, higher, etc.); geographical: place of long-term residence (if there are some, then indicate periods of residence); professional: work by profession/not by profession; auditory: state of hearing, presence/absence of pathology; medical: chronic/non-chronic diseases; voice: trained, singing, smoking, stressful, etc. voice; musical: musical information, etc.; hobby: sports profile, musical profile, etc. Thus, in the SND the number of communicants' characteristics determined by acoustic data in IP telephony, can include the following characteristics: social: by level of education; social status; sphere of activity; physical characteristics; emotional characteristics; regional characteristics: place of birth; place of long residence; nationality; additional information; psychological characteristics: mental pathologies; HNA type; character traits; types of intoxication (alcohol, drug); pronouncing characteristics: spontaneous speech; quasi-spontaneous speech; prepared/unprepared reading of text; emotional characteristics: positive, negative, etc. [1, 3, 4, 6, 9, 10, 15, 19, 21, 22].

As an example of personality profiling on the Internet is presented a speaker voice sample recorded hourly and analyzed by expert listeners on the basis of special instructions. It was found that the dynamics of prosodic features (pitch, tempo, and loudness variations) is a reliable diagnostic tool of acmeologic speaker profiling with regard to emotional state changes of this personality from a normal emotional state to agitation, anger, fury, etc. This experiment deals with speakers' emotional state acmeologic profiling during hourly communication on the Internet by means of perceptual-auditory analysis (step by step every ten minutes). The sentences used in the experiment were taken from speech communication dialogues on the Internet: a group of communicants (n = 10), male voices, speakers of 18–25 years old; a group of professional listeners (n = 10). The sentences were taken from a pre-election campaign debate in Russia on the Internet. The listeners were asked to evaluate the pitch, tempo, and loudness dynamics of all voice stimuli. The responses across all the stimuli examples are summarized as mean data. The Figs. 1, 2 and 3 show the dynamics of mean pitch, loudness, and speech rate data during one-hour recording. The experiment involved the perceptual-auditory evaluation of such voice features of the listener as: pitch (very low, low, lower medium, medium, high, very high); speech rate (very slow/slower, slow, slowed, moderate, fast, very fast); loudness (subaudible, very slow, low, middle loud, loud, very loud). All conversations data were recorded, copied to a CD and sent to listeners with instructions to define pitch, speech rate, and loudness characteristics of every subject.
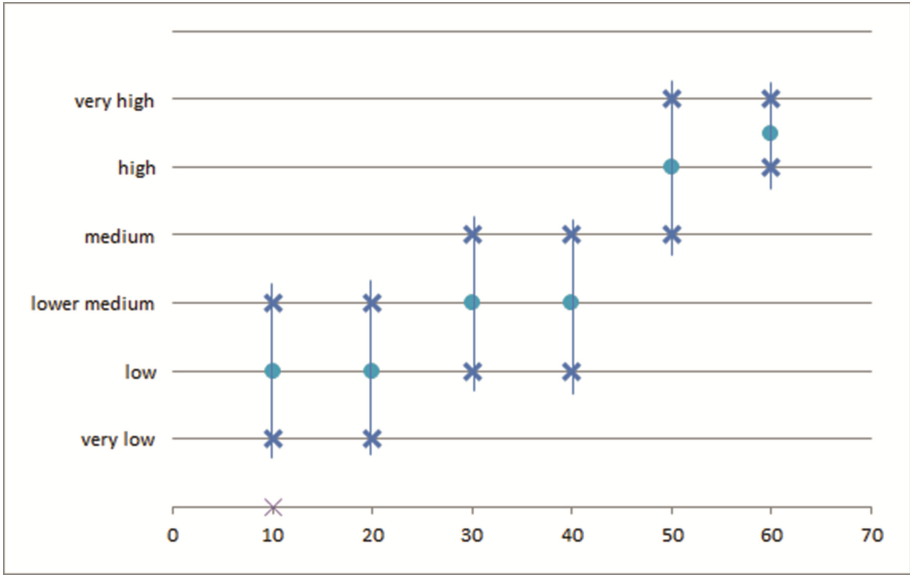
Fig. 1. Mean perceptual-auditory data of the pitch evaluation and data scope zones ($\omega = F_{0max} - F_{0min}$) regarding hourly dynamics of the speech characterization features (in ten-minute steps).
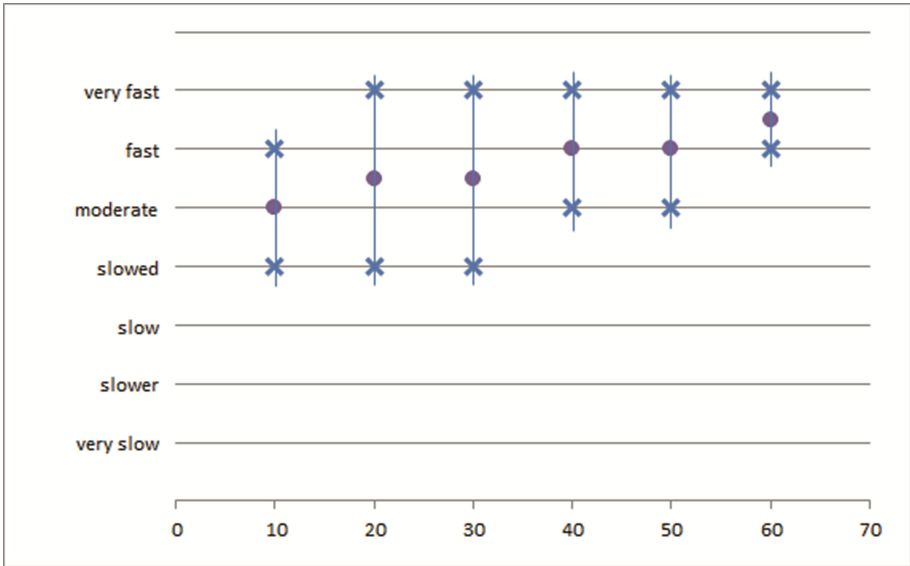
Fig. 2. Mean perceptual-auditory data of the speech rate evaluation and data scope zones ($\omega = t_{max} - t_{min}$) regarding hourly dynamics of the speech characterization features (in ten-minute steps).
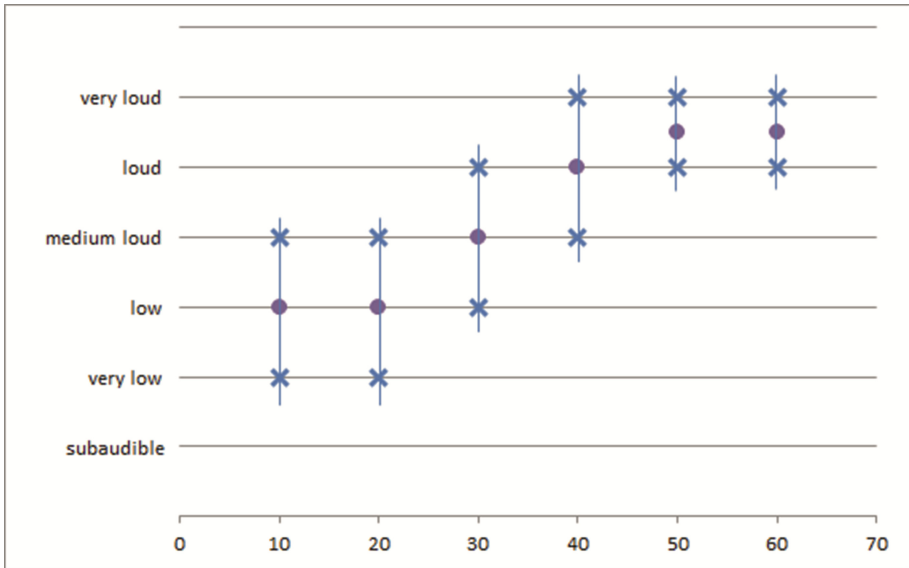
Fig. 3. Mean perceptual-auditory data of the loudness evaluation and data scope zones ($\omega = I_{max} - I_{min}$) regarding hourly dynamics of the speech characterization features (in ten-minute steps).

It can be concluded that every ten minutes acoustic stimuli had enough information available to draw distinction between those with no aggressive behavior dynamics of speech and those with some ones. It is known that pitch, intensity, and speaking rate are affected, e.g., by aggressive emotions and perceptual-auditory corresponding features. The speech acoustic correlations of this aggressive behavior invoke some challenges in defining the real emotional state. But the acoustic data measurements could not be always reliably interpreted with regard to acoustical speech signals. The optimization of the experimental methods lies in looking at combination of perceptual-auditory and acoustic analysis on the basis of fundamental sciences in the field of interdisciplinary speech research, regarding acmeologic personality profiling on the Internet. As example for emotional personality profiling vector, a speaker voice sample is presented which was recorded during one hour and analyzed by expert listeners on the basis of special instructions. It was found that the dynamics of prosodic features (pitch, speech rate, and loudness variations) is a robust diagnostic tool of acmeologic speaker profiling with regard to emotional state changes of the personality regarding changes in personal emotional characteristics with connection to psychological, social, physical, etc. factors.

5   Conclusion

Thus, the study of the speech variability process on the Internet is a task of immense complexity connected, on the one hand, with the articulatory-acoustic specifics of spoken speech and its perceptual auditory and acoustic characteristics, on the other hand,

with the specifics of constructing any utterance taking into account the prosodic-semantic variability of the speech product itself. At the same time, the transmission of a high-quality speech signal with regard to IP telephony (more precisely, Voiceover IP (VoIP), etc.), due to the specificity of encoding, compression and packaging of the speech signal into IP packets, may be a kind of obstacle for the successful solution of the task [5]. An analog voice signal digitized by the PCM method and compressed by codecs to eliminate redundancy, undoubtedly undergoes certain changes at the output. As the results of the preliminary study [20] have shown, the prospect of using special software for establishing acoustic and perceptual-auditory equivalence with some degree of probability is quite promising in solving the problems of an "electronic personality" profiling in the Internet information and communication environment. The above-described correlations between the SND-characteristics of the speakers on the Internet and his/her speech reactions on the communication stimuli in Internet dynamics make it possible to undertake further research in the field of the acmeologic personality profiling on the Internet and other speech communication transmission devices.

References

1. Braun, A.: Sprechstimmlage und regionale Umgangssprache. In: Braun A. (ed.). Beitraege zu Linguistik und Phonetik. Festschrift fuer Ioachim Goeschel zum 70 Geburtstag, pp. 453–463. Stuttgart (2001)
2. Brown, R.: Auditory Speaker Recognition. Helmut Buske Verlag, Hamburg (1987)
3. Brown, W.S., Morris, R., Hollien, H., Howell, H.F.: Speaking fundamental frequency characteristics as a function of age and professional singing. J. Voice **3**, 310–313 (1991)
4. Kuenzel, H.J.: How well does average fundamental frequency correlate with speaker height and weight? Phonetica **46**, 117–125 (1989)
5. Mikhailov, V.G.: Features of the formation and analysis of voice signals transmitted by means of IP-telephony. Theor. Pract. Forensic Exam. **3**(7), 129–140 (2007). (in Russia)
6. Oksaar, E.: Idiolekt als Grundlage der variationsorientierten Linguistik. Sociolinguistica **14**, 37–41 (2000)
7. Potapova, R.K.: Linguistic and paralinguistic functions of prosody (On the experience of searching for prosodo-semanthema). In: Kedrova, G.E., Potapov, V.V. (eds.) Language and Speech: Problems and Solutions, pp. 117–137. MAKS Press, Moscow (2004) (in Russia)
8. Potapova, R.K.: Some observations on artificially modified speech. In: Ideas and Methods of Experimental Speech Study. I.P. Pavlov Institute of Physiology (Russian Academy of Sciences), State University, Sankt-Petersburg, pp. 124–135 (2008). (in Russia)
9. Potapova, R.K.: Speech: Communication, Information, Cybernetics, 4th edn. Book house "Librocom", Moscow (2015). (in Russia)
10. Potapova, R.K., Potapov, V.V.: Language, Speech, Personality. Languages of Slavic Culture, Moscow (2006). (in Russia)
11. Potapova, R., Potapov, V.: Kommunikative Sprechtaetigkeit. Russland und Deutschland im Vergleich. Boehlau Verlag, Koeln (2011)
12. Potapova, R.K., Potapov, V.V.: Speech Communication: From Sound to Utterance. Languages of Slavic Cultures, Moscow (2012). (in Russia)

13. 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). https://doi.org/10.1007/978-3-319-01931-4_9

14. 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). https://doi.org/10.1007/978-3-319-23132-7_19

15. 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). https://doi.org/10.1007/978-3-319-43958-7_61

16. 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). https://doi.org/10.1007/978-3-319-43958-7_65

17. Potapova, R., Potapov, V.: Cognitive entropy in the perceptual-auditory evaluation of emotional modal states of foreign language communication partner. In: Karpov, A., Potapova, R., Mporas, I. (eds.) SPECOM 2017. LNCS (LNAI), vol. 10458, pp. 253–261. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66429-3_24

18. Potapova, R., Potapov, V.: Human as acmeologic entity in social network discourse (multidimensional approach). In: Karpov, A., Potapova, R., Mporas, I. (eds.) SPECOM 2017. LNCS (LNAI), vol. 10458, pp. 407–416. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66429-3_40

19. Potapova, R.K., Potapov, V.V., Lebedeva, N.N., Agibalova, T.V.: Interdisciplinarity in the Study of Speech Polyinformativity. Languages of Slavic Culture, Moscow (2015). (in Russian)

20. Potapova, R., Sobakin, A., Maslov, A.: On the possibility of the skype channel speaker identification (on the basis of acoustic parameters). In: Ronzhin, A., Potapova, R., Delic, V. (eds.) SPECOM 2014. LNCS (LNAI), vol. 8773, pp. 329–336. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11581-8_41

21. Ryan, E.B., Capadano, I., Harry, L.: Age perceptions and evaluative reactions toward adult speakers. J. Gerontol. **33**, 98–102 (1978)

22. Sharp, D.: Personality Types: Jung's Model of Typology. Inner City Books, Toronto (1987)

23. Titscher, S., Meyer, M., Vetter, E., Wodak, R.: Methods of Text and Discourse Analysis. Sage, London (2000)