Moscow State Linguistic University

Institute for Applied and Mathematical Linguistics

Rodmonga Potapova
PhD, ScD, Prof.

Vsevolod Potapov
PhD, ScD, Prof.

On Individual Polyinformativity of Speech and Voice Regarding Speaker's Auditive Attribution

(Forensic Phonetic Aspect)

Budapest 2016

1. Introduction

Subjective (auditive) and objective (acoustic) events of speech behavior are of great interest for several professional groups. In the science of forensic phonetics investigations in the field of acoustic and auditory features of voice and spoken language of subjects under the intoxication by drugs are very important.

The aim of the present study is to obtain new auditive data in order to expand the knowledge of a variety of **personal characteristics of speech** of Russian native speakers. The term “personal characteristics of speech” expresses the well-known fact that speakers can be distinguished and recognized by their voices and speech.

Personal characteristics of speech may be described as a complex of those sound qualities which enable us to identify the speaker. Our research is focused on the problem of auditive definition of speech qualities resulting from the psychic or psychosomatic conditions of Russian-speaking individuals. It should be noted that in our investigations we use scientific principles of auditory analysis in the field of forensic phonetics [6, 7; 10: 81–138; 14, 18].

Perceptual analysis aimed at the determination of a set of **perceptual cues relevant to the description of the peculiarities of voice, segmental and suprasegmental characteristics of speech** of native Russian speaker introduced by the alterations of emotional, psychic, psychosomatic and physiological state [1, 2, 3, 4, 5].

Voice-based evidence is an important part of many criminal investigations and has commonly included such things as threats left on an answering machine, a robbery caught on videotape, or a confession recorded during a police interrogation.

In the technological age of mobile telephones, voicemail, and voice-recognition software applications, the potential for voice-based evidence continues to increase, on the domain of personality identification and attribution in the communication by means of Skype, You Tube, com. and in the case of telephone terrorism, Internet pranker communication, etc.

2. Method, Experiment, Results

The speech signal therefore contains at least two kinds of information. As a linguistic signal, it conveys the communicative content of the utterance.

The speech signal also conveys information about certain features of the speaker, such as his sex, age, regional origin, etc. An important distinction may be drawn here between whether these kinds of information are intentionally introduced or not. Laver [8, 9, 12] proposes a classification of the **different kinds of indexical information** present in speech: biological information (size, physique, sex, age and medical state); psychological information (personality); social information (mainly accent information of regional origin, social status, etc.).

The main reason for the specifics of the speech signal may be neurophysiological and psychological features of the phonation and articulation process, the implementation of which is controlled by the speaker’s central and autonomic nervous systems [14, 17]. It is important to distinguish between two types of speech signal variability:

- ***inter*individual** variability due to individual anatomical and physiological, psychological and social characteristics of speakers which is the basis of individually-significant attributes;
- ***intra*individual** variability caused by a number of non-semantic factors and expressed in spontaneous variation of voice and speech, even within an unchanged speech segment according to various uncontrollable factors related to multi-component vocal apparatus functioning.

Speech is both a **mechanism of intellectual activity**, which allows to perform operations of abstraction and generalization that provides the basis of categorical thinking, and a **mechanism of semantic programming** enabling the transition from the semantic level to the syntactic level with the help of psycho-physiological mechanism called “internal (implicit) speech” [12].

Human speech is characterized by an operating component, the first element of which is **physical or sound matter**, the analysis of which allows to determine the **relationship between individual voice production with an invariant and variants of sound and intonation patterns** on the basis of a specific language.

The next link of the operating component of the speech process is a lexical-semantic organization of verbal material including implementation of the lexical-morphological code of the language that converts images and concepts to their verbal forms.

The above determines the conceptual basis of the successful development of attributes and identification of the speaker in forensic phonetics [11,12,15,16,17,19].

Based on the premise that human speech is individually organized on the basis of individual phonational and articulatory gestures in close connection with the phonological representation of an utterance and its lexical and semantic features, it seems reasonable to build an acoustic-linguistic algorithm of the speaker identification analysis taking into account the following factors:

- acoustic (hardware and software) processing of the speech signal;
- anatomical and physiological-based decoding of the speech signal;
- social- and psychological-based decoding of the speech signal;
- intellectual and semantic decoding of the speech signal;
- tiered global linguistic decoding.

In this regard, all solvable problems can be roughly described as tasks of drawing up an individual “portrait” of the speaker, which includes phonational (voice), articulatory (segmental) and prosodic (suprasegmental) correlates of his/her speech. The basis of the acoustic-linguistic analysis are iterative speech wave processing procedures.

It seems reasonable to divide acoustic and linguistic features of the speech signal into primary and secondary ones.

The primary ones include:

- **phonational features** (typology of voice mimics, such as forced or gentle phonation / with correlation of the speech signal analyzed to one of the phonation types;
- **articulatory features** (articulatory typology of generating speech signal (e.g., tense or relaxed articulation) with correlation of the speech signal to one of the articulation types).

Primary features are directly dependent on the specific **anatomical and physiological nature**.

Secondary (prosodic) features are of conditionally superstructural character with respect to the primary ones and are implemented on their basis.

Suprasegment implementation of secondary features of the speech signal leads to formation of a kind of structurally-organized speech figures and their **concatenation of strictly individual character**.

According to recent data, voice features that characterize the speaker as well as the specificity of his/her individual character formation (i.e. **idiosyncrasy**) contain two types of information: **communicative** and **individual** one.

As a linguistic (verbal) signal, speech includes communicative content of a message, and as an extralinguistic (non-verbal) signal, it correlates with the information about such speaker’s features as gender, age, region of origin, etc.

In portrait attributes of the speaker by voice and speech there are three types of norms: **universal**, **group** and **idiosyncratic** ones.

A special role belongs to speech and voice information decoded at the level of auditory perception [18].

The purpose of our experiment was to identify key features of the perceptual-auditory perception of speech necessary and sufficient to answer the question: what individual features of the speaker may be used by an expert making up a “portrait” of the speaker. In addition, it was necessary to answer the question whether the information content of features was identical to establish the speaker’s “profile”.

Special questionnaires [13, 14, 18, 19, 20] were used for the experiment. Listeners were asked to listen to some phonograms and then record their answers in a special questionnaire.

The material (phonograms for each speaker) was played repeatedly. There were no restrictions in time and number of plays.

The listeners were to note those features in their questionnaires, which, in their view, matched the “profile” the speaker. The listeners of the experiment were represented by 4 groups of listeners: 2 groups of experts who were people from various business dimensions and had fundamental knowledge in the field of speaker identification by voice and speech (**n = 21**); and non-experts – students of Moscow State Linguistic University (**n=45**).

The subjects belonged to various age, gender, social and territorial groups.

One of the hypotheses put forward to test the empirical data, was the assumption according to which the subjects (in this case, the listeners) possess different levels of language competence and skills of listening, which affects the final results of the perceptual-auditory analysis.

Along with listening to the phonograms presented, the listeners were to fill special questionnaires giving their opinion on speaker's profile characteristics.

Speakers were represented by:

- males and females *(the subjects also had an option "transvestite" in their questionnaires)*;
- people of various social groups *(high school students, teachers, politicians, etc.)*;
- representatives of various age groups;
- representatives of regional groups *(residents of various regions in Russia)*.

Speakers’ speech was recorded under various conditions: physical (various rooms with varying degrees of noise insulation) and communication (radio interviews, spontaneous speech, lectures, reports, phrases from polylogues, dialogues, etc.).

According to the procedure of the experiment, the listeners had no information on speakers in advance. They were to fill in the questionnaire while listening, focusing solely on their auditory impressions.

The listeners were asked to analyze the acoustic part of the sounding material (expression plan) rather than specific semantic content of speech fragments (content plan).

Data obtained as a result of the perceptual-auditory experiment were analyzed and statistically processed. For each of the 4 groups of speaker’s personal characteristics, namely speaker’s phonetic characteristics, language characteristics; physiological and anthropometric characteristics of the speaker’s appearance, his/her physical and emotional state, tables were drawn containing the results of the perceptual-auditory analysis. Thus, for each group 2 tables were drawn showing the number of listeners’ reactions to the presence/absence of a characteristic proposed in the questionnaire (as well as the parameter of this characteristic, for example, voice pitch – medium) in absolute and relative units (%).

Further evaluation of the results obtained was carried out by two vectors: vertical vector – for perceptual-auditory definition of interspeaker features (and parameters) essential for each speaker separately; horizontal vector – for classification of parameters singled out for each speaker (intraspeaker section) on the basis of the statistical weight of each parameter. The “horizontal vector” gives an insight into the intraspeaker mechanism of perception by speech.

Classification of parameters within each feature is based on statistical weights (***W***, %) attributable to each parameter according to the following formula:

$$\boldsymbol{W = \frac{a \ * 100}{A}}$$

where $\boldsymbol{a}$ is the number of positive responses of the subjects received for a specific parameter for all speakers (i.e. how many times the listeners noted this parameter during the experiment), $\boldsymbol{A}$ is the total number of positive responses regarding specific features for the entire group of speakers.

Next, each parameter was assigned a rank value it takes with respect to the appropriate feature.

The parameters that are well perceived by the subjects by ear (most listeners noted their presence) have, respectively, a greater statistical weight (***W***) and, as a consequence, a higher rank.

The empirical evidence also showed that the features do have various weights and various significance for the completion of the task, that is drawing up the speaker’s “portrait”. In each of the 4 groups under consideration, the characteristics were assigned ranks according to their statistical weights in the group.

This ranking of the features can be interpreted as follows: the higher rank is assigned to a particular feature of any group of characteristics (phonetic, linguistic, physiological and anthropometric or physical and emotional characteristics), the more accessible and more important it is for the expert studying speaker’s characteristics.

This classification can be perceived as a kind of guide for an audio expert indicating which attributes of the speaker should be considered and analyzed in the first place, what indicators are reliable and meaningful to perform such a task as drawing up the speaker’s “portrait”.

It is seen from the obtained data that the listeners best perceived the following characteristics:

- generation in the process of speech breathing, strength of voice and specific features of pronunciation;
- temporal peculiarities, melodic patterns, distinguishing stressed and unstressed syllables, speech rhythm;
- language (native/foreign), language (standard vs. dialect) and a communicative act specificity (group of verbal features);
- gender, age and size of the speaker’s head (physiological and anthropometric features);
- physical state of a speaker (group of features that describe the speaker’s physical and emotional state).

The following features were **most difficult** for auditive speaker attribution:

- **voice timbre and strength** (group of phonetic features);
- **type of speech activity, functional style and language** (in opposition to the standard – vernacular; group of linguistic features);
- **speaker's height, weight, age and hair color, width of his/her chest** (physiological and anthropometric attributes);
- **defects in speech and pronunciation**;
- **emotional and emotional-modal state** (group of features that describe the speaker's emotional state).

3. Conclusion

Conclusions regarding the speaker’s attributes, which can be made in an intraspeaker (“**horizontal**”) analysis, have the greatest practical value.

The resulting information can be used particularly in forensic purposes in solving diagnostic tasks.

The purpose of this type of analysis is to identify a common mechanism of formation of listeners’ interpretation of the speaker (author of a spoken text) image.

The statistical analysis is used to determine which speaker’s attributes are perceived by the listeners, what personality characteristics are difficult to determine by ear, and what parameters are perceived by the listeners equally, etc.

With the “**vertical**” vector of data obtained in the course of the experiment, interspeaker identification features are determined.

Statistics show that it is possible to single out key features for each speaker who took part in the experiment.

For classification of parameters, the ranking method was used again. To this end, each parameter was assigned a numerical index, which reflects the number of positive responses to the presence of this parameter for each speaker (in %). Dominating parameters build the speaker’s “profile”.

Next, to assess the relevant parameters in the speaker’s personality profile, a sample of the maximum values (the highest values of the parameters) has been split into three intervals:

- [0–50] – parameters with values within this range, may be declared as those least perceived by the listeners. A small value of the parameter means that either the listeners are not able to auditive define what value of a speech parameter may be attributed to the speaker, or these characteristics are not strongly marked, which prevents the listeners from determining whether the speaker has this feature;
- [50–75] – parameters with values within this range are more perceptually significant. However, they still cannot be considered as “basic points” in a forensic expertise.

- [75–100] – parameters with values within this range are most perceptible. Considering the fact that most of the most listeners’ answers are identical on this particular subject, the speaker will likely be attributed these values of features.

Moreover, high values of the parameter indices of this interval indicate that the respective features may be included in a “palette” of an expert creating the speaker’s “portrait”. Next, values were considered from the upper range [75-100] including features whose values are both statistically and perceptually marked. It can be assumed that the parameters within of this interval, provide specific information about the speaker’s identity, and therefore they should be taken into account when drawing up his/her “portrait”.

Bar charts were built for these features (Fig. 1-4). Features that are included in the top interval are distributed along the X-axis; and the number of speakers whose “portrait” has this particular feature is shown on the Y-axis. Thus, according to the figures 1–4 the features attributed to most speakers (**n=15**) include:

- *pauses, speech rhythm, speech breathing* and *distinguishing stressed/unstressed syllables* (phonetic features);
- *language* (native vs. foreign), *communicative act, language* (standard vs. dialect), *education and economic status* (linguistic features);
- speaker’s gender and size of his/her head (physiological and anthropometric features);
- speaker’s overall condition (features of speaker’s physical and emotional state).

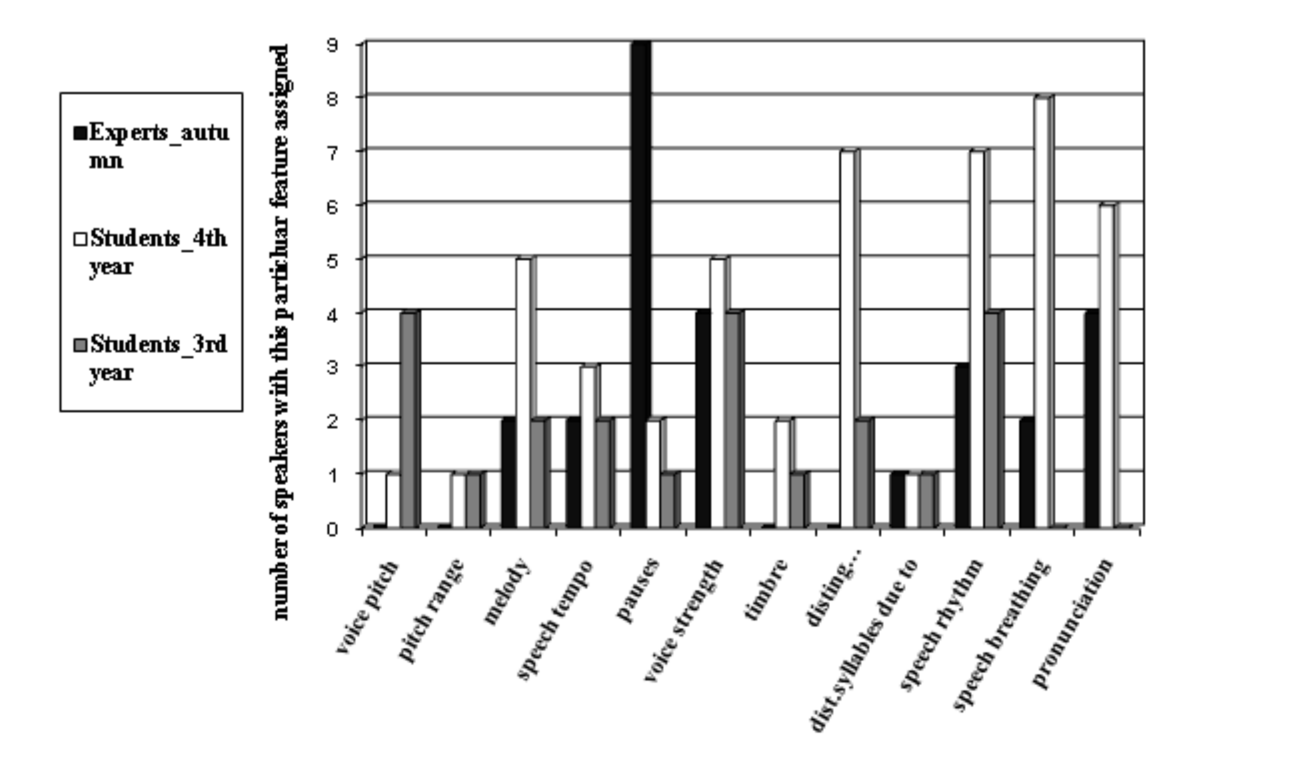


Fig. 1. Speaker’s phonetic features perceived by the listeners and included in the optimal range of values

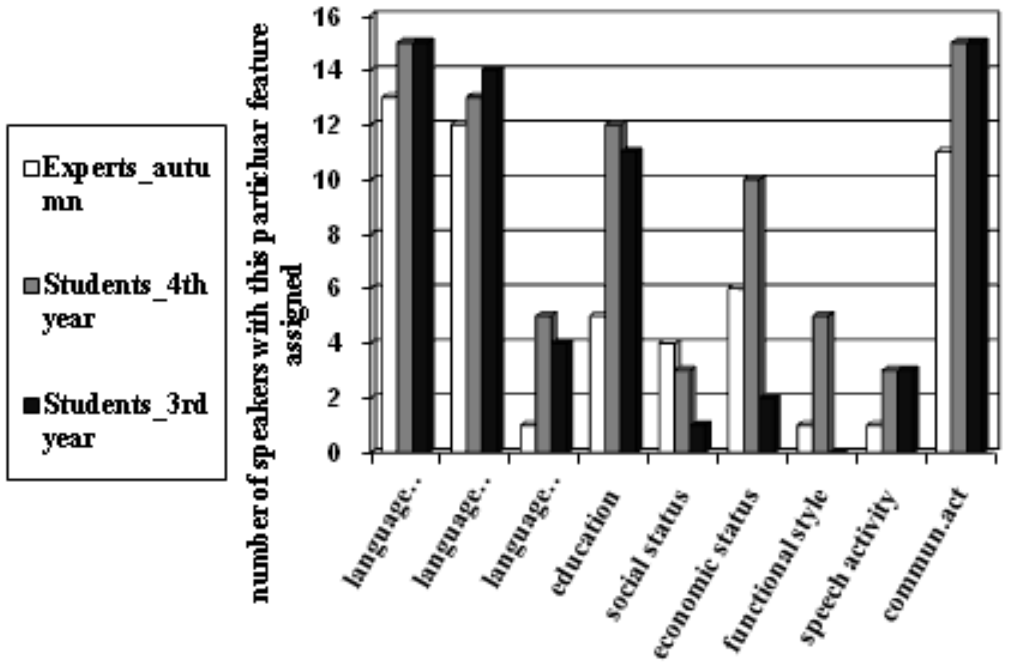


Fig. 2. Speaker's language features perceived by the listeners and included in the optimal range of values

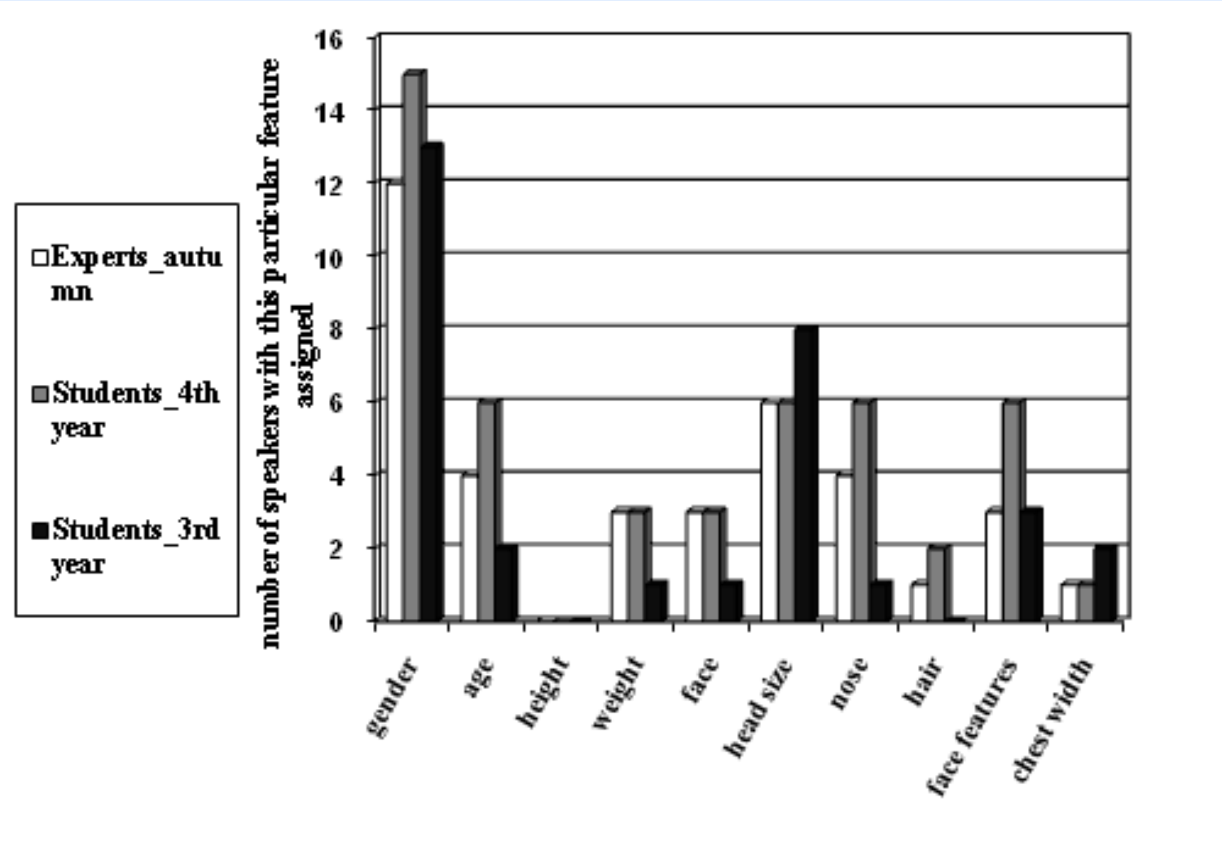


Fig. 3. Speaker’s physiological and anthropometric features perceived by the listeners on the basis of the speech material and included in the optimal range of values

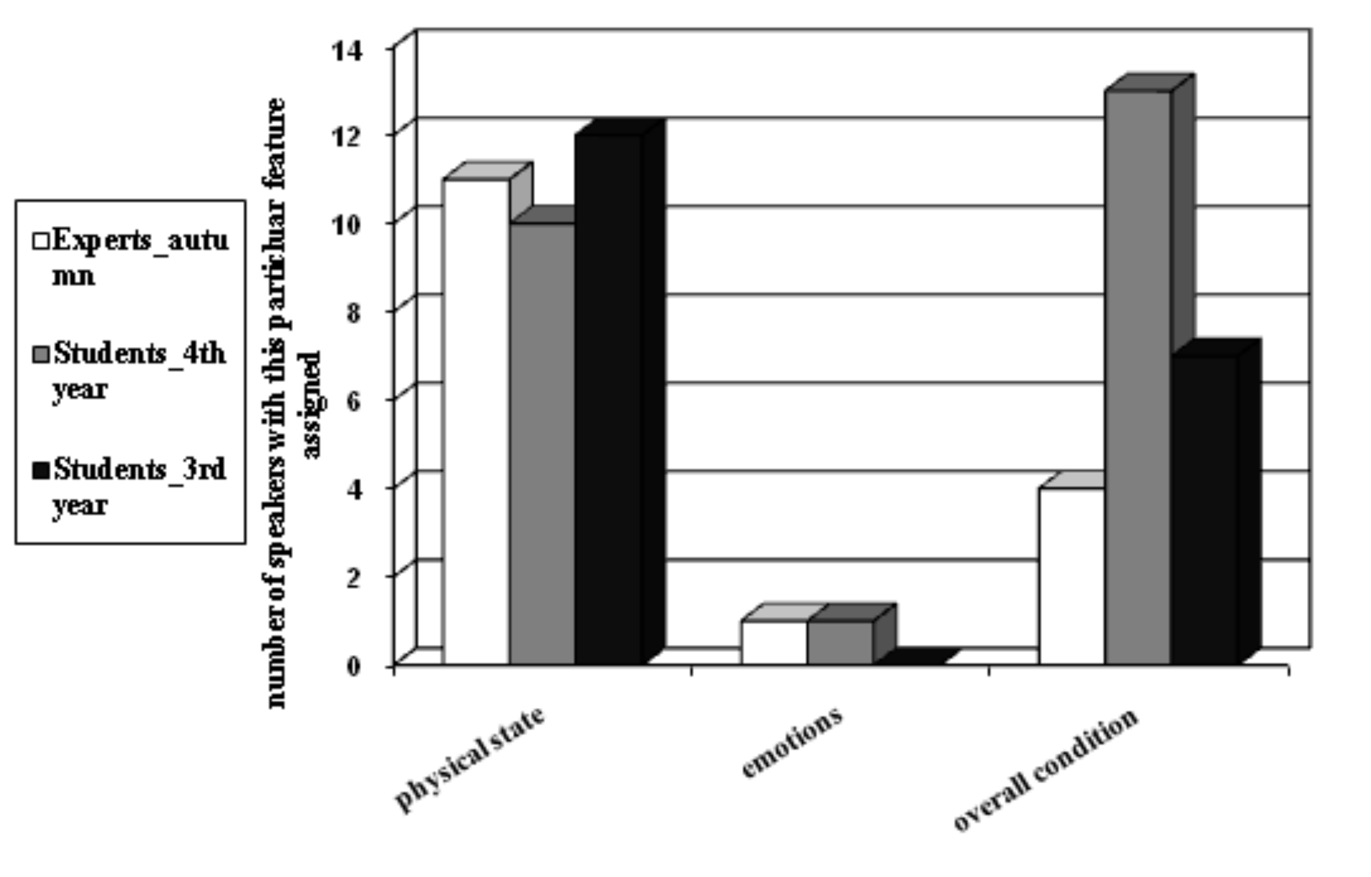


Fig. 4. Features of the speaker's physical and emotional state perceived by the listeners and included in the optimal range of values

Features that are “assigned” by the subjects to a minority of speakers (**n = 1–6**) are either purely individual and make the speaker’s voice and speech “exclusive” (these characteristics distinguish the speaker’s voice from all the others), or are difficult to determine by listeners. To answer these questions it is necessary to conduct an additional series of experiments to increase the numbers of both speakers and listeners.

Acknowledgments. This research is supported by Russian Science Foundation, Project N14-18-01059 (2014-2016).

References

1. Brown, R.: Auditory speaker recognition. Helmut Buske Verlag, Hamburg (1987)

2. Buzik, O.Z, Rychkova, O.V, Agibalova, T.V, Gurevich, G.L, Shalaeva, E.V, Potapova, R.K.: Emotional and cognitive disturbances in addictions: interactions and correlations. Zhurnal nevrologii i psikhiatrii imeni S.S. Korsakova, 79–83 (2014) (in Russian)

3. Fazakis, N., Karlos, S., Kotsiantis, S., Sgarbas, K. Speaker identification using semi-supervised learning. In: Ronzhin, A., Potapova, R., Fakotakis, N. (eds.) SPECOM 2015. LNCS, vol. 9319, pp. 389–396. Springer, Heidelberg (2015)

4. French, P.: An overview of forensic phonetics with particular reference to speaker identification. Forensic linguistics. International journal of speech, language and the law 1 (2), 169–181 (1994)

5. Hollien, H.: Forensic voice identification. Academic Press, London (UK), San Diego (California) (2002)

6. Jessen, M.: Phonetische und linguistische Prinzipien des forensischen Stimmenvergleichs. Lincom, Muenchen (2012)

7. Kuenzel, H.J.: Sprechererkennung. Kriminalistik Verlag, Heidelberg (1987)

8. Laver, J.: The phonetic description of voice quality. Cambridge University Press, Cambridge (1980)

9. Laver, J.: Voice quality and indexical information. British Journal of Disorders of Communication 3, 43–54 (1968)

10. Polzehl, T.: Personality in speech: assessment and automatic classification. T-Labs Series in Telecommunication Services. Springer, Heidelberg (2015)

11. Potapova, R.K.: Speech: communication, information, cybernetics. 5th ed. URSS, Moscow (2015) (in Russian)

12. Potapova, R.K.: The subject-oriented perception of foreign speech. Voprosy jazykoznanija 2, 46–64 (2005) (in Russian)

13. Potapova, R., Potapov, V.: Associative mechanism of foreign spoken language perception (forensic phonetic aspect). In: Ronzhin, A., Potapova, R., Delić, V. (eds.) SPECOM 2014. LNCS, vol. 8773, pp. 113–122. Springer, Heidelberg (2014)

14. 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, vol. 8113, pp. 62–69. Springer, Heidelberg (2013)

15. Potapova, R.K., Potapov, V.V.: Kommunikative Sprechtaetigkeit: Russland und Deutschland im Vergleich. Boehlau Verlag, Koeln; Weimar; Wien (2011)

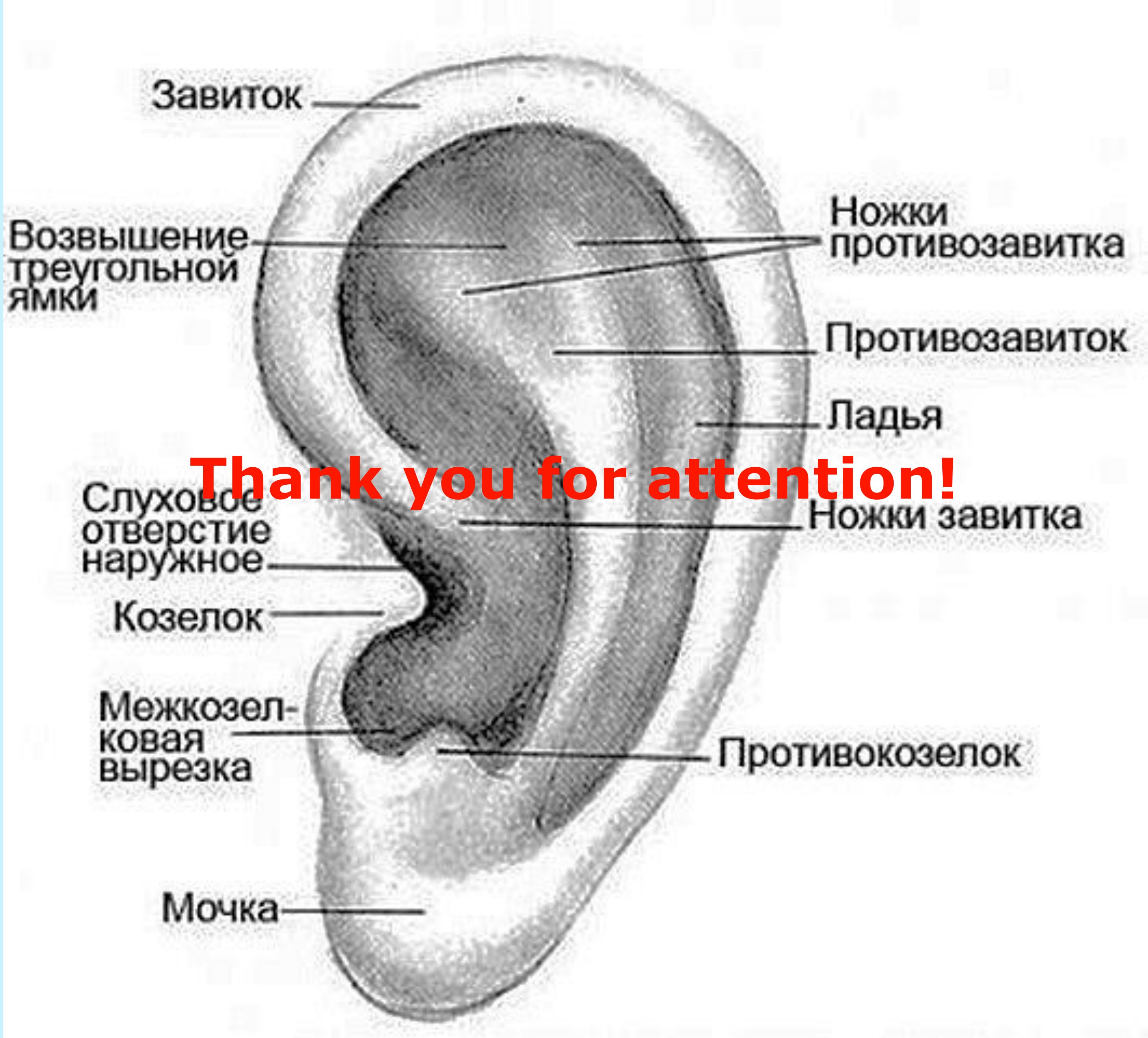
16. Potapova, R.K., Potapov, V.V.: Language, speech, personality. Publishing House "Languages of Slavic Cultures", Moscow (2006) (in Russian)

17. Potapova, R.K., Potapov, V.V.: On the correlation between attribute characteristics of a speaker and the speech signal. In: Proc. of the XVI International Scientific Conference «Informatization and information security of low and order bodies», pp. 330–336. Moscow, RF (2007) (in Russian)

18. Potapova, R.K., Potapov, V.V.: Speech communication: from sound to utterance. Publishing House "Languages of Slavic Cultures", Moscow (2012) (in Russian)

19. Potapova, R.K., Potapov, V.V.: Spoken language as an object of fundamental and applied linguistic investigation. In: Annual report of the Russian acoustical society "Speech acoustics and applied linguistics", pp. 6–28. Moscow (2002) (in Russian)

20. Potapova, R.K., Potapov, V.V., Lebedeva, N.N., Agibalova, T.V.: Interdisciplinarity in the investigation of speech polyinformativeness. Publishing House "Languages of Slavic Cultures", Moscow (2015) (in Russian)



Thank you for attention!