The choice of decisions at recognition of emotions on the speech

Victor P. Kalyan, FITs IU of RAS, Moscow
RUDN University, Moscow, kalyan@mail.ru

Anastasia V. Kalyan, RUDN University, Moscow,
nastya-kalyan@yandex.ru

Intelligent Data Processing: Theory and Applications
Barcelona 2016

Background

More than 40 years in the market of services in detection of a lie there are commercial devices positioning themselves as "stress analyzers in a voice".

It was claimed that these devices unlike a polygraph are capable to establish insincerity without connection to a body of the person of sensors, and by measurement of the changes in a voice caused by a stress which accompanies false statements.

The USA which is carried out by Institute of a polygraph the Ministry of Defence has shown the independent researches conducted by experts-polygraphologists, researches of the American association of a polygraph (MACAW), and also tests of the devices existing in the market that the accuracy of these devices is at the level of casual guessing.

Methods

Our work experience of the choice of decisions in system of **recognition of an emotional condition** of the person on the speech concerning **truthfulness and sincerity of speaking** is described. Informational content of measuring base of recognition on the basis of **paralinguistic**, **articulation** and **extralinguistic features of the speech** taking into account individual emotional and semantic connotations of the examinee is analyzed, algorithms of recognition of emotions according to the speech are described, the choice from a set of decisions and their verification concerning sincerity and truthfulness speaking taking into account a situational context is carried out.

That it is necessary for us:

For creation of really operating system of recognition of the emotional state speaking according to the speech the measuring base and an algorithmic basis of system of recognition of emotions according to the speech have to be worked thoroughly out.

The problem of automatic recognition of an emotional coloring of the speech is cross-disciplinary and constantly involves researchers of different specialties – not only linguists, but also mathematicians, programmers, psychologists, physiologists.

Directions of researches:

1. Modality of emotions. This traditional direction of works of psychologists on studying and classification of emotions, to identification of emotional and semantic connotations.

2. Finding of objective characteristics of manifestation of emotions in the speech, communications of emotions with paralinguistic, extralinguistic and articulation features of the speech. One of

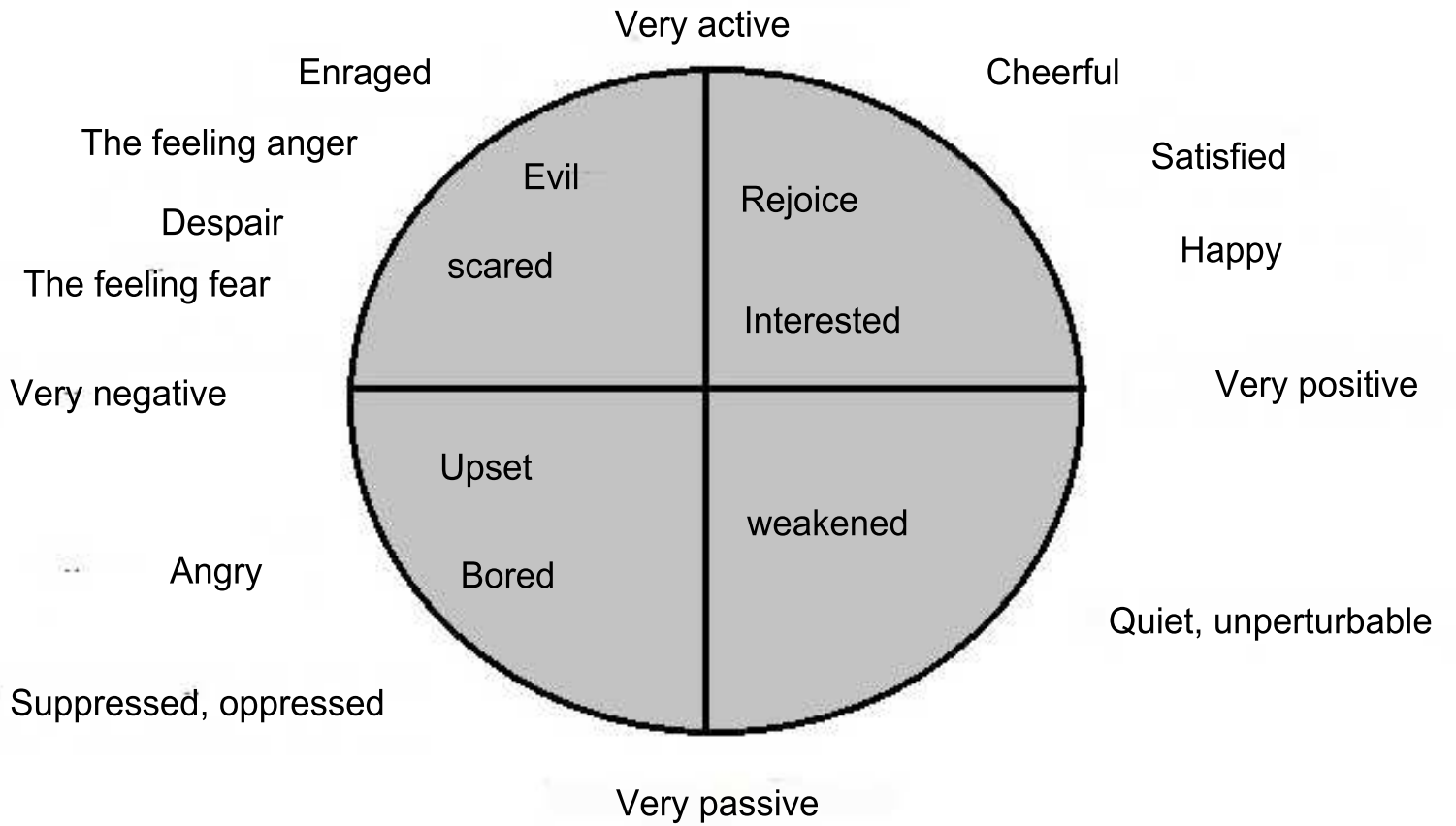
In the first direction

Psychologists subdivide all emotions into groups. It is accepted to distinguish from these groups primary and secondary.

Primary emotions are considered basic, congenital. They include generalized, close to a reflex ("automatic", or programmed) fear and instant reactions to the incentives constituting danger. They don't assume conscious reflections and include six basic emotions which are marked out with Darwin: fear, anger, disgust, surprise, grief and happiness; however, according to K. Izardu mark out 11 fundamental (basic) emotions: joy, surprise, grief, anger, disgust, contempt, unfortunate suffering, shame, interest nervousness, wine, confusion.

Secondary emotions are more difficult emotions, and they involve the highest centers of a cerebral cortex. They can comprise basic emotions of anger or fear, or to have more complex structure, for example, the regret, melancholy, shame, fault, envy, or jealousy will be added to them. Secondary emotions aren't automatic: they are made by a brain, the individual thinks of them and makes the decision that with them to do — what actions best of all to take in this or that situation.

Example of classification of a modality of emotions



Пример классификации модальности эмоций

In the second direction

In the second direction linguists and psychologists reveal emotional components of the speech analyzing her paralinguistic, extralinguistic and articulation features.

From applied linguistics and appellative phonetics it is known that many signs of an emotional state, sincerity and truthfulness speaking contain in a melodics, accentuation, change of speed and a rhythm of the speech, features of an articulating, trembling of a voice — especially in stressful situations, for example, at answers of the interlocutor to unexpected "inconvenient" questions.
In the second direction linguists and psychologists reveal emotional components of the speech analyzing her paralinguistic, extralinguistic and articulation features.

From applied linguistics and appellative phonetics it is known that many signs of an emotional state, sincerity and truthfulness speaking contain in a melodics, accentuation, change of speed and a rhythm of the speech, features of an articulating, trembling of a voice — especially in stressful situations, for example, at answers of the interlocutor to unexpected "inconvenient" questions.

In the third direction

The main objective of receiving features of an emotional component of the speech consists in transforming a sound wave to such feature space in which the set of objects of one class will be grouped together and the set of objects of alternative classes is most carried.

From all range of works at the present stage it is possible to allocate four groups of the objective signs and the corresponding methods allowing to distinguish speech samples: spectral and time features, kepstral feature, amplitude-frequency and features on the basis of nonlinear dynamics.

In the fourth direction

Developing of effective mechanisms and the strategy of recognition for creation of a speech polygraph. Creation of algorithms, scenarios, and, at last, systems of the recognition of truthfulness and sincerity speaking according to the speech. Verification of meanings of emotional speech reactions depending on a situational context, the choice of decisions.

It is necessary to recognize that the same phenomena of the emotional speech depending on a situation can be interpreted differently. When marking those moments in the statement where nervousness in the context of a situation is shown influence of circumstances on an overall picture of emotions and sense of the events at the time of the speech statement can be considered.

For example, depending on a situation the undisguised anger speaking (it is shown, for example, in characteristic change of speed and a rhythm of the speech, careful pronunciation of concordants in words) can testify about incorrectly suggested, contained in the question asked the examinee or about his relation to the situation of interrogation, to interrogating; the confusion and confusion which are shown in the uncertain speech can speak both about fear of exposure, and about misunderstanding of a question. Trembling of a voice depending on a situation can demonstrate offense, fear, anger or on the contrary - pleasures.

Morphology of a situation

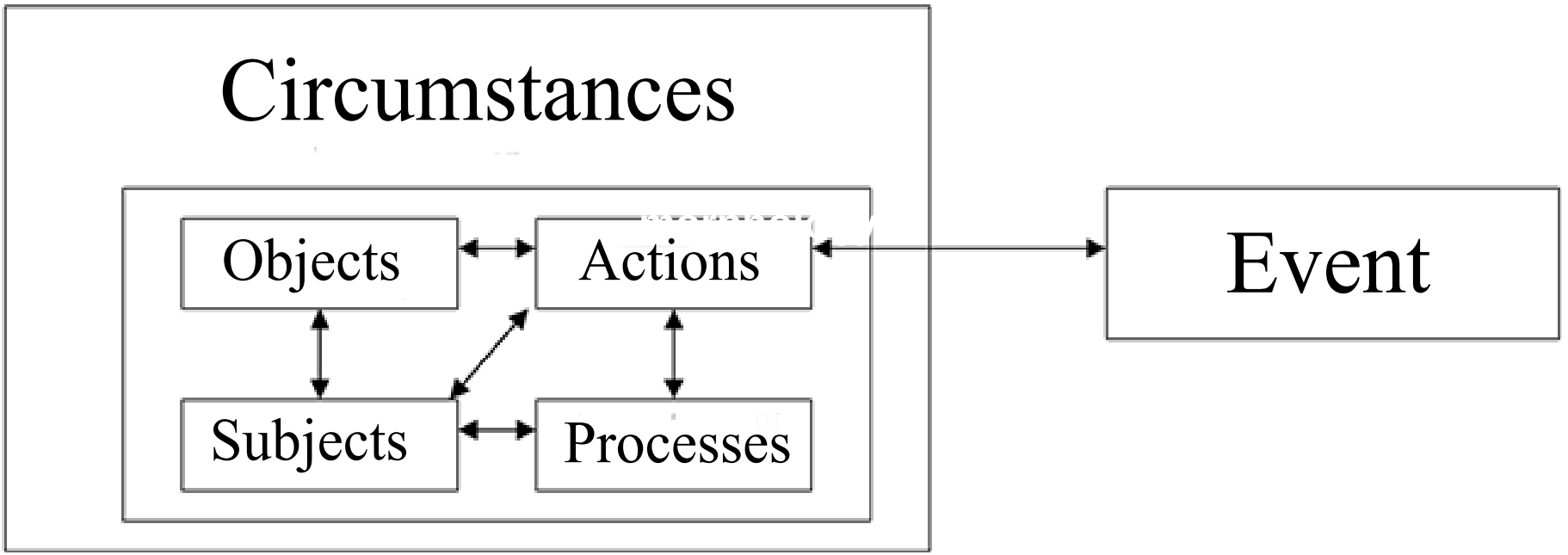
For comparison of speech reactions to a situation in which they were shown by us the standard language of the description of model of situations relying on our understanding of their morphology has been developed.

The situation is understood as some dynamic system of relationship of the **OBJECTS** and **SUBJECTS**, the related **PROCESSES** and separate **ACTIONS** developing in **CIRCUMSTANCES**, and the defining taking place **EVENTS**.

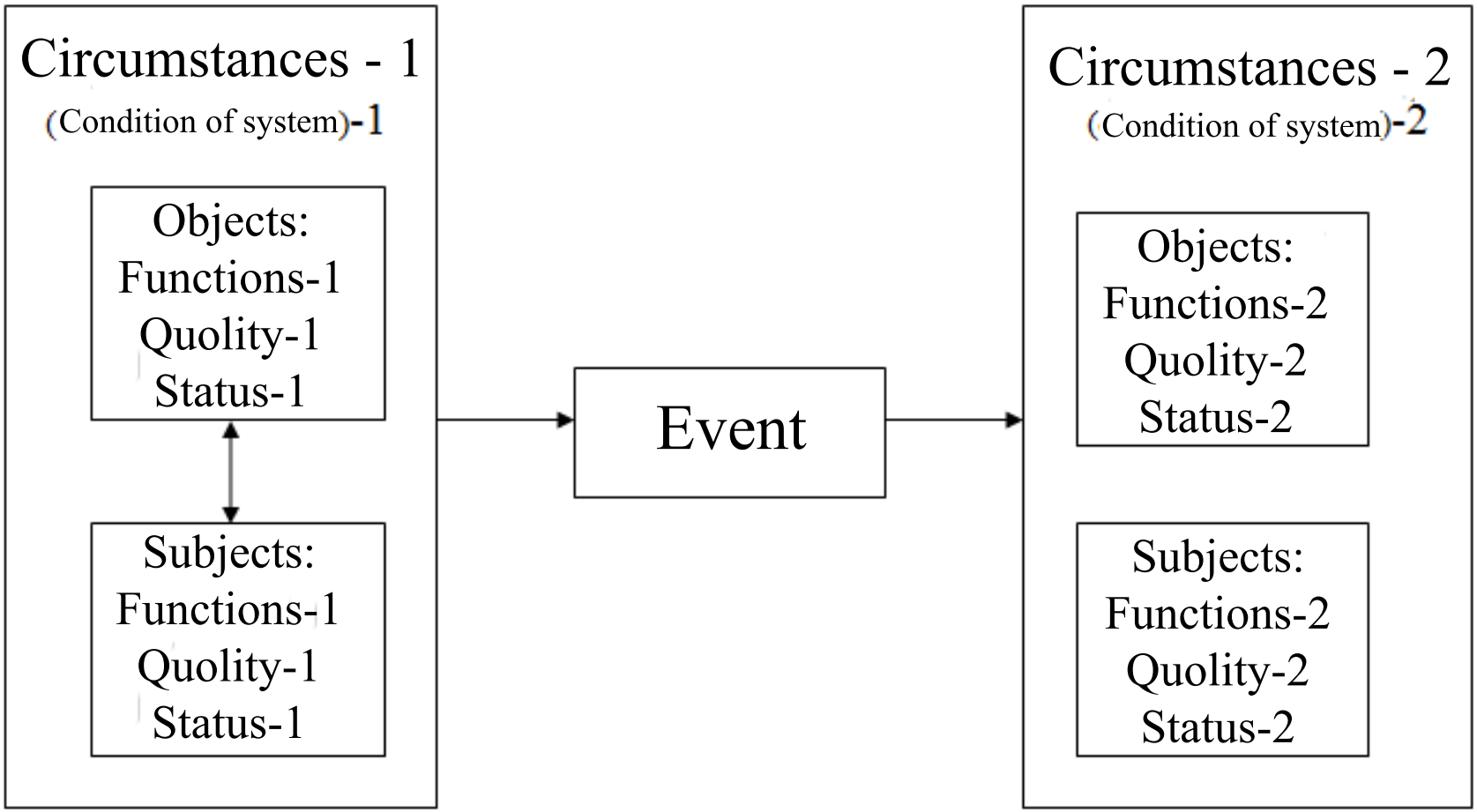
The OBJECTS and SUBJECTS involved in each situation are allocated with certain properties which define their **FUNCTIONS**, **QUALITIES** and the **STATUSES**. Development of the situation changes **FUNCTIONS**, **QUALITIES** and the **STATUSES** of objects and subjects.

Such changes can be desirable for one and are undesirable to other participants of a situation that leads to their counteraction to the happened changes and causes new changes. The reason and the moving mechanism of development of the situation consists in it.

Main components of a situation



Stage of development of the situation



Situational dynamics

Externally, the situation is shown by the sequence - or rather, CHAIN of events that unfold in time and space as a result of a single action or as discrete (landmark) displays some ongoing process.

These processes and actions define the dynamics of change in the circumstances of the situation and the clarification of the moment of communication with party events have consequences determine the behavior of participants in the dialogue (conversation on everyday topics, interview, interrogation, oral questioning, etc.) their emotional reactions.

In begining

Before the experiment on automatic or expert recognition of truthfulness and sincerity speaking according to the speech researchers as basic data, as a rule, have:

•the speech signal presented in the form of discrete function from time — the sequences of temporary counting (***k***)), where ***k***-numbers of counting of a signal on time axis with the fixed step,

• some information on character speaking, i.e. about cognitive, regulatory and communicative features of manifestation of emotions of ***E***, characteristic of him,

• primary characteristic of the investigated situation in the form of set of circumstances of ***G*** described by some standard language.

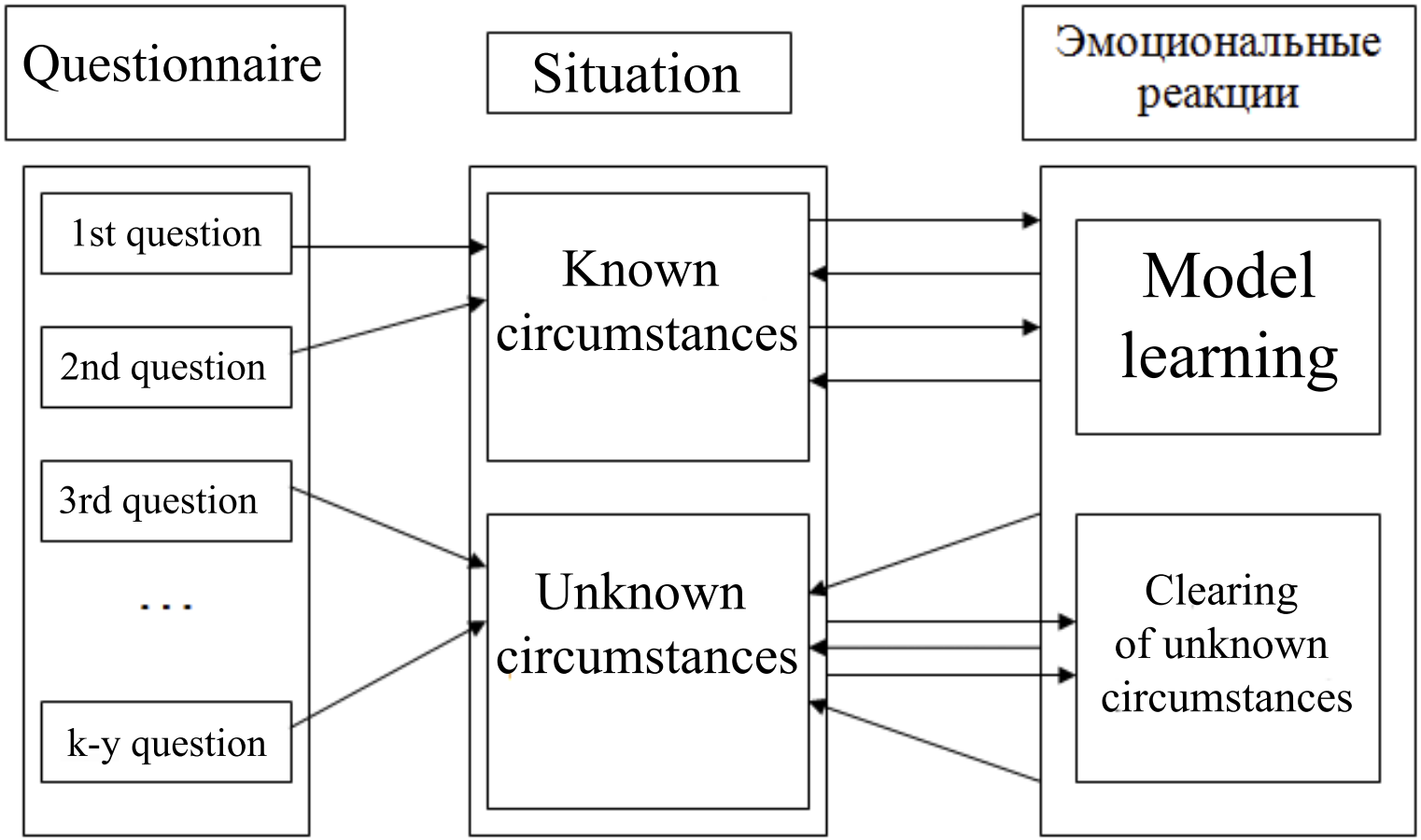
Task definition

We are faced by a problem of identification of an emotional component of the speech of paralinguistic, extralinguistic, articulation features of the statement and recognition of sense of emotion of the individual in the context of an inquiry situation. These emotions being involuntary reaction of the individual to attempt of the researcher in one way or another to clear the studied situation would have to allow to make the conclusions concerning told examinees and to clear — both the examinee's position in inquiry situations, and an overall picture of incident.

However ambiguity of emotional and semantic connotations in a projection to the reconstructed picture of incident can lead to essential mistakes and does necessary elaboration of special strategy for the choice of decisions in recognition of sense of speech emotions.

In this work we make an assumption that verification of meanings of emotional manifestations in the speech on the developed signs becomes possible by means of comparison of emotional and semantic connotations to a situational context on condition of application of special procedures of poll of the examinee in the course of reconstruction of the investigated event.

Learning of model and recognition of emotional reaction of the person in a situational context



Description of an experiment

In the real work recognition of emotions and the conclusion about truthfulness and sincerity of the examinee leaned:

- on paralinguistic features of the speech (i.e. her melodics, accentuation, a tempo-rhythm) characteristic of the individual;
- on specific features of an articulating;
- on extralinguistic features of the statement; treat them — pauses, laughter, a tussiculation, sighs, crying, low, stutter, trembling of a voice;
- on knowledge of the emotional and semantic connotations characteristic of the speech of the examinee;
- on correlation of emotionality of the statement with a situational context.

Measuring base of an experiment

At the first stage of data processing the speech signal of ***C(k)*** was exposed to the spectral analysis by means of bystry transformation of Fourier with consistently shifted weighed window.

The dynamic range in the form of the sequence of values of short-term power ranges of ***S(w,i)*** was calculated, measured in timepoints each ***20 ms***, trajectories of maxima three first a formant of ***F (j, i),*** intensity curves in the low, average and high frequency ranges of ***F (l, i),*** amplitude ***A(i)*** and a pitch contour of speech prosody of ***P(i)*** which is bending around the general intensity, having calculated for this purpose on special algorithms a trajectory of the main tone.

Calculations

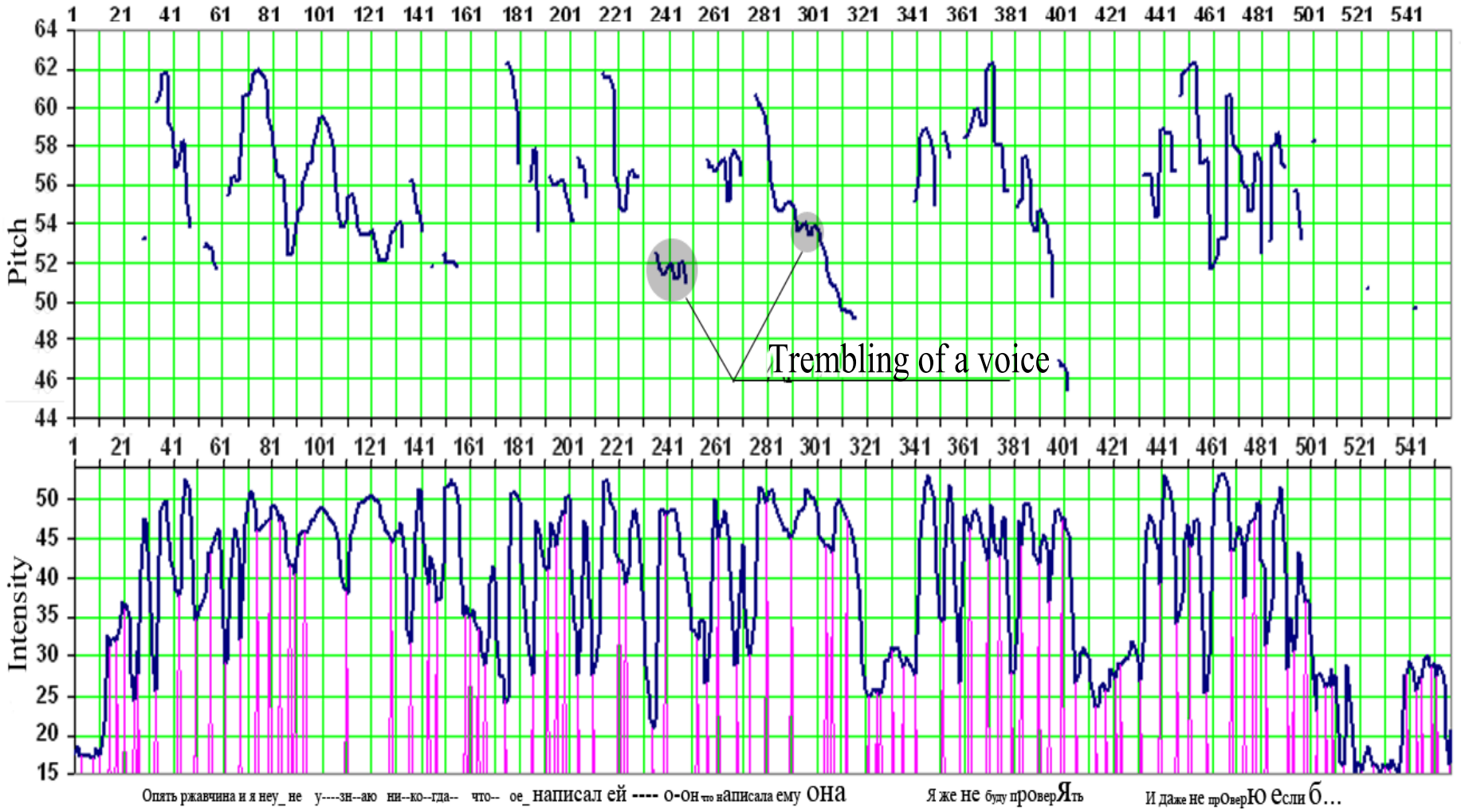
— primary segmentation on minima of intensity of a sound (sm.ris.5);
— marking of allophanes; A(m); segments identified on their spectral to characteristics and, on the basis of reference materials or expert estimates on type an allophane public/concordant (vokalizovanny, slot-hole, explosive, etc.);
— have corrected, have rearranged primary segmentation and have calculated duration of the sounds corresponding to vowels and concordants;
— have revealed prosody, i.e. paralinguistic features of the speech as that:
— height of a voice was led to a continuous musical scale of the MIDI standard where note ***"Do"*** of the first octave corresponds to the 52nd, ***"Re"*** of the 54th, etc., on this scale analyzed a melodics of the speech statement;
— dynamic tempo of speech;
— have distinguished rhythmic forms;
— features of an intoning, for example, have established presence of elements of a contrast and register intoning that is visually visible in fig. 5 and 6 in a time span of 273-321 counting - there is a voice height throw on an octave to exchange down, than in one and a half seconds.

The obtained data were used:

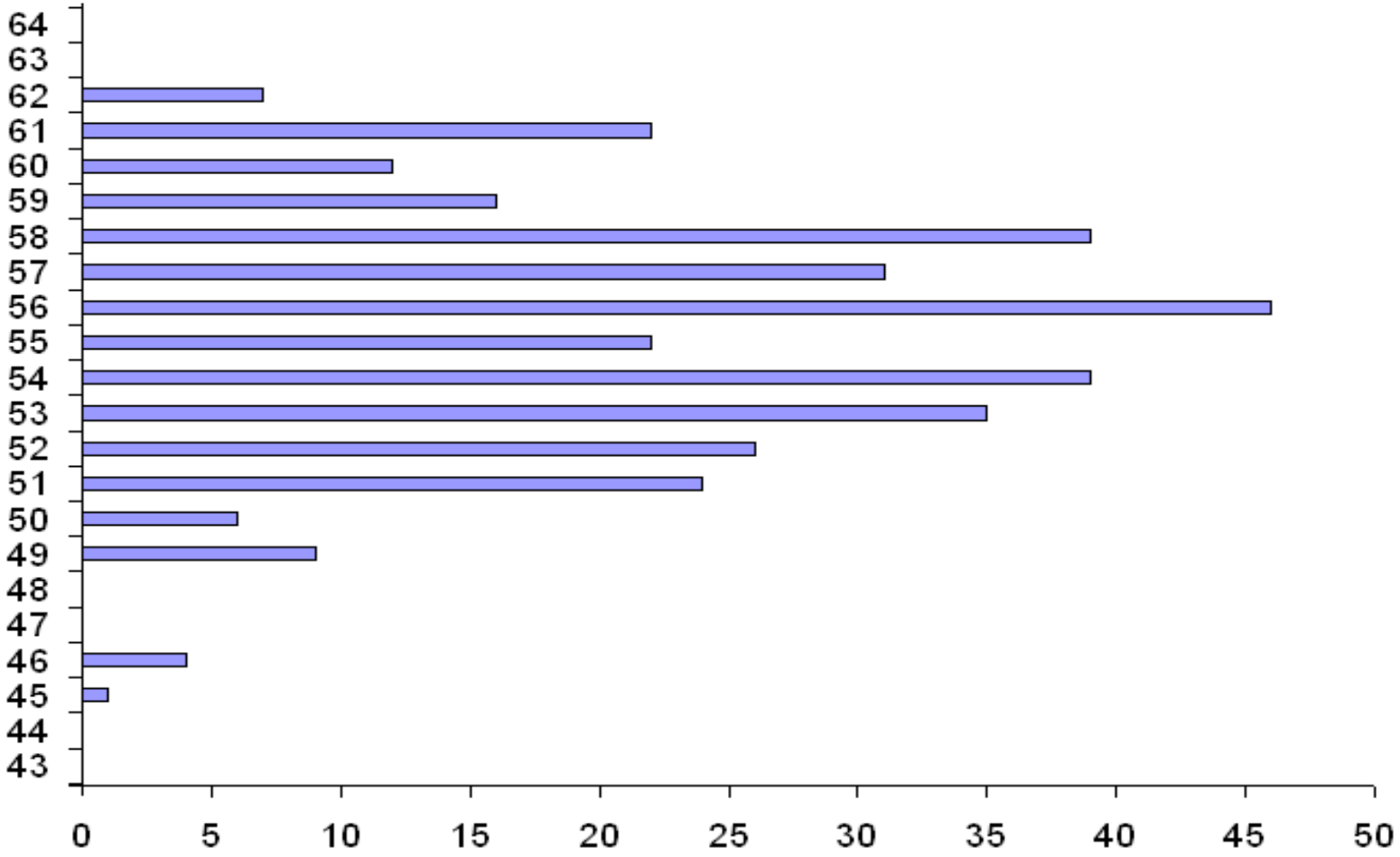
— at a model grade level at expert marking on the episodes testifying to emotionality of the speech and identification of emotional and semantic connotations, characteristic of the individual;

— at a recognition stage for the identification of an emotional component of the speech, the conclusion about sincerity speaking and truthfulness of told.

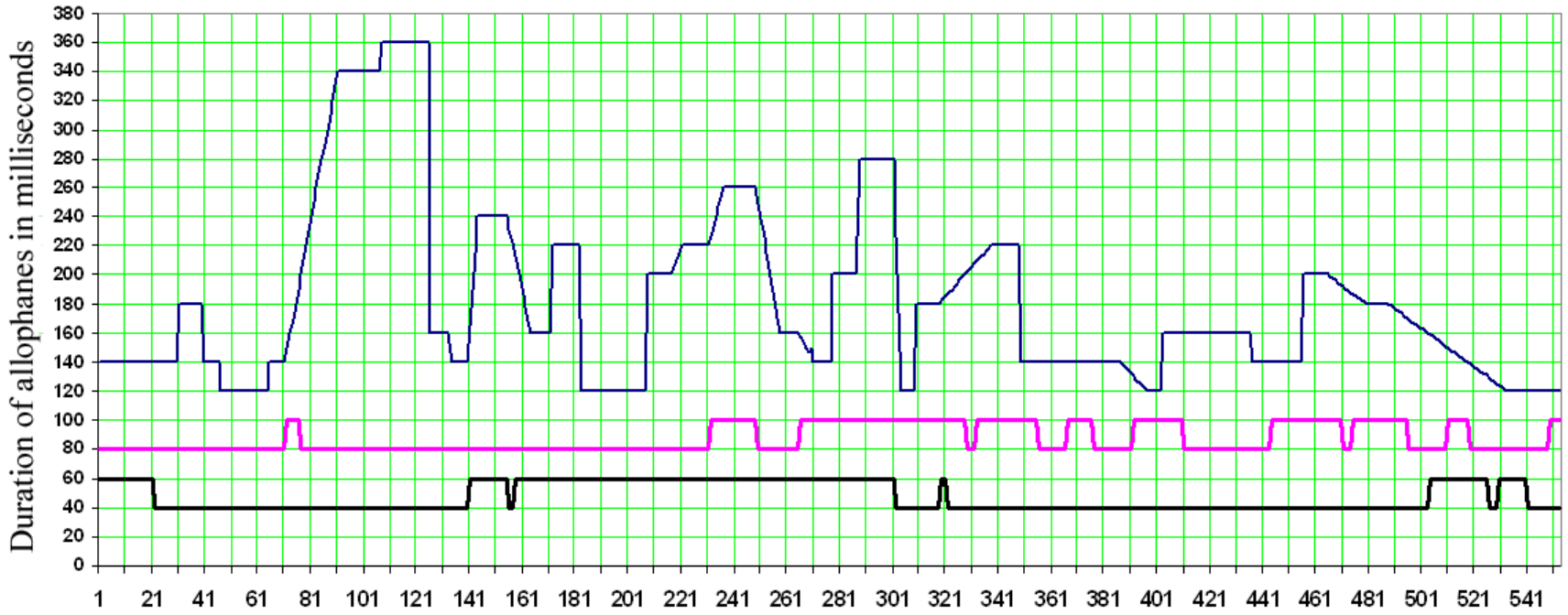
The primary segmentation



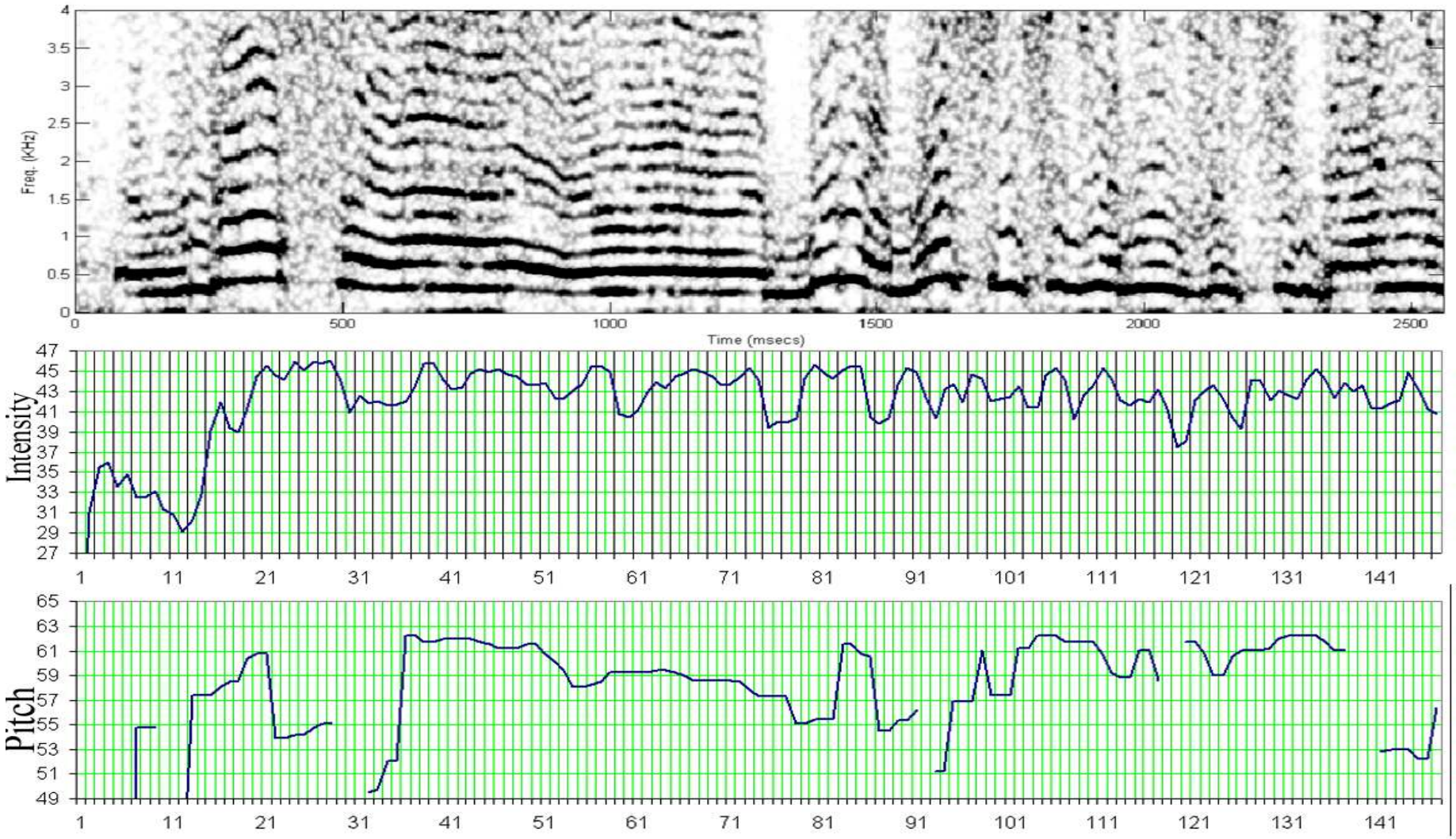
Histogram pitch of the voice in saying “Chto napisal” ("What is he wrote ...“)



Dynamics Charts durations of vowels (top), sonorous, slot (middle) and explosive (lower) according to the statement, "What is he wrote."



Intensity and height of a sound in the statement "We e-e-e-e had eight court sessions"



Identification of features of emotions in the speech

As a result of a series of experiences and expert opinions on emotionality of speech fragments the following signs were the most informative:

- in relation to the average duration of their pronouncing in the current episode phrase and emotional accents allow to reveal duration of stressed and unstressed vowels;

- lengthening pretonic slot-hole or the sonornykh of concordants are the means of emotional strengthening of accent speaking;

- the exaggerated accentuation of a shock syllable in the word due to intensity of a sound of a voice - the evidence of emotional excitement;

- accentuation due to increase in duration of stressed vowels and pretonic concordants - the evidence of desire to convince the interlocutor;

- the inverse value of dynamics of duration of vowels gives dynamics of tempo of speech;

- change of tempo of speech at the level of the word, the phrase, statements - demonstrate to intelligent manifestation the relation of the statement speaking to sense, desire to allocate or hide this relation;

- emotional strengthening of accent in the word, the phrase often is followed by "double accentuation" in stressed vowels and have two maxima of curve intensity, consist of two segments of primary breakdown;

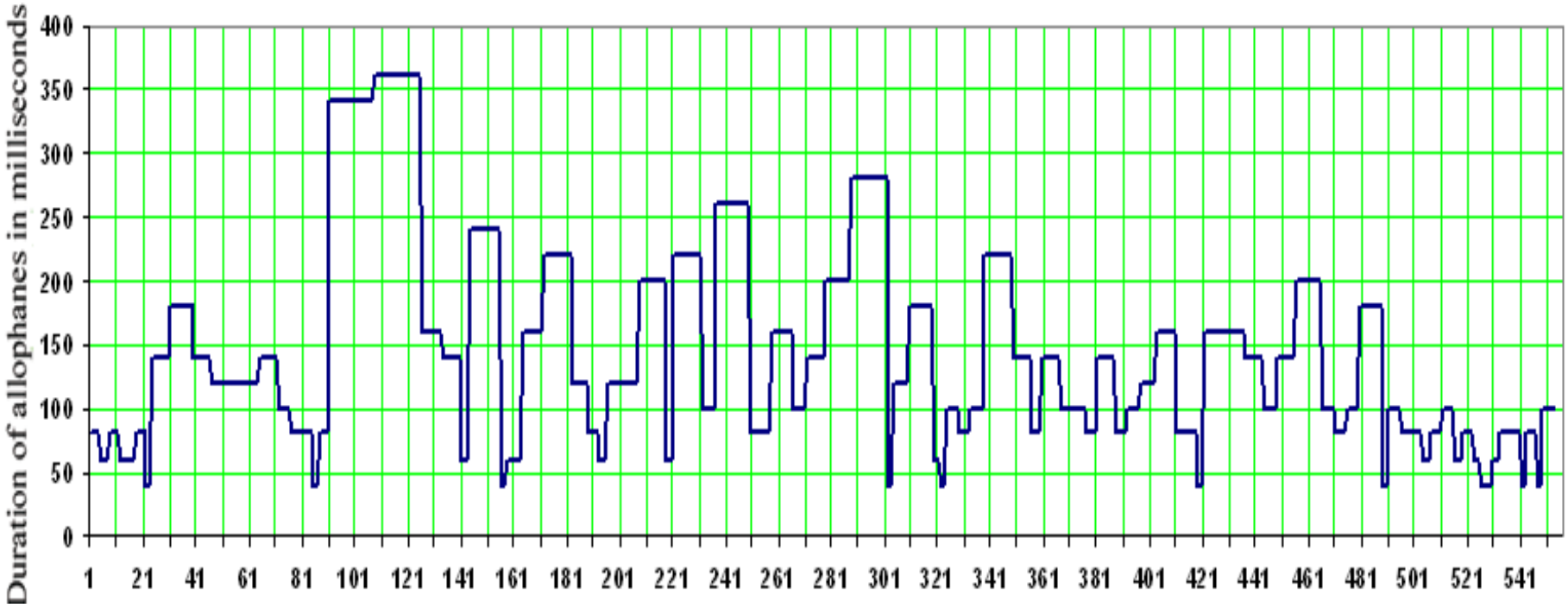
— the speech rhythm constructed on a ratio of dlitelnost of the next stressed and unstressed vowels besides that allows to distinguish accentuation, to specify existence structuring (verbal, phrase) and emotional accents, reveals the multiplicative forms (for example, a chant) most often having the emotional nature;

— trembling of a voice (treats extralinguistic elements of the speech, see fig. 5) demonstrates involuntarily shown nervousness - most often from indignation, fear, offense - or on the contrary - pleasures, delight; and if the indignation, joy and delight usually are followed by the increased level of intensity of a sound, then the fear and offense are shown by an average or the lowered intensity level;

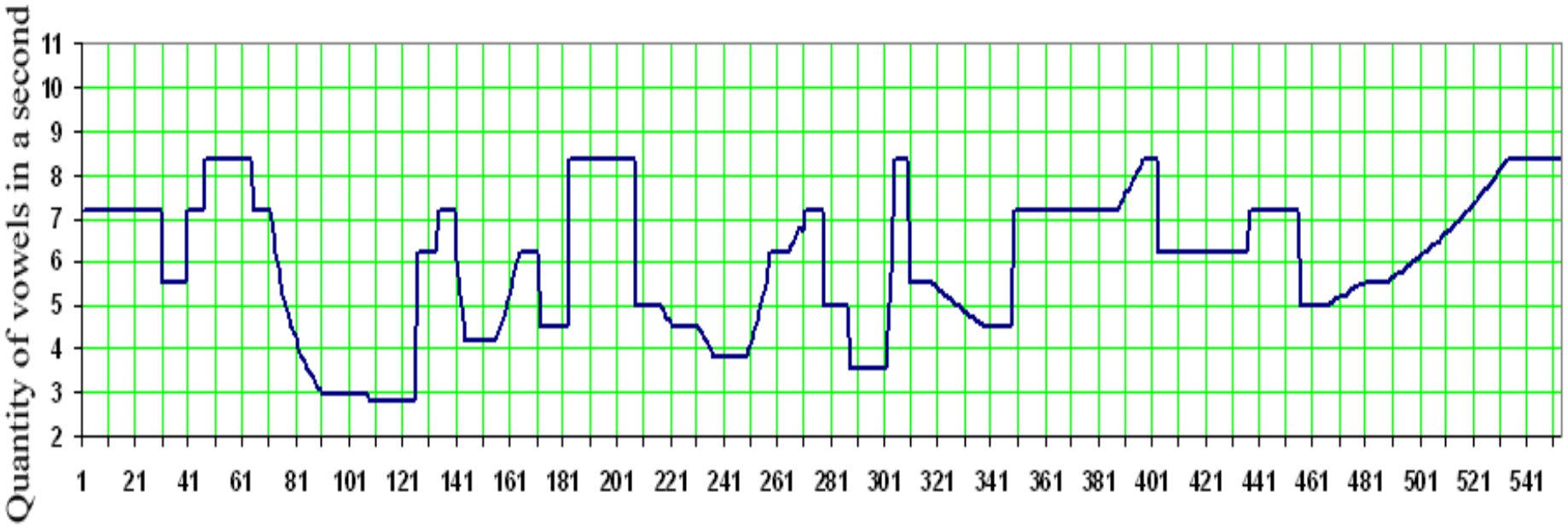
- characteristic periodic alternation of explosive sites and pauses (bow) demonstrate laughter, a tussiculation in the speech;

- long (about a second and more) vokalizovanny "And", "Э", "M" - demonstrate to the uncertain speech, unpreparedness of the speech statement;

Position the durations of allophones in saying “Chto napisal” (“What is he wrote“)



Dynamics of tempo of pronouncing vowels



New results

As a result of our researches earlier not studied communications of speech signs with a modality of emotions, such as have been revealed

— the contrast and register intoning meaning a fright, panic (see, for example, octava throw in a voice height trajectory in the word "it" in fig. 5);

— change of a rhythm with difficult on idle time, meaning irritation, anger;

— the double accentuation of vowels meaning indignation;

— substitution of vowels in the accented syllable, for example "and" on "ы" (an example in the phrase "itself you understand" the first vowel "and" sounds as "ы"), testifies to aggression, anger, rage, indignation; the first three formant in the recognizable segment is calculated on articulation models of vowels on the basis of values.

Besides have been taken into account of interdependence of height of a voice, intensity of a sound, speed, legibility and confidence of the speech received by other researchers:

— obviously high-pitched sound — enthusiasm, joy, the examinee is interested and shows interest;

— excessively high, shrill — concern;

— soft and muffled, with decrease in intonation by the end of each phrase — grief, fatigue;

— speeding up of a sound — tension, deception.

— fluent speech — obvious agitation - desire to convince or persuade someone;

— the slow speech — arrogance, fatigue, depression;

— the faltering speech — uncertainty;

— laconicism and determination of the speech — obvious confidence;

— stutter — tension or deception;

— indecision in selection of words — uncertainty in or intention to surprise suddenly with something;

— emergence of speech shortcomings (repetition or distortion of words, an obryvaniye of phrases stop short) — undoubted nervousness, but sometimes and desire to deceive;

— lowering of speech pauses — tension;

— too extended pauses — disinterest or disagreement.

Construction and training of model

For creation of model of reactions of the examinee the special investigation phase — creation of data array where data on speech reactions of the examinee collected has been allocated, allocation of significant parameters was carried out.

•For splitting the current values of speech parameters into classes (group of features) of such "continuous" parameters as the melodic contour, bending around intensity of a speech signal or the loudspeaker of tempo of speech, duration of pauses was used a clustering method, probabilistic approach. It was supposed that each object (emotional speech reaction) considered at a grade level belongs to one of k of classes of the training selections. For definition the centrode of a cluster was calculated a median and was made training of model.

•Accessory of a segment was determined after calculation of his metrics by group of the parameters stated above in correlation with the alphabet of the segments revealed and marked in the course of expert assessment at a model grade level; his multidimensional classification and respectively marking was carried out when performing a number of conditions.

Scheme

Let's designate a set of temporal and acoustic characteristics of the speech statement as ***M*** (from which subsets ***me*** demonstrates emotional coloring, so, that ***me(i)*** - a feature set of emotions in the speech where ***i =1,2... N*** of and ***N*** — quantity of the classes of emotional coloring which are reflected in speech parameters), set of emotional and semantic connotations of the individual as set ***E***, morphology of situations of inquiry and the reconstructed incident as the ordered sets of ***G1*** and of ***G2***.

In this work the task of the choice of decisions on sincerity speaking and truthfulness told them at recognition of emotions according to the speech in the related system has been set by { ***M, E, G1, G2*** }.

Before carrying out an experiment the model of possible emotional reactions of examinee has been constructed by ***G1-G2-M-E*** of which during the experiment studied according to the approximate scheme on slide 18.

Classification

So, we me have an ordered set from the M signs distinguished by experts as characteristics which confirm nervousness speaking or desire of the announcer to select the word, articulating sounds in him in a special way. At us this set is broken into N classes and it is presented by me data array (k, i, m, L) where each k-y an element from M is carried to i-mu to a class, and to each class is put in compliance of value from S - groups of partially ordered parameters presented by an array S (i, m, c, d),

where i - a name (number) of a class,

M - name (number) of parameter,

C - situation class centrode,

D - class i median.

Then the next distance between clusters of values (a vector difference) of parameters i-x classes of an array S and the corresponding values of parameters of the recognizable n+1st-go segment in space of signs, i.e.

$$\Delta\widehat{L} = \arg\min_{i}[\widehat{M}(k+1) - \widehat{S}(i)]$$

Emotion

- Emotional coloring of a segment can be considered as the probability following from the size of a deviation of his parameters from some "normal" values for this context. Here we lean on the available similar articulation positions on a context which can be distinguished as segments corresponding to "a quiet articulating".

- For fixing of existence of features of emotions in the speech, such as change of a rhythm with difficult on idle time (which established by means of autocorrelated function of duratuion of vowels), a contrast and register intoning (which was determined naploskost by distance between peaks of function of density of probability of height of a voice and temporary distance between their values in a speech fragment) и.т.п. the fact of existence of such sign was elicited, i.e. the binary opposition was used - is, isn't present (true, false).

Example of the choice of the decision

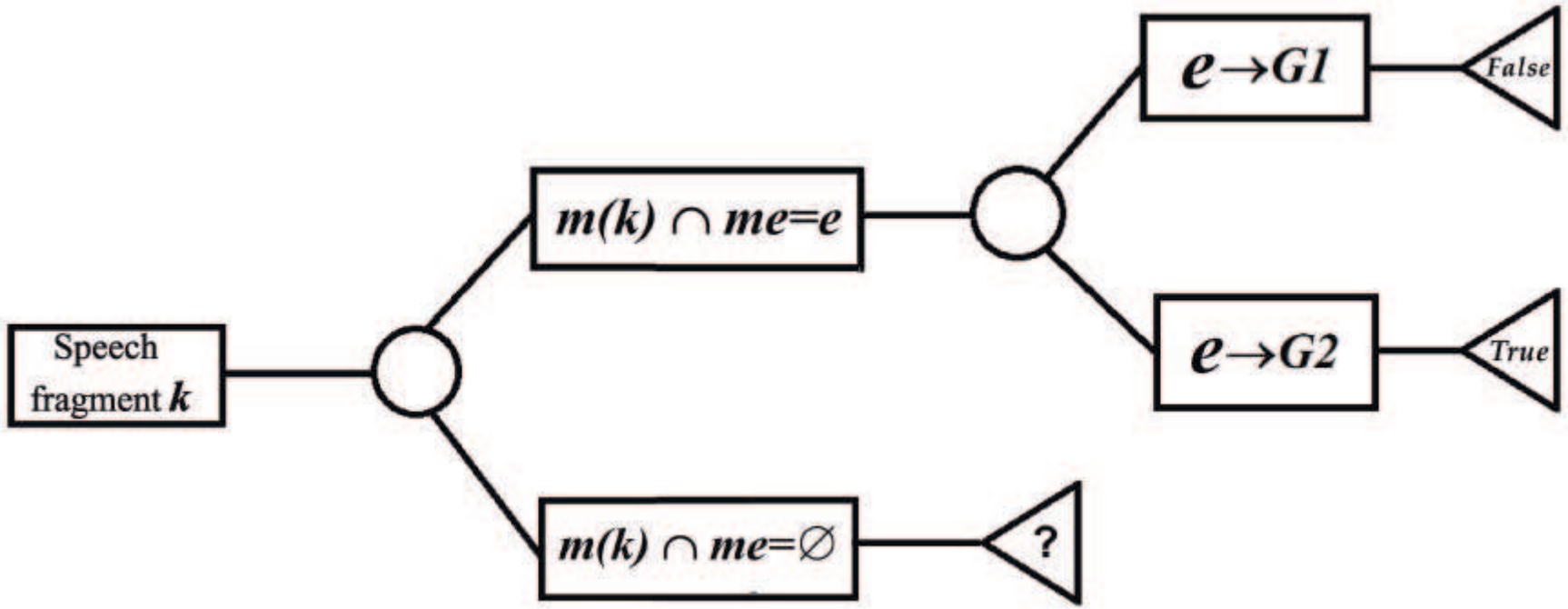
In the choice of the decision on true meaning and truthfulness of told correlation of three groups of signs was considered:

• acoustics-time — voice-frequency, the spektrodinamicheskikh and temporal characteristics of the speech and data on prosody and an articulation of the statement calculated on them;

• emotional and semantic connotations of the speech;

• situational — the statements given about a situational context; at the same time the morphology of two different, but connected among themselves situations — the current situation of inquiry and model of the chain of events reconstructed by the investigation was considered.

The scheme of calculation of a target variable on the basis of acoustics-temporal and situational features



- Questions of inquiry to the examinee at a model grade level along with discrete conditions of the description of a situation of inquiry which belong to a set of circumstances of ***G1*** are set on the known circumstances both ***G1***, and ***G2***. On reaction of the individual to questions on in advance known circumstances there is a training of model of recognition.
- At accumulation of sufficient presentability of the trained model to the examinee questions concerning circumstances of ***G2*** unknown to the investigation are set, emotional speech reactions of me of the examinee to questions from a set of ***G2*** correspond to a set of semantic connotations of ***E*** on the basis of what the conclusion about sincerity and truthfulness of the answer is drawn. At the same time are informative as sincere, truthful answers, and false since they confirm attempt of concealment of circumstances which are necessary for addition of the description of ***G2***. In this case regarding model of a situation of inquiry ***G1*** the additional scenario for clearing of circumstances which the examinee tried to hide can be created.
- Analyzing the initial speech statement of ***C(k)*** concerning the description of a sitution of ***G2*** among temporal and acoustic signs of ***M*** we allocate from the listed 28 signs 8 significant for this statement and we establish the emotional and semantic connotations of ***E*** connected with elements of the studied ***G2*** situation at the same time revealing ambiguity of emotional and semantic connotations.
- So, for example, increase in tempo of speech in a time span 125-210 counting and decrease in speed in the range of 210-250 counting can demonstrate as desire to convince the interlocutor, and uncertainty speaking, his nervousness.

Conclusion

- We have described the experience of choice-making in the system of recognition of human emotional state by speech.
- Analysis of emotional displays on the basis of correlation of steam and extra-linguistic features and patterns of speech articulation with their emotional and semantic connotations shows the ambiguity of the connotations that the projection on the incident now under upgrade and inquiry the situation could lead to a significant recognition errors.
- It is offered the strategy of the choice of solutions of recognition of an emotional condition of the person on the speech in the related system of temporal and acoustic, emotional and semantic and situational dependences.
- At the real approach verification of meanings of emotional speech reactions becomes possible thanks to comparison to a situational context.