JPRS L/10323

12 February 1982

Translation

SPEECH, EMOTIONS AND PERSONALITY

Ed. by

V.I. Galunov

FBIS FOREIGN BROADCAST INFORMATION SERVICE

NOTE

JPRS publications contain information primarily from foreign newspapers, periodicals and books, but also from news agency transmissions and broadcasts. Materials from foreign-language sources are translated; those from English-language sources are transcribed or reprinted, with the original phrasing and other characteristics retained.

Headlines, editorial reports, and material enclosed in brackets [] are supplied by JPRS. Processing indicators such as [Text] or [Excerpt] in the first line of each item, or following the last line of a brief, indicate how the original information was processed. Where no processing indicator is given, the information was summarized or extracted.

Unfamiliar names rendered phonetically or transliterated are enclosed in parentheses. Words or names preceded by a question mark and enclosed in parentheses were not clear in the original but have been supplied as appropriate in context. Other unattributed parenthetical notes within the body of an item originate with the source. Times within items are as given by source.

The contents of this publication in no way represent the policies, views or attitudes of the U.S. Government.

JPRS L/10323

12 February 1982

SPEECH, EMOTIONS AND PERSONALITY

Leningrad RECH', EMOTSII I LICHNOST' in Russian 1978 (signed to press 19 Jul 78) pp 1-198

[Proceedings and reports of an all-union symposium, 27-28 February 1978, edited by V.I. Galunov, Order of Lenin USSR Academy of Sciences, Scientific Council on Integrated Problems of Human and Animal physiology, and the Combined Scientific Council on the Integrated Problem "Physical and Technical Acoustics," 500 copies]

CONTENTS

ANNOTATION

This collection contains the abridged texts of reports and communications presented at the all-union symposium "Speech, Emotions and Personality" held in Leningrad in February 1978. The symposium was convened on the initiative of the speech sections of the USSR Academy of Sciences Scientific Council for Complex Problems of Human and Animal Physiology and the USSR Academy of Sciences Scientific Council for the Complex Problem "Physical and Technical Acoustics." The symposium's convocation was elicited by growth in interest toward problems associated with analyzing variability of spoken communication arising under the influence of the individual features of the speaker and changes in his emotional state.

Reports dealing with the following directions were discussed at the symposium:

the dependence of speech characteristics on the personality properties of the speaker;

the dependence of vocal manifestations of emotion on the personality characteristics of the speaker;

simulation of emotional and individual variability of speech.

SPEECH, EMOTIONS AND PERSONALITY: PROBLEMS AND PROSPECTS

V. I. Galunov

Researchers in linguistics have traditionally been interested only in the first element of the classical three-element formula: "What is being said?, Whom is it being said by?, In what state is it being said?" However, interest in the two other aspects of the spoken signal has noticeably increased in recent years. There are two reasons for this: First, a number of applied problems have arisen associated with the need for defining the personality and the state of the speaker on the basis of his spoken signals; second, most experts in spoken communication have recognized the inseparability of the three indicated aspects of the spoken signal, and the need for analyzing the latter, in all the complexity of this indivisible trinity, even when confronted by a classical problem which might appear to be simple--automatic recognition of speech. What is meant by indivisible is that as a rule both semantic information and information about the speaker's individuality and his state are encoded by the same parameters of the spoken signal.

This paper examines the complex of problems facing researchers attempting to establish a relationship between the characteristics of the spoken signal and the speaker's personality and state.

The Emotive and Indicative Function of the Speech System

The first problem, of course, is to accurately word the task itself, to realize that which we wish to find. Let us attempt to do so by starting with the sufficiently general scheme of communication represented in Shannon's well known scheme (see Figure 1). In the case of spoken communication, the information source is said to be some central cerebral mechanisms shaping the content and structure of a statement, the encoding unit is the articulatory system that transforms a statement into acoustic form through the movements of speech forming organs, the decoding unit is the organ of hearing, which translates the acoustic signal into a neural code "comprehensible" to the brain, and the information receiver is once again represented by central mechanisms responsible for comprehension of the meaning of speech, mechanisms which extract behaviorally useful information from this code. From the standpoint of ensuring maximum resistance to interference for the spoken communication system and simplicity of the speech decoding system, it would be desirable to make the encoding system constant--that is, to make the articulatory tracts of all people identical, and to ensure constancy of these characteristics with respect to time in everyone. Obviously, this is not so in fact. The parameters of the speech forming system vary within broad limits both from one person to another and in a given person depending on a number of causes, particularly his psychophysiological and emotional state. Presence of just this variability alone is enough to bring about individual and emotional peculiarities in speech. It is clear in this case that within the framework of the communication model examined here, in which the main goal of the communication system is to transmit semantic information, individual and emotional features manifest themselves only as additional variations in semantic parameters.

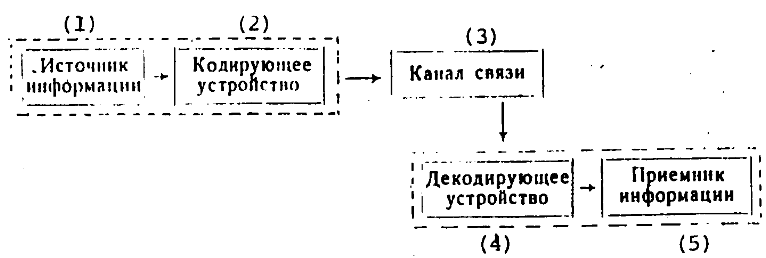

Figure 1

Key:

1. Information source
2. Encoding unit
3. Communication channel
4. Decoding unit
5. Information receiver

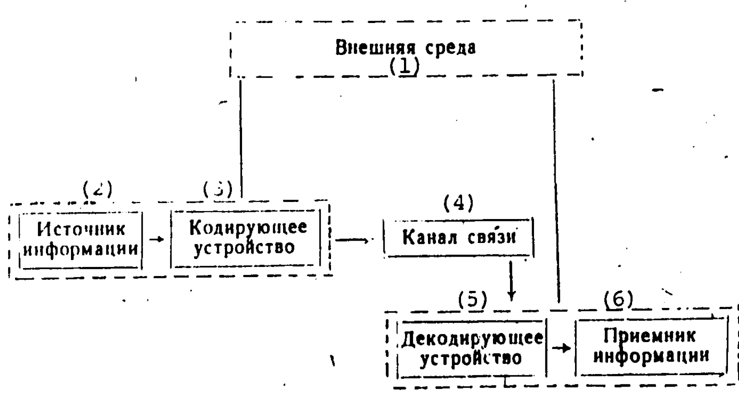

Figure 2

Key:

1. Environment
2. Information source
3. Encoding unit
4. Communication channel
5. Decoding unit
6. Information receiver

Let us examine a model of greater complexity. Following Buhler (1), let us add the environment to Shannon's communication scheme (see Figure 2). This addition is not as primitive as might appear at first glance. Besides introducing the "environment," we implicitly presuppose isolation, from this environment, some object connected with the information source, the range of action of which is not reducible exclusively to the information contained in communications but includes a broader complex of behavior. The concept of information receiver is broadened in similar fashion as well. Such broadening permits us to examine communication not only in its technical and organizational but also its functional plane--that is, to analyze why communication might occur, given sufficiently general hypotheses concerning the interaction of the communicating partners between themselves and with the environment. It would be impossible here to provide a detailed analysis of this system and to examine all functions of a communication system which may be discerned within the framework of the latter (concerning this topic, see (2-3)). We will examine only three functions of interest to us from the standpoint of the topic at hand. The main function of a communication system is to transmit information about the environment. This of course is the main function, and in Shannon's scheme, in which everything except the ideal (that is, not varying and not having any functions other than communication) information source and receiver are defined as the environment, this is also the sole function. The second function is emotive--the function of transmitting information about the internal state of the source. The third is the indicative function, indicating the individuality and the group or social status of the participants of communication. Clearly within the framework of the topic at hand, we are mainly interested in the emotive and indicative functions. But we cannot simply ignore the informative function, since all processes associated with realization of the emotive and indicative functions proceed on the background of processes supporting this main function.

It should be noted that the two functions of interest to us are supported in three ways. First, through verbal expressions ("I do not feel well," "I am happy," "My name is Ivanov," "I am your chief" etc.). Second, by nonverbal sounds (laughter, weeping, groaning etc.). Third, by variation of speech parameters (changes in loudness, in the characteristics of the principal tone, in rhythm and pitch, in the structure of the statements etc.). We should probably exclude the first way from our examination right away, since it is fully identifiable as the means for realizing the informative function, and it may be analyzed successfully within the framework of the classical methods of linguistics, automatic speech recognition and so on. The second of these ways is characterized by a rather narrow functional range, and it would probably elicit only limited interest. Thus our main attention should be concentrated on the third way of supporting the emotive and indicative functions; it will be the main topic of discussion below.

There is one more rather fine distinction between two levels in the support of emotive and indicative functions. These two levels can be seen more clearly in the emotive function. The first level is represented by changes in speech parameters that are realized by the speaker and which yield to his control (practically all such parameters can intentionally transmit indications of excitement, calmness, dissatisfaction and so on). Certain situations may require certain style of speech in order to realize the emotive and indicative functions (an "entreating" or "commanding" voice). The second level is represented by realization of these functions through uncontrollable changes in speech characteristics. Clearly these two levels overlap to a significant extent, but they also possess their separate elements (4). As a rule the level of controllable manifestations is beyond the interest of researchers dealing with applied problems. However, it should be considered that in extreme conditions a normally controllable system supporting one of the functions may also go into action automatically.

Thus examination of the communication schemes adopted here leads to a rather narrow problem: analysis of uncontrollable changes in the structure of spoken communications supporting the emotive and indicative functions of the communication system. Returning to the initial scheme, we should make two more remarks of methodological nature. First, differences in both the coding system (anatomical and physiological features or changes in the speech forming system) and in the information source itself (which would basically lead to change in the structure of a statement and not in its acoustic characteristics) can serve as a source of individual and emotional variability of speech. This allows us to define two pathways for analyzing variability, ones which may be called biophysical and psycholinguistic. Second, two approaches to analyzing the relationship of speech to emotional and personality characteristics are possible. The first boils down to initially analyzing the variability of speech and subsequently searching for individual and emotional characteristics eliciting this variability. The second boils down to initially isolating the personality or emotional characteristics of interest to the researcher and then searching for their correlates in speech.

The Alphabet of Emotional States

One of the fundamental difficulties the researcher encounters in analysis of the emotional variability of speech is the absence of a satisfactory theory of emotions. The researcher in linguistics faces the problem of determining what it is he wishes to find reflected in a spoken signal--a problem outside his competency. A purely pragmatic approach is possible in principle: A list of states and the means of their formation can be given from without, and then the linguist solves the purely applied problem. Clearly the research would suffer in the generality of its application. Nevertheless we could try to draw up a more-general list of states using some particular conception. One such general and promising conception is P. V. Simonov's theory (see the present collection). We will examine two other approaches that also permit us to draw up sensible lists of states.

The first is associated with the well known classification of emotions suggested by Wundt (5), who characterized states in relation to three sets of properties: positive-negative, arousal-inhibition, anger-fear. This old conception has recently enjoyed support among authors using the so-called method of the semantic differential or, in other words, the method of semantically opposite pairs (6). It has been found that upon its perception, every stimulus is evaluated on the basis of a limited number of affective characteristics. According to Osgood there are three such characteristics (general evaluation, activeness, strength), ones which are in full agreement with Wundt's system. We distinguish four independent characteristics. 1) general evaluation ("good-bad"), 2) activeness ("active-passive"), 3) degree of domination ("suppressive-subordinate," "strong-weak"), 4) degree of predictability ("commonplace-odd," "stable-changeable"). The capability of the sensory system for determining the values of these characteristics of all stimuli may be called its evaluation function. In principle we can hypothesize that emotions are a certain subjective behavioral equivalent of a general evaluation of a life situation in the given moment, or simply of an evaluation of internal state. In this case we can now draw up the final list of states. Thus in our four-dimensional model the extreme states on the individual axes would be as follows, given that the states assume neutral values on the other axes:

1. Pleasure-repulsion,

2. arousal-inhibition,

3. anger-fear,

4. Interest (attention)-indifference.

This list may be lengthened in due course by considering different combinations of the values of the evaluation function in relation to all four dimensions.

The second approach to compiling the alphabet of states is for practical purposes a generalization of the purely pragmatic position in which the list of emotions subject to analysis is determined by the applied problem facing the researcher. Is there some way to represent all such problems and correspondingly list all states that are of interest from a practical point of view? It may be hypothesized that all emotional states having significance in practical life should be contained in a dictionary. On this basis we examined a dictionary of the Russian language, and we wrote down all terms which define emotional state to one degree or another. There were more than 500 such terms in all. Obviously most of them were synonomous. An analysis of a thesaurus produced 22 sets of terms defining different states. Each of these sets could be represented by a most typical and most widely used word: 1. indifference, 2. calmness, 3. concentration, 4. tension, 5. tiredness, 6. anxiety, 7. doubt, 8. embarrassment, 9. excitement, 10. inspiration, 11. frenzy, 12. joy (pleasure), 13. grief, 14. dispair, 15. anger, 16. fright, 17. shock, 18. depression, 19. aggressiveness, 20. satisfaction, 21. revulsion, 22. melancholy.

As with the list of states obtained by the first method, this list is somewhat extensive from a practical point of view, but on the other hand it probably covers all states of interest to the researcher. To conclude, here is an abbreviated list of emotions which from the author's standpoint represent all states of interest in the applied aspect: 1. joy, 2. grief, 3. excitement, 4. depression, 5. rage, 6. fear, 7. apathy, 8. the norm.

Classification of Personality Characteristics

Given the great amount of confusion about the problem of classifying personality characteristics, the researcher in linguistics is in a somewhat better position here than with emotions (perhaps precisely owing to this excessive confusion). First of all it is clear that a speaker's identification by voice depends on a number of anatomical and physiological features of the speech forming system of the given individual, features which are typical of him alone and which make it possible to distinguish him from all others. Because in this case the matter boils down to studying just the spoken signal and the means of its generation, the researcher in linguistics remains within his element. On the other hand it is obvious that the spoken signal also reflects a number of group characteristics, though once again associated basically with the anatomical and physiological features of individual groups (based on sex, age and so on). The obvious list of such group characteristics is not very long, though it would be difficult to imagine a definite procedure by which it could be compiled.

There is a rather vast literature on the psychology of the personality. Irrespective of that classification of personality traits which we would decide to adopt, it would be rather difficult to expect these traits to be reflected in the spoken signal, except for those associated with emotionality, and even more likely with the emotional reactivity of the speaker (such as, for example, general activeness, impulsiveness, emotional activeness and so on, as defined by Guilford (7)).

Concluding this section, we should note yet another reason why researchers in linguistics show little interest in a classification of individual personality traits. The fact is that the principal applied problems associated with analyzing the individual characteristics of speech are associated with identifying the speaker as such--that is, with determining which concrete individual uttered a particular passage of speech (this is true of criminology, systems limiting access to documents and facilities on the basis of speech patterns, military intelligence and so on (8)). It is only recently that problems requiring quick testing of group psychological traits of the personality have come into being.

The Model of Emotional States, and Their Control

One of the principal stages of research on the emotional variability of spoken signals is selection of a sufficiently convenient experimental model simulating manifestation of emotional reactions in natural conditions. This choice becomes necessary when it is practically impossible to record speech accompanying naturally experienced emotions. An emotional state model should: be based on a broad range of subjects so as to ensure sufficient statistical significance; ensure acquisition of the required speech reactions; provide a possibility for obtaining a broad spectrum of emotional states; permit current evaluation or control of the state of the informant.

These requirements are satisfied most simply in the model provided by acting. Of course, we do need to mention some negative traits of this model right off. First of all when reflected by an actor, all states, even the most emotional ones, proceed on the positive background of an actor's inspiration. Owing to this the model would produce the most authentic results when positive emotions are involved. Moreover rather than experiencing an emotion, an actor more than likely simulates some traits of its manifestation in ordinary life, or he may even use a certain stereotype to symbolize an emotional state. This makes it difficult to distinguish controllable from uncontrollable variations in the spoken signal (specifically, they are all controllable for the actor). Whether or not purely specific traits, inherent only to theatric speech and performing an emotive function, are present in an actor's simulation of emotion is not very clear. No such traits have been revealed as yet. The sensory system of the audience may be influenced by music or color as supplementary means for deepening the emotional state simulated by the actor. The model of emotions induced in a hypnotic state is very close to the acting model. Today this is one of the most promising models. Application of stimuli to specific body points used in acupuncture may be combined with hypnotic suggestion in order to reinforce the latter, or this may even be done independently.

Consideration of persistent shifts in emotional states observed in the psychiatric clinic is a separate direction. Unfortunately this model possesses two significant shortcomings. First of all it is rather difficult to obtain a broad spectrum of emotional states from a single patient within a suitable time frame. Second, side-effects of drugs upon the speech behavior of subjects, ones which are difficult to control, are practically always present.

Concluding this review of models of emotional states, a little should be said about so-called "natural" states. Rather often, researchers studying applied problems use states arising in situations that are either identical or sufficiently close to practically important ones (for example firing an ejecting seat, experiencing accelerations in a centrifuge, parachute jumping and so on). When analyzing and using the results of such studies, we must always remember the intense influence factors dependent on purely physical accelerations and a number of other incidental factors have on speech production. This influence makes it practically impossible to use the obtained data beyond the limits of the narrow situation under study, though in relation to the latter the resulting data are highly significant and most valuable for practical purposes.

Control of the speaker's state is a separate problem. Were we to look at the models presented above, we would usually find that the state of the speaker is determined by a purely subjective method by the director, the hypnotist or the psychiatrist. There is interest in obtaining additional physiological or biochemical data describing the speaker's state. But unfortunately the data that have been obtained pertain to only one of the dimensions--"inhibition-arousal." Nonmodal activation of the ascending reticular system manifests itself as desynchronization of the EEG. Arousal of the sympathetic nervous system manifests itself in the GSR, in growth of arterial pressure, in dilation of the pupils, and in respiratory activity, pulse, muscle tone and skin temperature. Correspondingly, this same scale correlates with release of epinephrine and norepinephrine, and it may be tied in with the appropriate biochemical analyses. Unfortunately the other dimensions of emotions are not reflected by any physiological indicators that have been discovered thus far. Analysis of electrophysiological indicators at special points of the body may offer certain promise in this area. In principle, however, we can assert today that the spoken signal is one of the most informative indicators of emotional state.

Two Approaches to Analyzing the Relationship of Speech Parameters to Individual and Emotional Characteristics*

Going on to the problems of speech, we will return to the two approaches, mentioned at the beginning of this paper, to seeking relationships between speech parameters and the individual and emotional characteristics of the speaker. We will examine, as an example, the relationship between speech and emotional states.

*The problems to which we now turn, ones which are associated specifically with speech, are given a rather cursory examination, and only in the methodological aspect, since the concrete results of speech research are the object of analysis of most non-review reports presented at the symposium.

The first approach begins with analysis of states and changes in the body elicited by changes in emotional state, and ends with the spoken signal. As was noted in our discussion of physiological control of states, it is rather difficult to describe change in biophysical characteristics in the presence of changes in state. This makes analysis of possible changes in a speech signal much more difficult. However, we can list a number of rather obvious factors that may influence speech formation. First of all if emotional states deviate from the norm, we might expect destabilization of the generation of spoken messages. At the psycholinguistic level (as defined above) this should lead to a certain regression (primarily a simplification) of the structure of statements, to use of simple, habitual syntactic structures that are easily made automatic. An extreme manifestation of such destabilization might be the arisal of speech failures, of mistakes in grammatical and syntactic formulation of statements. At the biophysical level, destabilization of the mechanism responsible for controlling speech formations manifests itself as growth in the scatter of the values of the typical parameters of individual sounds. A shift in the mean values of these same parameters may in principle be associated with presence of a particular emotion in the speaker, but an increase in scatter may probably be expected with any deviation of state from the norm. We can also examine subtler processes influencing the characteristics of the spoken signal and allowing us to treat, as suspicious, parameters such as formant width, the frequency of the principal tone and the characteristics of the melody curve of the principle tone, the relationship between the high frequency and low frequency parts of the integral spectrum and so on (see (9)).

The second approach to seeking the relationship between characteristics of the spoken signal and the speaker's state begins with analysis of the variability exhibited by the parameters of the spoken signal, and ends with revelation of the emotional factors responsible for this variability. That is, in the first stage we determine the limits for the variability of individual speech parameters, and only after this do we reveal whether or not this variability is associated with fluctuations in state. This approach is more exotic, and it is hardly ever encountered in research having a narrow practical orientation, though in principle examples of its successful use can be cited (10).

All that has been said about analysis of the emotional characteristics of speech may also be repeated in relation to analysis of the relationship between speech and individual and personal characteristics.

In conclusion we should point out one more direction of research in which active work has been done in recent years: analysis of the mechanisms responsible for perception of emotional and individual characteristics of speech. This research direction is attracting attention for two reasons. First of all the perception mechanisms and the corresponding distinguishing characteristics that are revealed may be used to develop automatic recognition systems. Second, in a number of practical cases it would be permissible to use auditory experts to determine the emotional state or personality of the speaker. However, before we can sanction the use of such experts we would have to learn the possibilities of the human auditory system.

BIBLIOGRAPHY

1. Buhler, H., "Sprachtheorie," Jena, 1934.
2. Jakobson, R., "Linguistics and Poetics," in Sebeok, Th. (Editor), "Style in Language," N.-J., 1963, pp 350-377.
3. Sebeok, Th. A., "The Informational Model of Language," in Garvin, P. (Editor), "Natural Language and the Computer," N.-J., 1963, pp 47-63.
4. Galunov, V. I., and Tarasov, V. I., "Natural Manifestations of Emotional States and Investigation of the Characteristics of the Spoken Signal," in "Rech' i emotsii" [Speech and Emotions], Leningrad, 1975, pp 55-61.
5. Vundt, V., "Osnovy fiziologicheskoy psikhologii" [Principles of Physiological Psychology], St. Petersburg, 1874-1881.
6. Osgood, Ch., Suci, G., and Tannenbaum, P., "The Measurement of Meaning," Urbana, 1957.
7. Guilford, J. P., "Factors and Factors of Personality," PSYCHOL. BULL., Vol 82, No 5, 1975, pp 802-814.
8. Beeh, B., Neuberg, E. P., and Hodge, D. C., "An Assessment of the Technology of Automatic Speech Recognition for Military Applications," IEEE TRANS., ASSP-25, No 4, 1977, pp 310-321.
9. Galunov, V. I., Koval', S. L., and Tampel', I. B., "Effect of Individual and Emotion-Dependent Changes in Parameters of the Articulatory Tract on Characteristics of the Spoken Signal," in the present collection.
10. Beskadarov, A. V., and Galunov, V. I., "Analysis of the Variability of the Melodic Contours of Speech," in the present collection.

THE PROBLEM OF CLASSIFYING EMOTIONAL STATES IN LIGHT OF THE INFORMATION THEORY OF EMOTIONS

P. V. Simonov

1. Definition of emotions as a phenomenon of higher nervous (mental) activity: According to the information theory of emotions (Simonov, 1964) emotion is an active state of a system of specialized brain structures stimulating the subject to change his behavior in the direction of minimizing (weakening, interrupting, preventing) or maximizing (intensifying, prolonging, repeating) this state. The quality, degree and sign of an emotion are determined by the need for its satisfaction and the predicted probability (possibility) of its satisfaction on the basis of inborn and previously acquired experience. Consideration is given in this case to how well perfected the organism's habits are, the energy resources of the organism, and the time necessary and sufficient to perform adaptive actions.

Information theory is valid in relation to the genesis of all emotional states, including the emotional tone of sensations. For example, an evaluation of food as being pleasant arises only when a hunger stimulus (a need) is integrated with afferentation from the mouth cavity signaling impending satisfaction of this need. To a sated subject, the same afferentation can elicit the negative emotion of revulsion, and an avoidance reaction.

The probability of goal attainment (need satisfaction) may be predicted at both the conscious and the unconscious level, for example through action of the mechanisms of intuition. In the latter case probabilistic prediction of goal attainment concludes with an emotional "presentiment" of the closeness of a solution or of the hopelessness of searching in the given direction.

2. The way emotions differ from other phenomena of higher nervous activity must be analyzed in connection with the fact that the terms "emotion," "motivation," "drive," "instinct" and so on are often used as synonyms. Many authors prefer to refer to "emotional behavior," "motivational-emotional arousal," "the emotional-volitional sphere" and so on. We now adhere to the following working definitions:

Need--selective dependence of living organisms on environmental factors significant to self-preservation and self-development; a source of activeness of living systems; the inducement and goal of their behavior in the surrounding world. Three basic groups of needs can be distinguished in man:

material-biological: for food, clothing and shelter, for self-preservation of the individual and the species, for perpetuation of the line and so on;

social needs in the strict sense (inasmuch as all human needs are socially mediated)--that is, the need to belong to a social group, to occupy a certain place within it, to enjoy the respect of its members, to correspond with ethical standards accepted by the given community and so on;

needs of the cognitive-creative type, so-called ideal or spiritual needs, satisfaction of which leads to positive emotions, produced by the process itself of learning about and transforming the realities surrounding man.

An important objective indicator of the quality of needs is the "goal postponement" parameter as defined by P. M. Yershov (A. S. Makarenko's "personal perspectives"). Satisfaction of material-biological needs (hunger for example) cannot be postponed for a time of any major length. Satisfaction of social needs is limited to the human life span. Ideal goals may be attained in the remote future. The individual set of needs and the hierarchy they assume make up the "core" of the given individual's personality, its most significant characteristic. It is namely the sphere of needs and the emotions arising on their basis that make up the "zone of overlap" in which research on brain activity in the natural sciences makes its most intimate contact with the complex of humanitarian sciences.

Motivation--a physiological mechanism activating engrams, stored in the memory, of important objects that are capable of satisfying a need of the organism, as well as engrams of those actions which are capable of leading to satisfaction. In order that a motive could be transformed into outwardly realized behavior, real signals heralding the appearance of target objects must be present. Engrams themselves may not serve as the triggering stimuli of behavior; otherwise the subject would be living in a world of hallucinations and illusions. The mechanisms of motivation promote selectivity of contact with the environment, as dictated by the needs important at the given moment.

Behavior--a form of a living organism's function which changes the probability of contact with the object of need satisfaction.

Will--activity motivated by the need to surmount an obstacle, by a need that is relatively independent and supplementary to the motive which first initiated a particular behavior. The inborn "freedom reflex" described by I. P. Pavlov is obviously the phylogenetic precursor of will. Reactions to "internal interference" (competing motives for example) and participation of the consciousness, which perceives freedom as a conscious necessity, are typical of the activity of volitional mechanisms in man. Deviation from this recognized need is perceived by the subject as "nonfreedom," and it activates the mechanisms of volitional effort. Will can serve as an indicator of a need which has held a dominant position in the structure of the given personality for a long period of time and which determines the choice of actions in a conflict situation, if any one of the subdominant needs generates an emotion that is stronger than this dominant need.

Consciousness [soznaniye]--knowledge [znaniye] which an individual can share with another individual (compare with the words so-chuvstviye [equivalent to "sym-pathy"], so-perezhivaniye [equivalent to "sharing of experience"], so-deystviye [equivalent to "co-operation"] and so-trudnichestvo [equivalent to "col-laboration"]). Unconscious forms of higher nervous (mental) activity would best be subdivided into the subconscious, which supports the protective tendencies of individual and specific defense in the broad sense, and "superconsciousness" (K. S. Stanislavskiy), which services the trends of development, creativity and progress. The mechanisms of superconsciousness surmount the known conservatism of the conscious--its rationalistic nature and its rigid dependence upon former individual and group experience. The superconsciousness promotes arisal of hypotheses contradicting formerly known hypotheses, while the consciousness retains the most important function of selecting only those hypotheses which correspond to the objective realities and which are confirmed by practice. On the other hand the mechanisms of the superconsciousness provide the individual the possibility for acting "unsensibly" but in a way that would be necessary on the scale of the development of civilization as a whole. As an example several fine people may rush to the aid of a person they do not know and perish while rescuing this person who turns out to be a desperate scoundrel.

3. If we are to classify emotional states, we would necessarily have to introduce some single valid principle applicable to all emotions. From the positions of information theory of emotions, emotional states should be classified in a system of three coordinates: The magnitude of the need, growth or decrease of the probability of its satisfaction in comparison with a former prediction, and the nature of the action in the course of which the given state arises. The classification of emotions we suggest is summarized in the table below.

Contact interaction is defined as interaction with the target object which may undergo weakening, interruption, intensification or prolongation but which is already proceeding and cannot be averted. The remote actions of taking possession, surmounting and defending are associated with the three fundamental emotions--joy (corresponding to grief), anger and fear. Thus the sphere of human emotions rests on a foundation of four basic states: pleasure-revulsion, joy-grief, confidence-fear, cheerfulness-anger. The specific features of the need impart qualitative uniqueness to the emotion.

When two or more needs are important simultaneously, each may generate its own emotion depending on the situation and on the probability of satisfaction. In such a case we arrive at mixed forms of emotions, observed so frequently in real life.

The assertion by some critics of the information theory of emotions that this theory is supposedly incapable of explaining the genesis of states such as "My sadness is elevating, my sorrow is sweet" rests on a misunderstanding. These critics forget that coexistence of several needs often generates an entire range of emotions, each of which is subordinated to the rule formulated by this theory.

The individual tendency to predominantly react with one of the fundamental emotions lies at the basis of the classification of temperaments (Simonov, 1968). The weak (melancholic) type of nervous system has a special relationship to the fear reaction, the strong, unrestrainable type (choleric) has a special relationship to rage, the sanguine type is related specifically to positive emotions, and the phlegmatic type is not generally prone to tumultuous reactions, though like the former he tends potentially toward positive emotions. In distinction from melancholic sadness, Classification of Emotions Depending on the Magnitude of the Need, the Probability of Its Satisfaction and the Nature of Actions (Simonov, 1966)

Magnitude of the Need	Evaluation of the Probability of Satisfaction	Contact Interaction With the Object	Remote Actions		
			Seizing, Taking Possession of the Object	Defending, Preserving the Object	Surmounting, Struggling for the Object
Increases	Exceeds the existing prediction	Sateation, pleasure	Delight, happiness, joy	Fearlessness, boldness, confidence	Triumph, exhilaration, cheerfulness
Small	High	Indifference	Calmness	Relaxation	Imperturbability
Increases	Decreases	Dissatisfaction, revulsion, suffering	Restlessness, sadness, grief, despair	Caution, alarm, fear, terror	Impatience, indignation, anger, wrath, rage

choleric grief always borders on anger, while choleric joy borders on cheerful aggressiveness, on ardor.

Although the fundamental emotions listed above may arise in the course of satisfaction of a need belonging to any of the three basic groups, fear is most typical of the biological needs associated with self-preservation, anger is most typical of social motivations, and the need for cognition and creativity definitely tends toward positive emotions.

Concluding this brief outline, it is my hope that the suggested classification of needs and emotions would be useful to the study of the emotional and personality characteristics of speech.

LINGUISTIC INVARIABILITY AND INDIVIDUAL VARIABILITY

L. V. Bondarko, V. G. Shchukin

The social nature of language presupposes mandatory presence, invariability and stability of the units forming a language system. These properties are precisely what make communication among people of the same linguistic collective possible, communication not only at the moment of a concrete act of spoken communication but also through the retention and transmission of all information in this language. However, we know that the utterance of language units in speech often fails to satisfy any of these requirements: Research on living speech, conducted especially intensively in recent years, has shown that the more the units of analysis are broken down, the more obvious it becomes that their properties correspond little to the initial concepts based on the study of that which linguists refer to as a language system. Changes in a language system in time may be explained mainly by failure to comply with the principles of mandatory inclusion, invariability and stability. Naturally the main element within which these principles are violated is the spoken activity of each individual speaking in the given language.

In this connection we are forced to raise the question as to how realistic is the specific object of linguistics--"the abstract, homogenous speech collective, all the members of which speak in the same way and learn the language instantaneously" ((1), p 100). From a linguistic point of view it would be important to determine precisely what characteristics of individual speech activity cause change in the language system and what conditions promote such change.

In this connection the idiolect (that is, the individual's system of speech resources) is to us, on one hand, a realization of the language system inherent to the given individual and, on the other hand, a cause of change in this system. Of course, the variability inherent to an idiolect far from always leads to such changes. It all depends both on the frequency with which a certain deviation is encountered and on its causes. From the standpoint of the frequency with which it is encountered, we can refer to four different types of variability:

1. Intra-idiolectic variability.

This type of variability is characterized by presence, in the idiolect, of variants of linguistic units used occasionally, irregularly, in addition to other variants of these same units.

2. Inter-idiolectic variability.

This type of variability is characterized by presence, in the idiolect, of regularly used, stable variants of linguistic units distinguishing this given idiolect from other idiolects.

3. Group variability.

Group variability comes into existence when the variants of linguistic units common to the idiolect of representatives of the same group do not agree with the variants of the same units in the idiolects of representatives of another group. As a rule this type of variability reflects differences between groups, each of which is homogeneous in one regard or another. The homogeneity of such groups may be based on the territorial factor (identical dialect and similar regional characteristics), the social factor (social variants), age-related and occupational factors (age-related characteristics and occupational jargon) and so on.

4. Mass variability.

Mass variability comes into existence when a variant enjoys broad dissemination in the speech of different strata of the population irrespective of age, occupation, territorial affiliation and so on.

The frequency itself with which variants are encountered depends on a large number of factors. Being extralinguistic by nature, these factors generate different forms of variability. Let us examine some of these factors from the standpoint of their mutual relationships with the language system and with the types of variability.

The division of the causes of linguistic changes into internal and external, which we encounter in special studies devoted to this issue ((2), pp 197-313), is not always justified in research on these causes in application to the idiolect: Realization of potential changes under the influence of external factors is essentially possible only in the event that internal factors do not oppose such realization.

The most general and universal factor is the desire for economy of effort in speaking. This factor should be interpreted broadly, to include not only the tendency to economize on the energy of pronunciation (3) but also the desire to express identical or close meaning by single form, the desire to limit the complexity of spoken communications and so on ((2), pp 241-250). This factor operates in the speech of every speaker, and the extent to which this economy manifests itself is regulated in many ways by linguistic factors: Thus, the high variability of reduced vowels in modern Russian is a result of "smoothing" of the characteristics of these vowels in unstressed syllables, of their convergence with the characteristics of "neighboring" consonants ("stopping short" in the articulatory movements necessary to achieve the "needed" rank, highness and labilization). Such smoothing is a product of the properties of the Russian language system, in which only two degrees of highness (open-close) are functional in relation to unstressed syllables, and the rank of unlabilized vowels is defined by the hardness-softness of the preceding consonant.

The next group of factors includes those associated either with arisal of new concepts or with the borrowing of words from another language to represent previously known concepts. They characterize not so much the speech behavior of the speakers (as is the case of the first factor) as the situation in which the speaker may find himself. The next group is made up of factors associated directly with the personality of the speaker--his sex, age, social status, education and occupation.

The effect of these factors on the characteristics of the idiolect have been discussed in sufficient detail in the linguistic literature. It might appear that these should be called specifically external factors, but even here the limitations imposed by internal factors on their influence is obvious, as is the influence of these external factors on internal ones.

The territorial affiliation of the individual, which is responsible for arisal of dialectal deviations in his idiolect, may be characterized as an external factor, especially if we broaden this concept such that territorial affiliation could also be taken to mean possible bilingualism, and not just presence of dialectal features.

The phenomena of speech pathology, which doubtlessly influence some characteristics of the idiolect and which may become widespread, also have special significance. This factor will be discussed in greater detail a little later.

The list of factors influencing the forms of speech variability is far from exhausted by those discussed here, inasmuch as our objective is not to systematize the factors but to systematize speech variability phenomena in relation to language.

Let us now examine the way in which the types of variability are brought about by these factors.

Intra-idiolectic variability, which we defined as irregular usage of different variants of linguistic units, is caused by almost all of the factors listed above, and in this case all we can say is that the influence of these factors varies and that the factors themselves interact with one another; thus the influence of territorial differences may depend on the sex of the speaker: Women retain dialectal traits in speech more consistently than do men, who consequently exhibit greater intra-idiolectal variability; higher education reduces intra-idiolectic variability associated with territorial affiliation, but it may increase the weight of factors associated with arisal of new concepts or words, and so on.

Inter-idiolectic variability is caused by the presence, in the given idiolect, of stable variants of units inherent only to the given idiolect. Arisal of these stable differences may be associated with any of the factors.

Group variability, which characterizes not the individual idiolect but a certain group, is caused primarily by factors such as sex and age, social status and occupation, territorial affiliation and education.

Mass variability, which for practical purposes characterizes the majority of the bearers of the given language (that is, it is represented in almost all idiolects), may be described as the highest stage of speech: We would rightfully expect at this stage that a phenomenon inherent only to speech would achieve the status of a linguistic phenomenon, and thus restructuring of the language system would occur.

Factors / Types of Variability	Economy of Effort	Arisal of New Concepts	Arisal of Borrowed Concepts	Sex	Age	Social Status	Path-ology	Edu-cation	Terri-torial Affil-iation	Occupation
Intra-idiolectic	+	+	+	+	+	(+)		(+)	(+)	(+)
Inter-idiolectic	+	+	+	+	+	+	+	+	+	+
Group	+	+	+	+	+	+	+	+	+	+
Mass	+	+	+	+	+	+	(+)	+	+	+

APPROVED FOR RELEASE: 2007/02/09: CIA-RDP82-00850R000500030032-8

As was said earlier, the influence of the idiolect on the language system may be represented as a combination consisting of different types of variability, characterizing the frequency with which a particular phenomenon is encountered, and the factors responsible for this variability. Such a classification may take the form of a table allowing us to account for participation of different factors in formation of different types of idiolectic characteristics (see table above).

As we can see, participation of different factors is almost always possible. This table should be interpreted in two directions: first, by establishing correlations between different factors, and second, by differentiating the factors in relation to different levels of language. This can be clarified with examples.

1. Speech pathology, which participates in formation of inter-idiolectic and group differences, may also stimulate changes in mass speaking styles. Thus pronunciation of single-focus fricative consonants with a lisp, typical of (shchiptsovyye) [forceps?] children and persisting in the pronunciation of adults, may lead to corresponding changes in mass speaking style, inasmuch as it is supported by both the tendency for economy of pronunciation effort and the absence of the corresponding functional contrast in Russian language. However, it would be difficult to believe that speech pathology may have a broader influence at the phonetic level, or that it may be in any way tangible at the level of grammar, syntax and word use.

2. Reduction of unstressed vowels in inflections leads to indistinguishability of the grammatical forms of a word. This is a realization of the tendency toward economy of effort, and it is permitted by the Russian language system, in which grammatical information on a word may be deduced from the structure of the sentence in which it is used (4). In a number of cases however, this phenomenon, which characterizes the segmental composition of a word, may be the cause of stress changes penetrating through the idiolect into the language system: The impossibility of distinguishing between the forms "якори" and "я́коря" promotes arisal of the plural form "якоря́" (5).

Systematic description of the idiolect as a language realization and as a source of change in language presupposes analysis of the influence of different factors both from the point of view of the frequency of their occurrence (the type of variability) and from the point of view of their single or multiple influence upon different levels of the language system.

BIBLIOGRAPHY

1. Labov, U., "The Study of Language in Its Social Context," in "Novoye v lingvistike" [Advances in Linguistics], 7th Edition, Moscow, 1975.

2. "Obshcheye yazykoznaniye" [General Linguistics], Chapter III. Language as a Historically Developing Phenomenon, Nauka, Moscow, 1970.

3. Martine, A., "Printsip ekonomii v foneticheskikh izmeneniyakh. Problemy diakhronicheskoy fonologii" [The Principle of Economy in Phonetic Changes. Problems of Diachronic Phonology], Moscow, 1962.

4. Bondarko, L. V., and Verbitskaya, L. A., "Phonetic Characteristics of Unstressed Inflections in Modern Russian Language," V. L., No 1, 1973.

5. Gorbachevich, K. S., "Phonetic Prerequisites of Some Stress Changes in Modern Russian Language," V. L., No 6, 1975.

EXTRALINGUISTIC SIGNALS AND THE PROPERTIES AND STATES OF THE INDIVIDUAL

V. Kh. Manerov

1. Introduction

It is a universally recognized premise that spoken communication is a process in which resources at the linguistic and extralinguistic levels interact in complex fashion. But despite the fact that the latter have attracted the attention of scholars since antiquity, attention has always been focused on symbolic or verbal communication. It was not until our century, and especially in its second half, that systematic research was started on the extralinguistic resources of communication. Today, if we consider only the acoustic extralinguistic signals reflecting the states and properties of the individual, we can name several hundred papers devoted to this problem. However, this abundance of publications has not produced a qualitative change in our understanding of the essence of the phenomena of extralinguistic communication. This is associated with the lack of research generalizing the numerous but dissociated facts obtained, moreover, by representatives of different sciences. Exclusions in this aspect are the works of V. I. Galunov and the monograph by Nosenko (1).

This paper provides a definition of the problem based on an analysis of both the results of my own work and published data collected by psychological and psycholinguistic studies. I was interested in changes contributed by the emotions, expressed intentions and individual characteristics of the speaker to the acoustic and phonetic structure of a spoken statement--changes which do not themselves lead to change in the objective and logical content of the statement but produce information of another sort: information about the speaking individual (the communicator). On the other hand I was also interested in how perception of these acoustic phenomena is affected by the characteristics--the properties and states of the perceiving subject, the recipient.

An external analysis scheme developed within the framework of the systems approach to analysis of complex systems was employed. Such a systems approach is close to optimum in application to extralinguistic communication, since the overall picture of this complex, multicomponent phenomenon must be based on a large quantity of poorly structured, fragmentary data.

2. The Analysis Scheme.

An example of successful application of the systems approach to social phenomena can be found in Kagan's work (2). The author suggests the following basic components of systems analysis: 1. Structural component analysis, 2. functional analysis, 3. historical analysis.

The subsequent discussion will follow this scheme, and the content of the analysis will be interpreted within the framework of these three dimensions.

1. Structural Component Analysis

The task of structural component analysis is to reveal the components of the system and their mutual associations. While linguistic communication occupies the central and most important place in human communication in general, extralinguistic acoustic communication, which is facultative in relation to the former, occupies an intermediate position, being in a sense a shell separating the core--linguistic communication--from other forms of communication, to include visual, tactile and so on, forming in their sum total the system of human communication. Because extralinguistic and linguistic communication are relatively independent of one another, this shell can be isolated as a special system.

Let us imagine the simplest communication situation involving two communicants alternating their roles as communicator and recipient, and an acoustic signal transmitting the content of the process of communication in a case where the linguistic content is fixed or eliminated.

In principle, component analysis of the system in relation to all three components of this situation is possible. At the level of the communicator the analysis must include all potential forms of information embodied within the signal, and thus it must concern itself with the classification of emotional states, expressed intentions and individual manifestations. However, because little research has been conducted on these fundamental problems, for the moment this approach would be extremely unwieldy and difficult to carry out. A more attractive approach would be to reveal only those components of the system which, being encoded in the acoustic envelope of the signal, are perceived and interpreted by recipients with sufficient effectiveness. The entire inventory of extralinguistic acoustic signals was classified on the basis of the following criteria: controllable--uncontrollable, non-speech--quasi-speech, regular--situational. Uncontrollable and partially controllable signals include acoustic phenomena that can be classified as emotional non-speech sounds (groaning, laughter, weeping, sighing and so on). Partially controllable signals also include the emotional tone of speech. Controllable signals are acoustic resources for expressing the attitudes, intentions and desires of the speaker, and descriptive acoustic resources. Non-speech phenomena are those acoustic signals which do not interfere with speech sounds. These include the emotional sounds mentioned above, and the noises produced by the body's physiological functions. Quasi-speech phenomena pertain to the emotional coloration imparted to the acoustic structure of speech. The emotional and expressive acoustic phenomena listed above can be classified as situational phenomena.

Regular extralinguistic acoustic phenomena include the formal indicators of voice quality and the characteristics of articulation and intonation typical of the given speaker.

In what way are the extralinguistic signals listed above perceived and identified by the listener? In the case of emotionally colored speech, experiments performed with the purpose of identifying such signals can provide a direct answer to this question. Research conducted by many authors has shown that the principal emotions (fear, anger, joy, grief) and the state of astonishment can be encoded by acoustic resources quite adequately. The recipient is also able to determine the intensity of each of the principal emotions--that is, for example, he is able to make a differentiation between vitality, joy and ecstacy. Thus at this level the system is discretely continuous: Qualitatively different emotions fall into a discrete series, and within the limits of each of the members of the series, continuous change in intensity is possible.

There is a subsequent stage of analysis possible, involving a search for components at the second level--that is, inspection of the internal structure of acoustic expressions of the principal emotions. A highly convenient tool of analysis in this case is the methods of multidimensional psychological scaling, particularly the method of the semantic differential. When the data acquired as a result of such inspection are subjected to factor analysis, we find that the principal emotions form the foundation of a geometric model with inhibition-arousal serving as the principal dimension (4).

In terms of their internal structure, these factors are components at the second level of human perception. Thus fear can assume characteristics such as trembling, gasping, constrained and plaintive; anger may be dull, sharp or stark, and so on. It should be noted that second-level components fall within the first-level subsystem together with weight factors, and the correlations they exhibit are statistical in nature.

Perception of 25 sounds uttered by actors simulating 25 different states was studied in experiments with non-speech emotional sounds. It turned out that listeners were capable of correctly identifying all of the principal emotions by ear (astonishment and emotionally colored reactions such as pain and suffering). The method of the semantic differential resulted in factors similar to those obtained for the emotional coloration of speech. Thus at the level of the recipient, the components of the system of emotional non-speech sounds and the system of emotional colorations of speech are highly similar.

The question as to the structural components of the system of expressive acoustic phenomena has not been studied sufficiently. It is illuminated to some extent in Tseplitis' book (5), in which the author did of course use different terminology. However, if we assume that this system is composed of expressions of relationships such as tenderness, contempt, embarrassment and so on, volitional phenomena such as orders, requests and complaints, and descriptive phenomena, our experiments concerned with identification of such expressions simulated by an actor provide an approximate answer to the structure question. We found that the group of listeners was incapable of correctly and unambiguously determining most of the emotional expressions contained in standard phrases out of context. Thus tenderness is identified as pleasure and joy, and an order is perceived as perturbation and anger. Thus these signals are identified in the terms of emotional expressions.

As far as the components of the system of individual colorations of voice are concerned, experimental research conducted by Voier (6) and Galunov by the method of the semantic differential, applied to the reading of standard phrases, showed that this system also has a discrete-continuous structure with four or five of the following discrete qualities (1,6): 1) articulatory activeness-passiveness of the speaker, 2) voice volume and dimensions, 3) timbre, 4) general tone evaluation. The fifth quality (dimension), which did not appear for all authors, is usually associated with the pitch of the speaker's voice.

Continuous change of the corresponding quality is possible within the limits of each of these dimensions, and if we assume that the human ear can distinguish seven gradations in each of the four or five dimensions, an extremely rough estimate of the number of distinguishible shades of vocal tone would be from several thousand to 15,000 variants.

However, a person perceiving certain features of a voice is capable of using them to reconstruct some of the details of the speaker's appearance. Addington's work answers the question as to what the components of this appearance are. Using different variants of voice tone during reading, the author obtained the following data by the method of the semantic differential (7):

Perception of Male Voices

1. Appearance (fat, old--lean, young)
2. Passive--active
   Depressed--happy
3. Social and physical status
4. Peace-loving--cruel

Perception of Female Voices

1. Introversion--communicativeness
   Sullenness--happiness
2. Passiveness--activeness
   Obedience--aggressiveness
3. Nonemotional--emotional
4. Polish--vulgarity
   Prosperity--poverty
5. General evaluation (bad--good)

The results of our research are contained in (8). Each of the component-factors breaks down into second-level components, which may have to do with both sound and personality. Thus the negative evaluation factor is applied to nasal sound, which is associated in women with "stupidity, homeliness." In men, tension in the voice is associated with "old age, nervousness, callousness," and so on. It would be interesting to note that a speaking male is perceived to a greater extent in relation to his physical attributes, while social characteristics are more important in the perception of a woman's personality on the basis of her voice.

2. Functional Analysis

Functional analysis presupposes examination of the function of extralinguistic acoustic resources of communication. In the first phase we analyze the external function of the system in comparison with other communication resources. R. Yakobson lists six principal functions of verbal communication: intellectual, expressive, conative, factual and so on, depending on which element of a text is said to be most significant. Many of these functions can obviously be performed by extralinguistic resources; this is especially true of the expressive (emotive in our definition) and conative functions.

In living spontaneous speech, the functional load is constantly alternated between linguistic and extralinguistic signs; moreover the informativeness itself of the latter is situational in nature as well. Thus for example, verbal resources usually play the dominant role in neutral situations, while in a situation of extreme emotional arousal of the speaker, when he is capable only of uttering unintelligible sounds, their role is reduced or eliminated.

Shibutani (9) mentions another class of situations in which extralinguistic signals assume priority. These are situations that standardize verbal resources (for example at a first meeting). In these cases extralinguistic signals induce real but concealable emotions and attitudes in the communicants, and make a significant contribution to the "first impression" phenomenon.

A comparison of the effectiveness of acoustic and nonacoustic extralinguistic signals (see (10)) would reveal that signals produced by facial expressions have an advantage over intonational signals in terms of transmitting the attitudes of the communicants. It cannot be doubted that facial expressions and acoustic resources are inferior to verbal resources in terms of the accuracy with which the particular emotion is represented, but they do have their advantages as well. Experiments have demonstrated their greater resistance to interference in comparison with spoken signals, the directness of their expression, associated with the limited possibilities for voluntary control of signals at this level, and their capability for eliciting emotions, their "contagiousness." This provides the grounds for some researchers to believe that real attitudes may be expressed in communication only by nonverbal means.

Extralinguistic signals (a smile, an affectionate intonation) are resources of averting or softening aggressive behavior.

Another important feature of extralinguistic emotional signals is that they are a motor expression of internal states, and they create a dynamic picture of the function of a certain center of activity and ensure a more-sensitive reaction.

In psychology, an external expression of emotions, including an acoustic expression, is interpreted as an instinctive reaction. Many authors feel that the capability for perceiving expressions is a phylogenetically confirmed capability, for the realization of which mastery of linguistic communication is not mandatory.

There are several theories on perception of expressive movements: the inference theory, the role theory and the empathy theory (see (11)). The last theory, which belongs to T. Lipps, appears to be most plausible to us in application to this form of perception. The theory is based on three basic premises:

1. Perception of an emotional expression elicits emotional reactions within the recipient himself.

2. These reactions arise owing to realization of the need for motor imitation of another's expressive movements.

3. The emotional reaction of the recipient is ascribed by the latter to the subject being perceived.

These premises of Lipps' theory make plausible the suggestion that the actual emotional state and emotionality of the recipient have an influence on the way extralinguistic signals are perceived. Several studies taking this direction have been published in the literature, and the overall but unconfirmed conclusion they lead to is that the recipient has a tendency to project his own state upon the communicator (12). We arrived at a similar result in our research, in which students experiencing anxiety just before taking a test had to evaluate the state of speakers on the basis of the emotional coloration of their speech (13). There are data, however, indicating that it is possible to project one's state having a "minus" sign. In this case recipients are the least sensitive to states similar to the ones they themselves are experiencing at the time.

At the level of the human mind, the content of which is mediated by verbal experience, perception of nonspoken signals is dependent upon this experience and upon the personality of the communicant as a whole.

However, there are also significant differences at the level of the nonverbal animal mind. Marler (15) notes that some animals are able to generate expressive movements well, but they perceive them poorly. Others on the other hand are good recipients of state signals and poor transmitters of such signals. It is interesting that these authors revealed similar groups in an experiment with people. Various researchers have mentioned that signals of state are perceived more successfully by sensitive, subservient subjects (14). There are data indicating that individual experience has an influence on reception of emotional signals. Ramishvili (16) writes that blind subjects are able to deduce the state of a speaker from the coloration of his speech more successfully than sighted subjects.

In our experiment we wanted to follow the process of forming an impression of the speaker, and arrive at a description of this process. By interviewing subjects who had to describe the appearance and internal make-up of the speaker on the basis of the way the latter read a standard text, we were able to distinguish two extreme strategies. Having listened to the speaker's voice, some subjects fashion his visual appearance, and on the basis of this appearance they attempt to fashion his personal characteristics. Other subjects take an analytical route: They single out different characteristics of the voice and associate them one at a time with personality characteristics. Insufficient information about the speaker forces the recipient to make extensive extrapolations. Thus, hearing emotional tones in the voice, he may interpret them as a manifestation of a permanent quality of the communicator--his emotionality, while in reality they may be a situational phenomenon. A listener attempting to evaluate the personality of a speaker often employs metaphoric elements, transferring properties of the voice to properties of the possessor of this voice. Thus a person with a voice having a pleasing timbre is also assessed positively in terms of his appearance, while active articulation creates the impression of a person who is energetic in general.

Inasmuch as voice characteristics afford a possibility for considerable arbitrariness in the interpretation of the speaker's properties, such an interpretation often carries more information about the listener than about the speaker. In this case the voice may perform the role of a unique Rorschach "inkblot."

Without a doubt the voice also carries information of greater objectivity: The pitch of the voice, for example, particular features of articulation and timbre, and the hollowness-clarity of the voice provide indications of the speaker's age. General characteristics such as the psychodynamic features of the individual that make up the basis of his temperament (activeness and emotionality) are represented rather fully in the voice. As we had seen earlier, these characteristics are partially represented by those criteria-factors which listeners use to evaluate the personality of the speaker. On the whole, however, individual voice characteristics cannot serve--we agree with A. A. Bodalev in this regard--as dependable indicators of the personality as a whole (11).

3. Historical Analysis

The next direction of analysis presupposes examination of the historical roots of the phenomenon--its origin and development. Without a doubt this topic requires special and lengthy analysis. We will cite only the raw data used in such analysis.

C. Darwin (17) is the author of the first theory of expressive movements which, in particular, compares the external expression of the principal emotions in man and higher animals and proves the existence of a phylogenetic relationship between them. Without a doubt this general premise is valid in relation to the resources of acoustic expression of emotions. Confirmations of Darwin's theory can also be encountered in modern studies. Man's experimentally proven capability for identifying the state of animals through auditory perception of their acoustic signals is such a confirmation, though of course indirect (18). References to the capability of animals, especially domestic ones, of sensing the state of an individual are often encountered in the literature (19). All of this attests to presence of similar factors in the external, and particularly the acoustic expression of the principal emotions of man and higher animals. However, if this is valid in relation to non-speech sounds, how do matters stand with the emotional coloration of speech? Experiments involving identification of the emotional state of a speaker on the basis of stressed syllables or even vowels extracted from a passage of emotionally colored speech show that these elements transmit a significant amount of information on the state of the speaker. At the same time, these elements are also the result of interference between phonetic and psychophysiological processes. A number of researchers have noted, in relation to both vocal and conversational emotionally colored speech, presence of the "audible smile" phenomenon, or tearful, plaintive intonation, sighs of annoyance or moans of pain directly within the structure of the spoken statement (20,21). Moreover the factor structure of listeners' evaluations based on emotionally colored phrases of standard content is also similar according to our data (22). Therefore there are grounds for assuming that the emotional coloration of speech is a derivative phenomenon of emotional non-speech sounds. The question as to the presence of non-speech emotional sounds in speech is extremely interesting. On one hand they are present in speech in their first-existing form and in the form of emotional coloration of speech. On the other hand, if we assume the position of many scientists who believe affective sounds to be among the raw material for creation of language, they are present in language in transformed, removed form. It would be interesting to find their traces in the phonemic content and in the nonemotional intonation of language, and in particular to try to examine the phenomenon of phonetic symbolism from these positions.

As far as expressive acoustic resources are concerned, their inventory is diverse, and it includes phenomena existing at different planes, ones which may have different historical roots. The one thing that is sure is that they all have a later origin in comparison with emotional resources. There is almost no mention of the evolution of expressive acoustic resources in the literature. This question is touched upon only in Bubrikh's work (23). From this author's point of view we can distinguish three stages in the development of language: The first is typified by visual-objective thinking and by speech having a signaling function. In the second stage thinking assumes a visual-descriptive nature, and representational and expressive resources come into being.

Discussing the problem as a whole, we can assert that man typically uses more-ancient acoustic resources when emotionally aroused--that is, his speech typically undergoes acoustic primitivization, using E. L. Nosenko's term as he applied it to the semantics and syntax of speech in stressful conditions.

3. Conclusion

Two points require consideration in the conclusion. First of all the initial scheme of analysis, which contains three basic dimensions, should be supplemented by a fourth dimension associated with semeiotic analysis of extralinguistic signals, inasmuch as the system under discussion here is a communication system. But this approach has not been fully worked out yet in application to such signals, and it still requires extensive work prior to its implementation.

The second point is associated with the problem of determining the type of emotional state on the basis of the acoustic characteristics of speech. The characteristics which researchers have at their disposal still do not permit determination of the type of state; nor do they even permit differentiation between positive and negative emotions. Despite the fact that some authors have reported the possibility of making such a diagnosis, the characteristics they suggest are contradictory (24,25).

We suggest another approach to solving this problem based on the considerations discussed above concerning the nature of the emotional coloration of speech. In the first stage of this approach we would need to subject the acoustic correlates of an external expression of emotions (emotional sound) to meticulous analysis, and in the next stage we should seek these characteristics in a spoken signal having an emotional coloration.

BIBLIOGRAPHY

1. Nosenko, E. L., "Osobennosti rechi v sostoyanii emotsional'noy napryazhennosti" [Characteristics of Speech in a State of Emotional Tension], Dnepropetrovsk, 1975.

2. Kagan, M. S., "Chelovechaskaya deyatel'nost' (Opyt sistemnogo analiza)" [Human Activity (An Experiment in Systems Analysis)], Moscow, 1974.

3. Galunov, V. I., "Speech, Emotions and Personality: Problems and Prospects," in this collection.

4. Manerov, V. Kh., "Investigation of the Spoken Signal to Determine the Individual's Emotional State," Candidate Dissertation Abstract, Leningrad, 1975.

5. Tseplitis, L. K., "Analiz rechevoy intonatsii" [Analysis of Speech Intonation], Riga, 1974.

6. Voier, W., "Perceptual Bases of Speaker Identity," J. ACOUST. ASS. AMER., Vol 36, No 6, 1964,

7. Addington, D., "Voice and Perception of Personality," N. Y., 1968.

8. Alekseyev, V. I., Manerov, V. Kh., and Ustinovich, Ye. A., "Analysis of Voice as a Source of Information of Properties of the Speaker," see this collection.

9. Shibutani, T., "Sotsial'naya psikhologiya" [Social Psychology], Moscow, 1969.

10. Mortensson, C., and Sereno, K. K., "Advances in Communication Research," N.Y., Harper, 1973.

11. Bodalev, A. A., "Vospriyatiye cheloveka chelovekom" [Perception of Man by Man], Izd-vo LGU, Leningrad, 1965.

12. Kvasovets, S. V., "Opyt izucheniya emotsional'nykh sostoyaniy. Problemy neyropsikhologii" [Experience in Studying Emotional States. Problems of Neuropsychology], Moscow, Nauka, 1977.

13. Il'in, Ye. P., Manerov, V. Kh., Katygin, Yu. A., and Shatalova, T. N., "Effect of Pretesting Arousal on Evaluation of Speech Tone" (in press).

14. Korneva, T. V., "Some Factors Defining the Accuracy of a Listener's Evaluation of Emotional States," see this collection.

15. Marler, in Krames, L., et al. (Editors), "Nonverbal Communication," N.Y.-London, 1974, X, p 202 (Vol 1, "Advances in the Study of Communication and Affect").

16. Ramishvili, D., "K prirode nekotorykh vidov vyrazitel'nykh dvizheniy" [The Nature of Some Forms of Expressive Movement], Metsnireba, Tbilisi, 1976.

17. Darvin, Ch., "Vyrazheniye dushevnykh volneniy" [Expression of Spiritual Agitation], St. Petersburg, 1996.

18. Gershuni, G. V., Bogdanov, B. V., Vakarchuk, O. Yu., Mal'tsev, V. P., and Chernigovskaya, T. V., "Human Identification of Different Types of Acoustic Signals Emitted by Monkeys," FIZIOLOGIYA CHELOVEKA, Vol 2, No 3, 1976.

19. Lorents, K., "Kol'tso tsarya Solomona" [King Solomon's Ring], Moscow, 1977.

20. Morozov, V. P., "Biofizicheskiye osnovy vokal'noy rechi" [Biophysical Principles of Vocal Speech], Leningrad, 1977.

21. Kotlyar, G. M., "Analysis of Acoustic Resources for Expressing Emotional States in Vocal Speech," Candidate Dissertation Abstract, Leningrad, 1977.

22. Manerov, V. Kh., "Analysis of Emotional Non-Speech Sounds," (in press).

23. Bubrikh, D. V., "Origin of Thinking and Speech," in Anisimov, A. F., "Istoricheskiye osobennosti pervobytnogo myshleniya" [Historical Characteristics of Primitive Thinking], Nauka, Leningrad, 1971.

24. Blokhina, L. P., and Gomina, T. G., "Significance of Prosodic and Spectral Parameters of Spoken Signals Expressing Different Emotional States," see the proceedings of this symposium.

25. Taubkin, V. L., "Identification of the Emotional State of a Human Operator Using Spoken Signal Parameters," Candidate Dissertation Abstract, Moscow, 1977.

USING SYMMETRICAL BIOLOGICALLY ACTIVE POINTS TO MONITOR CHANGES IN HUMAN PSYCHOPHYSIOLOGICAL STATE

A. S. Abduakhadov, V. I. Galunov

The goal of our study was to develop a method for objectively monitoring the psychophysiological state (PPS) of an individual. An analysis of the available data would show that traditional objective physiological indicators such as pulse, respiration, EEG, GSR and so on are not sufficiently informative when it comes to determining a number of psychophysiological states (PPS's). Thus there is interest in studying the biologically active skin points (BASP's) used in acupuncture therapy to elicit change in psychophysiological state (1). Ours is an attempt to study the characteristics of bioelectric reactions (BER's) of biologically active skin points (BASP's) in the presence of different positive and negative emotional states.

Research Methods

The research was conducted on 10 healthy subjects 21-31 years old in a comfortably appointed soundproof room. To simulate emotional states of different signs, the subjects were asked to act out or, for practical purposes, autosuggest negative and positive emotional experiences. Thus realistic emotional states were produced without their active motor component.

BER's were recorded with a 16-channel "Al'var" electroencephalograph (GDR) using nonpolarizing (platinum) electrodes with a diameter of 2 mm. The inter-electrode distance was 3 mm. Symmetrical general-action BASP's (designated Gi-4 by international convention) and symmetrical inactive points on the palms of both hands were the object of research. A cardiogram was recorded in parallel.

The timing of the experiment was as follows: First we recorded the initial BER of symmetrical BASP's and inactive points, the cardiogram and the pulse (15 minutes). After a little while we recorded the physiological parameters indicated above for 10 minutes, during which the subject imagined a negative emotional situation accompanied by the appropriate state (fear, melancholy), during the time he remained immersed in this state (10 minutes), and during his emergence from this state (20 minutes). Following this, the entire course of the experiment was repeated with the subject imagining a positive emotional situation accompanied by joy and pleasure.

The PPS indicators noted above were then compared with the particular state and with the properties exhibited by symmetrical BASP's. The coefficients of asymmetry (K_{as}) of the parameters of symmetrical BASP's and of symmetrical inactive points in the presence of different emotional states were calculated using the formula:

$$K_{as} = \frac{A-B}{A+B},$$

where A--parameters of points on the left hand, B--on the right hand.

Research Results

The BER recorded from symmetrical BASP's in response to voluntarily induced emotional states of different signs may be described as a slow oscillatory process having a wide frequency range (from 0.2 to 2 Hz) and an amplitude range from 50 to 900 μv. BER's recorded from symmetrical BASP's (Gi-4) in response to the different emotional states studied are described below.

Complete emotional rest was adopted as the initial state. In it, BER's recorded from symmetrical BASP's were unstable in amplitude, asynchronous, irregularly arising oscillations consisting of primary biphasal components lasting about 1 sec and secondary, negative monophasal late components lasting 3-4 sec. In all cases, potentials of significantly higher amplitude were recorded from the right Gi-4 point.

When the subject was in a state of emotional relaxation (probably similar to a state somewhere between drowsiness and sleep) the potential oscillations exhibited a tendency toward synchronization. Oscillations recorded from the right Gi-4 point were biphasal in shape, with the amplitude of the primary components being 400-800 μv and their duration being 1 sec. The primary oscillation was followed by a secondary monophasal negative oscillation with an amplitude of about 200 μv and a duration of 3-4 sec. Biphasal oscillations of irregular shape lasting 1 sec and having a lower amplitude (300-500 μv) were recorded from the left Gi-4 point. The secondary component, which had a duration of 2-3 sec, had a rounded peak.

The nature of the BER potential oscillations changed dramatically in amplitude and frequency when the subject imagined a state of fear. The oscillation frequency of potentials recorded from symmetrical BASP's was greater than normal. On the background of a general increase in frequency, statistically significant asymmetry of this indicator was observed, expressing itself as a higher oscillation frequency at the left Gi-4 point in comparison with the right. Later on, as the subject emerged from his negative emotional state, the frequency of BER's recorded from symmetrical points gradually fell. It should be noted that in addition to this, we recorded low amplitude monophasal negative oscillations with a duration of 1 sec and secondary long-lasting positive oscillations (2 sec) from the right Gi-4 point.

At the left Gi-4 point, amplification of the frequency characteristics of the BER is also accompanied by change in the amplitude of the potentials. They acquire an approximately sinusoidal biphasal shape. Secondary slow oscillations disappear. Alternation of oscillations of average size with shorter, asymmetrical oscillations is observed.

The changes described above in BER's recorded from symmetrical BASP's assume a different form when we proceed to a positive emotional state (joy). Amplification of the frequency characteristics of BER's recorded from symmetrical BASP's can be noted; however, this increase is expressed to a lesser degree than with a negative state. The asymmetry of BER frequency characteristics was predominantly right-sided. Arisal of two-component oscillations, with an initial fast negative component and a late biphasal positive-negative slow component (4-5 sec) was noted at the right Gi-4 point. At the left Gi-4 point the oscillations were biphasal, positive-negative and slow; their shape was irregular, and their duration was 3-4 sec.

Thus these data reveal differences in the frequency and amplitude characteristics of BER's recorded from symmetrical Gi-4 BASP's in response to emotional states of different signs.

The changes indicated above in BER's recorded from symmetrical BASP's correlate with the dynamics of electric resistance (total resistance, to include its active component and its capacitive component) recorded in similar experimental conditions. They also agree in part with the dynamics of pulse changes occurring in the corresponding states. Pulse grew faster with positive and negative states, and decreased in the emotionally relaxed state.

The high effectiveness of using BASP's having general restorative action to detect emotional states of different signs can be demonstrated by comparing BER's recorded from biologically active and inactive points on the skin. Our experiments showed that the size and shape of the electric potential and the asymmetry exhibited by the dynamics of amplitude and frequency characteristics recorded from symmetrical BASP's are significantly more precise indicators of psychophysiological state than are similar characteristics recorded from inactive points, for which the electroenencephalographically recorded BER's either do not change at all in response to similar situations, or they fluctuate insignificantly about the zero point.

Considering our present knowledge of the mechanisms responsible for the work of BASP's, it may be hypothesized that amplification of the amplitude and frequency characteristics of BER's recorded from symmetrical BASP's of general tonic action in response to emotional states of different signs reflects nonspecific activation of the central mechanisms of the states examined here. Asymmetry of BER indicators recorded from symmetrical BASP's probably reflects activation of cortical mechanisms of the studied emotional states, and it may be associated with functional lateralization of positive emotions in the left hemisphere and of negative emotions in the right hemisphere (2).

The thus the research permits the following conclusions:

1. Biologically active points that cause change in psychophysiological state when affected in a certain way may also serve as a source of information on state.

BER's and the resistance to electric current recorded at BASP's of general tonic action are promising indicators of qualitatively different changes in emotional state.

3. The amplitude and frequency characteristics of BER's recorded from symmetrical BASP's differ in relation to emotional states of opposite sign.

Asymmetry in the dynamics of the amplitude and frequency characteristics of BER's recorded from symmetrical BASP's is the principal indicator of emotional states differing in sign.

BIBLIOGRAPHY

1. Chzhu Lyan', "Rukovodstvo po sovremennoy chzhen'-tszyuterapii" [Handbook of Modern Acupuncture Therapy], Moscow, Medgiz, 1959.

2. Gazzaniga, M. S., "The Bisected Brain," N.Y. 1970.

ANALYSIS OF VOICE AS A SOURCE OF INFORMATION ON PROPERTIES OF THE SPEAKER

V. I. Alekseyev, V. Kh. Manerov, Ye. A. Ustinovich

Being a means of communication, speech is a system of signs organized in a particular way. The principal sign of this system is the word. Thus speech is traditionally studied as a means of verbal communication. But recently researchers have shown increasingly greater interest in parameters of speech having to do with nonverbal communication. Acoustic phenomena (prosodic characteristics, sonorousness, pronunciation) which accompany speech and which may bear information supplementing the meaning of a statement are now becoming an object of study.

A researcher studying perception of these acoustic phenomena is able to distinguish a group of speech elements that are constant characteristics of speech formation which a listener can use to form an idea as to the age, education, appearance and the psychophysiological and characterological features of the speaker. In other words this group of speech parameters generates an impression about the speaker--that is, it has an impressive function. Voice has special significance among these parameters. Our study was devoted to the influence of vocal characteristics on the listener's resulting impression of the speaker's properties.

Psycholinguistic studies of voice characteristics (4,5,6) distinguish the following descriptive qualities of speech sounds: pitch, loudness, speed, rhythm, timbre, melody, sonorousness, intensity. The impression an audience develops about the speaker influences the content of his message--that is, together with other factors it predetermines the effectiveness of mass media. It was for this reason that scientific investigation of the role played by these speech characteristics in formation of an impression about the individual features of a speaker began with the development of radio broadcasting. Thus the first experiments were conducted in this area by the British Broadcasting Company in 1931 (3).

A researcher addressing this problem must answer the following questions:

1. What impressions does the speech of a speaker induce in his listeners?

2. How is the speaker's voice described, and what characteristics used in its evaluation are correlated with opinions about the speaker?

3. How do opinions about the speaker, arrived at as a result of listening to his speech, correlate with data obtained by other means (for example with personality inventories)?

4. What physically recorded parameters of a spoken signal predetermine the opinion of a speaker?

Thus four sets of data must be compared: the physical description of the spoken signal, its subjective description (the voice), opinions about the speaker's properties and the psychophysiological indicators of the speaker's individual features.

Our research dealt with problems associated with the first two groups. Voice evaluations and opinions about speakers based on such evaluations can be obtained by the well known method of semantically opposite pairs suggested by C. Osgood (1), which is broadly employed today in research on auditory perception.

Psycholinguistic studies that have made use of this method (2) show that the "field of meanings" is determined by two systems of evaluations. The first system, referred to by Osgood as affective, contains three factors: evaluation, strength and activeness. The second system is formed out of denotative characteristics, and it reflects the physical properties of the object. Inasmuch as the objective of our work was to isolate those vocal characteristics which predetermine opinions about the properties of the speaker, denotative evaluations are of the greatest importance to us. In addition to characteristics used in the affective evaluation (good-bad, strong-weak and so on), the list of characteristics intended for voice evaluation included terms describing the voice qualities listed above--three or four terms for each quality. For example high-low, loud-soft, fast-slow and so on. In all, the list consisted of 36 pairs of adjectives.

The second list we used in the listening sessions consisted of 54 characteristics describing different properties of the speaker. It included some indicators of appearance (tall-short, thin-fat) and age, and terms standing for emotional, volitional, intellectual and characterological features (emotional-unemotional, anxious-serene, smart-stupid, willful-unwillful). In this case each of these features was described by approximately the same number of terms.

Subjects referred to these two lists as they listened to tape recordings of speech excerpts by five male speakers. The excerpts, each about 2 minutes long, were fragments of undirected speech. During the recording sessions, the speakers were asked to speak freely about some interesting event in their life.

Subjects listened to the recordings in groups. In all, 60 persons listened to each voice. The instructions did not limit the listening time, and they required the subjects to pay no attention, to the extent possible, to the content of the excerpts.

Ten evaluation intercorrelation matrices were obtained on treating the results: five matrices (36×36) of voice evaluations and five matrices (54×54) of evaluations of the speaker's individual properties. They were subjected to factor analysis by the main factor method. The calculations were performed with an M-222 computer. In addition we obtained combined matrices of voice evaluations and evaluations of individual characteristics of the speaker. These matrices were compiled on the basis of evaluations of five voices listened to by 20 audiences. Thus we subjected (54×100) and (36×100) matrices to factor analysis. As a result we obtained the invariant factor structures of evaluations shown in tables 1 and 2.

Table 1. Invariant Factor Structure Obtained Through Factor Analysis of Composite Matrices of the Evaluations of Individual Features of the Speaker (Characteristics With Maximum Weights Are Shown)

Factor	Characteristics	Weight
I	Decisive-vascillating	+0.73
	Stern-gentle	+0.71
	Fearful-fearless	-0.71
	Compliant-competitive	-0.71
	Willful-unwillful	+0.70
	Delicate-hard	-0.70
	Confident-unconfident	+0.68
	Imperious-quiet	+0.68
II	Slow-fast	+0.82
	Gay-depressed	-0.80
	Hurried-slow	-0.80
	Passive-active	+0.78
	Woeful-joyful	+0.76
	Energetic-inert	+0.76
	Introverted-communicative	+0.75
III	Conscientious-irresponsible	+0.67
	Serious-flippant	+0.66
	Rough-delicate	-0.64
	Intellectual-primitive	-0.64
	Smart-stupid	+0.61
IV	Lean-stout	+0.60
	Fat-thin	-0.55
	Short-tall	+0.50
	Age	+0.50
	Stalwart-dwarfish	+0.47
	Old-young	-0.43
V	Anxious-serene	+0.47
	Nervous-calm	+0.44
	Relaxed-taut	-0.44

The first factor that was isolated by factor analysis of the combined matrix of evaluations of the individual properties of the speaker included the characteristics decisive-vascillating, delicate-hard, purposeful-spontaneous, willful-unwillful. These characteristics describe volitional qualities of the individual, and the factor may be defined as firmness or strength.

The second factor includes the characteristics passive-active, woeful-joyful, slow-fast, inert-energetic, quiet-talkative. This can quiet definitely be interpreted as an activeness factor.

The third factor of this invariant structure brought together the characteristics conscientious-irresponsible, serious-flippant, smart-stupid. This can probably be interpreted as the intelligence factor.

The fourth factor consisted of characteristics describing appearance (weight and heigt) and age. It is interesting that the characteristics fat-thin and short-tall are positively correlated, as is the case with all individual matrices. The fifth factor, which includes the characteristics anxious-serene, nervous-calm, can probably be interpreted as a tension factor which is associated with anxiety and impulsiveness.

Individual factor matrices of the evaluations of speaker properties are combinations of these factors. In some cases we revealed special factors in addition to these. As an example the final factor matrix of evaluations of speaker No 1 contains factors uniting the characteristics rough-delicate, dreamer-doer, confident-unconfident.

Table 2. Invariant Factor Structure of Voice Evaluations

Factor	Characteristics	Weight
I	Fluent-stumbling	0.83
	Uniform-nonuniform	0.76
	Rhythmical-nonrhythmical	0.50
	Tense-relaxed	-0.47
	Free-constrained	0.40
II	Squeaky-deep	0.80
	High-low	0.74
	Thin-thick	0.74
	Bright-dark	0.63
	Fast-slow	0.62
	Sharp-dull	0.61
III	Rich-dry	0.67
	Bad-good	-0.66
	Deep-cold	0.69
	Hollow-full	-0.59
	Unpleasant-pleasant	-0.57
IV	Dull-clear	0.76
	Monotonous-modulated	0.72
	Muffled-deafening	0.69
	Unpleasant-pleasant	0.53
	Bright-dull	-0.51
	Hoarse-clear	0.42
V	Loud-soft	-0.60
	Mobile-inert	-0.58
	Fast-slow	-0.56
	Hurried-unhurried	-0.48
	Harsh-mild	-0.40
VI	Distinct-inarticulate	0.75
	Comprehensible-incomprehensible	0.81
VII	Nasal-non-nasal	0.51

The invariant factor structure of voice evaluations, shown in Table 2, contains seven factors. The first factor combines characteristics describing the rhythm of speech and its intensity. This means that the "intensity" characteristic is determined by the rhythmical pattern of speech. The second factor may be interpreted as the voice pitch factor. The third factor includes characteristics describing timbre (rich-dry, warm-cold), and valuational characteristics. The fourth factor may most likely be interpreted as a sound fullness factor. The fifth factor basically involves speech rate evaluations. The sixth and seventh factors contain the largest number of characteristics, with the sixth most probably characterizing the recording quality. It can be noted that evaluations having to do with loudness did not compose a single factor, instead falling within the fifth and sixth factors.

The next step in the analysis was to obtain correlations between voice evaluations and evaluations of the individual properties of the speaker. Characteristics having the maximum factor weights were selected out of the factor matrices of evaluations of the first speaker. In all we selected 52 characteristics, which we subjected to correlation analysis (20 reflect voice qualities while 22 describe individual characteristics of the speaker). The most significant coefficients were obtained for correlations between the following characteristics:

Fluent-stumbling, woeful-joyful	0.40
Fast-slow, energetic-inert	0.43
Full-hollow, willful-unwillful	0.54
Deep-squeaky, willful-unwillful	0.53
Dull-clear, fearful-fearless	0.43
Hoarse-not hoarse, willful-unwillful	0.54

The following conclusions can be made:

1. The invariant structure of voice evaluations consists of five basic factors that may be interpreted as rhythm, rate, timbre, pitch and fullness.

2. The invariant structure of evaluations of individual properties of the speaker includes five factors: Activeness, will (firmness), intensity, intelligence and appearance.

3. Fast speech creates the impression that the individual is active and energetic. A low, dull, full voice is associated with a person who is purposeful, willful and decisive. Rhythmical speech is evaluated as a sign of an elevated mood.

BIBLIOGRAPHY

1. Osgood, C. E., Suci, G. H., and Tannenbaum, P. H., "The Measurement of Meaning," Urbana, 1957.

2. Tzeng, O., and May, W., "More Than E.P.I. Semantic Differential Scales," INTERNATIONAL JOURNAL OF PSYCHOLOGY, Vol 10, No 2, 1975.

3. Pear, T. H., "Voice and Personality," N. York, 1931.

4. Voier, W. D., "Perceptual Bases of Speaker Identity," JASA, Vol 36, No 6, 1964.

5. Addington, D. W., "Voice and Perception of Personality," N. York, 1968.

6. Farmann, R., "Die Deutung des Sprechaus drucks," Bonn, 1960.

THE SEMANTIC SPACE OF IDEAS ASSOCIATED WITH EMOTIONALLY COLORED SPEECH

Ye. F. Bazhin, G. A. Krylova

Speech occupies an important place in the communicative function belonging to expression. Speech is one of the sources of information indicating presence of a certain emotional state and an emotional relationship. It is commonly accepted (1) that this information is contained in three levels of speech: acoustic-phonetic, lexical-grammatic and semantic. Researchers are especially interested in the first level--that is, one having to do with almost direct transmission of so-called affective language contained in emotional intonation--in changes in the physical characteristics of the voice that are perceived and decoded by the listener. The task of searching for and discovering concrete prosodic characteristics bearing information on emotional state is in general a solution to part of the theoretical problem of man's perception by man. In the applied sense, on the other hand, the objective of this task is to develop an automatic system for recognizing emotions on the basis of the spoken signal.

This problem was studied in our previous work from the standpoints of both finding the objective correlates of emotional states in the melody of voice (2,3) and determining the possibilities of expert (listener) evaluation--that is, identification--of the emotional color of speech (4). To a certain extent these two approaches are opposites of one another: Thus on one hand we employed instrumental analysis based on analyzing purely physical characteristics, while on the other hand we studied the capability for communication associated with psychological features of a given individual as a personality. This paper describes an attempt to find something that would bring these two approaches closer together, something that could serve as a link between them. One such linking tool, we believe, is language, and namely its semantic wealth, which is used, among other purposes, for transmission of information on emotional state.

The immediate objective of our work was to reveal the semantic space of Russian language containing terms describing the emotional characteristics of voice and speech. The fullest and most reliable source of such terms is creative literature, in which the ideas of the author, his philosophy and his outlook on the world are expressed by means of various descriptive resources, to include character descriptions or expressions. A character description sometimes provides a direct indication as to whether or not a certain emotional state is inherent to the hero, and it may indicate the attitude of the narrator (the author) toward that state. Manufacturing an emotion-inducing situation by means of his creative imagination, the author serves as a unique sort of transformer, translating his ideas about the way the speech of his heroes sounds into specific terms, the adequacy, clarity and completeness of which define the artistic expressiveness of the work and promote transmission of the desired information, making it understandable to a large number of people (readers). In other words the objective of the author is to transmit acoustic characteristics at the verbal level, to place them in semantic space in such a way that they could be decoded--understood and felt. In this sense each writer acts as an experimental psychologist on one hand, reconstructing the behavior of people in different situations, and on the other hand he is a linguist-philologist, having an abundant lexicon at his disposal.

Our method required the study of the creative works of 16 Russian and Soviet writers (I. Bunin, N. Garin-Mikhaylcvskiy, M. Gor'kiy, F. Dostoyevskiy, A. Kurpin, N. Leskov, D. Mamim-Sibiryak, K. Paustovskiy, S. Sergeyev-Tsenskiy, A. Tolstoy, L. Tolsoy, I. Turgenev, K. Fedin, A. Chekhov, M. Sholokhov and I. Erenburg). Three excerpts from the works of each writer were analyzed. Each excerpt contained a standard number of characters--200,000; in all, we studied 48 such excerpts with a total volume of 960,000 characters. Expressions used to describe voice and speech were extracted from the text. For the convenience of analysis these expressions were given in adjectival form, for example "loud," "languid," and so on.

Table 1 contains the text analysis data. We can see from the table that the number of times different writers make references to the color of voice and speech varies broadly. A certain trend is evident, however: For about half of the authors analyzed this number did not exceed 250-270, while for the rest it was significantly higher. Consequently the frequency with which references are made to voice and speech in standard excerpts from creative works of like genre may vary.

A similar situation was also revealed on analysis of the terminological structure of these references--that is, the writer's lexicon (see Table 1). Here the difference was rather large--69 terms for N. Leskov and 237 for A. Kuprin.

Table 1

Writer	No. of References	No. of Terms	Ratio of No. of References to No. of Terms
1. I. Bunin	270	147	1.84
2. N. Garin-Mikhaylovskiy	239	108	2.21
3. M. Gor'kiy	605	171	3.53
4. F. Dostoyevskiy	236	128	1.84
5. A. Kuprin	719	237	3.03
6. N. Leskov	168	69	2.43
7. D. Mamin-Sibiryak	379	134	2.82
8. K. Paustovskiy	410	148	2.78
9. N. Sergeyev-Tsenskiy	220	121	1.81
10. A. Tolstoy	558	142	3.92
11. L. Tolstoy	159	85	1.87
12. I. Turgenev	267	118	2.26
13. K. Fedin	536	202	2.65

[continued on following page]

14. M. Sholokhov	527	185	2.85
15. A. Chekhov	306	109	2.8
6. I. Erenburg	371	118	3.14
	373	139	2.61

There is another interesting indicator in Table 1: the ratio between the total number of references to voice and speech encountered in a text, and terminological diversity--that is, the total number of terms used by a writer. This indicator illustrates the average frequency with which some term is used throughout the entire analyzed text--that is, 600,000 characters. As we can see from the table, the size of this indicator increases as the number of references in the text to voice and speech increases, while on the other hand it decreases as the lexicon becomes relatively less rich--that is, as relatively fewer terms are used to define the characteristics of voice and speech.

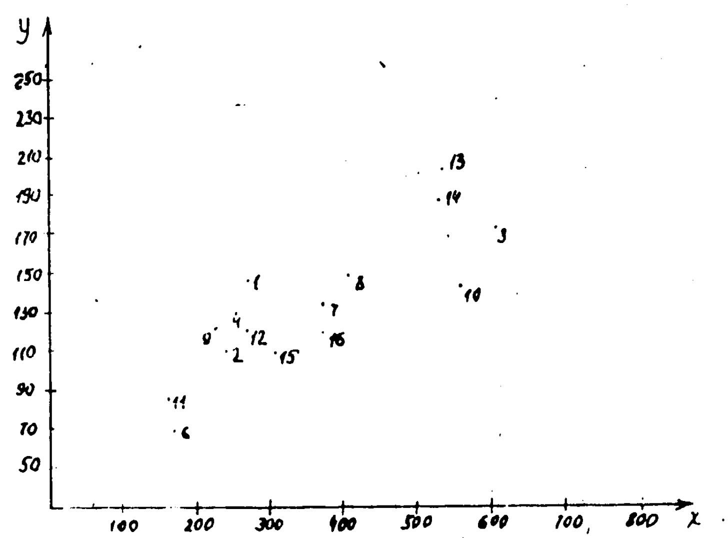

Figure 1

X axis--number of references in the text to different characteristics of voice and speech. Y axis--number of different terms used to describe voice and speech. 1--I. Bunin, 2--N. Garin-Mikhaylovskiy, 3--M. Gor'kiy, 4--F. Dostoyevskiy, 5--A. Kuprin, 6--N. Leskov, 7--D. Mamin-Sibiryak, 8--K. Paustovskiy, 9--N. Sergeyev-Tsenskiy, 10--A. Tolstoy, 11-L. Tolstoy, 12--I. Turgenev, 13--K. Fedin, 14--M. Sholokhov, 15--A. Chekhov, 16--I. Erenburg.

The laws revealed by analysis of the texts may be visualized in Figure 1, in which the total number of references in the text to voice and speech are plotted on the X axis and the number of terms (definitions) describing voice and speech is plotted on the Y axis. We can see from the figure that between these indicators there exists a dependence that would best be described as linear--that is, as the number of references to voice and speech increases, the number of terms employed--that is, the terminological diversity--grows as well.

The total number of different terms referring to voice and speech encountered in texts written by the 15 analyzed authors was 611. Analysis of the semantic content of these terms, used to describe voice and speech (we used a 16-volume modern Russian language dictionary published by the USSR Academy of Sciences), showed that it (this content) may be divided into three basic categories.

The first included terms describing the acoustic-phonetic characteristics of voice and speech directly: for example "whining, melodic, shrill," and so on. We revealed 150 such definitions, making up 25.5 percent of the total number of terms.

The second category contained terms metaphorical in nature, for example"urbane, thick, reedy, oily" and so on. These totaled 45--that is, 7 percent.

And finally, the third and largest category contained terms used by the authors to provide relatively direct information as to the presence of a concrete emotion: "hopeless, anxious, elated, melancholy, angry" and so on. We discovered 416 such terms--that is, 67.5 percent.

The next stage of analysis involves more-detailed classification within each of the categories of terms isolated above. Thus for example, six different factors were revealed in relation to acoustic-phonetic terms directly characterizing voice and speech (terms in the first category): intensity (60 terms)--that is, 40 percent of the total number of 150 terms, pitch (24-16 percent), speed (19-15.4 percent), rhythm (10-12.6 percent), timbre (10-6.7 percent) and distinctness (14-9.3 percent).

The terminological lexicon resulting from this sampling procedure was also subjected to analysis from the standpoint of the frequency with which individual terms were used. By studying weight factors we were able to arrive at a "mandatory" set of terms--that is, one common to all writers; obviously, this would also be a unique sort of semantic summary of Russian-language terms used to describe voice and speech.

On the other hand this terminological lexicon (containing 611 terms) permitted us to reveal clusters of synonyms, within which we determined, by analysis of weighted evaluations, the central terms and the distances by which they were separated from their synonyms. This in turn allowed us to represent these laws in the form of three-dimensional concept models for key terms such as, for example, "loud," "joyful" and so on.

In general the obtained data provide an impression of the semantic space of Russian language, within which key concepts describing voice and speech are located as individual points surrounded by synonymous terms, ones which obviously emphasize certain shades of meaning. This material will be used in the future to develop a specialized semantic differential that could be used to compile a terminological lexicon for voice, this time expressing concrete emotions--melancholy, joy, anger and so on.

The research method we used, which is essentially one of the variants of content analysis, appeared sufficiently adequate and promising to us. And in fact, we are dealing with writers whose works possess remarkable realistic strength of description and influence upon the reader's imagination. It is not difficult for us to imagine the visual picture of the heroes of I. Bunin, L. Tolstoy or F. Dostoyevskiy--their appearance, voice and manner of speaking. Therefore we can assume that the terms used by these authors are a unique source of information on ideas about emotionally colored voice and speech typical of man in general.

BIBLIOGRAPHY

1. Zhinkin, H. I., "Mekhanizmy rechi" [Speech Mechanisms], Moscow, 1958.

2. Bazhin, Ye. F., Galunov, V. I., Gorskiy, G. D., Manerov, V. Kh., and Khvilivitskiy, T. Ya., "Analysis of Prosodic Characteristics of the Speech of a Speaker Experiencing Different Emotional States," in "Analiz i sintez kak vzaimo-obuslovlennyye metody eksperimental'nykh foneticheskikh issledovaniy" [Analysis and Synthesis as Mutually Dependent Methods of Experimental Phonetic Research], Minsk, 1972.

3. Bazhin, Ye. F., Galunov, V. I., Gorskiy, G. D., and Manerov, V. Kh., "Objective Diagnosis of Emotional State in the Psychiatric Clinic on the Basis of Speech," in "Rech' i emotsii" [Speech and Emotions], Leningrad, 1975.

4. Bazhin, Ye. F., Vuks, A. Ya., and Koriyeva, T. V., "Possibilities for Recognizing Emotions on the Basis of an Isolated Spoken Signal," in "Psikhologicheskiye problemy psikhogigiyeny, psikhoprofilaktiki i meditsinskoy deontologii" [Psychological Problems of Mental Hygiene, Preventive Psychology and Medical Deontology], Leningrad, 1976.

A PACKAGE OF TESTS TO STUDY PERCEPTION OF EMOTIONAL SPEECH

A. V. Beskadarov, L. I. Vasserman, I. M. Tonkonogiy

Study of the emotional coloration of speech has recently been attracting the interest of specialists in different fields. On one hand it is an object of interest of researchers working on systems intended to monitor operator states. On the other hand the emotional coloration of speech is being subjected to detailed study in psychiatry, neurology, medical psychology and so on. Researchers attempting to determine the state of an individual on the basis of his speech are encountering certain difficulties in doing so. They are associated primarily with the absence of informative characteristics that would allow reliable differentiation between emotional states.

The role played in communication by speech formed out of the words of a concrete language is universally recognized. The characteristics of such languages are being studied in numerous linguistic, psychological and sociological research programs. But the significance of the language of emotions and of other paralinguistic forms of spoken communication continues to be significantly outside the field of view of the researchers, though these nonverbal forms of communication play a significant role in the individual's activity, in his recognition of a situation, in decision making and in evaluating the results of action and behavior. From our point of view emotions are among the simplest and most meaningful languages used by the human brain as it receives, stores and transmits information. This language is apparently limited to about 20 key concepts which the individual can use to arrive at a rough assessment of most situations and results of action. This significantly limits the number of classes of such situations that need to be identified, and it significantly facilitates and abridges information processing by the human brain, especially when frequently encountered, repeating events are involved. A significant advantage of this language is its genetic substrate--the possession of the same language of emotions by all people. It is a unique Esperanto permitting communication of one person with another, of a mother and a child in the first days of its life, of people of different nationalities. It also facilitates communication between people using conventional verbal language by imposing a general value judgment on events being described by verbal messages.

However, the information which we possess from research on human emotions is highly limited, even in regard to identification of emotions on the basis of facial expressions and voice characteristics. This article describes a method aimed at studying these indicators in patients suffering local brain lesions, since the study of patients with focal pathology may provide significant help in gaining an understanding of the cerebral mechanisms of emotional language. In order that other forms of nonverbal speech could be studied as well, the method was supplemented not only by tests aimed at studying perception of the emotional characteristics of speech but also tests oriented on perception of intonational and individual features of speech. Because the main objective of our work was not to analyze the capabilities for identifying emotional states but to study the cerebral mechanisms of these capabilities, the tests were written on the basis of not only tape recordings of the speech of mental patients in different emotional states but also recordings of the speech of actors simulating emotions and the speech of normal subjects to whom different emotional states were suggested under hypnosis. Our lexicon of emotional states included only five basic ones, encountered most frequently in the clinic and characterized by concepts describing positive and negative emotions on a more-general plane, without isolating the different variants of the states, analysis of the perception of which is an independent problem that was not within the objective of our work. These five types of emotional states included: 1. Normal. 2. Alarm. 3. Joy. 4. Melancholy. 5. Anger. The speech tests used in this method for studying perception of emotional, intonational and individual characteristics of speech are described below.

Test 1: Identification and paired comparison of emotional states (based on material gathered in the clinic).

There are two parts to this test: 1) Identification of emotional states, 2) paired comparison of emotional states.

a) Identification of emotional states: One sentence, "It was an early spring," was chosen as the starting material. It was tape-recorded while read by patients experiencing different emotional states. The procedure begins with a training series intended to teach the patients how to recognize and name emotional states on the basis of voice characteristics (the recording of each state is repeated three times).

The principal research series contained samples of emotional states presented in random order. The subject was first given a training series and then the principal series. On listening to the principal series, he had to recognize and name the appropriate emotional states.

b) Paired comparison of emotional states: Following preliminary training, in this experiment the subject was asked to successively compare two stimuli. In his response he had to declare whether the paired emotional states were different or identical. The experimental sentence was the same as in the first part: "It was an early spring." The training series consisted of the following seven pairs: 1) normal-normal, 2) alarm-alarm, 3) joy-joy, 4) melancholy-melancholy, 5) anger-anger, 6) normal-joy, 7) anger-melancholy. Of course, it would not have been suitable in this experiment to use training pairs consisting of all five states.

The principal series was based on 30 pairs of comparisons recorded in random order. It was unsuitable to include a large number of comparisons because this could tire the patients. On being presented each pair of stimuli, the subject had to say whether both stimuli were uttered in the same or in different emotional states.

Test 2. Identification and paired comparison of emotional states (actor simulation and suggestion under hypnosis).

This test made use of spoken material obtained: 1) by recording the words spoken by an actor (actor simulation) and 2) by recording the words of a speaker under hypnosis.

a) Actor simulation: An actor read the sentence "This is so simple that I have to say it" in one of the states he was asked to simulate: 1) Normal, 2) anger, 3) melancholy, 4) joy, 5) alarm. The design of the experiment was the same as with test 1 (paired comparison). The training series consisted of four pairs of stimuli: 1) Normal-normal, 2) melancholy-melancholy, 3) normal-melancholy, 4) joy-alarm. The principal series consisted of thirty pairs.

b) Suggestion under hypnosis: There were two parts to this test: a) paired comparison of emotional states, b) identification of emotional states. One of the states indicated above was suggested to the speaker, who then uttered the control sentence "This is so simple that I have to say it." The training and principal material used for paired comparison was arranged in the same order as with actor simulation.

Test 3: Identification and paired comparison of different intonational structures.

The experiment required analysis of seven different intonational structures based on the same sentence: "Mommy bathed Man'ya." Here is a sample of the training text offered to the subject for emotion identification:

No 1. Mommy bathed Man'ya (neutral advisory intonation).
No 2. Mommy bathed Man'ya (logical stress on the first word).
No 3. Mommy bathed Man'ya (logical stress on the second word).
No 4. Mommy bathed Man'ya (logical stress on the third word).
No 5. Mommy bathed Man'ya? (questioning intonation).
No 6. Mommy bathed Man'ya! (exclamatory intonation).
No 7. Mommy bathed Man'ya... (incomplete intonation).

These structures were arranged in random order in the principal text.

For paired comparison, the subject was offered a training sample of five pairs of stimuli, and he was asked to respond whether the stimuli were the same or different (the pause between pairs was 5 seconds). There were 30 pairs of stimuli in the principal text, with the number of pairs of identical stimuli being equal to the number of pairs of different stimuli.

Test 4: Identification and paired comparison of individual characteristics of pronunciation.

The objective of this test is to reveal the particular way subjects perceive different voices in a sample and their capability for distinguishing different voices presented in pairs.

In this test 15 male speakers uttered the control phrase "Everything was blanketed by dark clouds." The subjects are not acquainted with the voices of these speakers prior to the experiment. The training text consists of five pairs of stimuli (two pairs of identical voices and three pairs of different voices). In the principal series the subject is required to successively listen to 30 repetitions of this sentence, and venture a conclusion as to the similarity or dissimilarity of two similar repetitions--that is, the similarity or dissimilarity of each successive stimulus in relation to the previous one. The subject used the symbol "±" to denote similar stimuli and "-" to denote different stimuli.

This method was applied in a preliminary experiment to 11 patients with local lesions of different divisions of the cerebral cortex. The impression is that recognition of emotional, intonational and individual characteristics of speech in the tests is associated in a number of cases with focal pathology of the temporal divisions of the cerebral cortex, and predominantly in the right hemisphere of right-handed patients. Incidentally this research has only just begun, and more observations will have to be accumulated before the results could be summarized and analyzed.

ANALYSIS OF THE VARIABILITY OF THE MELODIC CONTOURS OF SPEECH

A. V. Beskadarov, V. I. Galunov

A large amount of applied problems have recently been found that can be solved successfully by the methods of factor analysis. In our work we had the possibility to examine one of these methods, namely a computer program method of isolating main factors. We used this method to detect individually variable parameters at the prosodic level (the melodic outline of speech), associated with analysis of different emotional states reproduced in the speech of an actor (the so-called "actor model").

Basically, factor analysis reveals the concealed laws of numerous measurements through analysis of correlation (or covariate) matrices. It is based on the assumption that observed variables may be expressed through concealed independent parameters or factors. In other words analysis of variations in some phenomenon can reveal the concealed laws of these variations.

1. Materials and Methods

In our experiments we used a tape recording of a single test sentence ("Это так просто, что хочется сказать" [Eto tak prosto, chto khochetsya skazat'; This is so simple that I have to say it]) spoken by an actor. The speaker was instructed to utter this sentence in one of 44 prescribed states (normal, arousal, languor, joy and so on). Correspondingly, 44 melodic contours were obtained by processing the oscillograms. Then we selected out the frequencies of the fundamental tone of vowel sounds within the test sentence (11 frequencies) (see Table 1).

The material in Table 1 was subjected to factor analysis in two variants: 1) factor treatment of the characteristics of emotional states (a 44×44 correlation table) and 2) factor treatment of the frequencies of the fundamental tones of individual vowel sounds (11×11 correlation table).

The first variant of treatment revealed four spectrums connected with the real data on emotional states. All of the states conditionally fell into four classes corresponding to meaningful factors, which were labeled: 1) "Relaxation" factor (14 emotional states), 2) "alarm" factor (18 states), 3) "aggressiveness" factor (6 states), 4) uncertainty factor (2 states). There were one or two states in each class which, generally speaking, were not suited to the label characterizing that class (thus the states of joy, tension and inspiration were such an exception in class 1). It may be concluded that factor treatment of the test sentence in relation to different

Table 1. Frequencies of the Principal Tone in the Phrase "This is so Simple That I Have to Say It" for Speaker M. (In 44 Emotional States)

№	1 э	2 о	3 а	4 о	5 о	6 о	7 о	8 'э	9 а	10 а	11 а	States
1	215	220	220	275	205	200	250	215	205	195	235	Normal
2	210	210	210	265	190	190	240	215	195	200	225	Concentration
3	215	205	205	230	200	195	235	220	195	195	240	Relaxation
4	250	270	280	330	230	235	230	270	250	240	285	Confidence
5	250	300	305	340	290	295	335	310	285	295	330	Arousal
6	220	220	230	260	195	200	260	235	210	210	245	Languor
7	235	230	240	300	220	225	270	252	230	240	260	Tension
8	275	285	290	315	295	280	325	305	385	295	318	Perturbation
9	260	350	330	375	305	295	350	310	325	320	345	Frenzy
10	250	270	260	300	205	225	305	260	250	222	297	Inspiration
11	330	330	330	440	230	265	390	330	295	270	307	Joy
12	230	295	290	272	272	295	360	320	312	307	295	Delight
13	210	235	240	290	250	215	270	245	220	220	265	Timidity
14	220	230	240	295	215	215	270	260	225	230	250	Embarrassment
15	190	195	210	255	180	210	260	235	225	215	235	Uncertainty
16	207	210	207	210	215	202	190	215	202	195	215	Doubt
17	245	220	222	282	190	197	237	220	225	212	235	Disenchantment
18	190	220	230	280	200	217	280	250	235	232	280	Insult
19	200	220	325	257	195	230	290	230	242	227	240	Displeasure
20	180	232	245	297	195	220	302	250	245	252	277	Revulsion
21	270	310	360	360	285	290	335	290	295	305	320	Aggressiveness
22	250	312	330	395	308	317	368	310	320	305	330	Anger
23	237	300	332	282	212	215	275	312	332	327	332	Indignation
24	200	207	240	350	275	310	347	237	217	210	257	Surprise
25	250	245	272	300	265	253	280	270	247	255	272	Confusion
26	265	302	340	410	310	305	370	320	290	295	350	Amazement
27	200	212	240	265	220	200	250	230	210	212	237	Shock
28	210	230	245	280	222	207	320	245	220	220	245	Relief
29	220	245	260	300	220	215	267	230	217	220	235	Pleasure
30	175	190	250	215	177	190	247	220	190	190	222	Satisfaction
31	210	222	245	285	240	230	285	240	235	250	270	Anxiety
32	200	217	225	245	240	215	257	240	232	230	245	Fear
33	275	310	300	350	282	280	325	295	295	290	325	Terror
34	225	225	220	265	190	190	240	227	205	190	210	Woe
35	205	210	210	240	165	185	250	200	180	170	170	Melancholy
36	200	200	200	180	180	170	217	190	160	170	175	Depression
37	200	225	305	205	200	200	280	250	210	210	220	Suffering
38	250	305	370	310	290	280	345	310	305	300	310	Despair
39	250	270	225	385	320	300	352	315	320	315	320	Irritation
40	300	350	382	340	290	320	345	330	335	312	340	Resentment
41	287	307	305	390	270	295	375	310	325	330	250	Hatred
42	270	250	265	240	210	200	240	230	210	200	225	Tenderness (variation 1)
43	220	235	235	250	185	210	255	230	225	220	240	Tenderness (variation 2)
44	210	220	235	260	175	205	290	250	225	205	172	Love

characteristics (states) produces an approximate breakdown of the mass of data into several classes containing similar states.

The second variant of treatment involved analysis of the succession of vowel phonemes in the test sentence. We attempted to reveal certain correlations between the revealed factors and the character of the melodic contours of the phrases analyzed, and to establish certain types of controlling influences which would form the melodic contour of the sentence in each individual pronunciation. Factor analysis in relation to sound revealed two factors, the first of which was found to be correlated with stressed vowels in the phrase, the second being correlated with stressed vowels. In other words the computer program of factor analysis did establish certain laws in the formation of the melodic contour of a statement.

Conclusions: Factor analysis in relation to emotional states revealed four classes of states; their comparison with spectrographic data showed that these classes are typified by different degrees of unevenness and range of the fundamental tone (high for factors I and III, low for factor II) and differences in the location of the first melodic rise--namely at the vowel "a" in "так просто..." for factor III, and at the first vowel "o" in "так просто..." for factor II).

II. Investigation of Factor Loads and Analysis of Semantic Content in a Study of Speaker Individuality

Materials and methods: The initial sentence "Мама мыла Маню" [Mama myla Manyu; Mama bathed Man'ya] was pronounced by 26 speakers (24 stated it once, one stated it 29 times, and another speaker stated it 35 times). The speakers uttering this sentence were not asked to read it in the same way. Therefore they read the sentence differently each time, while remaining within the framework of a given type of communication. Comparison of the factor loads and the real melodic contours revealed four basic patterns for the melodic contour of the control sentence.

Type I was characterized by a relatively even melodic contour, one with no sharp rises and falls throughout the entire sentence. Type II was characterized by pronounced type I and II melodic rises separated by a not very large fall. The type III melodic contour differs from the type II contour in that the fall between the types I and II melodic rises is very significant, while in comparison with the type I melodic rise, the type II melodic rise appears more "massive" than is the case with the type II melodic contour. The type IV melodic contour is characterized by a clearly pronounced descending melodic pattern.

Factor analysis in relation to segments of the test sentence revealed the same laws for formation of the melodic structure and the "responsibility" carried for this process by the individual factors indicated in part I of this communication.

SIGNIFICANCE OF PROSODIC AND SPECTRAL PARAMETERS OF SPOKEN SIGNALS EXPRESSING DIFFERENT EMOTIONAL STATES

L. P. Blokhina, T. G. Gomina

Isolating from an emotionally colored spoken signal the acoustic parameters bearing information on a given emotional state, and determination of the significance of each of these parameters is one of the central problems of research on emotionally colored speech.

Up until now, this problem has mainly been attacked from the standpoint of acoustic analysis of prosodic characteristics (1-5). Attempts at spectral analysis of emotionally colored spoken signals have been undertaken in a much lesser quantity (6-9).

Our research objective was to determine the significance of individual acoustic parameters (prosodic and spectral) in spoken signals expressing different emotional states by applying the analysis-synthesis-analysis method.

We used English-language material for our intonographic and spectrographic analysis of emotionally colored speech. We selected emotional states falling within the classes joy, anger, fear and melancholy in comparison with neutrally colored speech, which we referred to as normal. This report discusses the results of the synthesis stage and of subsequent listener analysis.

A two-syllable nonsense signal simulating primary and secondary syllables in experimental sentences uttered by English speakers was used as the material for synthesis. The experimental material was synthesized at the LEF [not further identified] of the First Moscow State Pedagogical Institute of Foreign Languages imeni M. Torez. The synthesizer was part of a complex of apparatus intended for primary analysis and synthesis of speech, fashioned out of an "Ural-14" computer complex. The number of temporal intervals in the programmer was 45. The Тинт. parameter is described by a five-digit code (from 5 to 160 msec with a spacing of 5 msec). $F_{ОТ}$ is described by an eight-digit code (from 78 msec to 50 msec with a spacing of 78 msec). Parameters F_1, F_2, F_3 are described by a seven-digit code with variations in the following ranges: 200-1,470 Hz, 500-3,040 Hz, 1,000-6,080 Hz. Parameters $F_{асп}$, $F'_{фр}$, $F''_{фр}$ are described by a four-digit code in the following ranges: 600-2,400 Hz, 1,000-1,050 Hz, 2,000-9,950 Hz. The amplitude parameters of all types of formants are described by a two-digit code. The synthesizer can work with four amplitude values: A_o, A_{min}, A_{med}, A_{max}.

The program fed into the synthesizer was written on the basis of conclusions made as a result of instrumental analysis of the experimental material contained in statements made by 11 speakers. The synthesis program was written with a consideration for only those emotional state characteristics which were most common--that is, those which were noted among all or most speakers. (In view of the limited technical possibilities of the synthesizer we were unable to include in the program a number of acoustic characteristics isolated in the course of spectral analysis).

In comparison with normal, the following characteristics were found to be the most common: For the state of joy--expansion of range and a rising-falling contour for the ChOT [frequency of the fundamental tone]; the vocal saturation of phrases; shifting of spectral energy into the high frequency range; expansion of the spectrum of stressed and unstressed vowels, and higher intensity of stressed vowels at high frequencies; for the state of melancholy--narrowing of range, and a falling contour for the ChOT; consonant saturation of phrases; shifting of spectral energy toward lower frequencies, narrowing of the spectrum of stressed vowels; for the state of anger--expansion of range and a falling contour for the ChOT; consonant saturation of phrases, shifting of spectral energy into the high frequency range, expansion of the spectrum of stress vowels, and growth in their intensity; for the state of fear [sic; the word "anger" was probably intended]--expansion of range and a rising contour for the ChOT, consonant saturation, shifting of spectral energy into the high frequency range, expansion of the spectrum of stressed vowels, and growth in their intensity; for the state of fear--expansion of range and an ascending contour for the ChOT, consonant saturation, shifting of the energy spectrum into the high frequency range, insignificant expansion of the spectrum of stressed vowels.

The program was written for the neutral state and for each of the emotional states named above. The program written for the neutral state was subjected to change in stages. The first stage entailed variation of the frequency characteristics of the signal with all of the rest of the parameters remaining unchanged. In the second stage the changes in the ChOT were supplemented first by modification of the length of consonant sounds and then modification of the length of vowel sounds. In the third stage the high frequency components of the vowel spectrum were introduced (during simulation of intense emotions), after which their amplitude was increased by one order of magnitude. Each separate modification of the initial program corresponding to a normal state was recorded on ferromagnetic film. Thus we obtained: a) signals in which only one of the prosodic or spectral characteristics was varied; b) signals with variations in prosodic characteristics (ChOT and length); c) signals with modifications of spectral characteristics; d) signals with simultaneous variations in prosodic and spectral characteristics. The model of a particular emotional state was a summational signal including both prosodic and spectral modifications typical of the given emotional state. Synthesized signals were analyzed by means of a spectrum analyzer with the purpose of testing the adequacy of spoken signals to program-simulated signals. Analysis of the obtained spectrograms indicated that the simulated and synthesized signals were identical, making it possible to go onto the next stage of the research--listener analysis.

Two series of listener analysis were conducted, with seven experienced listeners participating. The same listeners participated in both analysis series. The second series was conducted with the purpose of checking the reliability of data obtained in the first series. Comparison of the results of the two series demonstrated their sufficiently good match. The instructions required the listeners to perform two tasks: 1. classify a nonsense stimulus in relation to the following characteristics: a) neutrality-emotionality, b) positive emotion/negative emotion, c) intense emotion/weak emotion. At first this test was performed as the subjects listened to complete models of emotional states, and subsequently while listening to signals in which certain acoustical parameters were modified. Each emotionally colored nonsense stimulus was paired with an emotionally neutral nonsense stimulus, the pause between them being 2 seconds. 2. Determine emotional state on the basis of the nonsense stimulus provided. In this case the subjects listened only to the complete models of the emotional states and the normal recording.

The results of listener analysis permit the following conclusion, tentative for the moment.

In the first half, all listeners without exception were able to distinguish the normal stimulus from stimuli simulating emotional state. Judging from the responses of the listeners they were able to distinguish strong from weak emotional states and positive from negative emotional states rather easily (see Table 1).

As is obvious from this table, the listeners are able to classify joy and melancholy with sufficient adequacy. The most difficulty occurred in identification of signals simulating anger and fear. In general the listeners confidently related these signals to strong and negative emotions, but separation of the latter from one another was a complex task. Thus the signal simulating anger was identified as anger by 57 percent of the listeners and as fear by 43 percent. The signal simulating fear was classified as fear by 31 percent of the listeners and as anger by 69 percent. These difficulties in identifying these models compelled us to conduct one more series of listener analysis, in which signals simulating fear and anger were presented to the listeners in pairs.

The listeners were told that two successive signals they were to hear belonged to different classes--one to anger and the other to fear. The task of the listeners was to classify the presented signals appropriately. In this stage of analysis 80 percent of the listeners correctly evaluated the synthesized signals presented to them.

One interesting fact that should be noted is that signals with modified spectral characteristics (expansion of spectrum and growth in the intensity of vowel formants) which did not undergo corresponding change in their prosodic parameters were classified by the listeners as manifestations of weak negative emotions (caution, mild anger, reproach), though according to the results of spectrographic analysis these spectral modifications were typical of emotional states such as anger, fear and, to the greatest degree, joy. In all probability the spectral modifications typical of the state of joy cannot by themselves (without participation of prosodic characteristics) transmit this emotional state. Moreover we should also consider the still inadequately high technical possibilities of the synthesizer we used, making it impossible to simulate finer modifications of the spectrum.

An analysis of the listener responses revealed parameters promoting adequate identification of the corresponding emotional states. The principal parameter in identification of joy is the ChOT (its contour and range). When only the frequency contour was modified (narrowing of the range and depression of the frequency level),

Table 1. Data From Listener Analysis (Percentages)

Presented Stimulus	Neutral State	Emotional State	Intense Emotion	Weak Emotion	Positive Emotion	Negative Emotion	Emotion Named
Normal	100	-	-	-	-	-	Normal (100)
Joy	-	100	100	-	85	15	Joy (50) Delight (35) Terror (15)
Melancholy	-	100	20	80	-	100	Melancholy (50) Depression (50)
Anger	-	100	65	35	8	92	Anger (30) Malicious joy (10) Insult (10) Amazement (8) Fear (42)
Fear	-	100	100	-	-	100	Fear (35) Alarm (10) Displeasure (15) Anger (40)

the listeners were able to classify this zone only as a negative emotion (anxious remorse, caution, fear, reproach, displeasure). Additional increase in the length of consonants and vowels permitted the listeners to establish the membership of the given signal to the melancholy class with sufficient reliability.

Table 2. t_C/t_V Ratio Within Syllables (in Relative Units)

State	First Syllable	Second Syllable
Anger	2.88	4
Fear	3.56	5.11

Although auditor analysis of the emotional states of "anger" and "fear" was conducted in special conditions, we can still isolate the dominant parameters participating in identification of these emotional states (though this conclusion requires further testing more so than do the others). These parameters include the frequency contour typical of the given emotional states: Rising in the presence of fear and falling in the presence of anger, together with an increase in the length of the first consonant and a decrease in the length of the second consonant coupled with a simultaneous decrease in the length of vowels. Redistribution of the length of a consonant and a vowel within a syllable may be significant in this case (see Table 2).

BIBLIOGRAPHY

1. Vitt, N. V., "Expression of Emotional States in Speech Intonation," Candidate Dissertation, Moscow, 1965.

2. Uldall, E., "Attitudinal Meanings Conveyed by Intonational Contours," LANGUAGE AND SPEECH, Vol 13, 1958.

3. Lieberman, Ph., and Michaels, S., "Some Aspects of Fundamental Frequency and Envelope Amplitude as Related to the Emotional Content of Speech," JASA, Vol 34, No 7, 1962.

4. Bonner, M. R., "Changes in the Speech Pattern Under Emotional Tension," THE AMERICAN JOURNAL OF PSYCHOLOGY, Vol 52, No 2, 1943.

5. Fairbanks, G., and Pronovost, W., "Vocal Pitch During Simulated Emotion," SCIENCE, Vol 88, 1938.

6. Williams, C., and Stevens, K., "Emotions and Speech: Some Acoustical Correlates," JASA, Vol 52, No 4, 1972.

7. Nikonov, A. V., and Popov, V. A., "Structural Characteristics of the Speech of a Human Operator in Stressful Conditions," "Rech' i emotsii" [Speech and Emotions], Leningrad, 1975.

8. Frolov, M. F., and Taubkin, V. L., "Influence of a Speaker's Emotional State on Some Parameters of the Speech Signal," in "Rech' i emotsii," Leningrad, 1975.

9. Tishchenko, A. G., "Dynamics of Formants in the Spectrum of Audible Speech as an Objective Indicator Distinguishing Positive From Negative Emotions," KOSMICHESKAYA BIOLOGIYA I MEDITSINA, No 5, 1968.

INFORMATION CONTENT OF THE TIMBRE CHARACTERISTICS OF SPEECH

A. P. Varfolomeyev

Despite the fact that the overwhelming majority of works describing the system of prosodic characteristics of speech make mention of its timbre, the question as to timbre as an element of speech (and all the more so of language) possessing a certain information content for practical purposes remains open. As a rule we see the term "timbre of voice" used, and not "timbre of speech." And when discussion of the timbre of speech is found to be unavoidable, authors are compelled to limit themselves to general statements as to the possibility of expressing the emotional properties of speech through timbre, forming a subtext and so on. The fact that a listener's consciousness associates the timbre of the speaker with his emotional and mental state has been mentioned more than once (1,2).

We did not come across any attempts at classifying the timbre of speech in relation to the particular content it may express. The universally known classification of voices (alto, soprano, tenor and so on) is hardly applicable to the timbre of speech, since it does not correlate with any semantic elements of the speech act.

We obviously cannot base a classification of the timbre of speech on how content descriptions of timbre group together and contrast with one another. Instead, such a classification should be arrived at by seeking the laws governing change in the timbre of speech in connection with the change in its information content. It is not our objective here to arrive at such a classification, because to record changes in the timbre of speech, we must first of all find a means of describing the information content of timbre--that is, a means of accounting for the semantics of a speaker's timbre within a certain excerpt of speech.

By describing a speaker's timbre in relation to its information content, we would be able to reflect in a certain way the individual characteristics of speech. From our point of view the timbre characteristics of speech are the most individualized, and apparently this is precisely why it is so difficult to describe the content of these characteristics. Those aspects of the content of the timbre of speech that yield to description and which may be correlated with expressions of the speaker's emotional and mental state will find their use in general analysis of the emotional characteristics of speech.

We attempted to study the information content of the timbre of speech by the known method of the semantic differential (3), based on recording the verbal evaluations made by subjects of a certain aspect of a spoken stimulus. The use of the human auditory system as a sufficiently sophisticated analyzer of individual and emotional characteristics of speech has already recommended itself positively (4).

Methods of the Experiment

Subjects were asked to evaluate the timbre of the voice of a speaker (tape-recorded) uttering a certain phrase, using scales such as "good-bad," "large-small," "happy-sad" and so on. About 80 such scales were used in all. The scales were divided into five ranks, each having a score from 1 to 5. As an example the scores for the "good-bad" scale reflected perception of the following ranks of timbre: 1--"very good," 2--"good," 3--"neither," 4--"bad," 5--"very bad." The lexical and grammatical content of the phrases serving as stimuli was the same for all analyzed timbres; consequently the main element that underwent variation was the timbre of the speaker's voice.

College students of different majors and specialties and senior high school students served as the subjects. Each stimulus were evaluated by an average of 50 subjects.

Printed questionnaries bearing a list of the scales of characteristics and an explanation of the scoring system were used.

Statistical treatment of the data entailed finding the average scores for each stimulus in relation to all scales, and correlation and factor analysis of these averages.

Two series of experiments were conducted (with 11 and 17 stimuli).

Results and Discussion

Obviously average scores falling between 1 and 2 and between 4 and 5 would be significant. Averages such as these would persuasively indicate that the timbre characteristics of speech can be differentiated by the perceiver.

Thus for example, the information content of one certain timbre is expressed by the characteristics "sad," "old," "inaccurate," "rough" and so on. The information content of another is expressed by the characteristics "clear," "accurate," "strong," "good," "calm" and so on. One can easily be persuaded of the qualitative consistency of each of these descriptions.

Correlation-factor analysis of the average scores revealed the principal factors defining the way the content of the timbre characteristics of speech is perceived. One of them is represented by the characteristics "calm-irritable," "happy-sad," "unconstrained-tense," "smooth-rough," "good-evil" and so on--that is, by characteristics reflecting perception of the speaker's tone.

The second factor is represented by the characteristics "clear-distorted," "good-bad," "beautiful-not beautiful," "pure-hoarse," "usual-unusual." This factor combines perception in relation to subjective assessment and clearness criteria.

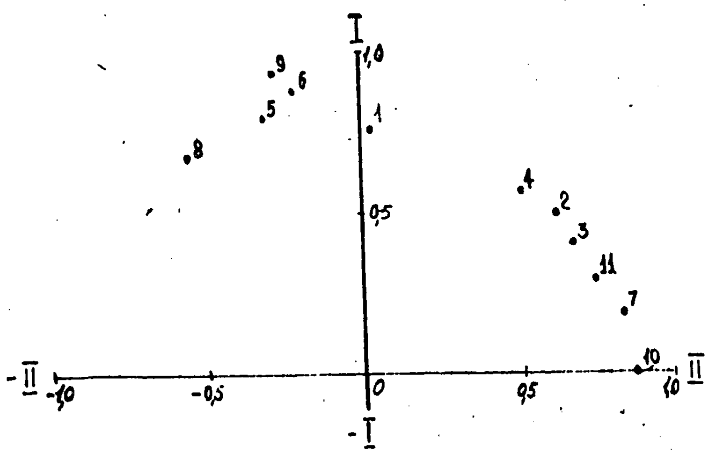

Figure 1. Intersection Plane of Factors I and II

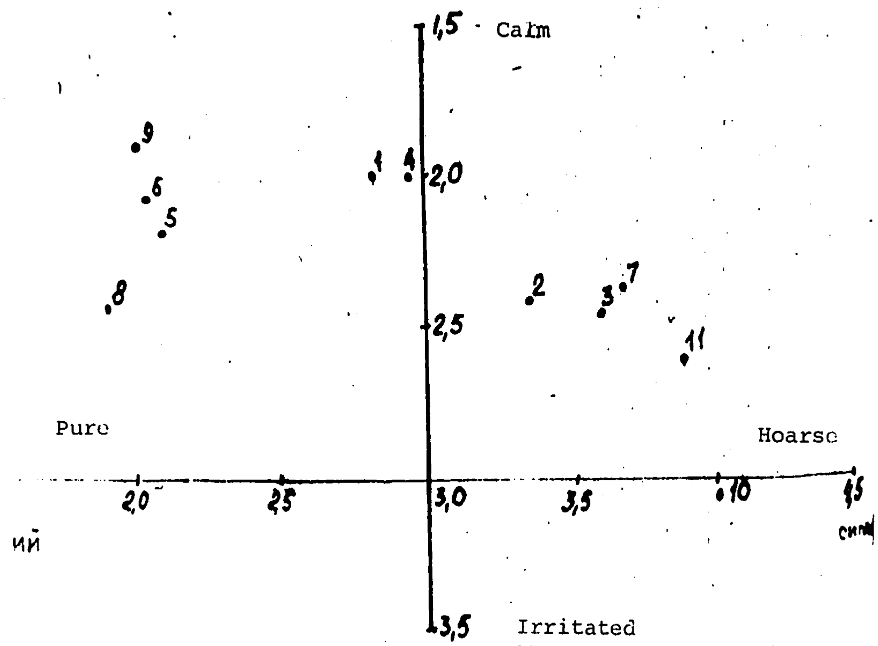

Figure 2. Intersection Plane of Base Scales

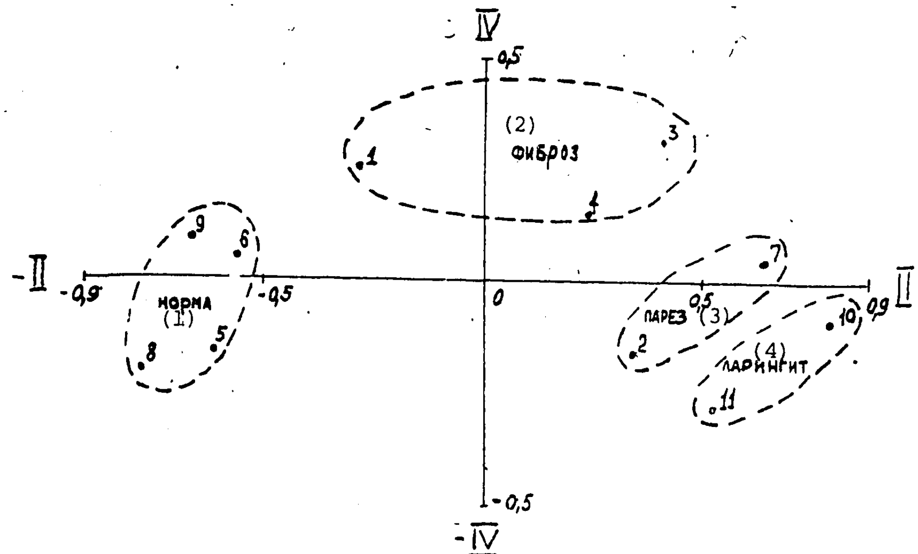

Рис 3

Figure 3. Intersection Plane of Factors II and IV

Key:

1. Normal
2. Fibrosis
3. Paresis
4. Laryngitis

Figure 1 shows the position of the timbre stimuli in the intersection plane of these two factors (determined for the first series, consisting of 11 stimuli). That the factors were interpreted correctly can be seen from a comparison of this space with the disposition of these stimuli in the intersection plane of the so-called base scales (the scales which most fully express the semantics of particular factor groups). The stimuli fall within this space in correspondence with their average scores on scales defined as being base scales.

For one factor group we adopted "dominant" as the base scale, and "hoarse" as the base scale for the other. The similarity in the distribution pattern of the stimuli in both spaces indicates, in our opinion, both that the factors are interpreted correctly and that the base scales are selected correctly (see figures 1 and 2).

Among the samples of speech studied, some belonged to speakers with a normal speech system (stimuli 5, 6, 8 and 9) while others belonged to speakers with certain speech disorders: fibrosis (1,3,4), paresis (2,7), and laryngitis (10,11). The use of normal and pathological speech is motivated by the fact that differences in the timbre characteristics of speech are highly significant in relation to such a sample, and manifestation or nonmanifestation of these differences has fundamental significance to determining the possibility of using it for diagnostic purposes.

The position of timbre stimuli in the intersection plane of the factors and in the intersection plane of the base scales indicates a clear tendency toward grouping in relation to the criteria for normal and pathological, and toward differentiation of stimuli in relation to pathology, depending on its nature. Thus the second factor also represents what is normal.

Our work permits the conclusion that it is possible to analyze the timbre characteristics of speech by the method of the semantic differential. That the characteristics of the "tone" factor managed to reveal themselves attests to the promise of studying the emotional and expressive content of speech represented by timbre. A description of the information carried by the characteristics of the timbre of speech, when compared with an acoustic description of the same stimuli, may lead to creation of automatic systems that could determine the individual's emotional and mental state on the basis of the timbre of his speech, and to creation of systems able to diagnose speech pathology.

BIBLIOGRAPHY

1. Torsuyev, G. P., "Fonetika angliyskogo yazyka" [The Phonetics of the English Language], Moscow, 1950.

2. Vaarask, P. K., "Tonicheskiye sredstva rechi" [The Tonic Resources of Speech], Tallin, 1964.

3. Osgood, Ch., Suci, G., and Tannenbaum, P., "The Measurement of Meaning," Urbana, 1957.

4. Galunov, V. I., Manerov, V. Kh., and Tarasov, V. I., "Auditory Analysis of Speech Recorded in the Presence of Emotional States Simulated by Different Methods," in "Rech' i emotsii" [Speech and Emotions], Leningrad, 1975.

THE SIGNIFICANCE OF PERSONAL MEANINGS TO REALIZATION OF THE PHYSICAL CHARACTERISTICS OF A SPOKEN STATEMENT (ACCORDING TO CLINICAL OBSERVATIONS)

Ye. N. Vinarskaya, A. S. Nikiforov, S. A. Soldatova

Research Objective

To be realized, the physical characteristics of a speech signal require externally determined energy outlays by the body, which change depending on the complexity of both the structure of the signal itself and the operational structure of the act of communication. The question is, would consideration of just the objective complexity of these structures be enough to arrive at a conclusion as to the level of energy to be expended?

Research Methods

To answer this question we used the method of clinical observation of a group of patients (150) suffering focal lesions of the mesencephalic-diencephalic division of the brain, manifested as symptoms of insufficiency of the ascending activating effects of nonspecific structures (1).

Research Results

Dynamic observation of our patients showed that as pathology proceeds, they must gradually experience a greater personal interest if they are to enter into spoken communication. Their speech becomes slower, quiet, monotonous, emotionally unexpressive, and lexically and grammatically simple; speech becomes increasingly more tiring to the patients, and they exhibit an increasingly greater need for supplementary motivation or volitional effort. In a more-pronounced phase of illness patients cease to use speech at all in situations not having direct personal meaning to them. However, if a patient is made emotionally interested in the topic of discussion and in the results to which such a discussion might lead, he arrives at the necessary motivation for speech, and he begins to speak. The more personally meaningful the topic and purpose of discussion become to the patient, the less constrained, louder and more expressive the patient's speech becomes and the richer is his use of segmental and supersegmental phonetic resources and of lexical and grammatical resources. Thus one of our subjects, patient Z., 42 years old, who was unable to make any sort of speech contact with medical personnel and even his wife, unexpectedly revealed the ability to advise his favorite daughter on discipline in public prior to an examination: In the course of an hour he essentially answered all of her questions with a quiet, monotonous voice. Just prior to his death, another subject who was even more gravely ill, patient K., 57 years old, suddenly entered into spoken contact with his wife when she broached the subject of writing his will.

As a rule speech of patients improved in all respects when the discussion turned to their health, to the prospects of a certain method of treatment and to the prognosis of illness. As an example patient B., 43 years old, who exhibited a pronounced lack of motivation to speak, reacted to questions with monosyllabic, quiet, unintelligible and intonationally unexpressive replies, but when it was remarked that the color of her face was good that day, she showed signs of interest, and clearly uttered in a sonorous voice with an intonation of prideful joy: "I never use any sort of creams, just cold water and soap."

It may be concluded that focal lesion of the mesencephalic-diencephalic division of the brain, coupled with selective disturbance of ascending activating effects of its nonspecific structures upon cortical and subcortical neurons specific to speech behavior make these neurons functionally lacking. While in the normal individual the excitability of these neurons is raised to the necessary level automatically, in our patients such regulation loses its automatic nature and therefore becomes accessible to analytical study. The method of clinical observation is enough to permit the conclusion that imparting meaningfulness to the topic of conversation in relation to a given personality has fundamental significance to making speech behavior possible irrespective of the complexity of its structure. In some cases personal interest assumes exceptional forms in relation to regulation of the energetic support of the individual's speech behavior. The following observation is an example.

Valeriy I., 24 years old, learned that he was going to die soon. This fate was postponed by his wife Marina, who took the risk of subjecting her husband to a new method of treatment. From that moment on, and until his death 4 years later, Valeriy devoted his entire life, in all of its manifestations, to the service of Marina, to her happiness and welfare. Everything in which Marina was interested acquired personal meaning to Valeriy; all else had no meaning to him, and was unconsciously ignored by him. As Valeriy's physical strength drained away, the restrictions he imposed upon his own personal activity became increasingly more severe. Several times in the course of these years Valeriy was on the brink of death, but despite the severity of his physical state, he unexpectedly recovered once again, because "he could not bear to anger Marina," because "he had no right to abandon her in such a state" (she was expecting), because "funerals are difficult in winter, and Marina's labor had to be easy."

In the middle of the fourth year the patient's state deteriorated dramatically once again, excessive mental and physical exhaustion developed, and various symptoms of organic affliction of the central and peripheral nervous system appeared, to include in the mesencephalic-diencephalic division. It is noted in the disease history that the patient responded to questions with a quiet, unexpressive and monotonous voice, often pronouncing only the first syllable of words making up very short sentences. Soon the patient once again developed pneumonia, severe, painful contractures of the arms and legs arose, and convulsive spasms appeared in individual muscle groups. His speech became almost inaudible and incomprehensible, and he almost stopped talking. However, on the day following the birth of Marina's child, Valeriy was not to be recognized. All who entered the ward were greeted by a smile, and using a loud, emotionally expressive voice and clearly articulated words, he informed them of the weight and length of his daugher and the contents of his telephone conversation with his wife. Despite cachexia of the greatest severity and worsening organic affliction of the brain, the patient was able to read and to work out the schedule for his daughter's care, and he was able to maintain an interest in talking with visitors about Marina's future and, in this connection, the image of women in Russian literature. He was unable to maintain any sort of conversation on topics not having a relationship to his health, his wife and daughter: His voice would die down, the rhythm, intonational structure and articulation of sounds would become irregular, phrases would become simple, and Valeriy would fall altogether silent.

Conclusions

When taken all together, the clinical facts suggest that energetic support to the individual's speech behavior, and the physical characteristics of the spoken signal as well, are dependent primarily on the personal meaning of this signal (2,3) within the structure of the activity going on. This premise should obviously be accounted for when creating technical systems interacting with man and automatically identifying his emotional state on the basis of the spoken signal.

BIBLIOGRAPHY

1. Vinarskaya, Ye. N., Nikiforov, A. S., and Soldatova, S. A., ZH. NEVROPAT. I PSIKH. IM. KORSAKOVA, No 9, 1977, p 1347.

2. Leont'yev, A. N., "Deyatel'nost', soznaniye, lichnost'" [Activity, Consciousness, Personality], Politizdat, 1975.

3. Vilyunas, V. K., "Psikhologiya emotsional'nykh yavleniy" [The Psychology of Emotional Phenomena], MGU, 1976.

SPEECH RECOGNITION SYSTEM RECOGNIZES SPEAKERS BY VOICE

T. K. Vintsyuk, A. I. Kulyas, A. G. Shinkazh

A learning phonemic speech recognition system operating at the Ukrainian SSR Academy of Sciences Institute of Cybernetics is capable of recognizing, with high reliability, the words and coalescent sentences of only that speaker to which the system had been tuned in its learning mode (1,2). The reliability of recognizing the spoken signals of another speaker dropped to 70 percent, given a lexicon of 300 words. These facts permit the assertion that both the spoken signal description employed and the recognition and learning method account for voice individuality. If this is so, then the description and the recognition method may be applied, without any special changes, for recognition of the individuality of the speaker in relation to some key phrase or password.

A mathematical model of a spoken signal, the recognition method, the method by which the system learns to recognize the speaker by voice and the results of experiments are briefly described below.

Spoken Signal Description (3)

A spoken signal is a sequence of vectors (elements) read out uniformly in time:

$$X_l = (x_1, x_2, \ldots, x_i, \ldots, x_l),$$

where l is the length of the realization. Elements x_i are 48-dimensional vectors with binary components 0 or 1. Component $x_{i\nu}$, with element x_i having number ν, is equal to 11, if at the i-th moment in time the energy in the ν-th spectral band is above a certain threshold $\Theta\nu$, and simultaneously greater than the energy in the neighboring or $(\nu-1)$-th band. The $\Theta\nu$ thresholds are chosen such that the pauses of vectors x_i would be zero in 90 percent of the cases.

Thus elements x_i represent the sign of the frequency derivative in relation to the frequency of the current spectrum of speech. Obviously elements x_i define the position of the maximums and minimums of spectral energy on the frequency axis and, equally so, the quality of the formant maximums. The sequence X_l contains information on change in the position of the maximums and minimums of the current spectrum with respect to time.

Sequence x_i does not depend explicitly on the intensity of pronunciation.

Mathematical Model of a Spoken Signal Accounting for Voice Individuality

Now let us describe a mathematical model of word signals applicable to phoneme recognition. This must be done before the method of speech recognition can be explained.

Observed realizations X_l are interpreted as the results of random distortions in standard signals of the same length l. The set of standard signals for one speaker is determined as follows.

Let there be a finite ordered set E of standard elements e(j), where j is the number (name) of the element. Let us assume that standard elements e(j) have the same physical meaning as elements x_i--factors with binary components. Elements e(j) are elementary segments of speech with a duration of 15 msec, and they represent phonemes, or usually parts of phonemes.

Each word or word combination with number k is defined by phonetic-acoustic transcription R_k--a sequence of element names from set E:

$$R_k=(j_{k1},\ j_{k2},\ \dots,\ j_{ks},\ \dots,\ j_{kq_k}), \tag{1}$$

where q_k is the number of symbols in the transcription of the k-th word.

Transcription R_k is defined as an operator which, being applicable to E, generates the initial standard signal for the k-th word:

$$R_kE=(e(j_{k1}),\ (e(j_{k2}),\ \dots,\ e(j_{ks}),\ e(j_{kq_k})). \tag{2}$$

Next we introduce transformation v of the initial standard spoken signal. Operator v, when applied to R_kE generates a standard signal with length $l \geqslant q_k$ with the following structure:

$$vR_kE=(\underbrace{e(j_{k1}),\ \dots,\ e(j_{k1})}_{v_1\ \text{раз}},\ \underbrace{e(j_{k2}),\ \dots,\ e(j_{k2})}_{v_2\ \text{раз}},\ \dots,\ \underbrace{e(j_{ks}),\ \dots,\ e(j_{ks})}_{v_s\ \text{раз}},\ \dots,$$

$$\dots,\ \underbrace{e(j_{kq_k}),\ \dots,\ e(j_{kq_k})}_{v_{q_k}\ \text{раз}}=e_1,\ e_2,\ \dots,\ e_i,\ \dots,\ e_l), \tag{3}$$

where v_s are the components of operator v--that is, $v=(v_1,\ v_2,\ \dots,\ v_s,\ \dots,\ v_{q_k})$. The following limitations are imposed on whole numbers v_s:

$$m(k,\ s)\leqslant v_s\leqslant M(k,\ s),\ s=1,\ 2,\ \dots,\ q_k \tag{4}$$

$$\sum_{s=1}^{q_k} v_s=l. \tag{5}$$

Relationships (4)-(5) determine the set $V(k,l)$ of operators v generating, from R_kE, all possible standard signals vR_kE with length l.

Assume that the first and last elements in the word transcriptions are represented by an elementary standard pause signal--that is, $j_{k_1} = j_{kq_k} = 1$. In addition, let $m(k,1)=m(k,q_k)=0$ and $M(k,1)=M(k,q_k)=\infty$. For all other s it is assumed that $m(k,s>0)$.

Obviously the standard word signals vR_kE, $v \in V(k,l)$ are coarticulated signals distinguished by a nonlinear pronunciation rate and different (including zero) length of pauses at the start and end of a word. The rate is adjusted by the choice of numbers v_s, and coarticulation is accounted for by the presence of such s that $m(k,s)=M(k,s)=1$.

These constructs account for the fact that speech is generated out of elementary "bricks" common to all words, and they account for the main factor of variability in speech--nonlinear change in the rate of pronunciation.

Variables $E, \{R_k\}_{k=1}^{N}$ (N--number of words in the lexicon),

$$\{m(k,s), M(k,s)\}_{k=1}^{N}, {}_{s=1}^{q_k}$$

define the parameters of the grammar generating the standard spoken signals. These parameters are individual to each speaker. They are evaluated on the basis of a learning sample consisting of realizations of a word uttered by the same speaker (3).

Speaker Recognition Method

If we designate by $p(X_l/vR_k(d)E(d)$ the probability of signal X_l on the condition that the standard signal is

$$vR_k(d)E(d),\ v \in V(k,l,d),$$

where d is the number of the speaker, using the method of maximum likelihood we can write the criterion for detection of a speaker on the basis of pronunciation of one word of a lexicon consisting of N words as:

$$d(X_l) = \underset{d}{\operatorname{argmax}} \max_{k} \max_{v \in V(p,l,d)} p(X_l/vR_k(d)E(d)). \qquad (6)$$

Let us assume that observed elements x_i derived from standard elements e_i as a result of independent distortions by additive noise with binary, identically distributed and independent components. Then criterion (6) could be written in the form:

$$d(X_l) = \underset{d}{\operatorname{argmin}} \min_{k} \min_{v \in V(k,l,d)} \sum_{i=1}^{l} H(x_i, (vR_k(d)K(d))_i) \qquad (7)$$

where $(vR_k(d)E(d))_i = e_i$ --an element with number i in sequence $vR_k(d)E(d)$ defined by relationship (3); $H(x_i,e_i)$--the Hemming distance between x_i and e_i. Similarly as with phonemic word recognition (1,3), criterion (7) is realized as follows: The minimum with respect to v is sought by the dynamic programming method, and the minimum with respect to k and d is found by complete sorting.

In this case when the speaker is recognized by a password or a word combination k*, minimization with respect to k is not necessary.

Characteristics of the Model

Different parameters of the generating grammar reflect individual voice characteristics differently.

Set E(d) primarily expresses the geometric characteristics of the speaker's articulatory system and the means of articulation of the principal sounds. Phonetic-acoustic transcriptions $R_k(d)$ depend on E(d) and reflect the individual manner of pronunciation of coarticulated speech--words in this case. Variables m(k,s), M(k,s) express the individual rate of pronunciation.

Of the three sets listed above, E(d) would apparently be the least individualistic. In particular it may be suggested that E(d) does not depend on the speaker, and set E(d*) of one speaker d* may be substituted for E as being common to all speakers. Then only

$$\{R_k\}_{k=1}^{N} \text{ and } \{(m(k, s), M(k, s))\}_{k=1,\ s=1}^{N\quad q_k}$$

would be evaluated in the speaker recognition learning mode.

Experimental Results

In the first experiment speakers were recognized on the basis of the key phrase "I am a person."

Twenty speakers, including four women, took part in the experiment. The phrase was spoken in the presence of the noise of a BESM-6 computer room with the speech recognition system operating normally. The signal/noise ratio was 20 db. An MK-61 microphone was used.

Each speaker repeated the phrase 10 times with an interval of 1 minute between repetitions. Four out of 10 of the repetitions were used as the speaker learning sample. Set E(d) was common to all speakers. The latter was represented by the E(d) function of one of the male speakers, calculated in previous learning experiments in the recognition of spoken words. There are a total of 80 elements in E. Transcription $R_{k^*}(d)$ and the variable

$$\{(m(k^*, s, d), M(k^*, s, d))\}_{s=1}^{q_{k^*}}$$

were determined for a given $q_{k*} = 19$ during the learning process. An algorithm described in (3) was used for this purpose.

The tables below show the results of recognizing learning and control samples separately for male and female speakers. Criterion (7) was used in relation to a fixed $k = k^*$.

	Men			Women	
	Total Realizations	Total Errors		Total Realizations	Total Errors
Learning samples	64	1	Learning samples	16	0
Control samples	96	2	Control samples	24	8

As follows from the tables, the results of speaker recognition were satisfactory for men. The relatively poorer results for female speakers can be explained by the fact that set E(d) determined for male speakers poorly appoximates the signals of female speakers. Individual E(d) must be used for each speaker.

Then we introduced recognition refusals (1). The following results were obtained for a refusal threshold of 12:

	Men				Women		
	Total Realizations	Errors	Refusals		Total Realizations	Errors	Refusals
Learning samples	64	0	2	Learning samples	16	0	0
Control samples	96	0	3	Control samples	24	2	6

Thus the sum total of speaker recognition is characterized by the following data: 1 percent errors and 5 percent recognition refusals.

In the second experiment the optimum pair (d,k) minimizing criterion (7) served as the recognition response. In this case the number of the word and the number of the speaker are indicated--that is, the word recognition and speaker recognition problems are solved simultaneously. There was a total of 205 classes in the experiment--200 words recorded from one speaker and five recorded from male speakers uttering the key phrase "Listen, computer" for identification purposes. At first the process of learning to recognize 200 words of one speaker was performed--E, R_k and (m(k,s), M(k,s)) were evaluated. Then the process of learning to recognize speakers on the basis of the key phrase was carried out in the presence of a fixed E--R_{k^*}(d) and (m(k*,s,d), M(k*,s,d) were evaluated. The resulting significance level of recognition was 99 percent.

Conclusion

Resources for learning and phonemic recognition of spoken words, developed and investigated by the Ukrainian SSR Academy of Sciences Institute of Cybernetics, may be used successfully to identify speakers on the basis of their pronunciation of passwords or key phrases.

BIBLIOGRAPHY

1. Vintsyuk, T. K., and Shinkazh, A. G., "Phonemic Recognition of Spoken Words: Learning and Recognition Algorithms, and Experimental Results," in "Tezisy dokladov VIII Vsesoyuznogo seminara 'Avtomaticheskoye raspoznavaniye slukhovykh obrazov'" [Abstracts of Reports at the Eighth All-Union Seminar "Automatic Recognition of Auditory Patterns], L'vov, 1974, pp 19-24.

2. Vintsiuk, T. K., Gavrilyuk, O. N., and Shinkazh, A. G., "Phoneme-by-Phoneme Recognition of Speech Composed of the Given Vocabulary," in "The Proceedings of the 1976 IEEE International Conference on Acoustics, Speech and Signal Processing," Philadelphia, 1976.

3. Vintsyuk, T. K., and Shinkazh, A. G., "Automatic Transcription of Patterns on the Basis of a Learning Sample," in "Obrabotka i raspoznavniye signalov" [Signal Processing and Recognition], Kiev, 1975, pp 102-120.

EMOTIONALITY OF THE PERSONALITY AS RELATED TO PSYCHO-PHYSIOLOGICAL AND SPEECH CHARACTERISTICS

N. V. Vitt, L. V. L. B. Yermolayeva-Tomina

The emotionality of an individual has a dual origin--biological and social. Being a component of temperament--the dynamic aspect of behavior, man's emotionality, as outwardly expressed, is unavoidably subordinate to socially accepted norms, and this is manifested especially strongly in speech. All changes in the way the subject regulates his reactions to emotion-producing situations fall on a bipolar scale corresponding to the binary principle of the expression of emotions in speech--voluntary-involuntary expression (1). When studying the expression of emotions in speech, it is important to find characteristics which would in a sense break through the regulatory filters, irrespective of the level and individual structure of the person's emotionality (2).

Within the context of "speech-emotions-personality," there are at least four properties of human emotionality that are most significant: 1) emotional reactivity (equatable to V. D. Nebylitsin's impulsiveness); 2) emotional stability-lability, equatable to the frequency with which emotions arise; 3) intensity, which can be determined easily from the EEG and from speech characteristics; 4) duration, manifested as the persistence of an emotion.

The objective of our research was to comparatively analyze general emotional reactivity and its expression in speech.

We used a modified variant of P. P. Blonskiy's procedure coupled with simultaneous EEG recording. The subject was asked to recall, in his memory, and "relive" emotion-producing situations of different modalities. The latent time of the recall of emotion-producing situations and the intensity of changes in brain biorhythms in response to "reliving" such situations were recorded. Spoken communications were analyzed in terms of two basic factors--formal linguistic and semantic content. The verbal material included verbal statements by subjects in oral and written form, and a recording of oral responses to Rorschach inkblots. The experiments were performed individually with each of 40 subjects (3).

Indicators of general emotionality included the average latent time of recall and the cumulative changes in biorhythms in response to recall of emotion-producing situations. A matrix (see Table 1) was drawn up on the basis of an analysis of the distribution of the frequency with which rhythms increased, decreased or remained unchanged when the subject recalled each modality--joy, anger, fear and displeasure.

Table 1. Changes in Cumulative Rhythm Energy in Response to Recall of Emotion-Producing Situations of Different Modalities

Emotional State Reproduced	Dominant Hemisphere	Forehead	Occiput
Anger	Left	Delta Theta Alpha Beta-2	Delta Alpha Beta-1 Beta-2
Fear	Right	Delta Beta-2	Beta-2
Displeasure	Right	Delta	Delta Alpha Beta-2
Joy	Left and right	Delta Theta Alpha	Delta Beta-2

Only those changes in rhythm which were observed in more than 50 percent of the patients were included in this matrix. By using the matrix we were able to monitor the ease with which emotional states were reproduced and the intensity with which the emotions were experienced. Local changes in rhythms with respect to 1) left-right hemisphere and 2) forehead-occiput were superimposed over the concrete changes recorded in the rhythms of each subject. This made it possible to determine the expressiveness of emotions of different modalities exhibited by each of the subjets.

For the purposes of comparative analysis we considered the following data in the spoken communications: total length, number of verbs, the number of attributes, with subjective evaluations placed in a separate category, and the number of interjections. In terms of semantic content we considered the modality of the produced text (that is, whether it was stated categorically or unconfidently); the nature of the topic, defined as static or dramatized--that is, dynamics. Temporal characteristics included the time of the verbal reaction and the length of pauses.

The analysis showed that the level of emotionality and the dominant emotion in the individual structure of emotionality are expressed more distinctly in the oral form of verbal communications than the written form.

A comparison of general emotionality and the average time of verbal reactions revealed a V-shaped dependence. The same sort of V-shaped dependence was discovered between general emotionality and the indicators of the productivity of creative verbal communications (determined in relation to the number of associations).

The EEG indicators for emotion dominance correlated with pauses, with the number of subjective evaluations made and interjections used by the subjects, and dynamic--that is, dramatized--narration of the plots.

This approach to studying emotionality of the personality in relation to psychophysiological and speech characteristics and application of the corresponding procedure revealed that it would be possible to study controlled and uncontrolled expression of emotions.

BIBLIOGRAPHY

1. Vitt, N. V., "Simulation of Emotional Speech," in "Materialy V-go Vsesoyuznogo simpoziuma po psikholingvistike i teorii kommunikatsii" [Proceedings of the Fifth All-Union Symposium on Psycholinguistics and Communication Theory], Part 2, Moscow, 1975.

2. Ol'shannikova, A. Ye., et al., "Evaluations of Procedures for Diagnosing Emotionality," VOPR. PSIKHOLOGII, No 5, 1976.

3. Vitt, N. V., and Ermolayeva-Tomina, L. B., "A Procedure for Revealing Emotionality in Mnemonic Processes," in "Materialy V-go Vsesoyuznogo s"yezda psikhologov SSSR" [Proceedings of the Fifth All-Union Congress of USSR Psychologists], Moscow, 1977.

VARIABILITY OF SPEECH TEMPOS

L. T. Vygonnaya

A large number of linguistic questions (the causes of acoustic variability, the tempo scale of a particular language, the tempo components of intonation and so on) as well as questions associated with clarifying the dependence of the rate of speech on the speaker's state and on the situation, with evaluating the individual's capacity for voluntarily changing his pronunciation rate and for hearing speech signals transmitted at different tempos, with selecting an optimum ratio between the tempo of speech transmission and the speech tempo typical of the listeners, and so on, may be clarified by determining how differences in the tempo of speech differ among representatives of different languages.

Interpretation of the tempo of speech as a particular language trait is consistent with the present notion existing among the bearers of a particular language that speech tempo is a variable characteristic which may be used to distinguish among representatives of the same language, to compare one's own and foreign speech and to make rather accurate evaluations and self-evaluations from the standpoint of such unique features of speech. The limitations imposed on speech in relation to tempo are not rigid. Speakers can vary it significantly while still remaining comprehensible. These variations are an indicator of the speaker's individuality, one of the characteristics by which the listener can assess the nature and genre of the statement, the style of speech and the emotional state of the speaker, as well as the rhythmical and intonational structure of phrases, which reveals the content of the statements.

Our objective was to reveal the tempo characteristics of Belorussian speech in comparison with Russian. The experiment was performed on 27 native Belorussian speakers having facility with literary language (22 men and 5 women).

One of the main difficulties in establishing tempo differences that are significant to a particular language is the ambiguity of variability in tempo--the fact that characteristics corresponding to the text, to the individuality and to the emotional state of the speaker are all represented simultaneously in a spoken message. This is why special research on tempo requires a standard text which may be deemed homogeneous and neutral, which does not elicit clearly pronounced emotions (of varying sign and level) in the speaker and which does not motivate him to emphasize, while reading, speech's function of emotional expression.

In our experiment we used 20 narrative phrases (806 syllables, 1,870 phonemes) of varying length--from 21 to 71 syllables. Each phrase had the nature of an assertion, and it transmitted established facts having no relationship to the speaker (these were phrases taken from an encyclopedia). Specifically, we used a narrative statement which, according to published data, is read at a tempo maximally close to the average tempo of the speaker (1). The material was read by the speakers in a recording studio, where it was tape-recorded. After first becoming acquainted with the experimental situation, the speakers--persons with an advanced philological education--were asked to read the text in a tempo comfortable to them--in a normal tempo. After this, the same text was read by another speaker in what was, from the standpoint of each speaker, a slow and a fast tempo. As he read, the speaker had to keep in mind his own impressions of how different tempos would sound. In one experiment some of the speakers were asked to read the text at some particular tempo at different times (within the same month).

Analysis of the recordings revealed the average pronunciation tempo of each continuous speech excerpt of each speaker in three required tempos. Pauses between phrases were not considered, since according to (1) pauses in messages, and in narrations specifically, do not have an influence on the tempo of phrase pronunciation. Three variants of individual tempo were established for each speaker--normal, fast and slow.

Given the conditional nature of the speech (we analyzed reading, and not spontaneous speech, and so on), the differences in the normal tempo of Belorussian speech obtained for native Belorussian speakers having facility with literary language are found to be extremely indicative when compared with the corresponding data for the same three gradations of tempo established for native Russian speakers (2,3). For the latter, the average duration of sound in normal, fast and slow individual tempo was, respectively, 65-73 msec, 63-60 msec and 75-85 msec. Thus an individual tempo that is slow to native Russian speakers is fast to native Belorussian speakers.

In the case of slow pronunciation of text, the average duration of the sound of Belorussian speech spoken by most subjects (17 speakers) was 91-111 msec; smaller groups had an average duration of 143 msec (8 speakers) and 200 msec (2 speakers).

In the case of fast pronunciation, 17 persons maintained a sound duration averaging between 55 and 63 msec, 8 persons had an average from 67 to 71 msec and 2 persons averaged from 77 to 83 msec.

Comparison of the obtained data would show difference in the significance of the general tempo characteristics of Russian and Belorussian language, as well as difference in their significance within the framework of the same language spoken by different representatives of that language.

All three established gradations of speech tempo are equally typical of speakers using a given language, and given its great significance, variability in tempo cannot be thought of as random, as fully arbitrary and unpredictable.

The diverse causes behind variability in tempo require further systematization and clarification.

We were interested in determining how our speakers would respond to an instruction to speed up and slow down their speech tempo, and in how much of an increase or decrease in their speech tempo was perceivable by them. The instruction to "slow down" the speech tempo produced a rather individual reaction in different speakers. The speech tempo decreased by 6, 9, 14, 15, 18, 23, 25, 30, 33, 37, 41, 55, 59 percent. Most speakers reduced their tempo by 14-25 percent. The response to a request to read faster was less variable: All speakers increased their tempo by not less than 16 percent, with most doing so by either 20-27 percent or 40-41 percent. The maximum acceleration of tempo noted was 67 percent. For speakers whose individual normal tempo could be interpreted as slow, acceleration of the tempo by 50 percent elicited the same phenomena noted in a state of emotional tension: The speakers increased the number of falsely started words, they omitted certain sounds and syllables, they misplaced their accents, they made semantically justified word substitutions, they spoke more loudly and so on.

BIBLIOGRAPHY

1. Typolohiya intonatsii movlennya" [Typology of Speech Intonations], Kiev, 1978, pp 151-152.

2. Bondarko, L. V., Verbitskaya, L. A., and Pavlova, L. P., "Acoustic Characteristics of Russian Speech Depending on Different Pronunciation Tempos," in "Voprosy fonologii i fonetiki" [Problems in Phonology and Phonetics], Part 1, Moscow, 1971, p 47.

3. Paufoshima, R. F., "Speech Tempo in Some Russian Dialects," in "Russkiye govory. K izucheniyu fonetiki grammatiki, leksiki" [Russian Dialects. Analysis of Phonetics, Grammar and Vocabulary], Moscow, 1975.

MUTUAL CORRELATION BETWEEN PERSONAL AND SPEECH CHARACTERISTICS IN AN EMOTIONALLY TENSE SITUATION

S. S. Galagudze, G. V. Nikolayeva

Analysis of studies undertaken to establish the influence of an individual's emotional states on the lexical and grammatic characteristics of his speech would show that this problem has still not been studied adequately. Soviet and foreign authors have accumulated a sufficient quantity of experimental facts on the dynamics of the higher levels of speech in the presence of different emotion-producing situations (G. Mal', N. V. Vitt, E. L. Nosenko). Nevertheless it would have to be asserted that the overwhelming majority of experimental research in this area is plaqued by a narrow empirical approach and by conflicts in the obtained results. Thus Nosenko (6) attempts to provide physiological and psychological grounds for the laws behind changes occurring in speech characteristics under the influence of emotional tension. The author interprets the emotionally grounded features of speech in light of his analysis of the general psychological characteristics of the way activity is organized in an emotional situation, characteristics which manifest themselves as a tendency to simplify speech to permit its more-optimum regulation.

However, this interpretation is clearly in conflict with the facts indicating that a number of characteristics of the lexical and grammatic level of speech improve when certain subjects experience emotional tension (3,6). In particular, we find the following remark in Vitt's work cited above: Emotional states have different stimulatory or inhibitory influences upon speech."

Without a doubt the authors listed above have accumulated sufficiently representative facts, ones which doubtlessly have pragmatic value; however, they do require theoretical interpretation and generalization.

The objective of our research was to examine the dynamics of the lexical and grammatic levels of speech in response to emotionally tense situations, from the standpoint of the information-energy approach developed in psychology by Vekker (1,2; L. M. Vekker 1976).

With this objective in mind we conducted experiments in which 30 subjects provided samples of spontaneous oral speech in an emotionally tense situation (their first examination in their principal specialized subject) and in normal conditions (background). Then the recorded speech was transcribed into typewritten text, from which the following lexical and grammatic characteristics were then isolated and analyzed:

the overall length of narration, lexicographic diversity, ratio of the number of verbs to the number of adjectives, the ratio of the number of abstract words to the number of concrete words, the quantity of "weed" words, the quantity of grammatically and logically incomplete sentences, presence of complex sentence structures and complex subordinate phrases, and the quantity of words having a clearly positive or a clearly negative connotation.

In order to reveal the dependence of speech dynamics in an emotional situation on the personal qualities of the subjects, we examined certain personality properties--introversion-extraversion and neuroticism (Eysenck's method) and anxiety (Cattell's method).

Hand tremor was used as an objective indicator of mobilization of the body's "energy resources" in an emotionally tense situation.

The experimental data were analyzed in several stages. In the first we examined the characteristics of the energy indicators of the subject in a background situation and in an emotionally tense situation. In this aspect the subjects fell into two groups with respect to the tremor dynamics indicator: The first group consisted of subjects for whom the difference between background tremor indicators and the indicators of tremor in an emotional situation was above average; in the second group this difference did not reach the average level.

In the next stage of analysis we had to compare the energy characteristics of the subjects with their speech characteristics. Comparison of energy characteristics, as defined by tremor indicators, and the speech characteristics stated above revealed the following laws behind their dynamics in response to an emotionally tense situation.

In the first group of subjects, for whom the indicator of emotional tension exceeded the sample average, the background indicators of the productivity of speech (total length of narration, lexicographic diversity, quantity of complex sentences and complex subordinate phrases) were found to be higher than in an emotionally tense situation.

In the second group of subjects, for whom the energy indicators in an emotional situation do not exceed the average, the results were different in nature. Some of the subjects were typified by the same dynamics of speech activity observed in the first group: An emotionally tense situation shortens their narration, makes it more stereotypic and so on (see table), while the other subjects of this group improved their speech characteristics in a stressful situation in comparison with the background.

Thus in the final variant we were able to distinguish three groups of subjects in terms of "energy-information" mutual dependencies:

Group 1 (9 persons)--subjects in an emotionally stressful situation exhibit improvement of a number of speech characteristics on the background of low energy activation;

group 2 (12 persons)--subjects exhibit a decrease in a number of productivity indicators of speech, on the background of low activation;

group 3 (9 persons)--subjects in an emotionally stressful situation also exhibit a decrease in a number of productivity indicators of speech, but on a background of high energy activation.

Distribution of the Indicators of Lexical and Grammatic Level of Spontaneous Oral Speech in Background and Stressful Situations Experienced by Isolated Groups of Subjects

Indicator	Group 3 Low Activation (9 Subjects) Background	Group 3 Low Activation (9 Subjects) Stress	Group 2 Low Activation (12 Subjects) Background	Group 2 Low Activation (12 Subjects) Stress	Group 1 High Activation (9 Subjects) Background	Group 1 High Activation (9 Subjects) Stress
1. Tremor energy (arbitrary units)	111.5	124.9	100.0	112.8	101.2	188.5
2. Length of narration (words)	148.8	110.6	200.2	116.6	128.2	69.2
3. Lexico-graphic diversity (index)	2.8	3.5	3.8	3.1	4.0	3.1
4. Verbs (adjectives)	2.8	2.9	2.7	3.4	1.7	2.8
5. Abstract (concrete)	0.82	0.63	6.51	0.35	0.31	0.37
6. "Weed" words (index)	0.04	0.06	0.06	0.08	0.11	0.23
7. Complex sentences and complex subordinate phrases (index)	0.77	0.92	0.73	0.35	0.52	0.37
8. Words with positive and negative connotation (index)	0.06	0.09	0.08	0.18	0.08	0.23

As may be deduced from the characteristics of the lexical and grammatic level of spontaneous oral speech, an emotionally tense situation does not have a destructive influence on statements spoken by subjects in the first group. In general the speech of subjects experiencing a stressful situation remained at the same level in regard to grammatic, syntactic and semantic structure as in a normal situation; moreover certain subjects even exhibited greater smoothness, better structure and greater expressiveness of statements in an emotionally tense situation.

At the same time, analysis of the spoken statements of subjects in the second and third groups revealed certain characteristics in the organization of speech in an emotionally tense situation, such as reduction of the total length of narration, a decrease in lexical diversity, an increase in a number of "weed" words, more frequent use of verbs than adjectives, growth in the number of concrete words, an increase in the number of logically and grammatically incomplete phrases, a decrease in the number of complex sentences and complex subordinate clauses, and an increase in the number of words having positive and negative connotation.

These laws agree well with published data (5,6) on the dynamics of emotionally dependent speech.

The next stage of analysis of the obtained data presupposed comparison of the speech characteristics of the subjects with their personality features. We found it interesting to compare the first and second groups of subjects in relation to their speech and personality features, since totally opposite tendencies in the nature of changes in speech activity were observed in these groups on the background of low energy activation in an emotionally tense situation.

We attempted to reveal the dependence between the indicators of extraversion-introversion, neuroticism and anxiety on one hand and the successfulness of speech in an emotional situation on the other hand by the method of tetrachoric correlations.

We found that the successfulness of speech in an emotionally tense situation is negatively correlated with the anxiety indicator ($r = 0.52$ at $p<0.05$) and positively correlated with the extraversion-introversion indicator ($r = 0.57$ at $p<0.05$). The coefficient of correlation between the neuroticism indicator and the successfulness of speech in a test situation was nearly statistically significant ($r = 0.39$) and negative in sign.

Thus the first group of subjects was dominated by unanxious extraverts exhibiting low neuroticism while the second group was dominated by anxious introverts exhibiting high neuroticism.

The third group of subjects varied in the type of laws that were applicable; however, the highly "energized" state of subjects in the emotion-producing situation made it different from the first two groups. It would be natural to assume that this case reflects the curvilinear nature of the "energy-information" function: According to the hypothesis of curvilinear dependence (presence of an inverted U-shaped dependence), as the energy factor rises in value the productivity of activity rises, and then, after achieving a certain optimum, it begins to decrease. Hence we can conclude that the law discussed above represents the descending branch of the "energy-information" curve. In this case the intensity of emotional tension exceeds a certain optimum, which is why destructive changes occur in the speech of subjects in this group.

All of the above-noted characteristics of speech in an emotional situation imply simplification of the informational structure of spoken signals. Thus reduction of the density of propositions, achieved in particular owing to growth in the number of irrelevant repetitions of words, parts of words and so on in emotionally dependent speech, and owing to growth in the number of "weed" words, is a direct indication of a decline in the information level of the communications of the subjects.

Dominance of the number of verbs over adjectives in the speech of subjects under emotional tension, a decrease in lexical diversity and so on lead to inadequately precise transmission of the contents of the message or, in other words, to parallel ordering of different levels of generality. Speech reveals a tendency toward syncretism, to adjacent ordering of constructs that should have been hierarchically ordered.

Let us now examine the dependencies revealed for the first and second groups of subjects. At first glance the pattern appears contradictory, since on the background of identically low activation in an emotionally tense situation, the changes in the speech of these subjects resulting from this tension occurred at different planes. However, when we compare this planar variation with the personality characteristics of the subjects, this contradiction disappears.

As was shown in research by M. S. Zhamkoch'yan, inward direction of the individual's thoughts manifests itself most clearly in the presence of anxiety, such that behavior is typified by restlessness, concern for the individual's physical well-being, high nervousness and subjectivity, dependence of behavior on the individual's moods and physical condition, and shifting of attention on the self. Consequently as an energy characteristic, anxiety represents a certain potential of subjectivity. Neuroticism may be classified in similar fashion. Extraversion, meanwhile, can be interpreted as an indicator of an aggressive, public orientation, and in this sense it represents a potential for outward orientation upon another object.

Anxious introverts are dominated by the motive of personal well-being, which in stressful conditions elicits a dominant uneasiness about failure. As a result such an individual expends considerable energy on his emotional experience of possible failure, which reduces the amount of energy that can be devoted to the activity itself. A lower information level would also correspond to a lower energy level, and therefore low activation in subjects of the second group is accompanied by a decline in the productivity of speech under the influence of an emotionally tense situation.

On the other hand if an orientation toward activity dominates over a subject's motive of personal well-being, the energy outlays are redistributed in favor of the object of activity. This law is observed in the fact that in the first group of subjects, which was dominated by unanxious extroverts exhibiting low neuroticism, the productivity of a number of speech characteristics improved on the background of low activation.

Our work was an attempt at applying the information-energy approach to the study of emotionally dependent speech. It permitted us to explain certain facts that do not fit within the framework of traditional concepts about the nature of changes occurring in speech in stress-like and stressful situations. It stands to reason that the problem is not provided with an exhaustive solution by the conclusions suggested here. The obtained results can be thought of only as a tentative working hypothesis on the road to solving the multifaceted problem of speech, personality and emotions in their mutual relationship.

BIBLIOGRAPHY

1. Vekker, L. M., "Psikhcheskiye protsessy" [Mental Processes], Vol 1, 1974, Vol 2, 1976.

2. Vekker, L. M., and Paley, I. M., "Information and Energy in Mental Reflection," in "Eksperimental'naya i prikladnaya psikhologiya" [Experimental and Applied Psychology], Issue 3, 1971.

3. Vitt, N. V., "Speech and the Critical Situation," in "Predvaritel'nyye materialy eksperimental'nykh issledovaniy po psikholingvistike" [Initial Findings in Experimental Research on Psycholinguistics], Moscow, 1974.

4. Zhamkoch'yan, M. S., "Correlations Between Intelligence Characteristics and Orientation Toward the Self and Toward the Group," diploma project, Leningrad State University, Department of Psychology, 1973.

5. Leont'yev, A. A., and Nosenko, E. L., "Some Psycholinguistic Characteristics of Spontaneous Speech in a State of Emotional Tension," in "Obshchaya i prikladnaya psikholingvistika" [General and Applied Psycholinguistics], Moscow, 1973.

6. Nosenko, E. L., "Osobennosti rechi v sostoyanii emotsional'noy napryazhennosti," [Characteristics of Speech in a State of Emotional Tension], Dnepropetrovsk, 1975.

7. Mahl, J. F., "The Lexical and Linguistic Level in the Expressions of Emotions," in Knapp, P. H. (Editor), "Expression of the Emotions in Man," 1963.

USING SPEECH CHARACTERISTICS TO MONITOR EMOTIONAL STATE IN CHILDREN

V. I. Galunov, A. I. Kliorin, M. M. Chekodanova

The objective of our work was to test, in real conditions, the method of determining an individual's psychophysiological (emotional) state on the basis of the temporal and melodic characteristics of the spoken signal (1). The work involved parallel study of the parametric characteristics of spoken signals, adaptive reactions and a number of biochemical and physiological indicators widely used in clinical practice. The research was conducted on children 5-9 years old. The child's hospitalization was the emotion-producing factor.

Research Methods

The study was performed on 31 children, for 27 of whom this was the first visit to the clinic. It was the second visit for 3, and the third visit for 1. The children were examined twice--on the 1st-2d day of hospitalization and on the 9th-10th day. We analyzed blood sugar and cholesterol, measured arterial pressure, performed an associative experiment and recorded a speech test. The children were divided into two age groups--5-7 years (unable to read) and 7-9 years (reading fluently). Subjects were shown pictures of an object or situation (representing, for example: "I have a coat, a hat, boots and a scarf"). They were asked to answer a question about the given picture and tell a story relating to it. Subjects in the second group were asked to read a standard text (for example: "We bought a notebook, a book and a pencil") and to tell a story related to it. The spoken material was recorded simultaneously by two microphones. One was set up 40 cm from the subject's mouth while the other was secured to the jugular fossa. The signal from the first microphone was used in analysis of the spectral and temporal characteristics of speech, while that from the second microphone was used to analyze the parameters of the fundamental tone (FT) and particularly the melodic characteristics of the FT. The spectral and melodic characteristics were analyzed by standard methods (2).

Serum cholesterol was analyzed by the Mrakos-Tovarek method. Blood sugar was analyzed by the Hagedorn-Jensen method. Blood was sampled from the ulnar vein of the children, on an empty stomach 3 hours prior to study of speech indicators. Arterial pressure was measured by Korotkov's method immediately prior to study of speech indicators.

APPROVED FOR RELEASE: 2007/02/09: CIA-RDP82-00850R000500030032-8

Research Results

Two scales with distinct correlates in speech parameters were used to describe the emotional state of the children: "inhibition-arousal" and "normal-fear." It was hypothesized on the basis of existing data that an increase in the tempo of speech, growth in the average frequency of the FT, an increase in the range of change in the FT and an increase in its unevenness correspond to change in state from inhibition to arousal. Change in state from normal to fear produces an increase in the number of pauses in speech without a change occurring in the length of an individual element, which is reflected as an increase in the length of standard phrases without change in the duration of the vocalized (voiced) fraction.

Comparative analysis of spoken material obtained in the first and second recording sessions showed that for most children, fear was experienced less in the second recording session. Only three children exhibited the opposite picture. Greater depression in the second recording session was diagnosed for two children on the basis of their speech indicators. This result is consistent with the observed pattern of changes in the behavior of these children.

Comparative analysis of the informativeness of the speech parameters we studied showed that:

For practical purposes the mean FT frequency does not change under the conditions of this experiment;

change in the variance of the FT is rather irregular (which is probably aggravated by the inadequate stability of the voice control mechanism in children);

unevenness of the FT melody was found to be the most informative indicator.

The biochemical indicators we studied did not produce a stable picture permitting us to diagnose change in psychophysiological state as a stress reaction to hospitalization. It should be noted that a decline in cholesterol was revealed in both children who exhibited increased depression in the second recording session. Moreover there is a general tendency for cholesterol to decline as arousal, diagnosed on the basis of speech characteristics, decreases. Observation of the dynamics of arterial pressure revealed its normalization in the studied group of children by the 3d-4th day of hospitalization, which correlates well with the revealed pattern of changes in speech parameters.

Thus the research indicates that:

The existing methods of evaluating pscyhophysiological state on the basis of characteristics of the spoken signal may be used successfully as one of the indicators for diagnosing emotional stress in the children's clinic;

the corresponding indicators should be considered when organizing therapy.

BIBLIOGRAPHY

1. Galunov, V. I., and Manerov, V. Kh., "Ways of Solving the Problem of Creating Systems by Which to Determine a Speaker's Emotional State," VOPROSY KIBERNETIKI, Moscow, Vol 22, 1976, pp 95-113.

2. Sapozhkov, M. A., "Rechevoy signal v kibernetike i svyazi" [The Spoken Signal in Cybernetics and Communications], Moscow, Svyaz'izdat, 1963.

EFFECT OF INDIVIDUAL AND EMOTION-DEPENDENT CHANGES IN PARAMETERS OF THE ARTICULATORY TRACT ON CHARACTERISTICS OF THE SPOKEN SIGNAL

V. I. Galunov, S. L. Koval', I. B. Tampel'

The objective of our work was to determine the parameters of the spoken signal dependent on the psychophysiological state of the speaker and to reveal the limits of possible changes in these parameters. Such an objective is usually reached by analyzing spoken signals obtained from a speaker experiencing different states. Some standard analysis procedures are used, and the characteristics to be used in diagnosis are selected randomly. Because both the analysis methods and the characteristics they reveal are rather sizeable in number, it may take a very long time to sift through all of the acquired data. However, we could somewhat alter the plan of research: We could start by analyzing the speech formation process, and after this, using the standard methods of the acoustic theory of speech formation, we could isolate those parameters which reflect specific features of the emotional spoken acoustic signal.

In this paper we will illustrate this approach to the problem using as our example the changes occurring in the tone of the speech control muscles of the speaker as he experiences different states. Muscle tone is known to be one of the physiological parameters that changes faithfully in response to changes in the speaker's psychophysiological state. In order to make it possible to capitalize on this fact to diagnose a speaker's state on the basis of spoken signals, we would have to answer the following questions:

1. Precisely how is each concrete state associated with muscle tone?

2. In what way can muscle tone be translated into the language of mechanical tissue constants (coefficients of tension or moduli of elasticity, impedances) used in the acoustic theory of speech formation?

3. What sort of influence do changes in the mechanical constants of the tissues of speech forming organs have on the spoken signals?

Our end was to answer the third of these questions. It should be considered that the results of such research could have independent significance, since they establish the set of parameters of the spoken signal which are subjected to the greatest changes. Further research on the spoken signal may establish in what manner these parameters change, if the answers to the first two questions are known.

According to traditional ideas, the spoken signal is the result of filtration of a signal by the vocal tract; therefore we studied the influence changes in mechanical parameters of tissues have on pulsations in the volumetric rate of the voice source and on the transmission function of the channel. Among the mechanical constants controlling oscillations of the vocal cords, the coefficients of tension and the distance between the resting vocal cords can apparently vary to the greatest degree.

The transmission function of the channel is basically determined by the configuration or function of the particular channel area; however, the impedance of the channel walls, which clearly depends on the tone of speech muscles, has an influence on the position and width of the first formants.

Speech formation has been studied with a mathematical model of the vocal cords (1) and the vocal tract (3) using an EVM-222 computer.

In the first part of the research the authors examined the influence of the variations in the coefficients of tension and the distance between resting vocal cords on the volumetric rate of air just beyond the vocal cords. A model with two degrees of freedom accounting only for symmetrical oscillations was studied (1). The overall tension of the vocal cord is described in this model by five tension coefficients; all coefficients varied proportionately during simulation.

The volumetric rate of air flow just beyond the vocal cords at different cord tensions is graphed in Figure 1. Table 1 shows the basic numerical results. "K" stands for the ratio of the coefficients of vocal cord tension to the corresponding coefficient values adopted as normal. The period and amplitude of the oscillations were measured 20 msec after cord pressure was "turned on" so that unstabilized oscillations could be avoided.

It follows from analysis of the results that when the tension of the vocal cords increases, the oscillation frequency grows by about up to half an octave over normal. The oscillation amplitude drops at this time. Beginning with a certain value of K, rather than coming together, the vocal cords oscillate without touching each other. As tension increases further, the oscillations attenuate soon after vocal cord pressure is applied. As tension increases the oscillation spectrum becomes simpler, especially in the high frequency range. When the tension of the vocal cord decreases, the oscillation amplitude increases while the frequency drops by within two octaves. The duty factor (the ratio of the time the vocal cords remain in contact with one another to the length of the oscillation period) grows at first and then declines. As the duty factor increases, the high frequency oscillation spectrum grows, while the oscillations themselves retain their previous shape. Because this is so, the ratio between the levels of the spectral components of the voice source in the high and low ranges at first grows somewhat in response to an increase in tension, and then declines significantly. The relationship of these changes in the tension of the vocal cords to the individual's actual anatomical and physiological potentials is not exactly known; however, it may be hypothesized that tension may increase within rather broad limits, while its decrease is apparently limited.

Figure 2 and Table 2 show data illustrating the dependence of the volume flow rate beyond the vocal cords at different values for the distance between the vocal cords (the initial equilibrium distance separating the vocal cords in the absence of vocal cord pressure). A negative distance means that before starting their work,

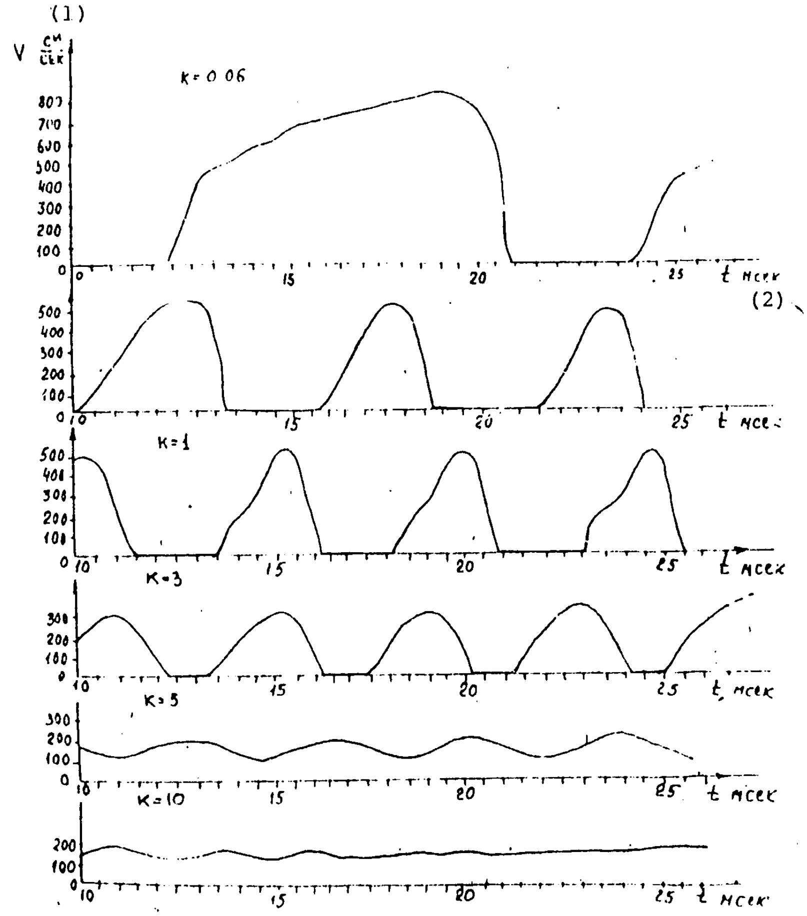

Figure 1. Volumetric Air Flow Rate Beyond Vocal Cords 10 msec Following the Start of Work: K = 1--normal tension; K>1--vocal cords more tense; K<1--vocal cords less tense

Key:
1. cm/sec
2. msec

the vocal cords are pressed against each other--that is, they invade each other's space by distance h. As h increases, at first the oscillation amplitude grows while the frequency remains practically constant. Then, after the vocal cords break contact (that is, when the oscillation spectrum changes abruptly) the amplitude and frequency decrease noticeably, and finally the oscillations begin to attenuate.

Table 1

«K»	(1) Частота основного тона (гц)	(2) Амплитуда скорости см³/сек	(3) Скважность	(4) Примечания
0,002	50	1823	0,29	
0,03	81	801	0,26	
0,25	146	455	0,55	
0,5	182	477	0,53	
1	217	483	0,48	
2	245	374	0,35	
3	260	317	0,28	
5	278	102	0	
10	393	17	0	колебания затухают (5)
20	551	14	0	
40	737	5	0	

Key:
1. Frequency of fundamental tone (Hz)
2. Rate amplitude, cm^3/sec
3. Duty factor
4. Comments
5. Oscillations attenuate

As h grows the high frequency spectrum becomes simpler. As h decreases, and then as the vocal cords press together more and more firmly, the oscillation frequency grows insignificantly while the amplitude declines. The speaker can easily compensate for this drop in amplitude by increasing vocal cord pressure; however, this would elicit a rise in the frequency of the fundamental tone and an increase in the duty factor of the oscillations. It should be noted in general that in terms of emotional speech, the amplitude of acoustic oscillations may be a significant parameter describing the state of the speaker. The speaker is able to compensate for amplitude changes, though apparently this changes the intensity ratios of different sounds.

Thus far we have been discussing the characteristics of the voice source. The influence the speaker's state has upon the transmission function of the vocal tract should be discussed as well. The fact is that changes in psychological state do more than change the tension of muscles working the vocal cords; they also cause change in the tone of the muscular walls of the articulatory tract, and consequently in a parameter of the mathematical model of speech formation characterizing this quality of the walls--the impedance of the vocal tract walls.

According to present ideas (2) the impedance of vocal tract walls is close to 1600+1.5 gm/sec·cm^2. The influence of variations in the real and imaginary fractions of impedance upon the position and width of formants was investigated. Calculations were performed using the algorithm suggested in (3). It was found that only the imaginary fraction of impedance has a significant influence on formants. As impedance grows the frequency of formants diminishes, and their width decreases. The

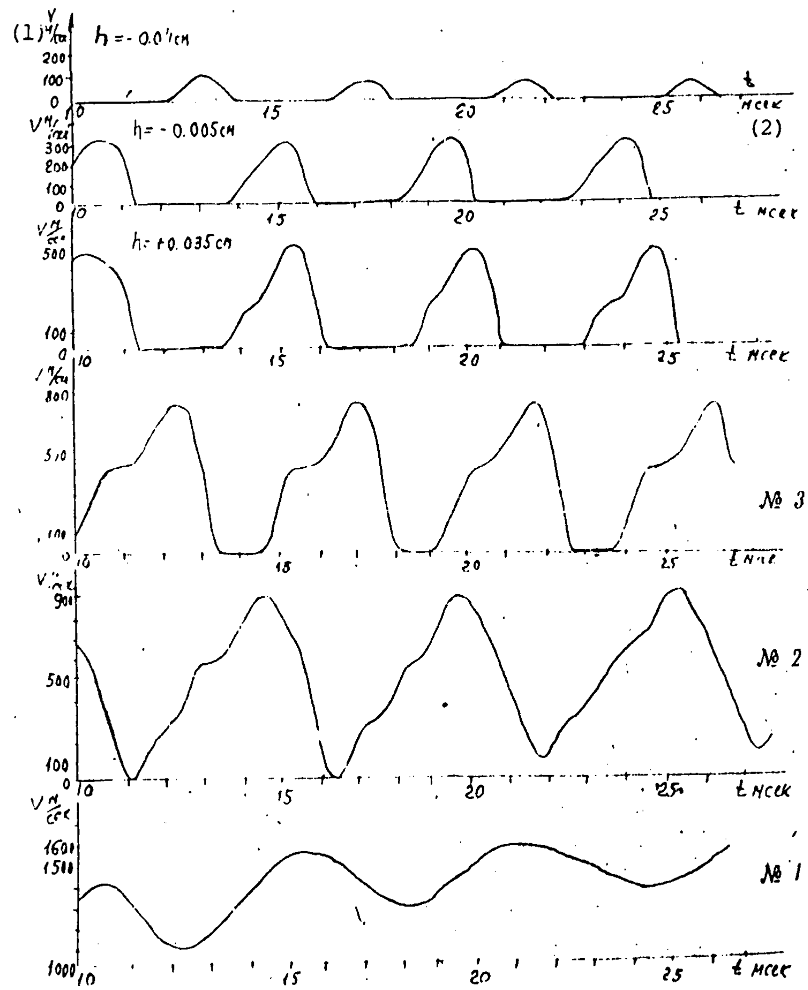

Figure 2. Volumetric Air Flow Rate Beyond Vocal Cords 10 msec After the Start of Work at Different Initial Distances Between Cords: No 1--h = +0.105; No 2--h = +0.18; No 3--h = +0.48

Key:
1. m/sec
2. msec

width of the first formant changes most significantly--by 2-2.5 times in response to a two-time change in the imaginary component. The calculation results are shown in Table 3. Here, F_1 and F_2 are the frequencies of the first and second formants, and B_1 and B_2 are their widths.

Table 2

h(см)	Частота основного тона (гц) (1)	Амплитуда скорости см³/сек (2)	(3) Скважность	(4) Примечания
0,48	130	300	0	
0,24	138	342	0	колебания затухают (5)
0,21	142	438	0	
0,18	161	860	0	
0,13	215	714	0,3	
0,105	215	728	0,28	
0,07	213	601	0,4	
0,035*	217	483	0,48	
0,015	222	400	0,51	
0,005	225	308	0,54	
0,015	228	260	0,57	
0,04	237	108	0,64	
0,05	252	17	0,76	
0,06	270	.	.	колебания затухают (5)

*Adopted as normal.

Key:
1. Frequency of fundamental tone (Hz)
2. Rate amplitude, cm³/sec
3. Duty factor
4. Comments
5. Oscillations attenuate

Table 3

(1) Гласные	(2) Импеданс г/сек·см²	Форманты (3)			
		F₁(гц) (4)	B₁(гц)	F₂(гц)	B₂(гц)
[а]	1600+0,8	682	74	1066	51
	1600+1,5	654	34	1052	36
	1600+3,0	636	20	1042	31
[о]	1600+0,8	545	84	853	45
	1600+1,5	520	36	840	20
	1600+3,0	500	17	830	20
[u]	1600+0,8	364	>200	650	94
	1600+1,5	321	102	620	41
	1600+3,0	282	42	610	24
[е]	1600+0,8	494	126	1924	77
	1600+1,5	464	54	1907	69
	1600+3,0	436	24	1897	66
[i]	1600+0,8	334	>200	2287	39
	1600+1,5	315	101	2271	34
	1600+3,0	280	42	2262	32
[ı]	1600+0,8	384	198	1522	120
	1600+1,5	357	77	1500	105
	1600+3,0	324	32	1488	101

Key:
1. Vowels
2. Impedance, gm/sec•cm²
3. Formants
4. Hz

And so, we can assert on the basis of the suggested approach and the preliminary trial calculations that specific features of speech associated with different psychophysiological states of the speaker may be revealed through the ratio of the levels of the high frequency and low frequency spectrums, through the frequency of the fundamental tone, the amplitude of acoustic oscillations, the ratio of formant widths and the dynamics of changes in these parameters from one sound to the next.

BIBLIOGRAPHY

1. Tampel', I. B., and Yurgenson, A. O., "Simulation of the Work of the Vocal Cords," TEKHNIKA SREDSTV SVYAZI, N. T. SB. MPSS, SERIYA TEKHNIKA PROVODNOY SVYAZI, Moscow, No 9, 1976, pp 64-70.

2. Flanagan, J. L., Ishizaka, K., and Slipley, K. L., "Synthesis of Speech From a Dynamic Model of the Vocal Tract and Vocal Cords," THE BELL SYST. TECHN. J., Vol 54, No 3, 1975, pp 485-506.

3, Garbaruk, V. I., and Tampel', I. B., "Dependence of Formant Structure on Variations in the Shape of the Vocal Tract," TEKHNIKA SREDSTV SVYAZI, N. T. SB., SERIYA TEKHNIKA PROVODNOY SVYAZI, Moscow, No 9, 1976, pp 44-48.

FORMANT FREQUENCY AS AN INDEX OF VOICE INDIVIDUALITY

V. B. Gitlin

Man is known to identify a speaker better by vowels (1,2), the spectrum of which is determined uniquely by the parameters of formants and of the voice source. We analyzed the correlation between individual frequencies of the first three formants of six sustained vowels а, о, у, ы, и, э.

Formant frequency was chosen as the object of analysis for the following reasons. First, it was one of the principal acoustic characteristics of the semantic content of speech, inasmuch as it is a polar parameter of the speech tract. Second, methods of isolating formant frequency are rather well developed today in comparison with, for example, the methods of isolating formant width.

Formant frequency was isolated by the method suggested in (3). Formant frequency was defined as the frequency at which the signal at the output of the formant filter intersects the zero line in synchrony with the fundamental tone. A description of the set-up for isolating formant frequency can be found in (4).

Phonemes were pronounced by speakers in a soundproof room. A total of 35 speakers participated in the tests--22 men and 13 women. The speaking time and the length of pauses were about 1 second. Twenty utterances of one phoneme were recorded during one testing session. Tests were performed for 2 months at irregular time intervals. Formant frequency was determined for each speaker as the average for all 20 utterances. One testing session was performed with each phoneme.

The measurement results are graphed in Figure 1. Here, Arabic numerals indicate the numbers of the male speakers (from 1 to 22) and Roman numerals indicate the numbers of female speakers (from I to XV). We can conclude from the appearance of the graphs that correlation exists between formant frequencies and the speaker's voice. The data in Figure 1 were used as a basis for calculating correlation coefficients using well known formulas:

$$r[F(x), F(y)] = \frac{\frac{1}{M}\sum_{i=1}^{M}[F_i(x) - \bar{F}(x)][F_i(y) - \bar{F}(y)]}{\sigma[\bar{F}(x)]\,\sigma[\bar{F}(y)]}, \qquad (1)$$

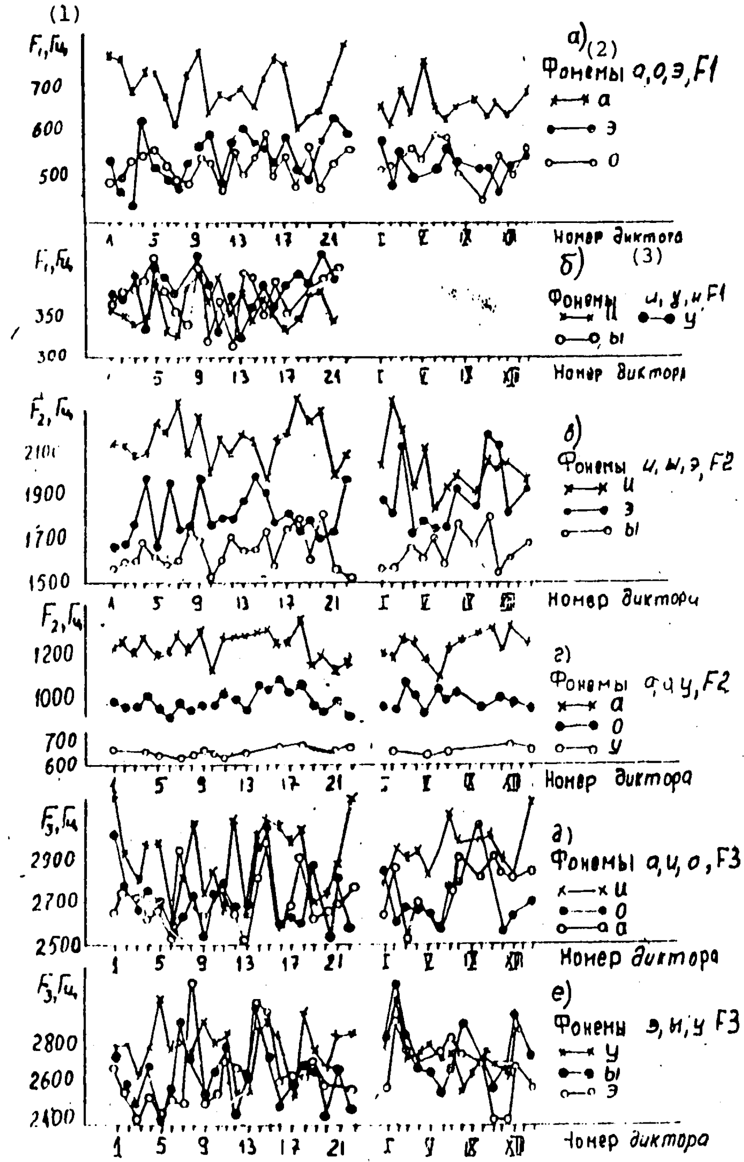

Figure 1. Dependence of Phoneme Frequency on Speaker

Key:
1. Hz
2. Phonemes
3. Speaker number

Table 1. Sample Standard Deviations $\sigma\{F(s)\}$, Hz

Formant	Phoneme					
	а	и	о	у	ы	э
	52,2	21,3	32,2	21,6	24,1	45,4
	56,4	113	42,2	15,1	80,4	127
	121	158	130	93,2	134	145

Table 2. Correlation Coefficients

Formant	Phoneme	Phoneme				
		а	и	о	у	ы
F_1	и	0,160	1			
	о	0,119	0,011	1		
	у	0,123	0,054	0,149	1	
	ы	0,211	0,076	0,147	0,108	1
	э	0,188	0,092	0,297	−0,152	0,082
F_2	и	0,299	1			
	о	0,279	0,190	1		
	у	0,227	0,054	0,242	1	
	ы	0,465	0,076	0,218	0,236	1
	э	0,341	0,066	0,194	0,299	0,127
F_3	и	0,415	1			
	о	0,293	0,353	1		
	у	[illegible]	0,132	0,170	1	
	ы	0,357	−0,051	0,468	0,412	1
	э	0,124	0,215	0,468	0,283	0,590

where $F_1(x)$, $F_1(y)$--formant frequencies for phonemes x and y respectively, averaged over 20 realizations, $i = 1,2,\ldots,$ M--speaker number,

$$\sigma^2\{F(s)\} = \frac{1}{M}\sum_{i=1}^{M}[F_i(s) - \bar{F}(s)]^2, \tag{2}$$

$$\bar{F}(s) = \frac{1}{M}\sum_{i=1}^{M} F_i(s) \tag{3}$$

Tables 1 and 2 show the results of calculations based on formulas (1)-(3). The calculations were made for the entire group of speakers (men and women together). The parameters of the first formant of the phonemes и, у, ы were calculated for men only. Inasmuch as the resulting matrix of correlation coefficients is symmetrical, the repeating elements are dropped from Table 2.

If we arrange phonemes in the order of decreasing standard deviations shown in Table 1, we get the following sequence:

a) for F_1--а, е, о, ы, у, и;

b) for F_2--э, и, ы, а, о, у;

c) F_3--и, э, ы, о, а, у.

The succession of phonemes in order of increasing error of identification of the speaker by another person, obtained in (2), was as follows: и, э, ы, о, а, у. It is consistent with the arrangement of phonemes in order of decreasing standard deviation for F_2 and F_3. Thus our data do not contradict the hypothesis that deviation of individual formant frequencies from a certain average is an index used to recognize a person's voice.

As follows from Table 2, a definite correlation exists between the individual frequencies F_1 for all phonemes and formants. This correlation is rather weak for the first formant, but it becomes significant for the second and especially the third formants.

BIBLIOGRAPHY

1. Ramishvili, G. S., "Automatic Voice Recognition," TEKHNICHESKAYA KIBERNETIKA, No 5, 1966.

2. Gitlin, V. B., Kuznetsov, P. G., Tikhonov, G. A., and Chepkasov, A. G., "The Informativeness of Russian Phonemes in Determining the Individuality of a Speaker's Voice," in "Avtomatika i opoznaniye obrazov. Avtomaticheskiye ustroystva ucheta i kontrolya" [Automation and Pattern Recognition. Automatic Accounting and Monitoring Units], Issue 4, Izhevsk, 1969, pp 54-60.

3. Campanella, S. I., and Coulter, D. S., "Formant Period Tracker," U.S. Patent, Cl. 179-1, No 3335225, 8 August 1969.

4. Gitlin, V. B., and Smetanin, A. M., "Formant Filter Design," IZVESTIYA VYSSHIKH UCHEBNYKH ZAVEDENIY. RADIOELEKTRONIKA, Vol 19, No 8, 1976, pp 98-100.

EFFECT OF DIFFERENT EMOTIONAL STATES ON CHANGE IN THE SPECTRUM OF ENGLISH VOWELS

T. G. Gomina

There are a large number of difficulties in instrumental analysis of emotionally colored, speech, stemming on one hand from the redundancy of the spoken signal and on the other hand from the absence of a clear classification of emotional states. Experimental research conducted in recent years showed that "the number of discontinuities in the spectrum of a vowel detectable by man is not only greater than two, but also, what is especially unpleasant, it differs for different vowels and depends on the concrete features of the stimulus" (1). What is implied here is that the vowel spectrum is segmented. It would be easy to imagine how much more difficult it would be to interpret a suprasegmental spectrum (2) dependent on the structure of the speech forming apparatus and the emotional state of the speaker, and in a sense superimposed over the segmental spectrum. This obviously explains the fact that almost no descriptions of a suprasegmental spectrum in acoustic parameters have been attempted in recent years. As of today, our information on the prosodic characteristics of emotional speech is much richer than our information on the suprasegmental spectrum. This is what motivated us in our choice of a research topic. Considering the fact that a tested and sufficiently reliable procedure for processing spectrograms of emotional speech does not exist yet, our first task was to find a procedure which would satisfy the objectives of our research to the greatest degree.

In our experiment we analyzed phrases uttered by 11 English speakers. The phrases were spoken in five emotional states: normal (neutral pronunciation), joy, anger, fear and melancholy. Two series of experimental material were employed. The first included four phrases with stressed vowels [e], [a:], [ɔ:], and the second included four phrases having [e], [ʌ], [a] as the stressed vowels.

One hundred twenty realizations were selected for spectrographic study as the result of listener analysis. Half of the selected phrases were analyzed with a 24-channel dynamic spectrograph at the Laboratory of Speech Pattern Recognition of the Kaunas Polytechnical Institute. The other half of the phrases were analyzed with the Kay-Sonograph spectrum analyzer produced by "Kay-Elemetrics." The recordings were made in contour block mode.

Processing of the spectrograms entailed dividing the spectrograms into bands coinciding with the formant ranges of the analyzed sound, allowing in this case for possible overlap of formant ranges. This method of dividing spectrograms into bands had been

APPROVED FOR RELEASE: 2007/02/09: CIA-RDP82-00850R000500030032-8

used in (3), the objective of which was to stud· the spectrums of Russian vowel sounds perceived by ear. In our experiment we found division of the spectrum into four bands having the following boundary frequencies to be the most informative: 250-980 Hz (the range of the fundamental tone and the first formant); 980-2180 Hz (the range of the second formant); 2180-3880 Hz (the range of the third formant) and 3880-8570 Hz (the range of higher formants). The total energy of spectral components in the appropriate band, expressed as a percentage of the total vowel energy in all four bands, was adopted as the energy characteristic of the spectrum of the given band. The energy characteristics were also compared in broader bands: The 250-2180 Hz band, treated as the low frequency range, and the 2180-8570 Hz band, treating it as the high frequency range. The analysis results showed that information indicating arisal of emotional tension can be obtained just by making this division into two bands (low and high frequency). Deformation of the spectrum in comparison with a neutral state, expressed as a shift in spectral energy into the higher frequencies, is observed when joy, anger and fear are expressed, and a shift into lower frequencies is observed when melancholy is expressed. More-detailed analysis accounting for four spectral bands revealed, besides the tendencies indicated above, a large number of individual traits in the spectrum of emotionally colored signals, typical of specific speakers and expressing themselves in most cases as compensatory processes, mainly between the first two bands of the vowel spectrum.

Additional characteristics of emotional speech were revealed by sonogram analysis, performed on the basis of a vocal signal classification suggested by Trojan (4). This classification is arrived at by contrasting the vocal signals in terms of sparing (forced) voice formation and a dilated (constricted) pharyngeal cavity. According to Trojan, constriction of the pharyngeal cavity and forced voice formation occur when fear, anger, melancholy, pain and revulsion are expressed. On the other hand dilation of the pharynx and sparing voice formations are typical of expressions of pleasure, joy and satisfaction. The author suggests acoustic correlates for each of these characteristics. Analysis of sonograms produced in our experiments showed that an emotional state such as anger can be expressed only through forced voice formation without constriction of the pharynx, while joy may be expressed through forced voice formation, and not just sparing voice formation alone. A comparison with normal sonograms revealed the following acoustic characteristics of emotional states. Joy is characterized by expansion of the spectrum of stressed vowels due to arisal of additional high frequency components, by a dominance of vocalism (5), by greater intensity of the high frequency components of the vowel spectrum, and by greater distinctiveness of unstressed vowels, expressed as expansion of their spectrums in comparison with both normal and other emotional states. Anger is also accompanied by expansion of the spectrum of stressed vowels, by greater intensity of high frequency components of the spectrum--though to a lesser degree than with expression of joy, and by insignificant dominance of vocalism. Slight expansion of the spectrum of stressed vowels and dominance of consonantism (5) are noted in the expression of fear, while an increase in vowel intensity is not observed. Pronunciation of consonants (such as -pl-, -ktr-) is typified by significant deadening of sonants, which reveals itself in recordings of the high frequency components of their spectrums. Expression of melancholy is accompanied by narrowing of the spectrum of stressed vowels--that is, by registration of only the low frequency characteristics, and by dominance of consonantism. Pronunciation of consonants and dull, explosive consonants, as in fear, is accompanied by registration of high frequency components of significant intensity.

This analysis of emotional speech at the spectral level confirmed the possibility for distinguishing among different emotional states on the basis of parameters such as shifting of spectral energy characteristics, expansion or narrowing of spectrums coupled with simultaneous enlargement or reduction of the intensity of spectral components, dominance of consonantism or vocalism in spectrums, and characteristic expression of consonants (especially in expressions of fear and melancholy). It should be noted, however, that the proportion contributed by each of these parameters varies in expressions of different emotional states. Moreover expressions of emotion manifest highly noticeable individual characteristics of the speaker, ones which are a consequence of the structural uniqueness of the speech apparatus, and of the individual's habits of pronunciation. Having all of this in mind, we would have to hypothesize that it is impossible to create a certain universal model of each emotional state. All we can do is reveal the most general tendencies, ones allowing for a certain degree of variability, inasmuch as signals associated with vocal expression--Trojan validly notes--are always located in a space delimited by two opposite poles: sparing (forced) voice formation at one end and dilation (constriction) of the pharyngeal cavity on the other.

BIBLIOGRAPHY

1. Chistovich, L. A., et al., "Fiziologiya rechi. Vospriyatiye rechi chelovekom" [Speech Physiology. Speech Perception by Man], Leningrad, 1976.

2. Tseplitis, L. K., "Analiz rechevoy intonatsii" [Analysis of Speech Intonation], Riga, 1974.

3. Varshavskiy, L. A., and Chistovich, L. A., "Average Spectrums of Russian Vowel Phonemes," in "Problemy fiziologicheskoy akustiki" [Problems of Physiological Acoustics], Vol 4, Leningrad, 1959.

4. Trojan, F., "Biophonetik," Zurich, 1975.

5. Potapova, R. K., "Temporal Organization of the Syllable as the Basic Component of the Temporal Emotive Characteristic," in "Tezisy dokladov ARSO-9" [Abstracts of Reports at the ARSO-9], Minsk, 1976.

ACOUSTIC ORGANIZATION OF SPEECH AS ONE OF THE MEANS OF ITS EMOTIONAL COLORATION

A. P. Zhuravlev

The sounds of speech, which impart phonetic value and phonetic symbolism to the latter, basically have a connotive and expressive nature, and therefore they may be used in speech as a means of intensifying emotional expressiveness.

Language makes broad use of the expressive possibilities of phonetic value in the acoustic make-up of words having an expressive coloration.

Let us examine the following synonym progression: лик [countenance], лицо [face], морда [mug], рыло [snout], харя [muzzle].

The synonyms in this progression are sylistic, and they are arranged in order of stylistic decline from the poetic "лик" through the neutral "лицо" to the coarser "морда-рыло-харя." With the exception of expressively neutral "лицо," all of the other synonyms in this progression are emotionally colored. This emotional coloration may be described by measuring the synonym progression against the attribute scale "beautiful-repulsive." Then the progression of synonyms given above would reflect a transition from the attribute "beautiful" (лик) to the attribute "repulsive" (харя).

The phonetic value (the meaningfulness of the acoustic form) of these words may be calculated by the formula:

$$S_t = \frac{\sum x_i k_i}{\sum k_i}, \text{ where } k_i = \frac{P_{max}}{P_i}$$

with $$k_i = 4\frac{P_{max}}{P_i}, \quad k_{ул} = 2\frac{P_{max}}{P_{ул}}$$

Here, S_t--calculated (theoretical) phonetic value; x_i--phonetic value of the i-th sound; P_i--normal frequency of the i-th sound; P_{max}--maximum frequency of the sound in the given word; i--each sound of the word.

In terms of the calculated phonetic value, all synonyms of our progression fall into the exact same sequence in which they arrange themselves with respect to the attribute scale:

Scale	Beautiful-Repulsive				
Words	лик	лицо	морда	рыло	харя
S_t	2.0	2.4	2.9	3.3	3.5

The following values are assigned to the degrees of the scale to calculate S_t:

Very beautiful	- 1
Beautiful	- 2
Neither	- 3
Repulsive	- 4
Very repulsive	- 5

Consequently the scores for the words "лик" and "лицо" may be interpreted as "beautiful," the score for "морда" may be interpreted as "neither," the score for "рыло" may be interpreted as "more likely repulsive," and the score for "харя" may be interpreted as "repulsive." These scores are also confirmed by the function of these words in speech--in particular, the possibility of forming diminutive affectionate forms. The words "лик" and "харя" do not allow for formation of such forms; this may be explained by the fact that they express the extreme degrees of the attribute "beautiful" and "repulsive"--that is, they are in a sense the ideal standards on the scale. The words "лицо" and "морда" may take diminutive affectionate forms--личико, мордочка, мордашка, which corresponds to their middle position on the scale. But the word рыльце, while being the diminutive affectionate form, it does not carry this value in language, instead acquiring what is more likely a degrading coloration: "ryl'tse v pushku" [he/she is up to his/her little nose in the dirty business].This reflects transition of the phonetic value of the word "рыло" into the "repulsive" half of the scale. Here are a few more examples of the correspondence between phonetic value and expressive coloration of words in synonym progressions:

Scale	Graceful-Coarse		
Words	очи [orbs]	глаза [eyes]	буркалы [peepers]
S_t	3.1	3.6	3.8
Scale	Safe-Terrifying		
Words	страх [fear]	ужас [terror]	жуть [horror]
S_t	3.6	3.7	4.1
Words	вампир [vampire bat]	вурдалак [werewolf]	упырь [vampire]
S_t	2.7	3.3	3.5

[continued on following page]

Words	ветер [wind]	буря [gale]	ураган [hurricane]	шторм [storm]	тайфун [typhoon]
St	2.8	3.2	3.4	3.9	3.9

Thus the acoustic form of expressive synonyms is clearly nonrandom. It imparts different content, assisting these words in carrying out their function and amplifying and emphasizing their emotional expressiveness.

Phonetic value may be seen to amplify expressive words not only in synonym progressions. As a rule, words having expressive acoustic form assume figurative meaning precisely owing to their phonetic value.

Thus the word "гад" once meant "reptile" (it still retains this meaning today in zoology). But the phonetic structure of this word emphasizes the attribute "coarse" (on the "graceful-coarse" scale, $S_t = 3.8$). It is due to its phonetic value that its meaning changed, and the word became coarse and abusive ["vile creature"].

A shift in meaning favoring the sound content of a word can happen in many other cases as well. Here are examples of words that have acquired "negative" emotional coloration owing to their "negative" acoustic content: выдра [otter], жаба [toad], жук [beetle], фрукт [fruit], лопух [burdock], крыса [rat], дуб [oak], рыло [snout], морда [mug], хряк [hog], паук [spider]. Examples of "positive" shifts in response to "positive" acoustic content are: бог [God], лилия [lily], мед [honey], мимоза [mimosa], орел [eagle], сокол [falcon].

The acoustic content of a word may even neutralize an emotional coloration that had been imparted to a word by its definition, and create a new, opposite emotional coloration corresponding to the sound. As an example the word лютик [buttercup] (derived from the root лют) was initially defined as "злючка" [a malicious creature], and correspondingly it had a negative emotional coloration. But the acoustic content of this word is characterized by entirely different attributes: "good" (1.9), "graceful" (2.0), "light" (2.3), "beautiful" (1.8), "kind" (2.2). It was in correspondence with this phonetic value that the word acquired a new emotional value.

Cases in which the semantic meaning of a word is fuzzy and ambiguous are especially indicative. In such a case the definition of a word is derived almost entirely from its phonetic content. Such words have an especially clear emotional coloration. Thus it would be difficult to interpret the semantic meaning of the word грымза, but its clear emotional coloration derives from its phonetic content, which assumes the following values on different scales: "coarse" (3.8), "dark" (6.4), "terrifying" (3.4), "awkward" (3.4).

Let us also compare these examples: дылда [hulk]--"big" (1.9), "coarse" (4.0), "slow" (3.5); карга [hag]--"coarse" (3.6), "terrifying" (3.5), "awkward" (3.7); фефела [gawk]--"bad" (4.1), "big" (4.2), "dark" (3.7), "passive" (4.0), "dull" (4.3), "puny" (3.9), "slow" (4.0); хам [cad]--"bad" (3.8), "coarse" (3.5), "dark" (3.9), "repulsive" (3.7), "lowly" (3.8); хиляк [runt]--"bad" (4.1), "passive" (4.3), "weak" (4.4), "repulsive" (4.0), "puny" (4.3); хрыч [harridan]--"bad" (4.8), "coarse" (3.7), "dark" (3.9), "repulsive" (3.7), "awkward" (3.5).

This property of words--that of emphasizing, amplifying and even creating emotional expressiveness in speech--is used by talented poets, who organize the acoustic fabric of their poetic works in such a way as to create an emotional tone corresponding to the general expressive content of the particular work.

An acoustic emotional tone is mechanically created in a poem by filling the text with sounds of the needed content and reducing the frequency with which sounds having opposite content are encountered. Knowing the numerical values for the content of individual sounds and the normal (average statistical) frequency of their occurrence in speech, we can calculate the phonetic value of a poem.

Here are the results of calculating the acoustic value of the content of some well known poems.

A. Pushkin, "A Winter Morning"--"bright," "light," "joyful."

M. Lermontov, "Mountain Summits"--"melancholic."

M. Lermontov, "Tedious and Melancholy"--"dark," "sullen."

F. Tyutchev, "A Spring Thunderstorm"--"bright," "strong."

F. Tyutchev, "We Met"--"exalted."

A. Blok, "Oh Spring, Endless and Boundless"--"joyful," "bright."

A. Blok, "Fatigue"--"sullen."

S. Yesenin, "I Remember, My Beloved, I Remember"--"wonderful."

V. Mayakovskiy, "Here"--"dark," "sullen," "terrifying."

V. Mayakovskiy, "Our Republic Grows and Stands on Its Feet" (an excerpt from the poem "Good")--"cheerful."

When in keeping with the general content of a poem, the subtle organization of its acoustic form influences the reader's consciousness, and it is one of the means by which the author creates a special emotional tone comparable to that produced by a piece of music.

Sometimes poets use a sound-color combination in their work; then the acoustic pattern of the text creates a color accompaniment in the reader's subconscious, intensifying the emotional influence of the poem.

It would be very important to emphasize that the acoustic tone of a poem is created by poets unintentionally, involuntarily. This acoustic tone comes about owing to their poetic talent, their "feeling for sound." Therefore when we analyze the acoustic organization of a work, we reveal not the emotional state which the poet tries to consciously shape, but rather the actual emotional tone to which the poet was tuned at the moment he created the work.

These two planes coincide in the examples given above. But this is not always the case. As an example for many of N. A. Nekrasov's poems that we analyzed, no matter what their "obvious" emotional content was, the attributes we arrived at were only those of a minor emotional tone--"melancholic," "sad," "dark," "sullen," "terrifying." This is a definite manifestation of the perpetual minor emotional predisposition of this "melancholic poet."

The possibility is not excluded that the capability sounds have for emphasizing and even creating emotional coloration in statements is also exercised in commonplace speech. (This is especially noticeable in the use of jargon and of affectionate and abusive words, which are phonetically very expressive as a rule).

This creates a possibility for using analysis of the phonetic structure of a speaker's speech (together with other criteria) to automatically diagnose his emotional state. It may also be possible that the special acoustic organization of spoken messages (once again, together with other techniques), may be used for directed emotional influence upon the listener.

USING SPEECH CHARACTERISTICS TO EVALUATE INDIVIDUAL PERSONALITY FEATURES

I. S. Zamaletdinov, R. B. Bogdashevskiy

Speech, which is a system of phonetic signs performing a communicative function on the basis of lexical and syntactical principles and employing sound modulation to raise information content (through connotation, emphasis and emotion), has recently attracted the attention of many specialists. Our purpose is to briefly illuminate certain directions taken in differential psychophysiology and pathological psychophysiology in speech research, and demonstrate the scientific grounds for approaching this subject from these angles.

I. P. Pavlov's suggestion that there are singular, common laws of nervous activity associated with the first and second (speech) signal systems opens up the prospects for experimentally studying, from an objective point of view, the typological features contained within a widespread form of general activity--speech. In addition, Pavlov's recognition of the relatively unrestrained nature of speech and its adaptive, creative orientation was the starting point for L. S. Vygotskiy's conception concerning the mutual relationship existing between spoken signals and the mind, and for his "cultural-historic" theory of the development and formation of the personality. We should also note that the present tendency for Soviet psychology to study the personality in terms of the activity category (N. A. Leont'yev) opens up highly promising prospects for studying individual personality features on the basis of parameters of the individual's speech.

It should be recognized that the work on typological differences in speech has unfortunately been extremely insufficient. Certain researchers have noted that speech intonations are definitely correlated with certain typological properties inherent to the personality (5,7). Thus persons with an inhibited-inert style of mental reaction speak predominantly with a monotonous tone and with a low range of tonal variations; speech is characterized by insufficient tonal expressiveness, disturbed rhythm, reduced tempo and longer pauses. On the other hand the speech of excitable-mobile personalities is characterized by diverse and rich intonation, a high tempo, and concurrent instability of tempo. Significant prevalence of paralinguistic, "kinetic" characteristics ("communicative gestures") was found to be more typical of subjects of the "mobile" type than of "inert" subjects experiencing emotional tension (9). Without diminishing the significance of these primary parameters of speech that reflect the typological properties of the personality, we feel it suitable to limit our discussion to works analyzing parameters of speech function having greater complexity--lexical, grammatical, syntactic, stylistic, semantic and motivational.

In our opinion these characteristics express the personality component of speech generation and oral behavior to the greatest degree. This conclusion is consistent with the need for taking an integrated systems approach to studying and evaluating speech, as is the case with research on individual typological differences, as had been emphasized many times by B. M. Teplov and V. D. Nebylitsyn, who cautioned against making premature conclusions from an analysis of the function of some single analyzer or a single neuromuscular action (8,11).

According to present ideas the strength of nervous processes, as related to arousal, is characterized by the maximum performance of nerve cells, lability is characterized by the rate of arisal and cessation of a nerve process, and mobility is characterized by the speed with which the signs of conditioned stimuli are transformed (B. M. Teplov, V. D. Nebylitsyn).

Recent research on the mutual correlation between typological properties of the individual and speed parameters of associative speech reactions revealed significant correlation between reaction speed and the lability and mobility characteristics. Moreover this correlation was significant only in relation to differentiated reactions associated with a combination of speech and mental processes and requiring the action of a feedback system (10). It was also revealed that "inert" subjects deal much more poorly with assignments testing speech and mental processes than do "labile" subjects (6). Although the speed parameters of activity were sufficiently high, "inert" subjects made more mistakes because they failed to think out the assignment fully and analyze changing conditions completely.

Research by L. B. Gakkel' persuasively shows that the "inertia of the stimulatory process," being a typical trait of a subject suffering an obsessive syndrome, is highly correlated in an associative experiment with difficulties in associations manifested as dramatic elongation of the latent period, echolalia, repetition and stereotypy of responses (4).

Hence it would be valid to expect, in the external expression of speech, a unique pattern of "marching in place," together with delays in speech and with expression of stereotypic grammatical structures.

Mention should be made here, however, of the role played by word associations in lexical construction of speech. It has long been known from a purely stochastic approach to speech generation that successive elements in speech are associated together by a statistical dependence defined predominantly by extralinguistic factors--that is, by the psychological facts of associative linkages (I. P. Pavlov, A. S. Vygotskiy, A. A. Leont'yev, P. A. Shevarev, Yu. A. Samarin). Much attention has been, and is still being, devoted to these realities.

The works of A. G. Ivanov-Smolenskiy, A. S. Vygotskiy, A. A. Leont'yev, A. R. Luriya, A. A. Brudnyy and others demonstrated that the consciousness contains the semantic fields of the associations typical of each word; the persistence of these fields is relative, and it depends on the frequency with which these associations are realized, on how widespread their use becomes and on how well they become entrenched in spoken communication.

It also appears possible to distinguish between the "artistic" and "thinking" type, using I. P. Pavlov's terminology, on the basis of speech content. Thus while associations made by the former type are abundantly represented by concrete associations reflecting superficial and descriptive aspects--that is, a content reflecting concrete situations, the associations of persons of the "thinking" type are predominantly generalizations and abstractions--that is, their associations are analytical-synthetic.

A. A. Brudnyy's hypothesis concerning the parasituational or systemic semantic state of a word as opposed to its situational state is interesting in this aspect. In the former case words have a certain semantic potential, and in principle they may be interpreted as having a direct association with internal speech, with subconscious speech activity. In the situational state, semantic potential is realized in the form of a set of "contextual meanings." While a situationally conscious use of words is associated as a rule with speech organized sequentially--"one word after another," systemic speech activity may entail synchronous use of several words--"one word simultaneously with another (others)" (1).

And in fact, it would be plausible to suggest that the speech of personalities with developed abstract thinking processes would be dominated by associations of the first type--semantic linkages with paradigmic content: relationships between synonyms, antonyms and generalizations of a certain class of statements. What we also frequently encounter in such speech is a unique sort of surface roughness due to absence of the fluent, coherent combination of words--"one word following another"--often encountered in the speech of personalities characterized by "artistic" or "descriptive" thinking. Mention of two styles of narration known in linguistics--practical and hypotactic--would be pertinent here. The former corresponds to the principle of written speech, where one meaning or event is followed sequentially by another--that is, elements in the content of narration are linked in time. This style is widespread in the speech of children and in the developmental years, while with age it is substituted by the more-developed hypotactic style, in which a large amount of semantic, subtextual, temporal, local and causal associations exist simultaneously. In its grammatical expression, such speech is therefore typified by complex branching, by various causal and subordinate relationships, by adverbs, and by various auxiliary grammatical resources.

Variations in the degree to which these styles of speech are represented and changes in these styles are often noted in the speech of adults experiencing various extralinguistic conditions making speech difficult (fatigue, emotional stress and so on).

The ideas suggested above have been confirmed by recent experiments performed by Kharash, who studied the influence of rigidity in speech on the psychological structure of meaning (12).

The author experimentally confirmed that the lower is the level of development of the paradigmic meaning of speech--that is, that side of meaning in which the individual's life experience and intelligence are structured and generalized, the higher is the rigidity (which we define as the reverse of mobility) of the individual's speech. Interpretation of the syntagmic meaning of the content of speech as its communicative aspect, and paradigmic meaning as the semantic or intellectual aspect is, in our opinion, in keeping with Vygotskiy's profound definition of the word as "a unity of communication and generalization, of communication and thinking" (3).

And so we have analyzed, to the extent possible in the space allowed, some pathopsychological and typological questions, dwelling in greater detail on "lability-inertia" and "mobility-rigidity" categories in application to speech. Two directions may be noted in speech-related research on the concept "nervous system strength," which is determined primarily by the strength of the stimulatory process--that is, by its functional endurance or performance, which is in turn characterized by the inhibition threshold. The first is represented by research on the stability of speech in the presence of strong interference (emotional stress, background noise, limitations on perception of speech generation by the speaker himself). The second is represented by research on stability associated with lengthy continuous speaking.

BIBLIOGRAPHY

1. Brudnyy, A. A., "The Problem of Semantic States," in "Soznaniye i deystvitel'nost'" [Consciousness and Reality], Frunze, 1964, pp 6,8.

2. Vitt, N. V., "Mutual Relationships Between Intellectual Processes and Functional State," in "Nauchno-metodicheskaya konferentsiya 'Lingvo-psikhologicheskiye problemy obosnovaniya metodiki prepodavaniya inostrannykh yazykov v vysshey shkole'" [Scientific-Methodological Conference "Psycholinguistic Problems of Validating Foreign Language Instruction Methods in the Higher School." Report Abstract], Moscow, 1971, p 81.

3. Vygotskiy, L. S., "Izbrannyye psikhologicheskiye issledovaniya" [Selected Psychological Studies], Moscow, 1956, p 52.

4. Gakkel', L. B., "Methods of Studying Directed Speech Reactions," FIZIOLOGICHESKIY ZHURNAL, No 5, 1951, p 547.

5. Zhigil', V. G., "Intonational Errors and Their Causes," in "Nauchno-metodicheskaya konferentsiya 'Lingvo-psikhologicheskiye problemy obosnovaniya metodiki prepodavaniya inostrannykh yazykov v vysshey shkole'," Moscow, 1971, p 134.

6. Kozlova, V. T., "Diagnosis of Lability in Speech and Thinking on the Basis of Tests and Self-Evaluation," VOPROSY PSIKHOLOGII, No 4, 1973.

7. Molonova, L. F., "Experimental Phonetics and the Psychology of Speech," in "Uchenyye zapiski MGPIIYa" [Scientific Notes of the First Moscow State Pedagogical Institute of Foreign Languages], Vol 20, Izd-vo MGU, 1960.

8. Nebylitsyn, V. D., "Osnovnyye svoystva nervnoy sistemy cheloveka" [Basic Properties of the Human Nervous System], Moscow, Izd-vo Prosveshcheniye, 1966.

9. Nosenko, E. L., and Velichko, L. F., "Some Characteristics of a Speaker's Kinetic Behavior as Indicators of Emotional Tension," in "'Rech' i emotsii'. Materialy simpoziuma" ["Speech and Emotions." Symposium Proceedings], Leningrad, 1975, p 23.

10. Suzdaleva, V. A., and Chuprikova, N. I., "Rate Parameters of Associative Reactions and the Typological Properties of the Human Nervous System," VOPROSY PSIKHOLOGII, No 3, 1974, p 137.

11. Teplov, B. M., "New Data From Research on Properties of the Human Nervous System," in "Tipologicheskiye osobennosti vysshey nervnoy deyatel'nosti cheloveka" [Typological Features of Human Higher Nervous Activity], Vol 3, Moscow, Izd-vo APN RSFSR, 1963.

12. Kharash, A. U., "Rigidity in Speech and the Psychological Structure of Meaning," in "Aktual'nyye problemy psikhologii rechi i psikhologii obucheniya yazyku" [Pressing Problems in Speech Psychology and in the Psychology of Language Instruction], Izd-vo MGU, 1976.

ACOUSTIC CHARACTERISTICS OF THE PHONETIC WORD IN DIFFERENT TYPES OF EMOTIONALLY ORGANIZED TEXTS

L. V. Zlatoustova, M. V. Khitina

Our objective was to study the acoustic characteristics of speech reactions to specific subjective experiences. The material we analyzed included wholly organized, emotional texts--syllabic verses (works by A. Blok and B. Pasternak), and prose written in contrasting style and characterized by reduced spatial and temporal organization (L. N. Tolstoy's story "The Shark").

The results of acoustic analysis of the verses were compared with the prose; from our point of view the simplicity of the text, the maximum descriptiveness of presentation, and the contrast between the author's narration and the emotionally colored statements of the characters provided sufficient material for comparison.

These texts were subjected to acoustic analysis with a Kay-Electric sonograph; wideband and narrow-band sonograms in the 8-4 kHz range, recorded from 10 readers, were examined.

Discussion of Experiment

In our assessment of emotional expression we based ourselves on an analysis of the integral organization of the phonetic word. The reason for this is that parameters such as the time taken to utter syllables making up a phonetic word, the frequency of the fundamental tone, and its intensity are characteristics not only of the stressed syllable of a word, but also of the unstressed syllables, though it is true that the stressed syllable is the most informative. At the same time we need to note the importance of accounting for the relationship between the frequency of the fundamental tone in the part of a phonetic word preceeding the stressed syllable and in the part following the stressed syllable.

I. Persons not having an education in dramatic arts and not being professional speakers simulating emotionally colored, communicationally meaningful statements exhibited the following type of organization of fundamental tone frequency (within the framework of a phonetic word) as being most characteristic:

1) The maximum value of the first pretonic vowel is always less than the minimum value of the tonic vowel;

2) the start of a tonic vowel is characterized by a sharp rise in frequency over a relatively short segment of time--from one/sixth to one/third of the total duration of the sound, with the shorter segment dominating (for example, for speaker B. the duration of this segment was 50 msec while the total duration of the sound was 148 msec; for speaker S. the figures were 33 and 180 msec respectively);

3) of interest is the steep leading edge of the frequency of the stressed segment and its gentler trailing edge, as is equally true of both the stressed vowel and the posttonic syllable. This decrease in frequency is characterized by a sufficiently stable low point (90-110 Hz) for all male speakers. Thus the frequency range from the peak at the stressed vowel to the end of the decline in frequency at the posttonic syllable varies for different speakers within 275-125 Hz, 250-90 Hz, 170-100 Hz and so on;

4) in the majority of cases the total realization time of an emotionally colored phonetic word is greater than that of a neutral phonetic word; here, the increase in time is predominantly due to greater length of the tonic syllable, and mainly the stressed vowel and the posttonic syllables. When the temporal values of emotional and unemotional statements are close, the main difference is in the distribution of the time of the pretonic and tonic vowels, this ratio being about 3:1 for emotional statements and 1.5:1 for unemotional statements.

II. Actor simulation reveals the same characteristics, though in emphatic form--that is:

1) The leading edge of the frequency of a tonic vowel takes up one/sixth of the realization time in a frequency range from 150 to 200 Hz;

2) an actor's pronunciation is distinguished by the dynamics of the tone of the first pretonic syllable, for which the minimum value is 125 Hz and the maximum is 240 Hz. Thus with actor simulation, a tonic vowel begins not at a point higher than the maximum of the pretonic syllable, but at a lower point, which creates additional contrast in the frequency of the fundamental tone within the phonetic word;

3) the posttonic syllable is uttered similarly as described in the general model above;

4) another feature of an actor's pronunciation that may be noted is the greater contrast between the segment characterized by a sharp rise in tone and the segment of decreasing tone. The tendency to hold the frequency at its maximum for a certain length of time is also more noticeable in an actor's pronunciation than in that of untrained speakers;

5) a dramatic change in the spectral pattern of the vowel is typical of emotional statements made by actors: A fundamental frequency of 400 Hz causes change in the formant structure of the vowel "y";

6) the tonic vowe' of an emotionally distinct word is characterized by high total energy, which remains highly stable throughout the entire time of the vowel's pronunciation.

A wholly organized poetic text (a syllabic-tonic verse) never produces the time contrasts described above, while the frequency of the fundamental tone remains more constant. The structure of a tonal line of verse is unique as a rule in that the frequency at the beginning of the line matches the frequency at its end. Such a tone structure is a consequence of the symmetry of the verse, which is a product of all of the parameters imparting organization to the verse, to include semantic and syntactic parameters.

Conclusions

Thus in terms of contrast, a phonetic word in prosaic text may be distinguished by the following: 1) The duration of the tonic vowel; 2) the specific contour of the frequency of the fundamental tone of the tonic vowel (a sharp rise in frequency during the first third of the sound to its maximum, and a smoother decline of the frequency of the fundamental tone over the remaining portion of the tonic vowel and the posttonic syllable); 3) the tonic vowel of a word distinguished by its contrast is characterized by total energy that is maximum for the text and stable for the entire sound; 4) sounded consonants, to include consonants contained in tonic syllables, do not contribute to the greater duration of emotional statements; 5) an integrally organized poetic text is characterized mainly by a fundamental tone frequency of smoother structure, without a sharply delimited tonic syllable, with a rather even increase in the time of pronunciation of tonic and pretonic vowels, and an increase in the time of posttonic vowels at the ends of a line of verse.

This experiment confirms observations (1) indicating that the traces of certain emotional states in the long-term memory are stable and that their verbal expressions are adequate.

BIBLIOGRAPHY

1. Gromova, Ye. A., "Emotional Memory and Biogenic Amines," in "Strukturno-funktsional'nyye osnovy mekhanizmov pamyati" [Structural-Functional Principles of Memory Mechanisms], Nauka, Moscow, 1976.

DYNAMICS OF EMOTIONAL TENSION IN A SITUATION INVOLVING AN ANTICIPATED OUTCOME OF VARIABLE PROBABILITY

S. L. Zysin

In 1965 Grashchenkov and Latash (1) indicated the importance of accounting for the meaningfulness of a signal, and not only its probability, in experiments in which the GSR is recorded. Since that time the GSR has earned itself an important role in the psychological experiment as an indicator of emotional state. Thus we can rightfully assert today that Grashchenkov and Latash demonstrated the dependence of the magnitude of an emotional reaction to a stimulus on the latter's meaningfulness. But what are the relationships between the meaningfulness of a stimulus and its probability? This question was examined by Feygenberg (5), who arrived at the conviction that an unexpected stimulus--that is, one with low probability--is meaningful. But this point of view cannot but elicit some objections. Assume that an object O having biological value is distinguished by a certain set of signs w_1, w_2, w_3, ..., w_i. Would we, then, not agree that the more faithfully each of these signs defines O--that is, the higher the probability that event O follows event w_i (P_{Ow_i}), the more meaningful this sign would be? This implies that event w_i would be encountered along the path to object O more frequently (the probability of such encounters would be higher) if the event's objective meaningfulness is greater. Thus we arrive at a conclusion opposite to Feygenberg's viewpoint: The meaningfulness of a stimulus is proportional to its probability.

Our objective was to experimentally test this hypothesis.

The method was as follows. The subject sat at a console with four lamps of different colors that could be turned on and off from another console by the experimenter. There was a button beneath each lamp that turned it off. The subject had to turn off the lamp indicated in the experimenter's preliminary instructions and he was not to turn off the others. The lamps were turned on in random order in series of four, and the place of the test lamp in the tetrad was random as well. The subject's emotional reaction was evaluated with the assistance of a GSR recorded on encephalograph tape. The magnitude of the emotional reaction was assessed on the basis of the amplitude of the GSR wave.

Experiments performed with this method produced the following results:

1) In the first series a GSR was recorded only in response to flashing of the test lamp.

2) Following a certain number of experiments (different for different subjects) an increasingly more expressive reaction to preliminary flashes of other lamps arose.

3) When the test lamp turned on in the middle of a series, subsequent flashes of other lamps did not produce a reaction.

4) The amplitude of the GSR grew distinctly from the first flash in the series to the fourth as the number of alternatives decreased.

5) In the intervals between flashes (10 sec) the GSR returned to its isoelectric line.

Conclusions and Discussion

The flashing of each lamp in our experiments may be interpreted as a sign of approach of an anticipated outcome. We can assess the meaningfulness of these signs to the brain on the basis of the magnitude of the emotional reaction. Thus the results in 2), 3), 4) above show that the intensity of an emotional reaction, and consequently the meaningfulness of the sign, is proportional to the *a priori* probability of the anticipated outcome. If the magnitude of the emotion is E, the flashing of a lamp is w_i and the anticipated event is s_j, then this proportion may be approximated by the expression $E = f(Ps_j/w_i)$.

This dependence would not be complete without consideration of the influence of motivation, which depends on need. Inasmuch as the magnitude of an emotion is proportional to the intensity of a need N (3,4)--that is, to the tendency for state s_o to transform into a certain state s_j, we could validly write $E = N_{s_j - s_o} \cdot Ps_j/w_i$

As follows from the results examined in (5), this function is not continuous, inasmuch as the GSR returns to zero in the intervals between lamp flashes. In other words there is no emotional tension when there is no movement toward the goal, and it arises only at the moment the probability of the anticipated event changes.

Results 1) and 2) are also a good confirmation of the hypotheses, suggested here and earlier (2), concerning the relationship between the meaningfulness and probability of an event. We find in fact that the meaningfulness (weight) of sign w_i cannot be established by the brain before the probability that w_i would appear prior to the anticipated outcome s_j--or, to put it in another way, the probability of the transition $w_i \rightarrow s_j$--is clarified. It takes a certain amount of time to determine meaningfulness, and therefore at the beginning of the experiment, before meaningfulness is established, there is no emotional reaction, but as soon as the brain clarifies the probability characteristics of the signs, it ascribes a certain meaningfulness to them, and an emotional reaction appears, the magnitude of which corresponds to this meaningfulness.

We can see from this discussion that a relationship arises between meaningfulness and probability only within the space of signs of a certain object (situation, task). And inasmuch as there can be no emotions apart from meaningfulness, it would hardly make sense to consider the influence of the probability of a stimulus upon emotions in general, without indicating the particular task in relation to which the probability of the sign is evaluated. For example the probability of seeing a freshly broken branch in the forest would be negligibly small in comparison with the probability of seeing an intact branch. Were we to agree with Feygenberg, such a discovery should have a most violent effect on our emotions during a forest stroll. But in fact we would barely take notice of it, since in this case a problem, in relation to which the broken branch might serve as a sign of something, and be correspondingly meaningful, is absent. However, in a situation where a hunter is following the tracks of a wounded animal, possibly even hiding in ambush, the same branch would become a highly probable sign of the closeness of danger, and therefore its meaningfulness and, consequently, the emotional reaction to it would be greater.

BIBLIOGRAPHY

1. Grashchenkov, N. I., and Latash, L. P., "The Role of the Orientation Reaction in Organization of Action," VOPROSY PSIKHOLOGII, No 1, 1965.

2. Zysin, S. L., "The Mechanisms of Thinking Pathology in the Presence of Schizophrenia," in "sb. psikhologicheskoy b-tsy No 3, g. L-da" [Collection of Leningrad Psychological Hospital No 3], Leningrad, "Meditsina," 1978.

3. Siminov, P. V., "Teoriya otrazheniya i psikhofiziologiya emotsiy" [The Theory of Reflection and the Psychophysiology of Emotions], Moscow, "Nauka", 1970.

4. Siminov, P. V., "Vysshaya nervnaya deyatel'nost' cheloveka (emotsional'no-motivatsionnyye aspekty)" [Human Higher Nervous Activity (Emotional-Motivational Aspects)], Moscow, "Nauka", 1975.

5. Feygenberg, I. M., "Mozg, psikhika, zdorov'ye" [Brain, Mind, Health], Moscow, "Nauka", 1972.

PERCEPTION OF SPOKEN INFORMATION ON A NOISE BACKGROUND BY LISTENERS IN A STATE OF SENSORY MONOTONY

M. N. Il'ina, I. M. Lushchikhina

We may conclude rather confidently from published data that perception of speech in a state of emotional tension is characterized by shifting of this process to the sensory level, without active, purposeful recognition of the meaning of the message. This process, which may be called "perceptual resonance," is abundant with echolalic reactions such as simple repetitions, frequent unfinished replies, and arisal of false alarms.

We were interested in perception of speech at the opposite pole or, more precisely, in a state of sensory monotony, which is typical today of automatic control and monitoring systems encountered in many forms of work performed by operators. We know that in the presence of sensory monotony, subjectively experienced by operators as apathy, boredom, loss of interest in work, sleepiness and so on, the number of times important signals are missed increases, the performance curve (the U-shaped monotony curve) periodically falls and rises, the capability for performing complex assignments worsens while the rate of performance of simple assignments remains high, and so on. These data were obtained mainly from research on psychomotor functions, meaning that they cannot be extended to the case of sensory monotony without initial testing.

The activity of listeners receiving spoken information on a noise background over a long period of time is, in our opinion, a classical example of sensory mental labor (perception and registration of a certain form of incoming signals). We had full grounds for assuming that monotonous reception conditions, aggravated by a negative environmental influence, would elicit sensory monotony within the first hours of work. We suggested the hypothesis that the mechanisms of sensory and motor monotony are the same. This had to be confirmed both by experimental data and by parallel psychophysiological analysis of the state of the listeners.

Five professionals participated in the research as listeners, working on a regular daily work schedule.

The subjects listened to articulation tables of words and phrases on the background of a chorus of speaking voices. The speech-noise ratio was kept at a constant level throughout the experiment, equal to 5 db. Under these acoustic conditions average word intelligibility was 40-50 percent--that is, the reception conditions were difficult.

The subject wrote down the words and phrases he heard in their entirety with a mechanical pencil. This made it possible to record reaction times with an electronic stopwatch. Thus besides work accuracy (the correctness of signal recognition), we accounted for work speed.

The recordings were played in a free acoustic field (through loudspeakers) or through headphones.

The following techniques were used to monitor the psychophysiological state of sensory monotony experienced by the listener:

1) measurement of pulse by a pulsotachometer during the entire time of work;

2) measurement of simple and complex choice reactions to sounds of different intensities after each hour of work;

3) measurement of the number of symbols reviewed in a correcting test before and after the principal work;

4) dynamometry and tapping tests following each hour of work.

Analyzing the results, we found that objective quantitative work indicators recorded in the course of lengthy reception of spoken information on a noise background taking the form of a chorus of speaking voices can in fact reveal the presence of the sensory monotony syndrome. Evidence that this is so includes, first of all, absence of deterioration, or presence of insignificant deterioration, of the results of articulatory tests using tables of both words and phrases toward the end of a lengthy experiment, and a general increase in the number of omissions and lapses during work with tables of both words and phrases.

Listener performance follows a sawtooth curve, which in the opinion of most researchers is typical of the sensory monotony syndrome.

That the listeners were in a state of sensory monotony was confirmed by their subjective experiences (a decrease in interest in the work, boredom, a tired feeling, sleepiness) and changes in a number of psychophysiological functions. Examined separately, the latter reveal a complex pattern of changes occurring in listeners during lengthy reception of spoken information on a noise background. Thus indicators characterizing the state of levels of regulation associated with complex functions worsen: The choice reaction time increases; the number of symbols reviewed in the correcting test decreases while the number of mistakes increases.

A clear improvement in performance of simple assignments reveals itself on the background of deterioration in complex functions. The time of simple sensomotor reactions decreases, and the number of symbols reviewed increases (in the correcting test, in the absence of differentiation).

The changes are found to be insignificant at a more-remote level of regulation--motor regulation. The physical strength and frequency of movements remain practically unchanged, which confirms the conclusion that the state induced is monotony, and not fatigue.

Thus the techniques we used, some of which were related directly to the second signaling level of regulation while others were related to the motor apparatus, confirmed the previously revealed fact that in the presence of monotony, higher centers undergo inhibition while motor centers undergo arousal.

Perception of speech by listeners experiencing a state of sensory monotony, as is true for cases of emotional tension, is characterized by a sharp decline in active, purposeful recognition of the meaning of messages coupled with periodic falls and rises of the performance level and an abundance of omissions and mechanical repetitions, without attempts at reproducing an entire message on the basis of a few of its elements.

ANALYSIS OF THE FUNDAMENTAL TONE OF IMPERSONATED SPEECH

N. P. Kazantseva, V. Kh. Manerov, B. L. Murav'yev

Impersonation of human speech, which is limited predominantly to stage performance, is beginning to capture the interest of specialists in speech research. We are now finding that the capability for imitating the voice and speech of another person may be a serious interference to automatic systems verifying the identity of a speaker on the basis of spoken signals. But the process itself of speech impersonation, and mainly the question as to what characteristics of speech are imitated and with what successfulness, has not been studied, to the extent that our data show.

Inasmuch as the successfulness of impersonation may be determined only through auditory analysis, another question arises: What attributes are used in auditory perception to describe the sounds of a voice (to include an impersonated voice)? Research conducted by the method of the semantic differential (Voiers, Galunov) showed that a listener perceiving a speaker's voice makes use of four or five significant independent perceptual dimensions: 1) an overall evaluation of the voice, to include an evaluation of its timbre; 2) the magnitude or volume of the voice; 3) the dynamic nature of the voice; 4) the pitch of the voice.

As we can see, the first two dimensions are qualitative. They are multidimensional unqualified variables that do not yield to objective measurement. The third dimension is associated with objective characteristics of the signal such as "loud" or "nonmonotonous," which obviously can be simulated in trivial fashion and were not examined in our work.

We were mainly interested in voice pitch, which has a clear objective correlate--the frequency of the fundamental tone. Published data indicate that it is responsible for a certain degree of resemblance of voices perceived by the ear. Thus R. Colton notes a correlation between the probability with which one speaker may be mistaken by a listener for another and the similarity of the fundamental tone frequencies.

Our research answered the question as to whether or not the fundamental tone frequency of the impersonator's speech is tuned to the fundamental tone frequency of the speech of the individual being impersonated: The impersonated speech of USSR people's artists V. Merkur'yev, E. Garin and A. Papanov was tape-recorded for this purpose, using a standard text. Lenkontsert artist Yu. Garin was the impersonator. Samples of the speech of V. Merkur'yev, A. Papanov, E. Garin and Yu. Garin were used for comparison of imitated and original speech. Using a specialized instrument, we

Speakers	Mean FTF (Hz)		FTF Irregularity (Hz)	
	Original	Imitated	Original	Imitated
V. Merkur'yev	165	166	258	462
E. Garin	152	155	380	431
A. Papanov	145	130	409	477
Yu. Garin	122	-	352	-

measured the mean fundamental tone frequency (FTF) and the irregularity of the fundamental tone frequency curve, which we defined as the sum of moduli of frequency changes measured within each cycle in the course of playback of a speech excerpt. Then we determined and compared the overall means of the fundamental tone frequency and irregularity for the entire text (10-15 sentences). The results are shown in the table above.

As we can see from the table, the results of the experiments do not lead to categorical conclusions. In two cases we observe exact tuning of the fundamental tone of the impersonator's speech to the fundamental tone of the individual being impersonated (V. Merkur'yev, E. Garin); in one case such tuning does not occur, though an increase in the fundamental tone frequency in the direction of the fundamental tone of the impersonated individual (A. Papanov) is noticeable. No tendency to tune the voice to the irregularity of the fundamental tone curve was revealed. It is suggested that the work be continued with the purpose of increasing the volume of spoken material and the number of speakers and impersonators.

POSSIBILITY OF STUDYING EMOTIONALLY COLORED SPEECH BY THE SEGMENTATION METHOD

N. G. Kamyshnaya

Various methods of objective and subjective analysis are presently enjoying rather extensive use in research on perception of emotionally colored speech. Much attention is being devoted in this connection to developing the methods (and techniques) of subjective evaluation of the state of a speaker on the basis of spoken signals.

The results of many studies show that as a rule, speech segments of up to a phrase in length are the most appropriate stimulus.

Our objective was to establish the possibility of identifying a speaker's emotional state on the basis a segment stimulus--a sequence of sounds, vowels and consonants, organized in a particular fashion. It should be emphasized that our work is one in a series of experiments having the end goal of revealing the optimum segment, one allowing for the greatest degree of identification of the emotional state of a speaker on the basis of the stimulus he presents.

We used an electronic segmentator--a special device that divides speech, with a high degree of precision, into discrete members--words, syllables, sounds and so on.

The experiment was performed with phrases satisfying the requirements of sound quality and categorized rather definitely by listeners as representing a certain class of emotional states. The phrase "Eto tak prosto" ["This is so simple"] spoken by four subjects experiencing different emotional states--melancholy, sorrow, anger, wrath and joy--was subjected to segmentation. The experimental set of phrases also included emotionally neutral phrases (that is, phrases representing normal speech), spoken by each of the subjects, making it possible to analyze identification of normal speech in general and deviations from it more clearly.

The segmentation technique using a segmentator was described in detail earlier (1). We used it basically as follows:

1. We found the temporal location and duration of "windows"--segments containing the vowel sounds A, O, O. In view of the instability of the way the word "ETO" was spoken, the vowel segments of this word were not factored out.

2. The resulting set of vocal segments was transcribed on magnetic tape in a specific sequence: "A-O-O" and "A-O"--that is, we arrived at stimuli containing three "windows" and two "windows" respectively.

3. We subjected the sequence of consonants remaining after removal of the vowels to further segmentation in order to achieve the clearest sound for each of the consonant segments individually and for the entire sequence as a whole. Thus we arrived at the consonant segment-stimuli "T:KPRST" for each of the states. The total number of vocal stimuli used in the experiment was 26 (13 segments containing three "windows" and the same number of segments containing two "windows"), and the total number of consonant stimuli was 13.

Listener analysis was performed by a group of 10 inexperienced listeners. They were somewhat familiar with the research material because they had formerly participated in an experiment involving identification of emotional states implied by spoken phrases. The stimuli were presented to the listeners several times in accordance with a certain program.

Series A--the listeners were presented segment-stimuli containing two "windows" segmented out of texts spoken in one of the emotional states and out of normal speech.

Series B--in this stage of the experiment the stimuli were presented to the listeners in binary fashion. Normal speech was compared with some other emotional state. The experimental material included pairs of stimuli which differed or did not differ from normal speech (that is, normal-anger, normal-melancholy, normal-normal and so on).

Treating the data obtained during listener analysis, in each case we found the percentage of positive and negative evaluations relative to the total number of decisions made by the listeners (the total was adopted as 100 percent).

The vocal segment data demonstrated the following. The emotionally neutral state (normal) and weakly expressed emotions such as melancholy and sorrow are recognized the most distinctly. In the overwhelming majority of cases the definitions offered by the listeners were consistent with the primary names attached to these states.

When normal stimuli were presented and the definitions of the state offered by listeners were not consistent with the primary names of that state, as a rule such definitions fell into the class of names characterizing the physical state of the subject (depression) and the class of emotional states specifically (sorrow). It would be interesting to note that the percentage of correct identification of normal state increases when the segment-stimulus contains two "windows." The percentage of correct identification of the normal state also rises when stimuli are presented in pairs. In the case of false identification of the emotional states of melancholy and sorrow, the definitions offered by the listeners were either names characterizing the physical state of the subject (depression) or, to a significantly lesser degree, names in the class of emotional states specifically. The state names encountered most frequently were regret and insult. In a number of cases weakly expressed emotions (melancholy, sorrow) were associated with the normal state. This permits the suggestion that the norm is a relatively variable concept. The definition of normal must be qualified in relation to each subject in each concrete case.

A less consistent and less regular pattern is observed with identification of maximally saturated emotional states--anger, wrath and joy. In most cases the definitions offered by listeners evaluating the states named above did not agree with the primary names of these states. Thus joy was identified as anger or fear; anger (wrath) was identified by the listeners as joy, fear and even sorrow. In other words in the case of wrong identifications, the offered definitions exhibited considerable scatter within the class of specifically emotional states. Typically, in contrast to the situation with the normal state, segment-stimuli containing three "windows" were found to be more informative in relation to identification of the emotional states of melancholy, sorrow, joy and so on: The percentage of correct identifications of the stimulus grows. The analyzed material did not reveal any sort of definite trend associated with paired stimulus presentation in relation to identification of both strongly expressed and weakly expressed emotional states.

Data acquired with consonant stimuli were subjected to similar treatment.

A comparison of the results of both types of segments would show that the percentage of correct identification of the speaker's emotional state depends to a certain degree on the nature of the presented stimulus. The percentage of correct evaluations of the normal state was higher with vocal stimuli. A definite increase in the percentage of correct stimulus identifications is observed when consonant stimuli are presented. This is most typical of strongly expressed emotions (anger, wrath, joy), and less so of weakly expressed emotions (melancholy, sorrow).

A comparison of the data also reveals certain trends in the qualitative aspect of listener evaluations. Identification of emotional states tends toward greater definiteness in the case of consonant stimuli. In most cases the spectrum of state names is somewhat narrower than that for identification of emotional states on the basis of vocal stimuli. What we have said concerning differentiation of the frequency of consonants and vowels in emotionally colored speech confirms the point f view that the time of pronunciation of consonants and vowels may be a sufficientl, informative indicator assisting in recognition of a number of emotional states (2).

A repeat experiment performed a year later confirmed the previous results. However, these results should be viewed as tentative, requiring testing with more facts.

BIBLIOGRAPHY

1. Potapova, R. K., and Kamyshnaya, N. G., "The Syllable and Its Perceptual-Temporal Correlates," VOPROSY YAZYKOZNANIYA, Vol 4, 1975.

2. Potapova, R. K., "Temporal Organization of the Syllable as the Principal Component of the Temporal Emotional Sign," in "Tezisy dokladov ARSO-9" [Abstracts of Reports at the ARSO-9], Minsk, 1976.

EVALUATION OF SPEECH BY LISTENERS EXPERIENCING DIFFERENT STATES

Yu. A. Katygin

Increasingly more attention has recently been devoted to evaluation of states on the basis of speech. The expert evaluation method used for this purpose has still not been sufficiently developed from a procedural standpoint. The question as to whether or not perception of speech is distorted by different states experienced by listeners is still open. This communication deals with the preliminary results of research on this question.

Prior to evaluating the tape-recorded speech of speakers, each listener evaluated his state according to the scale: aroused--calm--inhibited. Subsequent analysis of the material took account of such categorization of the listeners.

Speakers' voices were evaluated according to a modification of Osgood's method of the semantic differential (V. Kh. Manerov, 1975) using 17 preselected characteristics (out of 31): Sorrowful-joyful, toneless-clear, calm-aroused and so on. Minus point scores for each characteristic meant perception of the voice as inhibited, while plus scores meant perception of the voice as aroused.

The experiment was conducted with nonprofessional experts--43 persons from 18 to 24 years old. The voices of five speakers experiencing different functional states (naturally the speakers do not know which) were presented in strict sequence: a calm state, followed by change in the initial state (mental supersaturation or monotony elicited in specially conducted experiments).

The research was conducted in two series. Twenty-four listeners experiencing a state of relative physiological rest took part in the first series; these listeners did, however, note different degrees of arousal-inhibition within themselves. Equal numbers of listeners for each of the tests and for each category of self-evaluated state were sampled for analysis of the results. (The test material consisted of the voice of a speaker experiencing a particular state, tape-recorded for listener analysis).

The data show that on the average the span of the evaluations given to the five speakers increases as we proceed form "inhibited" to "aroused" listeners (0.68, 0.76 and 0.80 points respectively. The span of the evaluations is defined as the number of points between evaluations of the voice in a relatively calm state and in a state of mental supersaturation or monotony). The nature of the distribution of evaluations relative to the zero point of the "aroused-inhibited" voice scale implies that in distinction from "calm" listeners, "aroused" and "inhibited" listeners are more emotional, being more sensitive to change in voice characteristics depending on the state of the speaker. But there also is a difference between "aroused" and "inhibited" listeners: The evaluations offered by "inhibited" listeners exhibit a broader span, and they tend more toward extreme positive or negative values. Of course 8 out of 15 evaluations (tests) of "aroused" listeners were higher ("more aroused") than the identical evaluations offered by "inhibited" listeners, the reverse was true for five evaluations ("more inhibited"), and two of the evaluations were equal. In other words "inhibited" listeners tend to evaluate a voice as more inhibited than do "aroused" listeners, and the latter evaluate a voice as being more aroused in comparison with the former.

The evaluations made by men and women from the same group of listeners (six each), who evaluated the tests in a calm state, were compared. The same female listeners also evaluated the tests in a relatively tense situation (prior to an examination). The evaluations were consistent for both situations. The results showed that women were able to differentiate the state of the speaker more clearly than men (the difference is statistically significant). This was particularly true for differentiation between monotony and fatigue. The span of evaluations given by women suffering tension prior to an examination was greater than that observed for women in a calm state. That is, evaluations obtained in a more stressful situation reflect changes in the state of the speaker more adequately.

Nineteen listeners took part in the second series of experiments. While in the first series only six female listeners evaluated tests while suffering pre-examination tension, in the second series first the entire group evaluated the tests in a more stressful situation--prior to an examination, and the same procedure was used on the same listeners experiencing a relatively calm situation. In the calm situation, all but two listeners assessed their state as "calm," while prior to the examination there were 12 "aroused," 4 "calm," and 3 "inhibited" listeners.

As in the first series, 8 out of 15 evaluations made by "aroused" listeners were higher than the evaluations offered by "inhibited" listeners, 5 were lower and 2 were equal. The average span of evaluations given of the five speakers was greater for "aroused" listeners (0.96 points) than for "inhibited" listeners (0.64 points) and "calm" listeners (0.80 points) when the listening was performed prior to an examination. In a calm situation (among "calm" listeners) the evaluation span was 0.88 points.

The evaluations offered by "aroused" listeners for the voice of speaker No 1 (supersaturation) have a larger span (1.9 points) than those of "calm" listeners (1.7 points), while the span for "inhibited" listeners is even less (0.8 points). However, in terms of speaker No 2 (supersaturation) and No 4 (diagnosis unclear), the evaluations offered by "inhibited" listeners (1.7 and 1.2 points respectively are observed to have a larger span than the evaluations of "aroused" listeners (1.2 and 0.2 points). The evaluations offered by "aroused" listeners in relation to speakers No 3 (monotony) and No 5 (supersaturation) are higher than those given by "calm" listeners, and the evaluations of the latter are higher than those given by "inhibited" listeners.

Because the number of "calm" and "inhibited" listeners was small, we singled out 10 listeners from the entire group who evaluated their state as "calm" in a calm situation and as "aroused" prior to the examination. Then we compared the evaluations made by this group in the two indicated situations.

In general, the evaluations offered by "aroused" listeners of all speakers but No 2 (supersaturation) exhibit a larger span (0.86) than that of evaluations offered by calm listeners (0.56). In one case the evaluations were even of opposite sign. There are also more statistically significant differences in the evaluations given by listeners experiencing a stressful situation (8 out of 10) than in the evaluations offered by listeners in the relatively calm situation (3 out of 10).

Comparison of the evaluations offered by men and women of this group (five each) also reveals greater expressiveness of the evaluations made by women than those made by men in both the calm situation and prior to the examination. For the women, there were more significant differences between evaluations, and the average group data were more similar. The span of evaluations made by women in a calm situation is 0.76 points, while that of evaluations made prior to the examination is 1.32 points; the figures for men are 0.42 and 0.52 points respectively. With both men and women, the span of evaluations is greater in the situation of pre-examination stress; there are more statistically significant differences between the evaluations offered by men and women prior to the examination (8 out of 15) than in the calm situation (5 out of 15).

The results demonstrate the contradictory nature of the dependence of an evaluation made of a speaker's voice on the state of listeners. It may be hypothesized that the differences indicated above in the dynamics of evaluations made by listeners experiencing different states arise due to individual features of the voices of the speakers. These same features of the voice of a previous speaker obviously also influences the evaluation given to the voice of a succeeding speaker as the listener goes on from one test to the next. If in general an increase (decrease) in the fundamental tone corresponds in most cases to an increase (decrease) in the evaluations given by listeners, these factors may also have an effect in those cases, few in number, in which an increase in the frequency of the fundamental tone is accompanied by a decrease in the evaluation, and vice versa.

Considering the unique conditions of the second series of experiments, we might imagine that the evaluation given to the speaker's voice was influenced not so much by the state of the listeners as by the attitude of the latter toward the experimental procedure. In comparison with the pre-examination stress situation, in the relatively calm situation the listeners performed on the background of an obviously negative attitude toward the experiment, which in all probability had the effect of decreasing the expressiveness (the span, the amplitude relative to the initial state in one direction or the other) and adequacy (change in sign, inconsistency with general group dynamics) of the evaluations.

Three basic conclusions can be made from the data. The first: The state being experienced by the listener does not hinder him from recognizing changes in the speaker's voice tending toward greater arousal or inhibition. The second: The initial state of the listeners influenced the evaluation given to the voice of the speakers. In most cases "aroused" listeners evaluated the speaker's voice as being more aroused (irrespective of whether it was toneless or clear) than did "inhibited" listeners. This means that the voice of a speaker in an aroused state was recognized to be more aroused by "aroused" listeners, while the voice of a speaker in an inhibited state was recognized to be more inhibited by listeners feeling themselves to be in an "inhibited" state. The expressiveness of this law may also depend on the attitude of the listeners to the procedure itself of voice evaluation. The third: Female listeners are more categorical than male listeners in their voice-based evaluations of state, and they are able to discern subtler changes in the speaker's state.

SOME FACTORS DEFINING THE ACCURACY OF A LISTENER'S EVALUATION OF EMOTIONAL STATES

T. V. Korneva

The emotional state of an individual is an important factor defining his speaking behavior. Research on the ways emotional states are expressed at different levels of speech has now become extremely important in connection with the task of monitoring the state of a human operator. Most research being conducted in this area has to do with word pronunciation, which is closely associated with physiological processes and which therefore is least open to voluntary regulation (control) in comparison with higher levels of speech organization (1,2). The objective correlates of emotional states in speech must be sought not only with the purpose of creating automatic emotional state recognition systems but also with the purpose of training specialists that are to monitor the state of operators. Automatic systems that could satisfy practical needs have not been created yet, and conclusions are made as to the state of an operator by a group of expert specialists as a rule. Thus the task is to raise the reliability of expert evaluations. However, much less research has been conducted on the characteristics of speaking than on the characteristics of speech perception, and most of the research has been concerned with resistance to interference (2,3).

In this connection we undertook research with the listener as the object. We employed the following procedure: A subject was asked to listen to speech recorded from 23 mentally ill speakers experiencing different emotional states. The form and degree of expression of the latter were evaluated by a commission of psychiatrists on the basis of a complex of clinical psychopathological signs. The speakers uttered a set of specially selected phrases varying in syntactic structure and phonemic composition. The phrases were devoid of any content that would indicate the speaker's emotional state. Listening to the speech of each speaker, the listener had to select one of six equiprobable states: a depressed mood, anxiety, apathy, anger, an elevated mood and normal (an even mood), indicating the degree to which the given state is expressed. The list of possible states was written on the data sheets on which the subjects recorded their conclusions. The correctness of the listener responses was evaluated on a six-point scale. Theoretically, the highest score a subject could run up for the entire test was 138 points, and the minimum was 23.

One hundred persons (50 women and 50 men) from 25 to 50 years old took part in the experiment as the listeners. On the average, the subjects successfully handled their task of identifying emotional states on the basis of an isolated spoken signal devoid of any sort of lexical or semantic content. The average score for the group was

82 points. Thus the prosodic characteristics of speech are a dependable indicator of that individual's emotional state.

It has been established that people differ significantly from one another in terms of how well developed their capacity for adequately perceiving and analyzing expressive information is (4). Sex and age have been found to influence the successfulness of identifying emotional states on the basis of voice. However, significant scatter in the results for groups of subjects that are homogeneous in relation to age and sex led to the hypothesis that the strongest factor that basically predetermines the degree to which this capability is developed is the characteristics of the individual, and particularly the subject's personality features. Cattell's 16-factor personality inventory revealed correlations between certain features of the personality and the capability for listening, and it established that personality structure which would be associated with maximum development of the capability for adequately perceiving expressive information. The results of this research led to the conclusion that such a personality structure would include the following characteristics: dependence, sensitivity, conformity, anxiety and tension. Such a structure, which implies that the subject is greatly dependent upon his environment, seems to generally promote greater sensitivity in the perception of emotional information. And on the other hand a preponderance of such traits in the personality structure as dominance, authoritarianism, independence and confidence would indicate that such an individual depends much less upon his surroundings, and consequently that his personality does not promote development of an orientation toward other people. In this connection the capability for understanding emotional states is trained to a lesser degree, as is reflected in the lower scores of listening capability in our experiment.

Thus the capability for correctly evaluating an individual's emotional state and the dependability of expert evaluations are a function of the personality features of the expert himself. Inasmuch as personality structure is stable to a significant degree, the dependence upon its reliability in each case of evaluation is constant as well. This dependence may be significantly reduced through special training. Thus we demonstrated that psychiatrists do much better as listeners than do specialists with a nonmedical profile or physicians of other specialties (4). Identification of emotional states on the basis of verbal expressions is a necessary component of a psychiatrist's professional activity, meaning that he constantly trains the appropriate habits of identifying emotional states, and that the standards against which he measures expressive signs are more fully developed.

Consequently if we possess information on the individuality of an expert, particularly on his personality features and special training, we could make a very dependable assessment of the reliability of his evaluations. On the other hand we know that the activity of an individual, and particularly the impressive activity of an expert, depends on his own state. Inasmuch as the characteristics of an expert's mental state change from one act of evaluation to the next--that is, inasmuch as they are variables, the dependence of the reliability of expert evaluations upon these characteristics is more complex. Basically two techniques are used to study this dependence: The expert's state is altered by drugs (tizertsin) (5) or a state of emotional tension is elicited within him experimentally (2).

We studied the impressive activity of subjects experiencing different emotional states elicited by mental illness. With this purpose we examined 160 patients suffering manic-depressive psychosis (80 women and 80 men). In contrast to other mental illnesses, manic-depressive psychosis does not cause significant changes in the patient's personality, and it manifests itself as cyclic, more or less prolonged changes in mental state. Eighty of the examined patients were in a depressed state. We did not examine patients with extreme or serious forms of depression.

Analysis of the results showed that subjects in a depressed state basically retain the capability for adequately evaluating the emotional state of a speaker on the basis of his voice. The average score for the depressed patient group was 79.2; it was somewhat lower than the normative indicator (82.0), but the differences are statistically insignificant. Women managed better with the listening assignment: Their average score was 82.0, while that of men was 76.4 (the difference was statistically significant, $p<0.01$). Analysis of the structure of test fulfillment showed that reduced adequacy of evaluations is not uniformly distributed throughout the entire test. Subjects made the greatest number of mistakes identifying a depressed mood (37 percent correct responses out of the total number of responses pertaining to this state). The accuracy of identifying anxiety and apathy did not decrease below normal, and on the average it even grew somewhat. Subjects were most successful in identifying anger and an elevated mood, though in comparison with the normal state their accuracy was somewhat lower. Significant differences appeared in the structure of test fulfillment by men and women. The matrix for the distribution of the probabilities of errors made by women in a depressed state does not basically differ from the standard matrix. Thus for example, an even mood was identified by women correctly in 67 percent of the cases (compared to a standard of 63.8 percent), while men were correct in only 46.7 percent of the cases. An even mood was mistaken for a depressed mood by women in 12.5 percent of the cases and by men in 21.0 percent. The latter may be explained by projection--a well known form of manifestation of an expert's state in his evaluations of the state of other people. The reasons for the decrease in accuracy of identification of a depressed mood by subjects in a state of depression remain unclear. It may be hypothesized that reduced sensitivity to the emotional state of a subject experiencing the same state is a psychological defense mechanism. Evidence of this can be found in the fact that the accuracy with which a depressed mood is identified through verbal expression grows and reaches the standard level as the subject emerges from his state of depression. We established this fact by examining 80 manic-depressive psychotics in a state of intermission--that is, in the absence of disease symptoms. The structure of test fulfillment for this group of listeners corresponded to the structure of healthy subjects.

Thus we analyzed the dependence of the accuracy of listener evaluations on a number of factors. Having studied our group of expert listeners, we can conclude that the reliability of evaluations made by different experts varies, inasmuch as it is a function of certain characteristics of the expert's individuality such as his personality features, sex, age, presence of special training and so on. Because these characteristics are fixed in any one expert, differences in the reliability of his evaluations would depend primarily on variable factors, one of which is his emotional state. It was demonstrated that the accuracy of identifying emotional state on the basis of prosodic characteristics of an individual's speech decreases on the average when the subject is in a depressed mood. However, the influence of the expert's emotional state on the accuracy of his evaluations can vary.

BIBLIOGRAPHY

1. Luk'yanov, L. N., and Frolov, M. V., "Signaly sostoyaniya cheloveka-operatora" [Signals Indicating the State of a Human Operator], Moscow, "Nauka", 1966.

2. Nosenko, E. L., "Osobennosti rechi v sostoyanii emotsional'noy napryazhennosti" [Characteristics of Speech in a State of Emotional Tension], Dnepropetrovsk, 1975.

3. Pollack, I. et al., "Communication of Verbal Modes of Expression," in "Language and Speech," 1960, pp 121-130.

4. Bazhin, Ye. F., Vuks, A. Ya., and Korneva, T. V., "Possibilities for Recognizing Emotions on the Basis of an Isolated Spoken Signal," in "Psikhologicheskiye problemy psikhogigiyeny, psikhoprofilaktiki i meditsinskoy deontologii" [Psychological Problems of Mental Hygiene, Prevention of Mental Illness and Medical Deontology], Leningrad, 1976.

5. Kon'kova, O. V., Bazhin, Ye. F., and Uskov, F. N., "Some Features of Expert Evaluation of the Emotional State of a Communicant in a Communication Situation," in "Problemy meditsinskoy psikhologii" [Problems of Medical Psychology], Leningrad, 1976.

USING THE SEMANTICALLY CONTRASTING PAIRS METHOD TO EVALUATE PROFESSIONAL QUALITIES OF AN ACTOR'S VOICE

A. N. Kunitsyn, V. I. Tarasov

The question as to whether or not it is possible to study an artist's creativity by the methods of the natural sciences is still being widely debated in the arts. The principal objections to such application boil down to the idea that an actor's creative process is clearly subjective in nature, and that it does not yield to any sort of formal description.

Whether or not that is true, the procedure by which an actor transforms himself into another character has become an object of serious study. Research by Professor P. V. Simonov persuasively demonstrates the fruitfulness of using the methods of the natural sciences to study the emotional aspect of an actor's creativity (1).

We believe that the solution to this problem depends in many ways on the choice of the research method. The method selected must equally satisfy the demands of art criticism and the demands of the natural sciences, especially when study of the emotional substrate of an actor's creativity is involved. From this standpoint, methods based on studying subjective evaluations appear promising. A. A. Ukhtomskiy wrote: "...so-called subjective evaluations are just as objective as any others..." ((2), p 135). We believe that subjective evaluation methods are in keeping with the very essence of the creative process of an actor.

Among the subjective evaluation methods used to study works of art and the creative process, the method of semantically contrasting pairs (SCP) has earned the greatest popularity. Subjective evaluation of stimuli on the basis of certain characteristics, which is the basic principle of the SCP method, is also the basis for an actor's regulation of his emotions on stage; the actor's ability to parcel out the dominant characteristics of an object or a phenomenon on stage, and to organically experience them--that is, experience them in fact--is a typical feature of the school of experience which is now validly referred to as Stanislavsky's school. The education of an actor in Stanislavsky's school begins with teaching him to parcel out dominant characteristics; M. O. Knebel', one of the oldest teachers and one of Stanislavsky's colleagues, writes: "At first we try to dispell the notion of emotionality. The student sits with a pencil, a wall, a window, a window view or a notebook before him. He must examine them without leaving out a single detail. Stanislavsky calls this first exercise 'abstraction of characteristics.' This represents more the position of a scientist than an artist. And in fact, the exercise itself and its name are perhaps more suited to science than to art. Strict analysis of an object, an appeal to note only that which is, imagining nothing extra, but missing not even the most negligible detail..." ((3), pp 69-70). Abstraction of characteristics is becoming a general principle of initial training exercises. Students are taught to evaluate the most distant and the closest objects and the highest and lowest sounds, and they are taught to evaluate and experience all shades and intensities of sound. In other words characteristics at the basis of an actor's emotional regulation are also the object studied by the SCP method.

We studied subjective perception of the emotional speech of an actor. Acting apart from the experience of emotional states would be unimaginable, and therefore theatrical speech, being the principal resource of expression, is primarily emotional speech. This is why study of theatrical speech actually means study of the professional features of an actor's transformation. The following question naturally arises: How does a professional actor's speech differ from that of a first-year student of dramatic arts?

To clarify this question, we conducted experiments on subjective perception of emotional speech. Our subjects were students in first-year acting classes who had practically no acting skills and no facility with theatrical speech, and prominent masters of the Leningrad stage--USSR people's artists Ye. A. Lebedev, V. I. Strzhel'chik, and Ye. Z. Kopelyan.

The set of speech stimuli consisted of the phrases "This is so simple" and "This is so simple that I have to say it." These phrases satisfy two requirements: They do not have a definite emotional content, and they are easily uttered in any emotional state. These phrases were uttered in the following emotional states: melancholy, joy, anger, anxiety, fear, sorrow and normal.

Emotional states were simulated by means of actor transformation. The phrases were tape-recorded and played back for listener analysis. Thirty listeners evaluated the speech stimuli on the basis of 53 attributes using a seven-point scale (for polar attributes) and a four-point scale (for monopolar attributes). The listening results were subjected to factor analysis.

The emotional speech of students breaks down into four factors. The first factor was interpreted as a subjective correlate of acoustic expression of actively experienced joy. It included the following attributes:

Elevated	0.97	Bright	0.94
Avid	0.97	Gay	0.94
Lively	0.95	Sharp	0.93
Alert	0.92	Active	0.88
Aroused	0.86	Clear	0.86

The attributes associated with the second factor permit its interpretation as a subjective correlate of a general esthetic evaluation:

Good	0.97	Beautiful	0.82
Tender	0.89	Weakened	0.81
Pleasant	0.88	Pure	0.70
Mild	0.88		
(Not) malicious	0.83		

The third factor was interpreted as an "anger" factor:

Malicious	0.80	Harsh	0.58
Angry	0.83	Distorted	0.58
Resonant	0.63	Full	0.51
Strange	0.61		

The set of attributes associated with the fourth factor permit its interpretation as a "fear" factor:

Frightened	0.83	Unusual	0.61
Breaking	0.81	Subdued	0.60
Weepy	0.81	Encumbered	0.53
Plaintive	0.74	Unconfident	0.53

The emotional speech of the prominent actors breaks down into five factors. The content of the first four factors remains the same as with perception of the emotional speech of students; a new factor, the fifth, was interpreted as the "volume" or "saturation" factor:

Rich	0.94	Thick	0.80
Full	0.91	Weepy	0.76
Rattling	0.87	Shrill	0.76
Deep	0.80	Plaintive	0.64

This factor indicates that the voice of professional actors is more saturated by overtones, and that it is more flexible and expressive. This is a qualitative evaluation of voice, associated with experience acquired in many years of acting, in which individual features of the voice were used as a means of artistic expression.

BIBLIOGRAPHY

1. Simonov, P. V., "Metod fizicheskikh deystviy K. S. Stanislavskogo i fiziologiya emotsiy" [K. S. Stanislavsky's Method of Physical Actions and the Physiology of Emotions], Moscow, 1962.

2. Ukhtomskiy, A. A., "Physiology of the Motor Apparatus," in "Sobr. soch." [Collected Works], Vol 3, Leningrad, 1952.

3. Knebel', M. O., "Poeziya pedagogiki" [The Poetry of Pedagogics], Moscow, 1976.

DETERMINATION OF EMOTIONAL STATES ON THE BASIS OF SEMANTIC AND TEMPORAL CHARACTERISTICS OF SPEECH

M. V. Lasko, Zh. I. Rezvitskaya

This paper describes the possibility of using content analysis of speech to diagnose an individual's states arising in the course of highly meaningful activity, and it analyzes the dependence of the temporal characteristics of speech on the speaker's state.

A previously published report presented the results of research aimed at creating a method making it possible to use content analysis of speech, obtained in the course of interpretation of ambiguous stimulatory material in the laboratory, to reveal and numerically express states such as anxiety, depression and aggression, and psychoneural tone (1). Our attempt at measuring states by means of content analysis scales was successful in the first approximation. The results provided the grounds for suggesting that if a certain state is manifested in the semantics of the verbal behavior of an individual acting in a naturally formed emotional situation (apart from an experimental situation and without provocation of speech by specially chosen stimulatory materials), this state could be diagnosed by means of the content analysis scales we developed.

Comparison of the content evaluation of state with some structural and dynamic characteristics of speech revealed a stable negative correlation implying that the greater the expressiveness of anxiety, the greater is the rate of speech. According to published data the rate of speech may both increase and decrease when emotional tension rises (3). That this change can occur in either direction is usually explained by individual typological differences in reactions to emotion-producing stimuli (stimulatory and inhibitory reactions) (2). In our opinion this phenomenon is associated with more-general mechanisms. Inasmuch as the rate of speech is one of the formal characteristics of productivity, its dynamics should follow the Yerkes-Dodson law. In other words it may be hypothesized that the rate of speech is maximum at a certain average (optimum) level of emotional tension. Below or above this level, the rate of speech would fall. When the complexity (or difficulty) of speech and mental activity is increased, the level of emotional tension that is optimum in relation to the rate of speech shifts in the direction of lower values.

To test this hypothesis we analyzed the speech of students recorded a) during an oral examination (58 persons) and b) in the laboratory during unrestricted interpretation of ambiguous stimulatory material (30 persons). Speech that was first tape-recorded and then transcribed was subjected to content analysis. The average number of words uttered in 1 minute was the indicator for the rate of speech. Subjects participating in the research were also asked to evaluate their own state using scales specially developed in our laboratory.

The research revealed the following laws. Monologue speech--oral presentation of prepared responses--was found to be the least informative in the oral examination situation for state diagnosis by means of content analysis scales. The principal information on a state was obtained during dialogue speech, when the statements of the students were responses to the examiner asking for clarifications or additional information. It is namely in these speech excerpts that the verbal elements of anxiety revealed themselves. The verbal components of other content scales were practically absent from the speech of this group of students. Apparently the increase in the uncertainty as to the outcome of the situation, occurring in response to requests for clarifications and additional information, and the limited time available to prepare a response increased the student's emotional tension and caused him to reduce control over his speech generation, which is what made anxiety the dominant component of state in the content of speech. It may be possible that the peculiarity of the situation and the activity, and the differences in social roles left no room for manifestation of state components such as low psychoneural tone, depression and aggression.

Content analysis of speech generated during interpretation of ambiguous stimulatory material revealed verbal elements associated with each of the scales indicated above. The dominant state component for this group of students was once again anxiety, the average score for which was significantly higher than for students taking the oral examination (9.3 as compared to 3.7). However the reverse was found to be true from self-assessments of state: Emotional tension was lower among students participating in the laboratory studies (24.5 as compared to 27.9). At the same time positive correlations were revealed in each group between states defined by content analysis and self-assessments of state (on the average, $P<0.05$). Thus despite the contradictions between the absolute values for content-analysis state and self-assessed state in the laboratory and during the oral examination, the values obtained within the same situation were statistically consistent with one another.

The following was revealed by analysis of the average rate of speech and the nature of its statistical relationship to the particular features of emotional state. The rate of speech was higher in the laboratory than during the examination (252 syllables per minute, as opposed to 196). It might be concluded from just this fact alone that the rate of speech decreases as emotional tension rises. However, this conclusion is valid only in relation to the group average. The correlation between rate of speech and state may be either positive or negative within the group. Thus correlation analysis showed that higher anxiety (determined by content analysis) and emotional tension (determined by self-assessment) in the laboratory correspond to a higher rate of speech, while in the examination situation the reverse is true--higher values for these factors correspond to a lower rate of speech ($0.05>P>0.01$).

These research results provide the grounds for suggesting that the basic content of speech and the real conditions within which it is generated impose specific limitations on use of its semantic and temporal characteristics as indicators of emotional state.

When content analysis of speech is used to diagnose states in real conditions, such a limitation must be imposed due to reduction of verbal elements reflecting characteristics of state. However, the stable relationship we revealed between self-assessments of state and states determined by content analysis indicates that such reduction decreases the absolute numerical values of the assessments, and that it does not affect their diagnostic value. Inasmuch as the size of the reduction of verbal elements depends on the basic content and form of speech, absolute assessments of states determined by means of content analysis of speech differing in content would not be comparable. Therefore before we could use content scales to diagnose state, we would have to obtain normative data indicating the number of attributes of state to serve as content elements in each concrete form of speech.

In regard to the temporal characteristics of speech, such a limitation must be imposed due to the sign differences in the correlations between assessments of state and rate of speech. The fact that such correlations may vary in sign may be interpreted as evidence of a curvilinear dependence between rate of speech and degree of emotional and (or) intellectual tension. Consequently the absolute values of rate of speech are not reliable indicators of state. The obvious solution would be to use relative values of the rate of speech for this purpose, and namely the degree of its acceleration or deceleration. Under otherwise equal conditions (in terms of the content of speech and the situation), an increase in the rate of speech may indicate an increase in emotional tension or anxiety that has not yet exceeded the limits of optimum intensity. But if the rate of speech declines, we may presume that emotional tension or anxiety has exceeded the limits of optimum intensity and is having an unfavorable influence on speech productivity.

BIBLIOGRAPHY

1. Lasko, V., and Rezvitskaya, Zh., "Determination of Emotional State by Content Analysis of Speech," in "Rech' i emotsii. Materialy simpoziuma 11-14 noyabrya 1974 g." [Speech and Emotions. Symposium Proceedings, 11-14 November 1974], Leningrad, 1975.

2. Nosenko, E. L., "Osobennosti rechi v sostoyanii emotsional'noy napryazhennosti" [Characteristics of Speech in a State of Emotional Tension], Dnepropetrovsk, 1975.

3. Cook, "Interpersonal Perception," 1969.

RECOGNITION OF OPERATOR STATE BY MASKING CHARACTERISTICS OF THE SPOKEN SIGNAL

V. G. Lebedev

Our objective was to clarify the possibility for automatic recognition of the functional state of a human operator on the basis of certain characteristics of the spoken signal.

We studied a pair of states, "emotional tension" vs. "normal." Recordings of the voices of athletes experiencing different emotional states were used as the experimental material. Spoken signals were recorded as follows. The subject was shown a series of numbers written in the following order: 10, 12, 14, 16, 18, 20, 19, 17, 15, 13, 11--that is, even numbers from 10 to 20, then odd numbers from 10 to 20, and then odd numbers from 19 to 11. After reading this series of numbers the subject had to recite them from memory into a tape recorder's microphone.

Recordings were made in different situations:

1) in a state of rest;

2) 30 minutes before the start of the competitive event;

3) immediately prior to the athlete's performance.

In the first case all subjects handled their assignments rather easily. The recordings made in this situation corresponded to the "normal" state. In the second and especially in the last situation noticeable growth in emotional tension was observed among some athletes, expressed in the recordings as omissions of certain numbers and confusion of their order.

Recordings of those subjects who expressed the greatest emotional tension in the last two situations were selected for analysis. These recordings corresponded to the state of "emotional tension."

The recordings were analyzed with a "BESM-6" computer. The word to be analyzed was fed into the computer through an analog-digital converter with a quantization frequency of 20 kHz. It was then broken down into segments 16 msec long, and spectrum moduli were calculated for each segment by means of a BPF [expansion unknown] procedure. Then, using an algorithm simulating a masking effect, we isolated the masking characteristics [maskpriznaki] of this segment of speech (1). Each masking characteristic was represented by a pair of numbers: $F_i^{(j)}$, $E_i^{(j)}$, where $F_i^{(j)}$ is the frequency and $E_i^{(j)}$ is the amplitude of the i-th masking characteristic of the j-th segment. Thus the word under analysis was represented as a sequence of masking characteristics for 16-msec segments of a spoken signal. The similarity of two segments k and m was determined by the formula:

$$a_{km} = \frac{\alpha^2}{\alpha^2 + \rho^2_{km}} \qquad (1)$$

where ρ_{km} is the distance between segments k and m, being the sum:

$$\rho_{km} = \rho_{km}^{(1)} + \rho_{km}^{(2)} \qquad (2)$$

$$\rho_{km}^{(1)} = \sum_j \min_i [\ln F_i^{(k)} - \ln F_j^{(m)}]^2 \qquad (3)$$

$$\rho_{km}^{(2)} = \sum_j \min_i [\ln E_i^{(k)} - \ln E_j^{(m)}]^2 \qquad (4)$$

where $F_i^{(k)}$ and $E_i^{(k)}$ are, correspondingly the frequency and amplitude of the i-th masking characteristic of the k-th segment, i is the number of the masking characteristic of the k-th segment closest in frequency to the j-th masking characteristic of the m-th segment, and $\alpha = \text{const}$.

The transition from similarity measure a for segments to similarity measure A for words is achieved by a method described in (2). The maximum measure of similarity between two words (A_{max}) was determined by means of dynamic programming:

$$A = \frac{A_{max}}{L_{max}}$$
$$L_{max} = \max (L_1, L_2) \qquad (5)$$

where L_1 is the length of the first word and L_2 is the length of the second word.

The recordings of three speakers were selected for an experiment in automatic recognition of state. The learning sequence for the recognition algorithm contained one utterance each of the word "twelve" in the "normal" state and in the state of "emotional tension" by each of the three speakers--that is, there was a pair of standards, corresponding to the states to be recognized, for each speaker.

The control consisted of three utterances of the word "twelve" in the "normal" state and three utterances by each speaker in the state of "emotional tension." The measure of similarity (5) with each of the two standards corresponding to a given speaker was calculated for the control sample. Recognition of state was said to be correct if the control utterance had the greatest similarity to the corresponding standard of the learning sequence.

In our experiment the states of two out of three speakers were recognized correctly, and two mistakes were made in recognition of the third speaker's state. That is, 16 out of 18 utterances were recognized correctly.

For stricter testing of the effectiveness of the suggested algorithm for recognizing the state of a human operator, we would have to have recordings corresponding to different emotional states and confirmed as to authenticity by additional electrophysiological monitoring data.

BIBLIOGRAPHY

1. Zagoruyko, N. G., and Lebedev, V. G., "Effect of Masking and Automatic Analysis of Spoken Signals," in "Vychislitel'nyye sistemy" [Computer Systems], Issue 61, Novosibirsk, 1975, pp 103-111.

2. Velichko, V. M., and Zagoruyko, N. G., "Automatic Recognition of a Limited Set of Oral Commands," in "Vychislitel'nyye sistemy," Issue 36, Novosibirsk, 1969, pp 101-110.

A DEVICE MEASURING EMOTIONAL AROUSAL-INHIBITION

V. Kh. Manerov, R. Yu. Ramanavichus, G. A. Yakushenkov

Research by many authors has shown that the characteristics of the fundamental tone of speech carry a significant fraction of information on the speaker's emotional state. Evidence that this is true can be found, in particular, in experiments with auditory perception which have demonstrated that listeners are capable of identifying the state of a speaker on the basis of both the signal recorded with a "neck" microphone and a spoken signal passed through a low frequency filter.

Electroacoustic analysis has also demonstrated the diagnostic value of the fundamental tone of speech, and especially its frequency (1,2). Various authors recommend using, as informative characteristics, certain parameters of the curve of fundamental tone frequency plotted with respect to time--the average over a segment of particular length, the range (span) of changes in frequency, the autocorrelation function, the variance and frequency irregularity (1,2). The last indicator is defined as the sum of the moduli of the gain in fundamental tone frequency measured in each period,

$$\sum_i (f_{i+1} - f_i);$$

our research established that it is a highly informative indicator for the diagnosis of three states: inhibition, normal and emotional arousal. This indicator has been routinely measured by means of a special program written for the Nairi-2 digital computer.

This paper describes a specialized device measuring the average frequency of the fundamental tone of a phrase within an arbitrary segment of a recording of limited length, the irregularity of the melodic curve of the fundamental tone and the duration of the test phrase.

The device consists of the following basic blocks: a microphone amplifier, (MU), a level indicator (IU), a total segment isolator (VTU), a fundamental tone frequency isolator (VChOT), a speech segment duration measuring unit (IP), a switchboard (KOM), a unit determining the average frequency of the fundamental tone (OS), a unit determining the irregularity of the fundamental tone's period (OI), a digital display (TsI) and a print-out block (VPU).

The spoken signal is fed from the microphone (or the tape recorder) to the MU and the VChOT, after which it is converted into a sequence of pulses with a spacing corresponding to the periods of the fundamental tone. The output of the VChOT is connected to the appropriate terminal by a signal from the VTU. The OS uses these pulses to determine the average period of the fundamental tone within the phrase, and the OI determines the fundamental tone's irregularity. The work of the OS and OI is controlled by signals emitted by the IP and the VTU.

The output parameters of the device are: the average period of the fundamental tone, its irregularity and the duration of the phrase; these parameters are fed through the KOM to the TsI, and through the VPU to a teletype unit.

Let us examine the principle of operation of the instrument's basic blocks. The tonal segment isolator is intended for isolation of tonal segments within a phrase. The decision as to whether or not tonal segments are present within a speech excerpt is made on the basis of the ratio of energy in the high frequency and low frequency ranges of the spectrum of the spoken signal. A sixth-order filter in the low frequency channel isolates the energy of the spoken signal within a frequency range up to 1,000 Hz; a sixth-order filter does the same in the high frequency channel for freqencies above 3,000 Hz. Following full-wave detection and smoothing by low frequency filters, signals from both channels are transmitted to a comparison circuit, which outputs the logical result 1 when energy in the low frequency channel exceeds energy in the high frequency channel. The output of the comparison circuit is connected to an AND logic circuit, the second input of which receives signals from the analysis interval determining circuit. Signals pass from the output of the AND circuit to the OS and OI to control their work. The fundamental tone is filtered out in the VChOT. Four band filters span a range of fundamental tone variation from 80 Hz to 400 Hz (80-120 Hz, 120-180 Hz, 180-270 Hz, 270-400 Hz). The "most advantageous" frequency channel--that is, the channel in which the first harmonic of the fundamental tone is located at the given moment--is automatically turned on according to a certain logic. This logic helps to prevent low frequency noise, mistakes caused by jumps to the second harmonic and penetration of the first formant of the spoken signal into the particular range.

Signals pass from the VChOT block to the OS for determination of the average period of the fundamental tone of the phrase. The sum of all of the periods of the fundamental tone within the phrase and the number of periods and the average period are calculated in the OS. The OS contains AND and OR logic circuits, a delay circuit, a decoder, trailing and leading edge isolators and the appropriate pulse oscillators.

Calculated values of the average period of the fundamental tone are transmitted from the output of the OS via the switchboard to the digital display.

The procedure for determining irregularity of the curve describing change in the fundamental tone is carried out by the OI block. It contains four counters, six registers, two decoders and several AND-NOT circuits. This block determines the modulus of the difference between adjacent values of fundamental tone periods and sums all of these modules in relation to a single phrase. The calculated irregularity value is transmitted via the switchboard to the digital display.

The circuitry of the digital display permits representation of the parameter measurements in visual digital form in the decimal system. Four IN-2 cold-emission indicator lamps are used in the display. The circuitry includes three decoders based on 155ID1 microcircuits, three code converters based on 1LB551 and 1LB553 microcircuits, and three 2AND-NOT circuits based on 1LB558 circuits for potential matching.

Three parameters are displayed in succession. The succession is determined by the switchboard. The first lamp indicates the number of the parameter while the next three indicate its numerical value.

The device is designed as a table model powered by a 220 volt circuit.

The device was tested with a spoken signal recorded in different conditions causing emotional arousal or inhibition of the speaker. One of these situations was rock climbing, the program of which included a preplanned fall by the climber. The spoken signal was recorded with a portable Sputnik-401 tape recorder. Even though the recording quality was low, the resulting curves for the dependence between irregularity of the melodic curve and the frequency of the fundamental tone (averaged in relation to standard samples) revealed growth in the climber's emotional arousal well and were consistent with the evaluation of the climber's state made by the trainer. The device was also tested with other models of emotional states (mental illness, actor simulation). The tests proved its performance in real conditions.

BIBLIOGRAPHY

1. Manerov, V. Kh., "Investigation of the Spoken Signal to Determine an Individual's Emotional State," Dissertation Abstract, Leningrad, 1975.

2. Galunov, V. I., and Manerov, V. Kh., "Approaches to Solving the Problem of the Speaker's Emotional State," VOPROSY KIBERNETIKI, Moscow, No 22, 1976.

INFORMATION CONTENT OF THE EMOTIONAL CHARACTERISTICS OF SPEECH

V. L. Marishchuk

As has been demonstrated by many authors, intense emotions cause change in the frequency characteristics of speech, its tempo, the spacing between sentences, words and sounds, and so on. However, much less attention has been devoted to the information content of the speech. And yet, changes in semantic structure, in the logic of presentation and in the meaning behind the way phrases are put together bear especially important information allowing us to determine the onset of tension--a state characterized by a decrease in stability of mental processes and a decrease in reliability. We observed the behavioral reactions of students and pilots in testing units and in real flight, and when performing parachute jumps, laying special emphasis on novices. In addition to running various tests, we evaluated the information content of the speech of our subjects.

The procedures we used included recitation of a well-memorized poem, memorization, summarization of the meaning of a short story, tests of logic, associative thinking, interviews etc. (1).

The results showed that depending on their strength, duration, the degree of the subject's adaptation to them, his functional state, his psychological sets and his social motives, psychogenic effects may elicit different responses and states. We classified these reactions in relation to three conditional phases: A, B and C.

In conditional phase A, which is typified by general positive changes, we usually observed a certain increase in the overall tempo of speech, coupled with a slight lengthening of thinking pauses, improvement of diction and growth in expressiveness of the most informative phrases. Subjects usually did better on tests than in the initial background situation: The quatrain was recited accurately, the meaning of the control story was summarized more accurately, and the test for logical thinking was completed faster. Tests of associative thinking were completed more successfully, and the associations became generally richer. In interviews with the experimenter, subjects sometimes used figures of speech and similes. Increases in other test indicators and growth in the effectiveness of activity (professional or simulated) were noted as a rule. The isolated cases of transition to intonations, phonations and articulations atypical of the subject and use of unusual and inappropriate expressions were not enough to suggest presence of tension or to permit prediction of a tendency of deteriorating performance.

The next conditional phase, B, was usually characterized by dissociation of the external manifestations of tension (including structural and acoustic indicators) from the practical results of activity (to include the information content of speech). Thus we noted involuntary transition to atypical speech tones and timbre (for example the subject's voice turned either shrill or deep, atypical hoarseness and swallowing of occasional words appeared, and so on), but the semantic content of speech remained practically undisturbed, and the task at hand was performed effectively. In other (rarer) cases, while diction, articulation, and the tempo of speech remained clear and normal (special care was even taken to keep the structure of speech correct), we observed a certain decline in the semantic merits of the uttered phrases and deterioration of the performance of some elements of activity.

In phase C, concurrently with a general decline in test indicators and a worsening of activity results, we observed pronounced externally expressed structural and acoustic changes (all changes listed in phase B, and additionally a dramatic decrease in the length of pauses between words and sentences--"drum speech"--or inadequate pauses going as far as temporary stupor). Deviations in the content of speech were noted in parallel. Thus in comparison with the initial data subjects recalling the semantic content of the short story left out up to one-third of the semantic units contained within it, and the results of the test for logic decreased (by 15-20 percent).

The results of the associative thinking and analogy assignments deteriorated an especially great deal (by up to 50 percent of the initial level). In certain cases, tense students and some novice parachute jumpers were unable to interpret the figurative meaning of proverbs. Parachute jumpers in a state of extreme tension made certain errors in the well-learned quatrain:

A lonely sail hovers white
In mists above a sea of blue!
What does it seek in a land so distant?
What did it leave behind at home?

The parachute jumpers often left out the word "mists" or they substituted the word "sea" by "sky," and instead of "does it seek" they said "did it leave," in a sense anticipating the events.

A brief interview revealed difficulties in logic such as: transitions to topics having nothing to do with the question asked; unjustified and categorical deductions; unjustified conclusions; digressions from initial premises; refusals to reply ("I give up"); clear lack of organization in replies ("jumps from point five to point ten"); difficulty in solving analogies.

Among purely external speech characteristics accompanying intense emotions, we also observe inadequate intonation--bursts of loud speech or pauses, or transition to a whisper, not in keeping with the logic of presentation, but spontaneously, parallel to fluctuations in emotional arousal.

Type B and C reactions were also observed among subjects undergoing training in mountain climbing. The subjects usually increased their rate of speech, and missed words and phrases, which in a number of cases distorted the meaning of the transmitted information. This could be explained by tension combined with euphoria. In severe cases we observed stupor, a total inability to express thoughts.

It is quite understandable that analysis of information content of speech accompanying intense emotions is not meant to substitute for the methods of spectral analysis of spoken information, isolation of intonational characteristics and other procedures for evaluating acoustic changes in the structure of spoken signals.

Moreover it would be suitable to focus attention on the fact that an objective evaluation of state, and all the more so a sufficiently probable prediction of the successfulness of activity, may be arrived at only through integrated study of different sorts of information about the individual, to include physiological indicators and information on performance of his principal activity. This is especially important to consider because in the presence of extreme emotional tension, we may sometimes observe opposing trends in different functional and other indicators. For example some may decline (ones which are secondary to the activity at hand or which play a secondary role in the maintenance of overall performance) while others that are more important and more meaningful may stay at their previous level or even increase. The psychological set toward performance of the given activity and a high level of motivation play an important role in this case. Thus, investigating performance of complex flight tasks, we found that pilots suffering intense fatigue and mental stress were unable to complete relatively simple psychological tests, but they were able to perform the skills of aircraft control and make navigational calculations without reproach (2). We revealed almost no changes in spoken information pertaining to professional activity (radio commands and reports, internal communication, exchanges of flight information) that could distort the content of this information. On the other hand all of the changes we categorized as being in group C were observed in responses to psychological tests and in conversations on subjects other than flying (both in questions and especially in replies). It should also be noted that the total volume of spoken information decreased sharply.

BIBLIOGRAPHY

1. Marishchuk, V. L., Afanas'yev, Yu. A., Kuksa, I. I., and Mikhaylov, G. V., "Voprosy organizatsii i metodiki professional'nogo otbora v VVUZ" [Occupational Selection in the Air Force School--Organization and Methods], Izd-vo VDKFFK pri GDOIFK im. P. F. Lesgafta, Leningrad, 1972.

2. Marishchuk, V. L., Bondarev, E. V., Yegorov, V. A., Plakhtiyenko, V. A., and Frantsen, B. S., "Stability of Mental and Psychomotor Functions in Extreme Conditions," in "'Aviatsionnaya i kosmicheskaya meditsina'. Materialy III Vsesoyuznoy konf. po aviatsionnoy i kosmicheskoy meditsine" ["Aviation and Space Medicine." Proceedings of the Third All-Union Conference on Aviation and Space Medicine], Moscow, 1969.

CHARACTERISTICS OF ACOUSTIC RESOURCES FOR EXPRESSING EMOTIONS IN VOCAL SPEECH, AND SOME GENERAL ASPECTS OF THE PROBLEM OF THE LANGUAGE OF EMOTIONS

V. P. Morozov

The grounds for the choice of the object of analysis: Most studies of the acoustic resources for expressing emotions have involved ordinary conversational speech (1,2). The acoustic resources of emotional expressiveness in vocal speech were studied for the first time in our bioacoustic laboratory at the USSR Academy of Sciences Institute of Evolutional Physiology and Biochemistry imeni Sechenov (3-7).

Vocal speech (singing) differs significantly from ordinary conversational speech in acoustic structure, physiological mechanisms of its formation and its auditory perception. The intensity of the singing voice can reach 120 db (when measured 1 meter away), the range of variation of the fundamental tone is more than two octaves, and the average statistical spectrum exhibits a sharp increase in intensity of the third formant (high singing formant). Vocal speech is distinguished from conventional speech by vowels of tenfold duration, by presence of vibrato--modulation of the amplitude, frequency and spectrum (with a periodicity of 5-7 Hz), by the particular features of articulation (diction), by the nature of respiration, by unusually pronounced activity of the resonating system and so on.

We demonstrated that these features of vocal speech are significantly predetermined by the unique function of singing as a means for transmitting emotional and esthetic information primarily (6). The distinct emotional orientation of vocal speech is emphasized by the existence of such genres in singing as weeping (for example "Yaroslavna's Lament," which expresses the emotion of grief), various songs accompanying games and dancing (expressing the emotion of joy) and so on. Sometimes the same vocal work may contain an entire range of emotional experiences (for example M. I. Glinka's "Doubt" and others).

All of this makes vocal speech an extremely convenient object for studying the acoustic resources of expression of emotions by voice.

Our laboratory has been working in this direction for a number of years, since 1972. Considering that we have discussed the experimental results thus far rather completely in recently published works (3-7), we will limit ourselves in this communication to a brief commentary on them and to their extremely general evaluation in connection with the problem of the language of emotions.

The main objective of our research was to answer the question "What acoustic parameters of a singer's voice are responsible for transmission of information on an emotional state, and to what extent are they responsible for doing so?

To obtain the raw data for our research we used two modifications of the actor transformation method, which has earned a rather substantial reputation as a means of simulating the emotional states of man (8,9,11,12). The first variant of the method, which we described in a number of previous works and which we named the method of emotional-semantic inversion (3-7), essentially required a singer to sing the same phrase of a vocal piece, but with different emotional tones or in different emotional context (joy, grief, anger, fear, indifference). Then the samples of singing which were judged most adequate to the particular emotions by a sufficiently representative group of expert listeners were singled out (the probability of correct identification of the emotional context had to be not less than 0.8-1.0), and finally the phrases contained in these samples were subjected to acoustic analysis with the purpose of revealing structural changes in the signal associated with change in emotional context.

The second modification of the method, which was tried out in the work of one of the graduate students of our laboratory, G. M. Kotlyar (7), essentially entailed selection of samples of vocal phrases from pieces sung by an outstanding master of vocal arts--F. I. Chaliapin, samples which experts scored as having the most clearly expressed emotional context (joy, grief, anger, fear), and their subsequent acoustic analysis.

Acoustic analysis was performed by the methods of integral spectrometery, oscillography and intonometry, and by the methods of statistical computer treatment of the results. In distinction from our previous works (3,4) a significantly broader range of acoustic characteristics of the vocal spoken signal was subjected to analysis:

1) Average duration of a syllable within the phrase;
2) coefficient of variation of the duration of syllables in the phrase;
3) relative duration of pauses in the phrase (in proportion to the total duration of the phrase);
4) average loudness of the sounds of words in the phrase;
5) the coefficient of variation of voice loudness in relation to syllables;
6) average steepness of the front (the leading edge of sound loudness) of syllables within the phrase;
7) average steepness of the drop in loudness of syllables;
8) relative level of the high singing formant (according to integral spectrometry);
9) the peak frequency of the high singing formant;
10) vibrato frequency;
11) the depth of amplitude-frequency modulation of vibrato;
12) the nature of changes in the fundamental tone (the melodic orientation and the intonational stability of the voice).

It was established that all of the acoustic parameters of the singing voice listed above changed to one degree or another in response to change in the emotional context of the vocal phrase. In this case a unique set of distinguishing characteristics was discovered for each emotion--grief, anger, joy, fear. We adopted, as the acoustic characteristics of particular emotional states, the extreme values (in the mean statistical data) assumed by each of the acoustic parameters listed above in response to changes in emotional context. By doing so, we were able to determine the sets of characteristics that were associated with each of the emotions.* As a result we established that negative emotions--grief, anger, fear--possess a larger number of distinguishing acoustic characteristics (six or seven), while the positive emotion (joy) possesses fewer (two). We associated this finding with the greater biological significance of negative emotions, which had its influence on formation of the acoustic code of emotions in the course of evolution.

Acoustic resources for expressing emotions by voice are characterized as extralinguistic resources of communication (10). Our research showed that vocalists have the capability for transmitting information about an emotional state to their listeners not only by singing vocal phrases from different musical pieces, but also by singing vocal exercises (singing melodies without words, using one vowel) and even by singing just a single (!) vowel on a single note.

We also performed a special experiment having the purpose of establishing the possibility of expressing emotions by means of instrumental music. At our request, a professional violinist played a passage from the "Rondo Capriccioso" by Saint-Saens several times, imparting different emotional meaning to each execution--joy, grief, anger, fear, indifference. When these passages were subsequently played back to different categories of expert listeners, the percentage of correct identification of emotional context was extremely high (70-100 percent).

These data show good agreement with our previous experiments on the role of the dynamic characteristics of a signal as resources for transmitting the emotional context of a vocal phrase (4), and with our ideas about the extralinguistic nature of the resources of emotional expressiveness.

The general results of our research led us to the conclusion that the system of vocal speech contains a unique alphabet or set of acoustic voice characteristics which a singer uses to transmit certain emotional moods to his listeners. Despite the infinite diversity of shades of emotion and of the shades of emotion used in singing, and despite the individuality of renditions by different vocalists, there are grounds for suggesting that the acoustic code of emotions in singing is an invariant. Research has demonstrated the unity of the principles of emotional expression in both the singing of outstanding masters of vocal art (Chaliapin) and that of contemporary singers. Apparently it is precisely this unity of the acoustic resources of emotional expressiveness that makes the emotional content of singing universally intelligible to a broad audience.

Comparison of our data with the results of research on the acoustic resources of emotional expression in ordinary speech (1,2) also leads to the conclusion that the acoustic resources of emotional expression are common to both singing and speech. A talented singer selects the resources of emotional expression from life experience (that is, from speech). This is precisely why such resources are distinguished by plausibility and universal intelligibility.

---

*The procedure for determining these characteristics and the experimental results are described in greater detail in (3,4,6,7).

The physiological grounds for such universality and for the universal intelligibility of vocal resources of emotional expression can be found in the dependence of the nature of sounds emitted by man on the physiological state of the body of the individual experiencing a given emotion, and on the state of vocal organs in particular. Thus the acoustic characteristics of emotion contained in voice are determined by physiological characteristics of the individual's state. This fact apparently lies at the basis of not only the unity of the vocal resources of emotional expression but also the mechanisms by which emotional information is decoded by the listener--that is, by which he perceives emotions. Moreover the facts cited above also indicate that there is something in common in the vocal resources of emotional expression in people and animals--a phenomenon first recognized by Charles Darwin.

Finally, we have the grounds for suggesting that the resources of musical expression or, to put it more precisely, the resources of emotionality in music, also have something in common with the resources of emotional expression in speech and singing. Diverse though it may be today, instrumental music, which was born in man's distant past, more than likely as a means of imitating the human voice, remains emotionally effective and comprehensible to the listener to the extent that it reflects the resources of vocal emotional expression. A confirmation of this point of view can be found in the statements of many prominent musicians, and in their works as well.

The language of emotions is without a doubt evolutionarily older than logical speech. In the course of the individual's ontogenetic development, he masters the language of emotions long before logical speech. According to the theory of evolution this phenomenon reflects a phylogenetic law. However, with arisal of logical speech the language of emotions used by man's prehistoric ancestors did not lose its important function in communication. It makes the logical meaning of speech significantly richer, and it resumes its dominance in singing, music, theatrical speech and acting, acquiring the attributes of the highest form of art.

The most important general feature of the acoustic code of emotions is its independence of the phonetic acoustic code (logical speech), despite the fact that both are essentially characterized by the same acoustic parameters of the human voice. This feature, which we demonstrated experimentally with vocal speech using the method of emotional-semantic inversion, has a morphological explanation as well: It has been established by neurophysiological studies that man perceives the logical meaning of speech by the left hemisphere of the brain, and its emotional meaning by the right hemisphere (13).

The practical aspects of our research: We discovered during our experiment that far from all vocalists are capable of satisfactorily expressing different emotions by their voices, in the same way that far from all listeners can correctly identify the emotional context of the renditions played to them. Individual indicators varied from 30-50 to 100 percent for both performers and listeners. This provided the grounds for recommending special tests (created on the basis of the method of emotional-semantic inversion) to evaluate the emotional expressiveness of singing and the capability for perceiving the emotional content of singing. These tests may be used in conservatories and music schools as objective criteria in occupational selection and in voice instruction.

BIBLIOGRAPHY

1. "Rech' i emotsii. Materialy simpoziuma (11-14 noyabrya 1974)" [Speech and Emotions. Symposium Proceedings (11-14 November 1974)], Leningrad, 1975, 127 pp, illustrated.

2. "Rech', emotsii i lichnost'. Tezisy dokladov simpoziuma" [Speech, Emotions and Personality. Abstracts of Symposium Reports], Leningrad, 1978, 40 pp.

3. Kotlyar, G. M., and Morozov, V. P., "Temporal and Dynamic Characteristics of Vocal Speech as Acoustic Correlates of Its Emotional Expressiveness," in "Rech' i emotsii. Materialy simpoziuma," Leningrad, 1975, pp 29-37.

4. Kotlyar, G. M., and Morozov, V. P., "Acoustic Correlates of the Emotional Expressiveness of Vocal Speech," AKUSTICH. ZHURNAL, Vol 22, No 3, 1976, pp 370-376.

5. Morozov, V. P., "Measurement of the Emotional Expressiveness of Vocal Speech: Possibilities and Prospects," in "Nauchn. sessiya, posvyashch. 80-letiyu D. Ya. Andguladze. Tez. dokl." [Scientific Session Dedicated to the 80th Birthday of D. Ya. Andguladze. Report Abstracts], Tbilisi, 1976, pp 9-10.

6. Morozov, V. P., "Biofizicheskiye osnovy vokal'noy rechi" [Biophysical Principles of Vocal Speech], "Nauka", L. O., 1977, 232 pp, illustrated.

7. Kotlyar, G. M., "Investigation of Acoustic Resources for Expressing Emotional States in Vocal Speech," abstract of dissertation in pursuit of the degree of candidate of psychological sciences, Leningrad, 1977.

8. Simonov, P. V., "Metod K. S. Stanislavskogo i fiziologiya emotsii" [The Stanislavsky Method and the Physiology of Emotions], Moscow, 1970, 153 pp, illustrated.

9. Simonov, P. V., "Vysshaya nervnaya deyatel'nost' cheloveka. Motivatsionno-emotsional'nyye aspekty" [Human Higher Nervous Activity. Motivational and Emotional Aspects], Moscow, 1975, 150 pp, illustrated.

10. Bondarko, L. V., "Zvukovoy stroy sovremennogo russkogo yazyka" [Acoustic Structure of Modern Russian Language], Moscow, 1977, 176 pp, illustrated.

11. Kunitsin, A. N., "Experience in Using Stage Techniques to Simulate Oral Expression of Emotional States," in "Rech' i emotsii" [Speech and Emotions], Leningrad, 1975, pp 38-44.

12. Galunov, V. I., and Tarasov, V. I., "The Naturalness of Expressions of Emotional States and Study of the Characteristics of the Spoken Signal," in "Rech' i emotsii," Leningrad, 1975, pp 55-61.

13. Balonov, L. Ya., and Deglin, V. L., "Slukh i rech' dominantnogo i nedominantnogo polushariy" [Hearing and Speech in the Dominant and Nondominant Hemispheres], Leningrad, 1976, 220 pp, illustrated.

REMOTE CONTROL OF OPERATOR STATE IN CONNECTION WITH THE OBJECTIVES OF EXPERT CERTIFICATION

V. I. Myasnikov, O. V. Kon'kova, B. A. Popov, F. N. Uskov

Evaluations made by specialists representing the appropriate services are presently enjoying active use in some special areas of medical and psychological expert certification. The tasks of such experts include determining the mental status of operators and defining group psychosocial characteristics. The end goal of such specialists is to make a functional diagnosis--that is, not only to reveal the structure and degree of possible disturbances in mental activity, but also to make a prediction and determine what measures would promote psychological and occupational adaptation. An operator's speech and his verbal behavior are among the principal sources of the information such specialists need. Research and practical efforts in this area have shown that we are making far from full use of all of the possibilities afforded by dialogue speech. Thus relatively little attention is being devoted to spoken communication, or dialogue, as an independent source of information, even though some positive results have been achieved in this area (1).

But use of dialogue speech for diagnostic purposes does involve certain difficulties. First of all it should be pointed out that control of the parameters of dialogue speech always presupposes comparison with certain standards. As a rule indicators obtained from different people in a representative sample serve as such standards. However, deeper examination of this question has revealed that application of standards to dialogue speech is not entirely valid, inasmuch as the laws of speech generation are to a high degree individualistic for each pair of communicants. This leads to the practically important problem of obtaining standards applicable to each pair of observed persons. In this case we would be dealing with difficulties of another sort, namely choosing the method that would ensure reliability in the comparison of different passages of dialogue speech in the course of its generation. A number of studies have been performed in this area; the most significant are those performed under Jaffe's guidance (2). Jaffe's system provides the framework for tackling the tasks of operational diagnosis.

As we know, the state of an individual is characterized by a broad spectrum of behavioral manifestations. It should be emphasized that different aspects of behavior are closely interrelated. Each of them assumes greater or lesser diagnostic significance depending on the type of situation and on the state elicited by that situation. Consequently one-sided consideration of individual aspects of behavior would result in one-sided evaluations of states. In this connection behavior data must be subjected to integral analysis. Speech behavior must not be examined apart from other resources of communication.

In recent years specialists have focused attention on a group of behavioral elements that have come to be called nonverbal. It has been demonstrated that nonverbal behavior is related to speech behavior by certain laws, and in this connection it may serve as an additional source of diagnostic information. Elements of nonverbal behavior which should be considered include: the distance between the partners; the location of the partners in space relative to one another during discussion; the postures of the partners; gesticulations; visual contact. According to our data, directed consideration of these behavioral elements significantly raises the diagnostic value of information, and especially in regard to evaluating the emotional relationships between the communicants.

One example would be to record the number of visual contacts achieved during oral communication (by means of television apparatus). Establishment of the fact of visual contact would require the ability to determine, from the image on a monitor screen, that the gaze of one of the partners is fixed upon the other. Special experiments were conducted to reveal the accuracy with which an expert could establish the point of gaze fixation under these conditions and to determine the possibilities for developing this capability in an expert. The research results showed that the accuracy with which an expert determined the gaze fixation point is sufficient to establish the presence or absence of visual contact ($\pm 10°$, $p \leqslant 0.01$). We revealed a tendency to understate the true value of the gaze fixation point ($p \leqslant 0.001$); consequently the vector of arising errors is in the direction of the conditional zero point (a fixation point located right before the eyes). Thus a situation is created in which angular dimensions are systematically understated by the expert. In a case where the plane of the subject's face on the monitor screen is at an angle to the plane of the screen, the quantity of errors made in the evaluations is dependent on yet another variable--the angle between the axis of eye fixation and the axis formed by the expert and the image of the observed subject. As this angle increases, the number of errors grows. It was also established in this research that the expert's capability for determining the gaze fixation point by means of a televised image is a trainable function. Hence follows the need for special training that would increase the reliability of establishment of the fact of visual contact.

Another example of using nonverbal communication resources is the measurement of the metric distance between partners, which is used as one of the indicators of social distance. A special computational formula was proposed in this connection (3). It provides quantitative data (in the first approximation) for this indicator, and it requires elementary treatment of the televised images of the partners:

$$D^2 = \left(\frac{Ac}{a}\right)^2 + \left(\frac{Bd}{b}\right)^2 - 2\left(\frac{Ac}{a}\right)\left(\frac{Bd}{a}\right)\cos\alpha + \left(\frac{Ak}{a} - \frac{Bk}{b}\right)^2;$$

where D--distance between partners; A,B--true linear dimensions of the partners; a,b--linear dimensions of partners on the screen; c,d--distance, on the monitor screen, from the conditional center of mass of the partners to the center of mass of the partners and to the center of the television screen; k--empirical coefficient of the television system; α--angle between vectors c and d.

The next important step is to consider the ratio of communicative to noncommunicative gestures, especially if we wish to evaluate the stability of the group as a single whole in stressful and critical situations. A shift of this ratio in the direction of noncommunicative gestures indicates certain difficulties in interpersonal communication and interaction. On the other hand mutual correspondence between the postures of partners engaged in oral communication indicates group cohesiveness. High correspondence indicators are observed with high cohesiveness of the group, which is highly significant to expert practice. We developed special printed forms to be used in expert certification to obtain quantitative characteristics of nonverbal behavior.

We now have unlimited possibilities for automating control of the nonverbal parameters of behavior. In the opinion of a number of authors, application of Jaffe's system to analysis of nonverbal communication resources opens up certain prospects. From all that we can see, the potentials of this system are in fact much greater, inasmuch as it makes use of stochastic processes.

In conclusion we would like to turn attention to one more information source, namely proxemic parameters. Proxemic parameters--that is, characteristics of the spatial organization of human behavior--must be accounted for if we are to raise the accuracy of conclusions made by an expert on all noted behavioral facts.

Thus it is rather difficult today to evaluate the state of an individual on the basis of just speech alone. Obviously if progress is to be made in this area, an integrated approach would be required, one which would not permit the rendering of one aspect of behavior absolute at the expense of others. Integrated experimental research should demonstrate the validity of this approach.

BIBLIOGRAPHY

1. Gazenko, O. G., Myasnikov, V. I., and Uskov, F. N., "Behavioral Control as a Tool of Evaluating the Functional State of Cosmonauts in Flight," AVIATION, SPACE AND ENVIRONMENTAL MEDICINE, November 1976, pp 1226-1227.

2. Jaffe, I. I., and Teldstein, S., "Rhythms of Dialogue," Academic Press, New York, 1970.

3. Popov, B. A., "A Procedure for Working With Telecommunication Data in the Period of LKI [expansion unknown]," in "Materialy konferentsii molodykh spetsialistov IMBP MZ SSSR, g. Moskva, 1977 g." [Proceedings of a Conference of Young Specialists of the USSR Ministry of Public Health Institute of Biomedical Problems, Moscow, 1977] (in press).

4. Val'siner, Ya., personal communication.

VARIABILITY OF THE AVERAGED SPECTRUM OF VOWEL SOUNDS IN QUIET, NORMAL AND LOUD SPEECH

A. V. Nikonov

Experimental research on spoken communication by an individual experiencing different emotional states showed that the dynamics of psychophysiological states are accompanied by specific changes in the semantic, grammatical and acoustic structure of speech. Thus when the individual experiences pronounced emotional arousal the most typical changes include, in particular, growth in the intensity of speech and in the frequency of the fundamental tone of the voice, change in tempo and distribution of pauses in speech, and deformation of the intonational contour and the spectral structure of the spoken statement. The most typical changes associated with fatigue and depression include a decline in the intensity of speech, monotony, an increase in the length of pauses and of the latent period of the speech response, and decreases in the volume of speech produced and the verbal and stylistic diversity.

The expressiveness of behavioral speech reactions is rather closely associated with the general emotional tone of the individual and with the degree of activation of mental processes.

General arousal of the organism directly effects the nature of breathing, the rhythm of speech, the tone of speech forming organs, and particularly the vocal force applied and the dynamic range of speech intensity. An increase in the intensity of speech is invariably accompanied by significant changes in the rhythmic, intonational and spectral structure of the spoken signal. Changes in the intonational contour of speech may be elicited by forcing of the voice, or it may serve an independent communicative purpose (for example that of expressing irony, sarcasm and so on) while variations in speech intensity remain relatively insignificant.

Our objective was to study the dynamics of the averaged third-octave spectrum of vowel sounds in quiet, normal and loud speech (low, normal and high vocal force exertion by the speakers).

Twelve male speakers took part in the experiments. Isolated vowels sounds were pronounced many times (not less than 10) with varying vocal force. The intensities of the third-octave components of the spoken signal were measured with the assistance of a condenser microphone positioned a standard distance away from the speaker, a measuring amplifier and an acoustic frequency spectrum analyzer (type 3347).

Analysis of the experimental data showed that simultaneously with growth in speech intensity, there occurs an involuntary increase in the average frequency of the fundamental tone--F_0 (Table 1). Thus when we proceed from quiet to normal speech (when speech intensity increases by 8-12 db), F_0 increases insignificantly, by 5-20 percent, while when we proceed to loud speech (with the speech intensity increasing by 20-24 db), F_0 increases by 60-90 percent. An increase in the average frequency of the first formant, F_1, is also observed when the voice is forced. According to data from third-octave spectrum analysis, meanwhile, changes in the mean frequency of the second formant of vowel sounds are less pronounced, and the directions in which the changes occur differ.

Table 1. Mean Frequency of the Fundamental Tone and First Formant of Vowel Sounds Uttered Ten Times by Three Speakers With Different Vocal Force

(1) Звуки	F_0 (гц) (2) (3) норм.	(4) тихо	(5) громко	ΔF_0 (гц)	F_1 (гц) норм.	тихо	громко	ΔF_1 (гц)
А	112	106	170	64	770	710	850	140
О	125	105	190	85	590	450	570	120
И	135	142	240	98	270	285	480	195
У	142	138	265	127	270	270	530	240
Э	132	115	235	120	570	500	760	260
А	142	128	190	62	750	660	860	200
О	155	130	190	60	600	450	660	210
И	165	140	205	65	325	260	390	170
У	185	145	225	80	390	315	420	105
Э	172	130	195	65	510	420	630	210
А	170	145	175	30	700	680	710	30
О	168	135	200	65	600	580	610	30
И	175	140	225	85	385	270	450	180
У	175	132	235	103	390	390	470	80
Э	167	140	217	77	500	430	650	220

Key:
1. Sounds
2. Hz
3. Normal
4. Quiet
5. Loud

Figure 1 shows the dependence of the mean frequency of the fundamental tone, F_0, of the vowel sound "O" on its intensity A.

As the vocal force increased from minimal to normal (as A increased from 60 to 80 db) F_0 increased by 41 Hz (from 92 to 133 Hz), while when A increased from 80 to 100 db (that is, when the speaker went from normal to forced speech) the increment in F_0 was 186 Hz.

An increase in the dynamic range of the mean frequency of the fundamental tone is typical of higher vocal force. This nonlinear dependence of F_0 on speech intensity is typical of male and female voices.

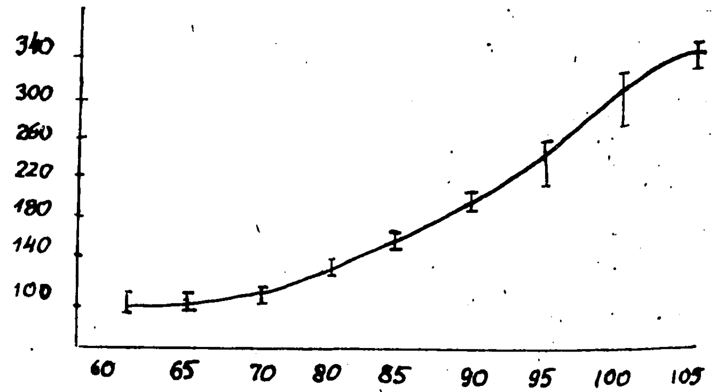

Figure 1. Dependence of Fundamental Tone Frequency F_0 on Sound Intensity: Points indicate mean values of F_0, lines indicate standard deviations of F_0

Table 3 shows typical values of the averaged third-octave spectrum of vowels uttered with different vocal force. We can see from Table 2 that when vocal force is low, the bulk of the energy in the spectrum of vowel sounds is concentrated in the low frequency range, 125-160 Hz (the range of the fundamental tone's frequency).

As the intensity of speech rises, the frequency of the fundamental tone increases, and the harmonics of the fundamental tone (in the 250-630 Hz range) have an intense masking influence upon the averaged third-octave spectrum, hindering measurement of formant frequencies.

At normal vocal force the spectral components in the range of the formant frequencies increase, while low frequency components decrease. When vocal force is increased, the high frequency spectral components (2,000-4,000 Hz) undergo additional increase while low frequency components (160-250 Hz) decrease.

Analysis of the results showed that third-octave spectral analysis would best be used to evaluate the dynamics of the fundamental tone's frequency and the extent of energy redistribution in the spectrum of vowel sounds.

The shape of the averaged spectrum of vowel sounds bears information on the individuality of the speaker and on the dynamics of his vocal force.

Table 2. Averaged Third-Octave Spectrums of Vowel Sounds Uttered With Different Vocal Force

Гласные (1)	Средняя частота 1/3 октавных голосовых фильтров в гц (2)																		
	100	125	160	200	250	315	400	500	630	800	1000	1250	1600	2000	2500	3150	4000	5000	6300
А тихо (3)	—	69	90	70	58	72	62	73	77	77	72	73	65	65	66	65	68	58	59
А норм. (4)	—	53	81	85	69	72	76	87	91	95	88	91	79	74	80	73	74	60	54
А громко (5)	—	—	52	80	71	59	83	75	93	88	87	91	78	81	91	90	86	65	56
У тихо	59	53	83	76	65	85	74	62	73	64	65	66	61	60	57	56	—	—	—
У норм.	56	54	72	90	78	74	89	75	86	74	62	56	55	60	64	69	70	53	52
У громко	—	56	67	78	93	85	81	87	85	95	75	64	61	78	72	84	71	55	70
О тихо	57	71	95	82	72	91	84	83	75	79	65	54	56	65	63	53	54	—	—
О норм.	—	53	82	86	66	87	90	91	85	82	83	65	63	71	74	80	72	62	—
О громко	—	57	65	75	85	72	84	97	91	93	94	82	89	54	87	83	65	64	63
И тихо	—	61	88	91	66	60	62	—	—	—	—	—	—	58	60	61	60	54	—
И норм.	—	—	64	82	70	59	78	60	60	—	52	52	59	75	91	93	89	72	55
И громко	—	—	54	55	75	92	82	63	75	65	75	77	81	99	104	98	100	84	83
Э тихо	50	74	89	66	70	84	68	69	62	62	57	57	65	76	71	53	63	54	—
Э норм.	50	70	64	69	90	75	81	57	65	66	79	94	94	79	—	65	—		
Э норм.	—	—	50	70	64	69	90	75	81	57	65	66	79	94	94	79	—	65	—
Э громко	—	—	—	—	72	61	66	89	76	73	65	69	80	85	92	86	83	62	54

Key:
1. Vowels
2. Average frequency of third-octave vocal filters, Hz
3. Quiet
4. Normal
5. Loud

POSSIBILITIES FOR EVALUATING INTENSITY OF A SPEAKER'S EMOTIONAL TENSION ON THE BASIS OF CHANGES IN CHARACTERISTICS OF HIS SPEECH

E. L. Nosenko

One approach to examining speech, in which it is viewed as one form of complex intellectual activity, presupposes representation of its organization as a "hierarchical multilevel substructure" (N. A. Bernshteyn, 1966), the levels of which differ in complexity and in the subject's awareness of them. As we know, the individual's awareness of individual components of polystructural activity increases in the course of this activity as we ascend from one level to the next. While the speaker is "actually" aware (using A. N. Leont'yev's term, 1947) of the dominant semantic level of speech and his awareness of operations associated with the lexical and grammatical structure of a statement is limited to conscious monitoring--that is, he is not fully conscious of them, motor realization of the statement is at an even lower level in the hierarchy of awareness of speech phenomena--the unconscious level, or the level of "unconscious control" (A. N. Leont'yev, 1965).

This communication will attempt to justify the possibility for evaluating the intensity of emotional tension experienced by the individual on the basis of a consideration of which components of speech are responsible for various difficulties in speech.

The hypothesis is suggested that arisal of mistakes and difficulties in speech of which the speaker is unaware (which he cannot correct) or of which he is aware but finds it hard to surmount, is an indication of a high degree of emotional tension in the speaker. This pertains to those elements of speech which are at higher levels of awareness in the hierarchy of the levels of organization of the spoken statement, and which can consequently be controlled more meticulously in speech proceeding in a normal state. This hypothesis is based on the experimentally established fact that in a state of emotional tension, conscious control over the quality of activity weakens.

Apparently the more intensively the speaker experiences emotional tension, the more his capacity for maintaining effective conscious control over the quality of his activity is disturbed, which is what leads to mistakes and difficulties not only at levels of organization of a statement requiring distribution of attention between the intent of the statement and its concrete linguistic realization, but even at the highest level of speech, the semantic level.

To test this hypothesis we subjected, to comparative psycholinguistic analysis, the spoken statements of the same subjects in a state of emotional tension elicited by different emotion-producing factors (an examination, questioning, anticipation of surgery and so on) and in a normal state.

Special attention was devoted to examining indecisiveness phenomena in speech and statements containing the speaker's self-assessment of the quality of his own statement. We also used the method of having a subject listen to a tape recording of his speech when he was in a state of emotional tension, and then asking him to comment on what he heard.

Our observations may be summarized as follows.

1. Mistakes and difficulties arise in the speech of subjects in a state of emotional tension primarily at the level of the grammatical structure of statements.

This may be explained by the following. Owing to the mechanism of "conscious control," a speaker in a normal state selects linguistic units and links them together in a syntactical scheme efficiently and without mistakes, all the more so because in performing this operation, the speaker need consider only the relationships between linguistic signs, and he need not relate them to extralinguistic objects or concepts, as in the case of selection of the syntactic structure itself, or in the process of choosing words adequate to a given goal of communication.

In a state of emotional tension of even an insignificant degree, it becomes more difficult to distribute attention between the semantic level of speech and its linguistic structure, which is what leads to errors in syntax, to "awkwardness" of composition and so on. As a rule the speaker does not even notice these mistakes. On hearing recordings of their own speech, subjects are bewildered by the fact that they may have, for example, declined a noun improperly without noticing this slip of the tongue. An example would be "...po sravneniyu s 1965 godu" (instead of "godom").

2. Arisal of mistakes (corrected by the speaker!) in the choice of words and selection of the syntactic scheme of a statement appropriate to the given goal of communication attests to greater weakening of conscious control over the quality of activity in a state of emotional tension, and consequently to greater intensity of this state.

The fact is that for the speaker to understand his choice of a certain word or syntactic scheme for a statement, he must consider the intent of the statement. This operation proceeds under a greater degree of control of the speaker's voluntary attention than does observation of the rules of grammar.

Therefore if even on the condition that the speaker is able to concentrate his voluntary attention on his speech he makes errors such as using an inappropriate adjective, and he fails to recognize such errors, we would have adequate grounds for suggesting that he is experiencing a state of severe emotional tension. The plausibility of this hypothesis is confirmed by the fact that even after becoming aware of the inadequacy of his choice of a particular word, a speaker in a state of emotional tension is unable to efficiently find an adequate substitute. Evidence of this can be found in the numerous "false starts" in speech, analysis of which would show that they unambiguously signal the arisal of difficulties in word choice. For example when asked the question: "What geometric figures do you see in Figure 3?", a traffic controller in a state of emotional tension replied: "... the pho... the photograph bears a a star, a circle and a triangle." The false start "The pho..." attests to the fact that the decision to use the word "photograph" instead of the word "figure" does not satisfy the speaker himself, but he is unable to effectively find the adequate substitute. A speaker's self-assessment of the quality of his own speech can make it especially clear that he is having difficulties in word choice. For example: "In the second row I see a (the speaker pauses for 1.8 seconds)...(the speaker stutters)... a watchamacallit (a pause of 1.2 seconds) a (a pause of 0.3 seconds) triangle, how silly of me!"

3. In a state of severe emotional tension (which we observe, for example, in subjects prior to surgery), speech changes occur even at the level of programming the intent of a statement.

The grounds for this assertion are that when subjects are permitted to hear recordings of their own speech, they note the overemphasized positive or negative connotation of the words they choose as being "unnatural," "atypical of them in a normal state." For example "...my lab results are very terrible."

Disturbances in the dominance of the names of things in units of speech beyond phrase length (in contrast to speech in a calm state, the speaker forgets whether or not certain objects and subjects presently under discussion had been mentioned earlier) also attests to weakening of control over the quality of speech even at its dominant level, the one of which the speaker is fully aware in the course of his speech.

This communication offers a classification of different changes in the characteristics of speech in a state of emotional tension from the standpoint of their usefulness in identifying different degrees of intensity of this state.

SOME FLOWCHARTS FOR ANALYSIS OF THE STATE OF AN INDIVIDUAL ON THE BASIS OF CHARACTERISTICS OF HIS SPEECH

E. L. Nosenko, O. N. Karpov, A. A. Chugay, G. N. Bordovskiy

Analysis of the state of an individual doing work presupposes effective acquisition of information about his state.

The state of a human operator is usually evaluated on the basis of measurements of a number of physiological parameters (pulse, heart beat, respiration and so on) made by contact sensors.

This report examines new set-ups for monitoring changes in the state of an operator, based on recording changes in the characteristics of his speech.

The authors obtained experimental material confirming the informativeness of a large number of speech parameters to be used as indicators of emotional tension that may arise in an operator in critical work situations and lead to work failure. In contrast to previous research in which the vocal communication channel was used as a source of information on the state of the human operator (P. V. Simonov, M. V. Frolov, L. N. Luk'yanov, V. A. Popov; C. E. Williams and K. N. Stevens, etc.), the authors of this communication have developed a means for monitoring changes in the state of a speaker requiring not comparison of the intonational contours of the same standard words or phrases, but analysis of the flow of coherent speech.

Changes in the characteristics of speech associated with particular features of the neurophysiological mechanisms of emotional state may be classified as follows:

1. Changes in the characteristics of speech in a state of emotional tension stemming from the characteristics of autonomic reactions inherent to this state.

2. Changes in the characteristics of speech reflecting the particular features of the sensory and mental processes occurring in a state of emotional tension.

3. Changes in the characteristics of speech associated with certain motor reactions occurring in a state of emotional tension.

Considerable tensing of the muscles of the speech forming apparatus, to include the vocal cords, in a state of emotional tension causes change in the frequency of the voice's fundamental tone. Consequently it would be suitable to use, as indicators of emotional tension, characteristics of the frequency of the fundamental tone such as the range of its variations, the swiftness with which zones appear in the flow of speech in which the frequency of the fundamental tone significantly surpasses the mean frequency typical of the given speaker, and so on.

Changes in breathing rhythms, which have an effect on the temporal characteristics of speech, are typical of emotional tension: The number of pauses in the flow of speech increases, their duration grows longer, and the locations of pauses change.

On this basis, the following could be used as objective indicators by which to identify states through speech characteristics:

1. Fluctuations in the frequency of the voice's fundamental tone.

2. Fluctuations in the loudness of speech (increases or decreases in comparison with speech in a normal state).

3. Changes in the tempo of articulation (in the absolute tempo of speech).

4. Fluctuations in the general tempo of speech from maximum to minimum (the range of variation of the speech tempo, the rate of change of speech tempo).

5. Change in the average length of a passage of speech uttered without pauses due to indecisiveness (a decrease in comparison with speech in a normal state).

Our objective was to create an electronic speech analyzer to be used in a diagnostic system recognizing an individual's emotional state on the basis of his speech characteristics.

The following were used as informative parameters: change in frequency of the voice's fundamental tone and changes in temporal characteristics of speech.

To simplify the circuitry of the analyzer, we settled on digital representation of amplitude and temporal parameters.

The speech analyzer is based on series K-155 integrated microcircuits assembled into a set of counters accumulating information on the temporal characteristics of speech within 10 seconds of current time, and information on changes in the frequency of the fundamental tone during each second of the tonal signal.

The fundamental tone analysis circuit represents a digital filter, and the parameter it measures is the per-second distribution of the number of periods in the spoken signal on a frequency axis.

Change in signal intensity is determined by comparison of the envelope of the wideband signal with the thresholds of normal intensity--that is, normal, loud and soft bands are isolated.

The parameter measured here is the number of times a given threshold is exceeded in a particular time interval.

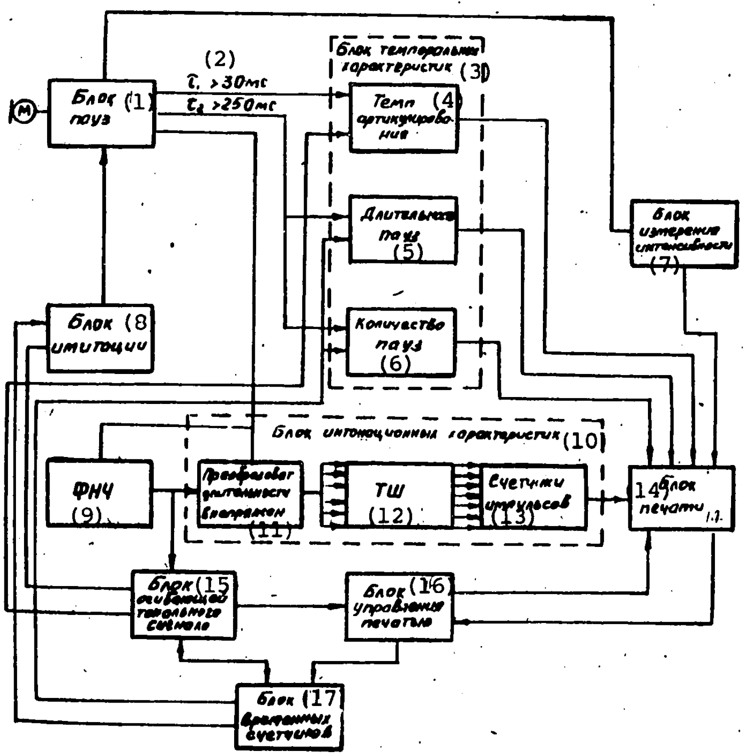

Figure 1

Key:

1. Pause block
2. Msec
3. Temporal characteristics block
4. Articulation tempo
5. Pause duration
6. Number of pauses
7. Intensity measuring block
8. Simulation block
9. Low frequency filter
10. Intonational characteristics block
11. Duration-to-voltage converter
12. TSh [not further identified]
13. Pulse counters
14. Print-out block
15. Tonal signal envelope block
16. Print-out control block
17. Time sensor block

Change in the parameters of speech with respect to time provides a complete dynamic picture of changes in the individual's functional state.

The functional layout of the analyzer consists of the following blocks (Figure 1):

temporal characteristics isolating block;

intonational characteristics block;

print-out control block;

print-out block;

pause simulation block;

time sensor block;

intensity measuring block.

The analyzer works as follows. In the pause block the spoken signal is transformed into an envelope indicating presence and absence of a signal; two types of envelopes are produced in this case: with a pause duration $\tau_1 \geqslant 30$ msec and with a pause duration $\tau_2 \geqslant 250$ msec. The envelope with τ_1 pauses is used to count the number of pauses and their duration within each 10 second interval, and the envelope with τ_2 pauses is used together with the envelope of the tonal signal to obtain the characteristics of articulation tempo. Concurrently the vocal signal passes from the pause block through the low frequency filter, with $f_{ave} = 400$ Hz, into the intonational characteristics formation block.

This block measures the duration of each period of the frequency-modulated signal of the fundamental tone, and depending on the duration of the period, a value of one is added to one of the eight accumulating counters. As a result the eight counters provide the pulse frequency distribution for a 1 second interval of the fundamental tone. Intensity is measured in the intensity measuring block and compared with eight thresholds. The parameter measurements are fed to the print-out block, which is controlled by the time sensor block and by signals from an MP-16 high-speed typewriter.

Independent verification of the analyzer is achieved by the simulation block, which generates pulses of variable frequency within the limits of the fundamental tone frequency, and tonal signal envelopes with pauses $\tau_1 \geqslant 30$ msec and $\tau_2 \geqslant 250$ msec.

SOME CHARACTERISTICS OF EMOTIONAL WHISPERED SPEECH

E. A. Nushikyan, T. A. Brovchenko, S. N. Kolymba

Whispered speech has been an object of analysis several times at both the perception level and the acoustic level. It has been studied mainly with respect to individual sounds, syllables and words (1,2,3). Interest in studying whispered speech has recently grown, but we were unable to find any descriptions, in the linguistic literature, of the acoustic characteristics of emotional whispered speech, which is of considerable interest to a number of disciplines, including linguistics.

The objective of our research was to determine the acoustic characteristics of emotional whispered speech. Our research material consisted of 76 emotionally colored phrases expressing anger, amazement, irony and approval, and neutral phrases corresponding to the former. In keeping with our objective, we had a speaker read each sentence four times--emotional and neutral phrases were read correspondingly in a whisper and at normal loudness. Emotionality was imparted to the phrases by their pronunciation in context. To achieve neutral phrases, we left out the emotionally colored context. Recordings of whispered and loud pronunciation of the phrases by the speakers were subjected to listener analysis to determine how identifiable the emotional states were by native listeners.

Only those phrases which, according to not less than 80 percent of the listeners, expressed the expected emotional connotations were selected for electro-acoustic analysis.

The next stage of the work was electro-acoustic analysis of the selected phrases, performed with an intonograph at the laboratory of experimental phonetics of Odessa State University. Electro-acoustic analysis was performed with respect to the following characteristics: the envelope of the fundamental tone frequency for the entire phrase, the frequency range of the phrase in pt [not further identified], the frequency interval of the principal stressed vowel in pt, the peak value of the fundamental tone frequency, phrase duration in msec, average syllable duration in msec, and the speech rate--the number of syllables uttered per second.

Phrases were subjected to comparative analysis in two planes: emotionality-neutrality, loud-whispered speech.

As far as the first opposition is concerned, it was examined in previous works by the authors (4) in relation to loud speech; our present objective was to study this opposition in relation to whispered speech.

As we know from the appropriate literature, the sounds of whispered speech are created by the passage of air through constricted passageways of the whisper triangle and the glottis with the vocal cords not vibrating. The arising noise is the principal source of the sounds of whispered speech, as opposed to vocal phonation (2).

Obviously when whispered speech supplants vocal phonation the envelope of the fundamental tone can still be isolated, as is confirmed by other studies. There are indications in the linguistic literature that not only can the fundamental tone be heard in whispered speech, but also it can be voluntarily modulated (3). Visual analysis of the fundamental tone frequency envelope of our intonograms revealed that in comparison with voiced speech, the periods of the fundamental tone of whispered speech, complicated by noise components, exhibit an irregular nature.

Calculations showed that the percentage of the signal falling within the fundamental tone channel was on the order of 30 percent, as compared to 95 percent for vocal speech. Graphs revealing the dynamics of the fundamental tone frequency distinctly showed retention of the configuration of the fundamental tone curve in relation to both whispered and voiced speech.

A decrease in overall frequency was found to be specific to whispering in both emotional and neutral pronunciation. Thus in regard to the opposition "whispered-voiced speech," the maximum frequency of whispered speech was found to be 5 pt lower in relation to neutral speech and 3 pt lower in relation to emotional speech. In whispered speech, emotional phrases were distinguished by a maximum frequency 5 pt higher than that of the neutral variant.

The van der Warden test, which does not require knowledge of the distribution function and which may be used with a small number of variants, was applied to reveal differences in the compared neutrality-emotionality oppositions of whispered speech.

Interpreting the peak frequency of the fundamental tone (for female voices) in whispered neutral speech as a manifestation of random variable X and the corresponding characteristic of emotionally whispered speech as a manifestation of a random variable Ψ, and identifying the latter with serial numbers r and considering that $n_1=n_2=14$ and that n=28, we calculate X using the formula

$$X=\sum_r \Psi\left(\frac{r}{n+1}\right)=8,27 \qquad (5)$$

Turning to the van der Warden test table, we find X_{05}, X_{01}:

$$\begin{array}{ccc} X > & X_{05} > & X_{01} \\ 8,27 & 6,09 & 4,69 \end{array}$$

The calculated value confirms the presence of significant differences between the average values of the peak fundamental tone frequency for emotional and neutral phrases uttered in whispered speech. It may be concluded that these two general sets are unconditionally different.

Presence of differences in the peak fundamental tone frequencies of emotional and neutral whispered speech can also be seen in relation to male voices, though it would be difficult to conclude that the general sets are unconditionally different, inasmuch as

$$\begin{array}{ccccc} X_{05} & < & X & < & X_{01} \\ 4{,}08 & & 4{,}50 & & 5{,}26 \end{array}$$

where $n_1=n_2=11$, $n=22$.

Analysis of the frequency interval of the principal stressed syllable revealed that the latter is typified by a lower value in whispered speech than in voiced speech, in relation to both neutral and emotional pronunciation (for voiced speech we have 6 pt for neutral pronunciation and 10 pt for emotional pronunciation, while for whispered speech the figures are 4 pt and 6 pt respectively). Emotional phrases uttered in whispered speech were characterized by an insignificantly larger interval between syllables in comparison with neutral phrases.

Analysis of the frequency range of the phrases showed that whispered speech was accompanied by a significant decrease in this characteristic in comparison with voiced speech (for voiced speech we have 14 pt for neutral and 16 pt for emotional pronunciation, while for whispered speech we have 4 and 9 pt respectively). Emotional phrases uttered in whispered speech were characterized by a broader range in comparison with neutral phrases (5 pt broader in phrases expressing an intense degree of anger and amazement, and 2-3 pt broader in phrases expressing approval and irony).

Hence it follows that this parameter is obviously the most meaningful of all examined frequency characteristics, and it may be said to be a distinct sign of emotionality.

Analysis of temporal characteristics revealed a general tendency toward slower tempo in whispered speech in comparison with loud speech, which may be explained in part by the fact that articulation can be stronger with whispering than with loud speech (1). The tempo of whispered neutral speech is 3.2 syllables per second, for loud neutral speech it is 3.8 syllables, for whispered emotional speech it is 2.6, and for loud emotional speech it is 3.9 syllables.

Emotional pronunciation in whispered speech was found to be slower than neutral pronunciation, the figures being 2.6 and 3.2 syllables per second respectively. Average syllable duration in whispered neutral speech was 300 msec, it was 270 msec in loud neutral speech, 380 msec in whispered emotional speech and 280 msec in loud emotional speech. Average syllable duration in whispered emotional statements expressing anger and approval was 380 msec, it was 300 msec for neutral phrases, it was 270 msec for phrases expressing irony and amazement, and in neutral phrases corresponding to the former it was 280 msec. Applying van der Warden's test to the figures for average syllable duration in emotional and neutral whispered speech, we come to the conclusion that differences do exist between them; however, it would be difficult to conclude that these general sets are unconditionally different, inasmuch as

$$\begin{array}{ccccc} X_{05} & < & X & < & X_{01} \\ 6{,}50 & & 6.98 & & 8.51 \end{array}$$

where $n_1=n_2=25$, $n=50$.

This means that temporal characteristics are not always constant in their manifestation, and that in comparison with other acoustic parameters they are the most susceptible to change, obviously connected with expression of different emotional connotations.

As far as frequency characteristics are concerned, despite a decrease in their value in whispered speech in comparison with voiced speech, the difference between them is enough to permit their use for identification of emotional and neutral whispered speech.

And so, whispered speech carries unique information on the frequency and temporal parameters of the acoustic signal, information which may serve as a definite criterion for differentiating between emotional and neutral speech.

BIBLIOGRAPHY

1. Zhnikin, N. I., "Mekhanizmy rechi" [Speech Mechanisms], Moscow, 1958.

2. Butyrskiy, L. S., "Whispered Speech," in "Trudy VII Vsesoyuznoy shkoly-seminara ARSO" [Proceedings of the Seventh All-Union ARSO School-Seminar], Alma-Ata, 1973.

3. Smol'yevskiy, A. A., "Intonational Resources of Whispered Speech," in "Materialy Vsesoyuznoy konferentsii 'Analiz i sintez kak vzaimoobuslovlennyye metody eksperimental'nogo foneticheskogo issledovaniya rechi'" [Proceedings of the All-Union Conference "Analysis and Synthesis as Interdependent Methods of Experimental Phonetic Study of Speech"], Minsk, 1972.

4. Kolymba, S. N., Nushinkyan, E. A., and Pirogova, A. A., "Acoustic Correlates of Emotionally Colored Phrases Expressing Anger, Approval and Irony in Modern English," in "Rech' i emotsii" [Speech and Emotions], Leningrad, 1975.

5. Brovchenko, T. A., Varbanets, P. D., and Taranets, V. G., "Metod statisticheskogo analiza v foneticheskikh issledovaniyakh" [The Statistical Analysis Method in Phonetic Research], Odessa, 1976.

CHARACTERISTICS OF HUMAN SPEECH BEHAVIOR IN STRESSFUL CONDITIONS

V. A. Popov, I. S. Zamaletdinov, N. V. Krylova, A. V. Nikonov

If we accept the premises of the adaptation and information theories of arisal of emotional states, we should find that in addition to physiological and acoustic parameters, psychological parameters should enjoy broader use in research on emotional stress, in view of their greater reactivity and the diversity of the forms of mental reactions to stressful conditions.

The objective of our research was to reveal the influence of emotional factors associated with real extreme conditions (parachute jumping) on the acoustic and psycholinguistic indicators of the speech behavior of operators experiencing different degrees of emotional tension. To permit evaluation of general mental state, in addition to measuring autonomic indicators we recorded parameters characterizing perception, attention, memory, thinking and so on.

Our subjects included persons making their first parachute jump, persons with some jumping experience and record-breaking parachute jumpers. This permitted comparative analysis of the levels of dependence exhibited by variations in speech indicators on the expressiveness of emotional tension. In addition to psycholinguistic methods such as the free association test, and directed and undirected reporting, our array of methods included tests of working memory using numerical material, and the recording of the parameters of motor activity and the basic physiological functions.

Comparative analysis of the experimental results permitted us to distinguish three basic groups of parachute jumpers, differing in terms of acoustic variations and psycholinguistic and behavioral characteristics.

The first group contained beginning parachute jumpers, the second contained persons with some jumping experience, and the third was made up of masters-class parachute jumpers.

The largest changes in characteristics of speech behavior and thinking were observed in reports provided by beginning parachute jumpers.

The reports were predominantly of assertive or descriptive content, and their expressiveness was limited to lexical resources. The reports were distinguished by brevity, by fragmentation coupled with digressions from the subject at hand, and by loose composition; syntactic relationships were violated.

Speech often included "feeler" words such as "eto," "kakoy-to," "etot samyy," "takoy" and so on, "parasitic" words such as "vot," "znachit" and "tak", and extended vowel sounds such as "e-e-e." We often encountered uncorrected slips of the tongue and instances of paraphasia. In verbal material there was an increase in the number of verbs and syllables having clear positive and negative connotations, such as "very," "completely," "none" and so on.

The psycholinguistic indicators of the speech materials representing the second group exhibited rather distinct differences from the indicators for the first group, to include broader use of creative and esthetic resources of oral speech. Rich content was noted in the communications, which possessed a descriptive, assertive and a demonstrated emotional orientation.

The speech behavior of parachute jumpers in the third group was stable and individualistic, and it exhibited a tendency toward a literary, creative poetic style. Originality in discussion of the subject and a somewhat melodramatic manner of speaking were noted in their reports.

In terms of the latent period of associative reactions, relative to background values the beginning group was distinguished by its shortening (by 20 percent on the average) and by its growth immediately after landing (by an average of 15 percent).

These changes were not as highly expressed in the second group of parachute jumpers, who had some jumping experience, remaining close to the typical reactions of the individual.

An increase in the latent period prior to jumping and a significant decline following the jump were observed among experienced parachutists (±40 percent).

It was established that pauses between phrases have the greatest influence on variability in the temporal characteristics of the report. Growth in emotional tension was accompanied in most cases by a 12-36 percent increase in the frequency of the fundamental tone and an additional 2-6 db increase in the loudness of speech.

Table 1 shows changes in the energy ratios for the third-octave spectrum of the word "gotov" [ready] uttered by a jumper with little experience in different stages of the jump. We can see from the table that the frequency of the fundamental tone assumes its largest value (160-250 Hz) immediately prior to the jump. Changes in the frequency of the first formant (400-630 Hz) were less expressive.

In the parachuting stage experienced parachute jumpers managed to utter up to 140-180 syllables in their reports, while beginning jumpers uttered 80-110 syllables. When the situation in the air was complex (or when the reserve parachute had to be opened), the reports given by beginning jumpers were highly abbreviated, or the jumpers refused to make their reports.

An analysis of the reports showed that during free-fall, speech intensity increases by an additional 4-12 db. Speech often assumes the attributes of screaming, and as a consequence it becomes less intelligible. In these cases the increase in loudness of speech is accompanied by growth in speech tempo (by 12-24 percent) and in the mean frequency of the fundamental tone (by 35-95 Hz).

Table 1. Third-Octave Values for the Word "Gotov" Uttered at Different Stages of a Parachute Jump: F_i--average values of third-octave filters, 0--level of the spectral components of the syllable "go", 0′--level of the syllable "tov", in decibels; boxes indicate frequency ranges of the fundamental tone and the first formant

Parachute jumper A. N.
Word: "gotov"

Fi (гц)	На земле	На взлете	На взлете	На боевом курсе	На боевом курсе	Перед прыжком	Перед прыжком
	0—0′	0—0′	0—0′	0—0′	0—0′	0—0′	0—0′
80							
100	27,5—29	18—23	20—20	20—20	20—20	21—21	21—21
125	37,5—38	24—24	24—24	28—28	28—28	30—30	28—31
160	37—38	30—35	34—32	34—41	37—38	34—30	36—36
200	23—24	33—34	34—34	40—39	40—40	35—39	38—40
250	26—34	21—27	27—21	38—45	38—39	40—40	37—39
315	34—32	24—28	28—28	36—30	26—36	28—36	30—34
400	29—31,5	30—31	30—31	31—36	36—35	30—36	32—34
500	39—35	36—34	34—33	41—39	42—40	42—44	41—41
630	25—32	35—32	29—27	38—37	35—40	37—39	38—36
800	16—23	28—25	20—18	35—32	33—28	32—37	32—26
1000	10—21	26—23	18—17	35—32	33—26	33—35	32—30

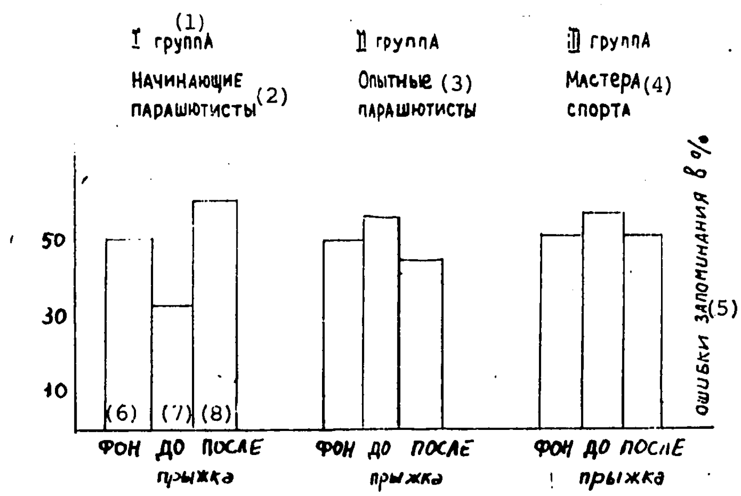

Figure 2. Memorization (Percent) of a Sequence of Six Two-Digit Numbers by Parachute Jumpers Exhibiting Different Degrees of Proficiency (Averages for Groups of Seven Persons)

Key:

1. Group
2. Beginning parachute jumpers
3. Experienced parachute jumpers
4. Masters of sports
5. Memorization errors, %
6. Background
7. Before jump
8. After jump

These changes in a number of acoustic parameters of speech are caused not only by high emotional tension but also by the influence of the intense noise created by currents of air about the free-falling jumper.

Significant changes in the cardiac rhythm of all parachute jumpers were recorded in the course of parachute jumps. However, they were most pronounced among beginners. Thus while the pulse rate of all jumpers did not exceed 74 beats per minute during a medical examination prior to the jump, after the jump (within the first minute after touchdown) the group average for beginning jumpers was 121 beats per minute, it was 104 beats per minute for the group having some experience, and it was 97 beats per minute for experienced jumpers.

Figure 1 shows the results achieved by different groups of parachute jumpers in a test requiring them to memorize a sequence of six two-digit numbers. We can see that emotional stress influences performance in this test in different ways: Stress has an intense mobilizing action upon beginners, raising their self-control and attention, as is expressed by the decrease in the percentage of errors made in memorizing numbers before the jump. Such a change was not typical of the experienced parachute jumpers taking this test, indicating that the emotional situation surrounding the jump had less influence on them.

Thus analysis of the results shows that the characteristics of speech behavior bear valuable information on the dynamics of mental state associated with different degrees of influence by emotional stress.

THE ROLE OF VERBAL PRESENTATION OF MATERIAL IN ONE METHOD OF DIAGNOSING THE EMOTIONAL CHARACTERISTICS OF AN INDIVIDUAL

I. A. Popova

The object of our study was the qualitative modal characteristics of stable human emotional characteristics, interpreted as the principal characteristics expressing the essence of emotional experience (4). Within the great diversity of such parameters, we can isolate three primary (basal) emotions: joy (J), anger (A) and fear (F). Modal parameters are manifested most clearly in individuals for whom one of these emotions is dominant (5).

The scarcity of methods for studying emotions provides motivation for seeking new approaches. We believe there is promise in developing not only "direct" techniques addressed specifically toward emotional phenomena, but also "indirect" techniques aimed at processes which cause emotional expression. One such "indirect" technique of diagnosing emotional characteristics is that of studying some aspects of color perception (color preference in particular), since the latter is closely associated with emotions (3,6,7,9, etc.). Persons differing in the mode of their dominant emotions share exhibit differences in color evaluation.

Many factors influence this evaluation: associations, the symbolic significance of color, the psychophysiological background of the organism, the type of "approach" taken to the evaluation, density, dimensions, materials, the object producing the color stimulus and so on (1,8 etc.). Obviously the means by which the color is presented for evaluation--verbally or by means of a color standard--significantly influences the results of the experiment. Both methods have their merits and shortcomings. When color standards are presented, the stimulus material is the same for all subjects--that is, they all evaluate the same colors. However, the density, dimensions and material of the standard become factors of influence. Preference can be shown only to the particular shades of color shown, which significantly complicates interpretation of the results. With verbal presentation of the color, the boundaries of the object of evaluation expand. Words represent a generalized impression of a color, ones which may be different for different people. Factors of the stimulus such as density, dimensions and so on are eliminated in this case.

Our objective was to study and compare two methods of presenting colors (verbal and standard) in regard to their differentiating capacity--that is, in regard to how well they divide subjects into groups differing in the modality of the dominant emotion.

Research Methods

The strategy of the work was as follows: Using "direct" methods on each subject, reveal the relationship between emotions of different modalities, combine the subjects into groups with dominant emotions of the same modality and compare these groups in relation to color preference (using the two methods of color presentation).

Seventy subjects--students of both sexes--participated in the experiment. They were 18-25 years old. Following analysis by the "direct" method, we distinguished the groups Jaf (seven persons), jAf (ten persons) and jaF (five persons), containing persons exhibiting dominance of an emotion of the same modality.* These groups were compared in pairs in relation to color preference (determined by different methods) using van der Warden's X-test.

Methods of Studying Color Preference

Verbal Variants

1. Numerical Ranking

The subjects were shown the names of eight colors (red, blue, yellow, green, gray, violet, brown, black), and they had to place a number from 1 to 8 beneath each color in order of preference (1--greatest preference, 2--lower preference and so on). A score corresponding to the number chosen by the subjects was awarded to each color.

2. SD (Semantic Differential) Method (Modification)

Preliminary research showed that Osgood's method of the semantic differential can be used with "abbreviated" scales to organize color preference. Four factor evaluation scales were employed. The experimental procedure required the subjects to evaluate eight colors on the basis of four scales in accordance with the instructions provided.

The results are represented in the form of a numerical value for each evaluated color (calculated as the mean score for all scales).

3. Lyusher's Test (Modification)

The experimental situation was as in the original test. Eight white cards with the names of colors written on them were placed before the subject. The procedure was repeated two times. The scale of values ranged from 8 to 1 in order of preference (the most preferable color received 8 points).

Variants in Which Color Standards Were Presented

1. SD Method (Modification)

The experimental procedure was as in verbal presentation, except that here the subject evaluated color standards similar to those used in Lyusher's test.

*Here and subsequently, upper case letters designate high values while lower case letters designate low values.

The subject had to arrange the same color standards in order of preference.

Results and Discussion

Groups of subjects differing in the modality of their dominating emotions were compared. Differentiation of the subjects with respect to their dominant emotions tended to be better with verbal description of the color than with perception of the standards. This was true for all colors. The one exception was the SD method: Not one of the variants revealed significant differences between the subjects.

Let us examine the possible causes of this fact. Significant differences between groups would confirm the possibility of using the color preference method as an indicator of emotional characteristics.

The differences between the verbal and standard methods prove that the way in which colors are presented for evaluation is a significant parameter of the evaluation--that is, it is not color preference in general but only certain of its attributes that are associated with emotional characteristics.

Absence of significant differences between subjects shown color standards may mean that on its own, the tone of a color is not informative enough for our purposes. Either color tone plays too minor a role in color sensation (for example it may be responsible for perception of a color as warm or cold), or it may exist in an indirect, more-complex relationship with emotionality, fitting within some integral characteristic of color including the other two components of color sensation--saturation and brightness.

When a color is named verbally to an individual, he evaluates his own general impression of the color for which the word stands. This general impression must be based not only on the tone of the color but also on the other components of the color sensation. Differences in emotionality revealed by this method of color presentation may be a consequence of the fact that these components make a more significant contribution to the general expression of the color. People are able to distinguish between different shades of a given color resulting from variations in its saturation and brightness.

The fact that no differences were revealed between the groups by the SD method is an important result.

We selected this method for color evaluation because it is believed that it may be used for fully objective experimental analysis of the evaluational component of pragmatic meaning--that is, because it is believed that the SD method can be used to determine a subject's attitude toward stimuli presented to him (2). For our evaluation purposes, however, the method turned out to be lacking. This fact requires further content analysis; at the moment we can only speculate on its possible causes. Apparently the evaluation scales (especially the "clean-dirty" scale) cannot be used for subjective color evaluation--that is, in this case they have somewhat different content. Moreover in the other methods all of the colors are presented together, while in the SD method the evaluations are made successively, which precludes comparison. All of this hinders the use of this method for our purposes.

Conclusions

1. Emotional characteristics can be differentiated better with the verbal variant of color preference methods than those in which color standards are presented. This allowed us to suggest the hypothesis that different components of color sensation play a role in color perception, and that they are associated with emotions.

2. The SD method is not suited to determination of color preference with the purpose of using it subsequently to diagnose emotional characteristics.

3. The results require further experimental testing and theoretical validation.

BIBLIOGRAPHY

1. Alekseyev, S. S., Teplov, B. M., and Shevarev, P. A., "Tsvet v arkhitekture" [Color in Architecture], Moscow-Leningrad, 1934.

2. Apresyan, Yu. D., "Modern Methods of Meaning Analysis," in "Problemy strukturnoy lingvistiki" [Problems in Structural Linguistics], Moscow, AN SSSR, 1963.

3. Dorofeyeva, E. T., "Changes in Color Sensitivity as an Indicator of Emotional States," in "Psikhicheskiye zabolevaniya" [Mental Illnesses], Leningrad, Meditsina, 1970.

4. Ol'shannikova, A. Ye., "Relationship of Some Adolescent Emotional Features to Physiological Characteristics," in "Problemy differentsial'noy psikhofiziologii" [Problems in Differential Psychophysiology], Vol 9, Moscow, 1977.

5. Ol'shannikova, A. Ye., and Rabinovich, L. A., "Some Individual Characteristics of Emotionality," VOPROSY PSIKHOLOGII, No 3, 1974.

6. Friling, G., and Auer, K., "Chelovek--tsvet--prostranstvo" [Man--Color--Space], Moscow, 1973.

7. Khausten, R. A., "Svet i tsvet" [Light and Color], Moscow-Leningrad, 1926.

8. "Tsvet v proizvodstvennoy srede" [Color in the Production Environment], Moscow, 1967.

9. Shvarts, L. A., "Change in Color Sensitivity in an Emotional State," in "Problemy fiziologicheskoy optiki" [The Problem of Physiological Optics], Vol 6, 1948.

COMPONENTS OF THE TEMPORAL EMOTIVE CHARACTERISTIC

R. K. Potapova

Research on the evolution of human language and speech has shown that a large number of factors have influenced the origin and development of particular features of phonation and articulation, among which biophonetic factors occupy one of the central places (1,2). According to the biophonetic conception the syllable is the elementary unit of spoken communication. On this basis it would seem right to suggest that changes affecting the temporal organization of a syllable and a train of syllables in a spoken statement may be a sufficiently informative characteristic by which to establish the particular emotional state to which the spoken signal under analysis belongs.

Our study was exploratory in nature. The end goal was to reveal the components of the temporal emotive characteristic (TEC) which would permit us to differentiate spoken statements in terms of the polarity "normal-abnormal." The TEC is defined in this case as a set of temporal components (3).

According to data acquired by objective and subjective analysis of emotionally colored speech, the TEC may be included among the basic emotive characteristics. When compared to other characteristics of this sort, the TEC is found to be a fully reliable objective characteristic; however, the main difficulty lies in revealing the concrete set of temporal parameters necessary and sufficient to the task at hand.

Our research made use of lexically and syntactically heterogeneous experimental material. Emotional states were simulated: a) in a hypnotic state; b) by an actor's performance. Experimental material of sufficiently representative volume was tape-recorded in the field and subjected to instrumental analysis following a particular program. The temporal data processing program entailed a certain number of parameters ($n = 25$).

The experimental material was analyzed with the purpose of revealing differences in relation to the polarity "normal-abnormal" (the definition of "abnormal" was previously determined by subjective analysis). Some parameters were found to be stably informative, while the informativeness of others was unstable.

By examining the syllable as the material basis of the speech code, we were able to reveal a certain set of temporal parameters that could be used to describe deformation of the temporal contour of a statement. A limited number of parameters were found to be the most informative ($n = 3$).

One of the principal informative parameters is the temporal relationship between consonants and vocal components of the speech segment under analysis: the tonic syllable in the phrase (word group); the rhythmic structure containing the tonic syllable; the phrase as a whole. The contrast between vocal and consonant time, which is characteristic of a number of emotional states indicated by speech, is sufficiently regular. This also may be explained from the standpoint of the bio-phonetic conception--by change in muscle tension, by constriction or dilation of the throat during articulation, by forced or sparing voice formation and so on (1).

The results showed that vocalism is dominant in spoken statements associated with some emotional states (for example the polarities "normal-sorrow" and "normal-melancholy") while consonantism is dominant in spoken statements associated with other emotional states (for example the polarities "normal-anger," "normal-wrath" and "normal-fear"). The TEC labels the second member of each of these polarities.

In a number of cases when an emotional state is compared with normal, changes in tempo provide adequate information. Such changes are described roughly by average syllable duration and more precisely by average sound duration, which is in this case a correlate of the rate of articulation. Analysis of tempo characteristics confirmed the notion that this parameter is informative. Moreover, changes in tempo had a dual nature, being positive or negative in comparison with normal tempo. The contrasts "normal-fear" (0-) and "normal-melancholy (sorrow)" (0+) were rather consistent in this regard.

Informative parameters also include temporal differences between vowels of the main rhythmical structure of a phrase (word group)--that is, of the rhythmical structure containing the principal stressed syllable of a phrase (word group). Thus for example a tendency toward isochronous stressed and unstressed vocalism was observed in spoken statements associated with anger (wrath). An opposite tendency toward temporal contrast in stressed and unstressed (especially posttonic) vowels was observed for the temporal contour of rhythmical structure in phrases uttered in a state of melancholy or sorrow.

We were able to confirm a number of our observations by studying the temporal structure of musical works. In particular musical speech confirms the validity of the conclusion that when states such as sorrow and melancholy are expressed, the duration of the vocal component of a syllable undergoes relative lengthening. If we turn to the works of M. P. Mussorgsky, who achieved astounding mastery in the transmission of the intonational wealth of human speech, we could illustrate our observation with the following example. In his "Folk Paintings," not only the metric aspect of the text but also the temporal organization of the phrases is indicative. For example in "The Orphan" Mussorgsky emphasized stressed syllables by increasing their temporal duration. By combining the number of words in the musical phrase (n = 3) with this increase in the duration of the first stressed syllable, he was able to transmit the intonations associated with grief, with sorrowful lamentation (4).

It should be noted that the weight of each component of the temporal emotive characteristic (TEC) may be verified with sufficient reliability by an analysis-synthesis-analysis method using musical resources. Naturally, the temporal parameters of the TEC would appear in a work of music as components of--in addition to the melody--timbre, its dynamic characteristics and the nature of articulation--that is, the means by which the sounds are elicited (staccato, legato and so on) (5). Preliminary experiments in this direction demonstrated broad possibilities for verifying the significance of TEC components.

BIBLIOGRAPHY

1. Trojan, F., "Biophonetik" Mannheim (Wien) Zurich, 1975.

2. Lieberman, Ph., "On the Origins of Language. An Introduction to the Evolution of Human Speech," New York (London), 1975.

3. Potapova, R. K., "Temporal Organization of the Syllable as the Basic Component of the Temporal Emotive Characteristic," in "Tezisy dokladov ARSO-9" [Abstracts of Reports at the ARSO-9], Minsk, 1976.

4. "Russkaya muzykal'naya literatura" [Russian Musical Literature], Issue 2, Leningrad, 1975.

5. Tyulin, Yu. N., et al., "Muzykal'naya forma" [Musical Form], Moscow, 1974.

A METHOD OF SUBJECTIVE ANALYSIS OF EMOTIONALLY COLORED SPEECH

L. V. Stat'yeva

The objective of our research was to find methods of auditory analysis of experimental material consisting of phrases spoken in different emotional states by actors and by hypnotized subjects.

An experiment consisting of four stages was performed in order to study the forms of expression of emotional states, and their identification by the method of subjective definition of emotions.

On comparing the subjective data acquired in the first stage of the experiment, we found it valid to group the emotionally colored phrases presented for evaluation as positive-negative and strong-weak; after this, the stimuli were distributed in relation to their saturation (by means of a ranking procedure).

Subjective evaluations of emotional states produced a list of state names that could be distributed into three classes: class I names reflecting the physical state of the subject resulting from a) inhibition and b) arousal; class II names reflecting the moral and mental state of the subject and modal elements, and being a) diffuse or b) compact by nature; class III names--emotional states specifically. The latter were found to fall into eight zones, each of which is represented by a general name (fear, anger, joy, sorrow, grief, irony, resentment, tenderness) and contains up to 10 names.

That the acoustic analysis method we used was effective was confirmed by objective analysis of emotionally colored spoken signals.

However, some expressions of the same emotional state by different speakers were identified differently by listeners--either as negative or as neutral, while some listeners believed them to be positive states. There was never any doubt as to normal pronunciation by all speakers; the associated state was designated as neutral.

Analysis of the results of the experiment described above led to the hypothesis that a number of the listeners categorized emotionally colored spoken stimuli not in relation to their negativeness, positiveness and neutrality but rather in relation to the degree of their expression--that is, strong-weak emotions. Therefore in the next stage of auditory analysis we felt it suitable to ask the group of listeners (the same ones) to determine the expressiveness of the signal they perceive--that is, they were asked whether or not the given stimulus could be categorized as a 1) normal statement or as a 2) strongly or 3) weakly expressed emotion.

Analysis of the experimental data made it obvious that emotional states such as sorrow and melancholy were perceived definitely as weakly expressed emotions. The following emotional states were strongly expressed: fear, horror, anger, wrath, joy and ecstasy. Ambiguous definition of emotions such as anxiety, tenderness, pleasure and love (the last three were interpreted by actors)--that is, association of the given stimulus with all three of the factors described above simultaneously--does not contradict the results of the previous experiment, and it affords the ground for suggesting that at the perceptual level, individual expressions of these emotional states are found to differ in terms of their expressiveness.

We compared the subjective results of listener analysis performed in the stages of the experiment described above; all emotionally colored phrases (except normal) broke down into groups of positive strong and positive weak emotions, and negative strong and negative weak emotions.

As a consequence it was deemed suitable to conduct an experiment to arrive at a subjective distribution of presented stimuli based on the degree of their saturation. These stimuli were distributed within the obtained groups of emotional expressions, wherever the quantity of experimental phrases was sufficient to permit this. The speakers were asked to rank a given quantity of emotionally colored phrases within the given group, for example the "negative strong" group, in relation to their strength. This made it possible to reveal the maximally/minimally saturated phrases and the stimuli falling in a middle position, and to determine equally saturated phrases. However, when the number of components was increased to $n = 6$ within a given group, the gradation was less stable, meaning that evaluations of emotional saturation are subjective.

While in the previous stages of the experiment attention was focused mainly on the way emotional states were expressed, the goal of the concluding state was identification of these states--naming emotionally colored speech stimuli on the basis of subjective definitions of emotion. This is a so-called open list method.

A list of names which listeners attached to a particular emotionally colored phrase was obtained in the course of listener analysis.

Together with data contained in the special literature (1-4), the results of logical and semantic analysis of our data made it possible to break down the entire list of names into three classes: class I names reflecting the moral and mental state of the subject and modality elements; class III representing the emotional states specifically. Classes I and II were differentiated into subclasses. Class I contained names reflecting physical states brought on by inhibition (fatigue, heaviness, weakness, sluggishness, inertness, depression, exhaustion, sleepiness) and arousal (agitation, nervousness). Class II contained names indicating diffuseness (confusion, uncertainty, pensiveness, sadness, indifference, trustfulness, compassion, inhibition, depression, hopelessness) and compactness (confidence, categoricalness, emotional enthusiasm).

Eight zones of emotional states were revealed in class III--fear, anger, joy, sorrow, grief, irony, insult, tenderness. Each zone contained a number of components (from two to nine names). Each of the eight zones could be represented by a general name that logically embraces all names within the particular semantic zone. The principle of growing emotional saturation is observed within each zone. Here is a list of the names contained in the zones named above:

1) anxiety, fright, fear, horror;
2) irritation, displeasure, anger, spite, perturbation, wrath, hatred, menace, sadness;
3) satisfaction, joy, ecstasy, delight, happiness;
4) sorrow, grief, melancholy, despair;
5) sympathy, distress, bitterness, grief;
6) derision, irony, taunting, malevolence;
7) resentment, disappointment;
8) tenderness, love.

In keeping with the objectives of this experiment, the results of listener analysis were processed with the purpose of revealing, for each emotionally colored phrase, the percentage distribution of state names in each class into subclasses and their components, and within each zone.

It should be noted that in most cases the names attached by listeners to weakly expressed emotional states, for example sorrow and melancholy, had to do more with a certain physical state of the speaker than with the emotional state expressed in the phrase. This confirms the hypothesis that weakly expressed emotional states are associated at the perceptual level primarily with the subject's physical state.

Identification of maximally saturated emotionally colored phrases was most consistent in that as a rule, spoken stimuli expressing anger, wrath, fear, joy and ecstasy were given the specific names of emotional states, or much more rarely, in cases where the listener found it difficult to classify an emotion, they were given the name of a physical state, namely arousal.

The list of names proposed by listeners for individual expressions of negative emotions such as anger, wrath and fear is somewhat longer than the list of names of positive emotions, and it embraces a larger number of zones, to include the joy zone, which does not contradict data obtained in the first stage of the research, in which listeners established them as having a positive nature.

The effectiveness of using the listener analysis method was confirmed by objective analysis of emotionally colored spoken signals.

BIBLIOGRAPHY

1. Vitt, N. V., "Expression of Emotional States in Speech Intonation," Candidate Dissertation, Moscow, 1965.

2. Rozhkova, G. I., "Perception of Emotional Intonations," Candidate Dissertation, Moscow, 1972.

3. Broadbent, D. E., and Gregory, M., "Perception of Emotionally Toned Words," NATURE, London, Vol 215, 1967, pp 581-584.

4. Fonagy, J., and Magdics, "Emotional Patterns in Intonation and Music," ZETSCHRIFT FUR PHONETICK, Berlin, Vol 16, No 1/3, 1963.

A METHOD FOR DESCRIBING INDIVIDUAL PROPERTIES OF VOICES EMPLOYING ANALYSIS OF SPOKEN SIGNAL SPECTRUMS AND BANDS

V. D. Serdyukov

Our main objective was to develop a description of the individual properties of voices based on analysis of the spoken signal spectrum and band. Such a description would be an aid to automatic identification of speech sounds and automatic voice recognition.

According to the idealized model of the acoustic theory of speech formation, the amplitude spectrum of the i-th sonorous sound has the following form:

$$P_i(\omega) = U(\omega) H_i(\omega) R(\omega) \qquad (1)$$

where $U(\omega)$--amplitude-frequency characteristic of the source; $H_i(\omega)$--amplitude-frequency characteristic of the speech tract; $R(\omega)$--emission characteristic (1).

A transformation of expression (1) is described in reference (1). This transformation combines all frequency functions that do not change from one sound to the next but which experience individual deviations:

$$U(\omega) R(\omega) = P_0 \frac{\omega/100}{(1+\omega/100)^2} \qquad (2)$$

where P_0 is a constant factor dependent on the particular acoustic pressure level. Under otherwise equal conditions it depends on the speaker. Taken together, the isolated source spectrum and emission characteristics make up a spectrum having an envelope that declines by about 6 db/octave (1).

To consider individual deviations from this rule, we rewrite expression (2) as:

$$U_k(\omega) R(\omega) = \alpha_{ku}(\omega) U(\omega) R(\omega), \qquad (3)$$

which is valid for frequencies ω for which the condition $U(\omega)R(\omega)>0$ is satisfied; here, $U_k(\omega)$ is the source spectrum of the k-th speaker, and $\alpha_{ku}(\omega)$ is the amplitude-frequency characteristic accounting for individual features of the source spectrum (including acoustic pressure) in comparison with the same characteristics of an "ideal" speaker. An "ideal" speaker is one having certain speaking habits typical of the given speech environment and being characterized by specific anatomical constants--averages for example. The ideal voice is programmed into the recognition device as a system of standard images stored in its memory.

The amplitude-frequency characteristic of the speech tract of an ideal speaker may be written, in correspondence with reference (1), as:

$$H_i(\omega) = r_3(\omega) \prod_{j=1}^{3} H_{ij}(\omega) \qquad (4)$$

where $H_{ij}(\omega)$--resonance characteristics corresponding to the first three formants of the i-th sonorous sound; $r_3(\omega)$--correction factor accounting for the influence of all formants above the third on the trend of the frequency characteristic in lower frequencies, and dependent not only on frequency ω but also on the total length of the speaker's speech tract.

To account for individual deviations in factor $r_3(\omega)$, we introduced the term $\alpha_{kr}(\omega)$; to account for individual features of the resonance frequency characteristics of the k-th speaker in relation to some ideal frequency characteristics, we introduced the term $\alpha_{kHi}(\omega)$. Then

$$H_{ki}(\omega) = \alpha_{kHi}(\omega)\, \alpha_{kr}(\omega)\, H_i(\omega) \qquad (5)$$

which is valid for frequencies ω satisfying the condition $H_i(\omega) > 0$.

(1), (3), and (5) give us an expression associating the amplitude spectrums of the i-th sonorous sound in the pronunciation of the ideal and the k-th speakers:

$$\begin{aligned} P_{ki}(\omega) &= \alpha_{ku}(\omega)\, \alpha_{kr}(\omega)\, \alpha_{kHi}(\omega)\, P_i(\omega) = \\ &= \alpha_k(\omega)\, \alpha_{kHi}(\omega)\, P_i(\omega) \end{aligned} \qquad (6)$$

which is valid for frequencies ω satisfying the condition $P_i(\omega) > 0$; $\alpha_k(\omega)$--amplitude-frequency characteristic reflecting differences between the voices of the k-th and the ideal speakers at frequency ω; it does not depend on the phonetic quality of the sonorous sound uttered:

$$\alpha_k(\omega) = \alpha_{ku}(\omega)\, \alpha_{kr}(\omega).$$

If we assume that voice differences depend mainly on parameters of the excitation source and the higher formants, we could write expression (6) in the form:

$$P_{ki}(\omega) \approx \alpha_k(\omega)\, P_i(\omega) \qquad (7)$$

It follows from (7) that the individuality of the speaker's voice may be described in the first approximation by changes in the amplitude-frequency characteristic of the communication channel.

A simple rule for describing the individual features of the voice of the k-th speaker at frequency ω follows from this same expression:

$$\alpha_k(\omega) \approx P_{ki}(\omega)/P_i(\omega) \qquad (8)$$

By subjecting speech to spectrum-band analysis using a comb of n filters, we can approximate the characteristic $\alpha_k(\omega)$ by the stepped function $\alpha_k(j)$, $j=1,2,\ldots,n$, the values of which are determined according to the rule:

$$\alpha_k(j) = P_{kij}/P_{ij} \qquad (9)$$

where P_{kij} and P_{ij} are the mean weighted spectral amplitudes of the learning patterns of the i-th sound in the base frequency band of the statements uttered by the k-th and the ideal speakers.

Considering deviations from rule (7), we can describe the individual properties of the voices by minimizing a functional of the form:

$$\sum_{i=1}^{I_0} \sum_{j=1}^{n} [P_{kij} - \alpha_K(j) P_{ij}]^2 = min \qquad (10)$$

where $I_0 \leqslant I$, and where I is the number of sonorous sounds.

The expression for determining $\alpha_k(j)$, obtained as a result of coordinate-by-coordinate minimization, appears as follows:

$$\alpha_K(j) = \sum_{i=1}^{I_0} P_{Kij} / \sum_{i=1}^{I_0} P_{ij} \qquad (11)$$

which is valid on the condition that

$$\frac{1}{I_0^2} \sum_{i=1}^{I_0} P_{Kij} \sum_{i=1}^{I_0} P_{ij} >> 1.$$

To evaluate the effectiveness of the equations we developed to represent information on voice individuality, we conducted a number of experiments establishing the conditions under which these equations are the most effective.

In particular we tested the effectiveness of normalizing individual differences among speakers according to the rule:

$$x^{H}_{\kappa ij} = \alpha_{\kappa}(j) x_{ij}, \; i = 1, 2, \ldots, l; \; j = 1, 2 \ldots, n,$$

where the values of $\alpha_k(j)$ were determined by expressions (9) and (11) at $I_0 = I$; $x_i = \{x_{ij}\} \varepsilon X$, where X--the recognition unit's set of standards; $x^{H}_{\kappa i} = \{x^{H}_{\kappa ij}\} \varepsilon X_{\kappa}$, $X^{H}_{H\kappa}$--set of sound standards corrected for the individuality of the k-th speaker.

The criterion for the effectiveness with which voice individuality was accounted for was:

$$S = (P_{H} - P)/(P_{c} - P),$$

where P_H, P and P_c--reliability of recognizing a standard test sample, correspondingly with amplitude-frequency correction of the standard sound patterns on the basis of one of the previously described rules (P_H), and using ideal standards (P) and individual sound standards (P_c). The patterns of identical sounds averaged for all of the speakers were used as the ideal standards.

The conditions of spectrum and band analysis were varied in the experiments. In the first experiment sounds in prescreened samples of sonorous speech were isolated as instantaneous spectral cross sections. Each spectral section was described by 12 readings taken off the envelope of the amplitude spectrum at 400 Hz intervals (B = 100 Hz, where B is the passband width of filters in the spectrum analyzer).

The experiment was run with 840 utterances of stressed and unstressed vowels and sonants (м, м', л, л', р, р', н, н') in the speech of five speakers. The test sample contained 690 utterances. The decision as to effectiveness was made on the basis of the minimum Euclidean distance from the standard patterns to the test utterances. Rule (11) was found to be best for determining the values of function $\alpha_k(j)$: S = 0.9, P_H = 68 percent, P_c = 69 percent; rule (9) was also found to be extremely effective, though only in relation to stressed vowel sounds--the sound "И" in our experiment.

This method of accounting for IRG [not further identified] using rule (11) was also tested in relation to vowel sounds in spectrum and band analysis of a signal using a comb of 20 band filters with B = 250 Hz, uniformly covering a range from 200 Hz to 4 kHz. The results of recognizing 336 control utterances in statements made by eight male speakers were as follows:

S = 0.78, P_H = 83.6%; P_c = 84.5%.

The third experiment was based on 2,400 utterances of stressed vowel sounds pronounced by 10 speakers; of these, 1,680 were test utterances. The spoken signal was subjected to spectrum and band analysis in the 180-8,200 Hz range using a comb of six filters having an average passband width B = 1,350 Hz.

This method of accounting for IRC using rule (11) produced the following results: $S = 0.3$, $P_H = 79\%$, $P_C = 93\%$. The experiments showed that as the passbands are widened, the work of the models worsens; correspondingly, change in the speaker is adequate to change in the amplitude-frequency characteristics of the communication channel in the first approximation. Our method of expressing the individuality of a speaker's voice was also tested with an SARG automatic voice recognition system (2). We found that by describing the voice as a stepped function $\alpha_k(j)$ we can significantly raise the reliability of recognizing a standard sample of 60 key phrases uttered by 10 male speakers (by 10 percent in the case of the piecewise decision making rule, and by 20 percent in the case of the linear decision making rule).

The experiments demonstrated the effectiveness of the proposed method of expressing the integral properties of voices in relative form, as compared to their absolute description, in spectral and band analysis of speech.

BIBLIOGRAPHY

1. Fant, G., "Akusticheskaya teoriya rechobrazovaniya" [Acoustic Theory of Speech Formation], Izd-vo "Nauka", Moscow, 1964.

2. Ramishvili, G. S., Serdyukov, V. D., and Tushishvili, M. A., "Experiments in Computer Voice Verification," in "Trudy VIII Vsesoyuznogo seminara 'Avtomaticheskoye raspoznavaniye slukhovykh obrazov'" [Proceedings of the Eighth All-Union Seminar "Automatic Recognition of Auditory Patterns" (ARSO-VIII)], Part 4, L'vov, 1974.

ANALYSIS OF THE VARIABILITY OF VOWEL FORMANT COMPOSITION ASSOCIATED WITH CHANGE IN PALATE SHAPE

A. I. Tarasov, A. A. Fedorov, G. P. Fisenko

While the speech forming process has characteristics common to all people, it remains deeply individualistic due to variations in the dimensions and shape of elements forming the speech tract. But it is difficult to study the influence of the geometry of tract elements upon the individual characteristics of speech due to poor accessibility to this system.

One of the methods that surmounts this difficulty is the use of a removable upper denture, which permits investigation of the influence exerted by the shape and depth of the palatine vault, the position of the teeth and so on upon formation of speech sounds.

Change in palate shape due to orthopedic interference has an obvious effect on formation of consonant sounds; however, the influence prostheses have on formation of vowels is unclear, and the available data are few in number and contradictory.

Methods of quantitative analysis of the quality of vowel sounds based on investigation of their formant structure have recently been developed. Individual voice characteristics are known to be the product of differences in the averages of the three first formants, as well as of the positions of formants above the third.

Our objective was to analyze the influence of removable dentures on formation of Russian vowel sounds.

We analyzed six Russian vowels uttered by 15 speakers. The speakers were patients of different age groups exhibiting no noticeable defects in pronunciation and making their first visit for the purpose of acquiring removable upper dentures. Some speech inadequacies associated with the loss of natural teeth were disregarded.

Simplified spoken material--the syllables ba, bu, bo, etc.--was analyzed. Speech was recorded according to the standard procedure with a "Zvuk-1" tape recorder. Each speaker uttered each syllable three times. We evaluated the dynamics of the three series of recordings. The first series was recorded before the subject was fitted for a prosthesis. Considering that the subject had no experience in using protheses, this recording was interpreted as normal, typical of the speaker due to prolonged adaptation and arisal of a stable form of articulation. The second recording was made immediately after the prostheses were manufactured so as to avoid the consequences of possible adaptation in pronunciation occurring as the speaker became accustomed to the prosthesis (that is, to the new tactile sensations), and arisal of compensatory forms of articulation. The third recording was made a month later, after the speaker adapted himself to the prosthesis. In this case we applied an additional layer of wax to the outside surface of the base of the prosthesis in the area of the middle third or the front teeth, or in the area of the side teeth, with the purpose of placing the speaker in radically unusual conditions by changing the actual resonant volume of the oral cavity.

To obtain the formant pattern of vowels, we processed the recordings with a 50-channel spectrograph connected to an M-222 digital computer. The spoken signal was processed within a constant range--75 Hz-5Khz. The entire range was broken down into 50 channels, with each channel 250 Hz apart. Following conversion to digital form, the output of each of the channels was fed to the computer. Each channel was interrogated 100 times per second. Next the computer was used to obtain spectral patterns taking the form of "visible speech"--that is, a dynamic picture of spectrum maximums.

Next the pattern of "visible speech" was analyzed: A stable vowel segment was selected out, and a spectral cross section was made through the central part of this stable segment. The results were in the form of a digital read-out and a picture of the spectral cross section. We obtained 3×6 or 18 cross sections for each speaker.

First of all, using published data on the formant structure of Russian speech, we determined if the vowels of speakers without prostheses corresponded to the obtained formant pattern. The results showed that the formant structure is within the range typical of the formant structure of Russian vowels, though individual variability in formant position was observed, within the permissible limits for each speaker. In other words on one hand the formant structure remains consistent with the informative structure of the vowel, while on the other hand it clearly reveals individual features inherent to the speech apparatus of the given individual.

Analysis of the formant structure of vowels uttered immediately after prostheses were fitted showed that disturbance of tactile habits by the introduction of prostheses which caused alteration of the acoustic signal (as well as acoustic feedback), did not cause change in the informative positions of formants inherent to the given vowel; moreover the formant positions characteristic of the speaker remained the same as well. The fact that the position of formants remained practically unchanged in comparison with the position of formants uttered without prostheses permits the suggestion that vowel formant structure is not disturbed by dentures.

Analysis of the third experiment, in which we determined the influence of admissible disturbances in the shape of the palate and alveolar processes on vowel formation, showed that in this case the formant structure does not change, and the individual formant positions characteristics of different speeches remain the same.

The results show that changes, within certain limits, in the normal anatomical structure of the palate and alveolar processes do not influence the individual properties of vowel formant structure due to the peculiar way in which vowels are formed.

SOME RESULTS OF RESEARCH ON INTONATIONAL CHARACTERISTICS OF PRINCIPAL STRESSED VOWEL SOUNDS OF EMOTIONAL SPEECH

V. L. Taubkin

Speech intonation is known to be an important means of transmitting information on the state of the speaker. Perceived subjectively as change in the pitch and intensity of the voice with respect to time, intonational parameters can also be objectively recorded and measured, which is why researchers are so interested in them.

Fundamental tone frequency is the intonational parameter which has been studied most fully thus far. The common conclusion of most works is that growth in the intensity of emotions causes corresponding growth in the mean frequency of the fundamental tone (1-5).

On the other hand development of depressive states causes a decline in the mean frequency of the fundamental tone (6-8).

The shape of the melodic curve undergoes significant changes as well. As a rule a speaker's neutral state is characterized by an insignificant range of temporal changes in the fundamental tone frequency, by smooth transitions from rises to falls and by absence of overpeaks among instantaneous frequency values. Growth in intensity of sthenic emotions usually causes an intense increase in the variance of the fundamental tone, and arisal of dissonances (overpeaks) in a melodic curve having sharply pronounced melodic rises and falls. Asthenic emotions are typified by a decrease, in comparison with the neutral state, in the variance of the melodic curve, smoothness of transitions from rises to falls, and intonational unexpressiveness (1,6,7,9).

A number of researchers have suggested that the shape of the melodic curve may be an adequate differentiating characteristic by which to recognize the emotions of the speaker, not only in terms of their intensity but also their quality (3,10).

However, it is very difficult to express the intonational characteristics of speech in formal terms with the purpose of using them to recognize the state of a speaker. The reason for this is the uncertainty contributed by the influence of the lexical-grammatical level of speech. Thus we must seek invariant intonational parameters or fragments, or we must subject the spoken text to parallel analysis.

Investigation of the intonational characteristics of the principal stressed vowel sounds of speech is of great interest in this aspect. These sounds bear significant information on the emotional coloration of speech (11-13), and they exhibit little dependence on the acoustic environment in quasisationary parts of the signal.

This communication presents some results of research on the melodic contours of the principal stressed vowel sounds of speech. Spoken material was obtained for the analysis by actor simulation. The faithfulness of the actor's experiences was monitored by electrophysiological indicators. Positive and negative emotions of varying intensity ("joy," "delight," "anxiety," "fear") were selected for analysis. Special attention was devoted to studying the behavior of speech parameters in response to intentional change in tempo and intensity of pronunciation by a speaker in a neutral state (we examined only one communicative type of sentences--narrative). Such speech distortions, which are ordinarily elicited by the real conditions of transmission--for example noises, lack of time and so on--must be differentiated from emotional speech reactions.

The experimental material was sampled and grouped in accordance with the results of electrophysiological monitoring and listener analysis. The intensity of emotions was evaluated by a group of experts using a five-point scale. The clearest emotions were awarded a score of 5, while the least expressive were awarded a score of 1; neutral states were scored from 0 to 1. Listener analysis of the states simulated in the actor experiment, which we arbitrarily designated as "anxiety"--A and "joy"--J, produced scores from 1 to 3, on the average for the whole group of listeners. The states "fear"--F and "delight"--D were characterized by average scores from 3 to 5. Utterances by speakers in a calm state speaking in a normal style--N, and with different variations in speech tempo and loudness (five groups: "dictation"--D, "rapid"--R, "normal loud"--N_L, "loud dictation"--D_L, "rapid loud"--R_L), made up separate groups. The overwhelming majority of averaged listener evaluations of these groups of speech responses were limited to values between 0 and 1. The one exception was several "loud rapid" reactions--R_L--evaluated by listeners as weak negative emotions (scores from 1 to 1.5).

The fundamental tone was isolated by an M-50 digital computer using the displacement method suggested in (14), which permits measurement of individual oscillation periods of the vocal cords.*

Our research revealed the following laws for the behavior of the melodic contours of the principal stressed vowel sounds of speech:

1. The state of rest, including cases of intentional distortion of pronunciation style (narrative sentences were used), is characterized by a practically horizontal intonational curve, with a slight downward slope toward the end of the sound; fluctuations in the instantaneous frequency values relative to the average frequency are highly insignificant, and in our cases they did not exceed 10 percent. In this case changes in speech tempo do not elicit significant changes in the average frequency of the fundamental tone ($\overline{F}_{FT}$, while an increase in pronunciation intensity causes it to rise significantly.

2. A typical feature of negative emotional states is a steep downward slope of the intonational curve from the beginning of the vowel to its end, coupled with a general

*The complete research program also included analysis of the behavior of the speech tract's first formant and the energy and duration of the principal stressed vowel. However, the results of this research are not discussed in this communication due to space limitations.

increase in $\bar{F}_{FT}$ proportional to the degree of emotional tension. The span for the slope angle of the intonational curve, when considered together with the increase in the average frequency of the fundamental tone of a given utterance of a principal stressed vowel (relative to the background frequency), may serve as a dependable correlate of emotional expressiveness.

3. Positive emotional states were also characterized in our experiments by an increase in $\bar{F}_{FT}$ proportional to the degree of emotional tension. In this case the melodic curve was characterized by a rise in the frequency of the fundamental tone in the initial phase (by up to 25-30 percent over the initial level), and by its decline at the end of the vowel. In this case the more clearly positive emotions were expressed, the higher and longer was the initial rise in frequency of the fundamental tone and the higher was the average frequency of the fundamental tone for the given utterance.

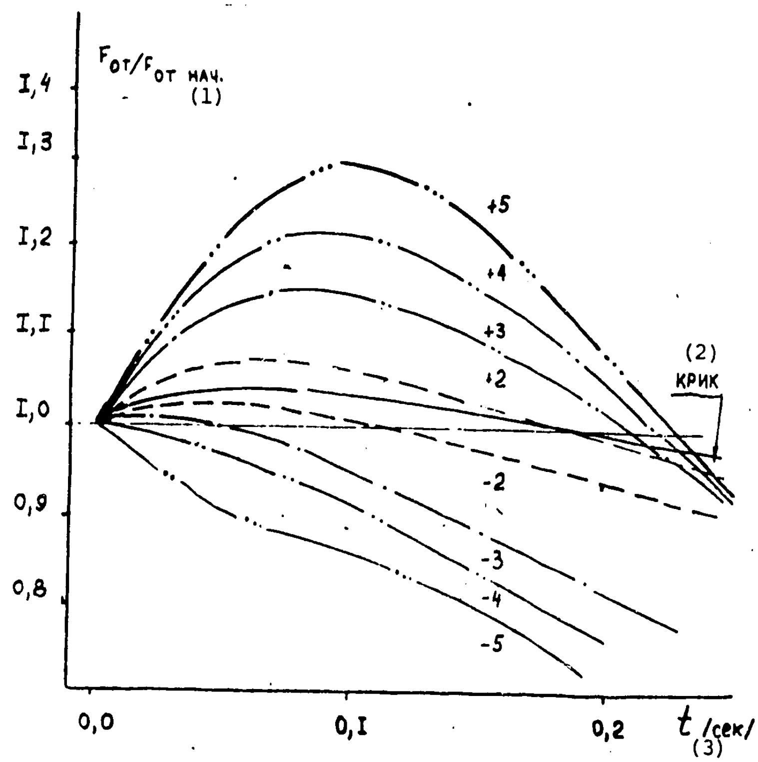

Figure 1. Example of Intonational Metrics (Obtained for $F_{FT} = 2 \pm 0.15$)

Key:
1. F_{FT}/F_{FT_i}
2. Shout
3. Seconds

Let us explain this with an example. We select, from among the available melodic contours for the main stressed vowel sound of one speaker, those for which the average frequency $\overline{F}_{FT}$ increased by about two times (2±0.15) in comparison with the state of rest--N. Normalizing, in relation to the initial frequencies, the melodic curves of selected utterances corresponding to positive and negative emotions of different intensities, as well as an unemotional shout, and superimposing all of the curves on the same figure (Figure 1), we get a graphical representation of intonational metrics suited to evaluation of the degree and sign of the emotions under analysis (given a fixed relative increment in $\overline{F}_{FT}$). We can see from the figure that transition from an emotion of one sign to an emotion of another sign occurs by way of a gradual decrease in its level (the results of listener analysis of the words from which the given sounds were isolated are indicated by the numbers next to the curves); this transition proceeds through the unemotional shout as well. We can find a zone in the figure within which the sign of emotions cannot be determined by ear (the score is below 2), though the listeners are able to discern emotional tension in the speaker's voice. Such representations of intonational metrics can also be obtained for other fixed increments in $\overline{F}_{FT}$ (to include the state of rest). Naturally the spectrum of listener evaluations would be different. It should be noted that the nature of the intonational metrics illustrated by this figure did not undergo qualitative changes when a different speaker was considered. We should also add that when we examined the melodic curves of stressed vowels occupying different positions in a word--that is, being within the first, second and subsequent syllables, we did not reveal significant changes in their form. Some differences were revealed in the melodic conclusion of the vowel of the last syllable; however, these differences (in the decline in frequency of the fundamental tone toward the end of the curve) are not fundamental. We should also note that in our examination of speech in a state of rest, we limit ourselves to just narrative phrases alone. Were we to examine the intonations of other types of communicative phrases, we should find unique differences in the behavior of the particular parameter under analysis (for example we might witness an increase in the average frequency and change in shape of the intonational curve when we analyze a questioning intonation). However, according to our preliminary data these changes would not be commensurate with emotionally dependent changes. When a speaker experiences emotional tension, the communicative characteristics of his statements fade as his degree of tension rises.

In our description of intonational metrics, it was not our objective to find the threshold of each parameter (average frequency of the fundamental tone, parameters associated with the form of the melodic curve) to be used in deciding the class to which a given utterance belongs. To find the boundaries between these classes, we would have to analyze pattern recognition in a multidimensional space of characteristics.

BIBLIOGRAPHY

1. Mourad-Krohn, H., THE JOURNAL OF MENTAL SCIENCE, Vol 103, No 9, 1957.

2. Nikonov, A. V., VOYENNO-MEDITSINSKIY ZHURNAL, No 9, 1973.

3. Galunov, B. I., and Manerov, V. Kh., "Correlation Between Psychophysiological State of the Speaker and Characteristics of the Spoken Signal," in "Tezisy dokl. VIII Vses. seminara po avtomaticheskomu raspoznavaniyu slukhovykh obrazov" [Abstracts of Reports of the Eighth All-Union Seminar on Automatic Recognition of Auditory Patterns], L'vov, 1974.

4. Frolov, M. V., and Taubkin, V. L., "Influence of the Speaker's Emotional State on Some Parameters of the Spoken Signal," in "Mater. Vses. simpoziuma 'Rech' i emotsii'" [Proceedings of the All-Union Symposium "Speech and Emotions"], Leningrad, 1975.

5. Simonov, P. V., Frolov, M. V., and Taubkin, V. L., "An Invariant Method of Recognizing the Emotional State of a Group of Speakers on the Basis of Their Speech," ZHURN. VYSSH. NERVN. DEYAT., Vol 26, No 1, 1976.

6. Moses, R., "Emotional Causes of Vocal Pathology," in "Psychological and Psychiatric Aspects of Speech and Hearing," New York, 1960.

7. Nikonov, A. V., and Popov, V. A., "Structure of the Speech of a Human Operator in Stressful Conditions," in "Materialy Vses. simpoziuma 'Rech' i emotsii'," Leningrad, 1975.

8. Bazhin, Ye. F., Galunov, V. I., Gorskiy, G. D., and Manerov, V. Kh., "Objective Diagnosis of Emotional State in the Psychiatric Clinic on the Basis of Speech," in "Materialy Vses. simpoziuma 'Rech' i emotsii'," Leningrad, 1975.

9. Kolymba, S. N., Nushikyan, E. A., and Pirogova, A. A., "Acoustic Correlates of Emotionally Colored Phrases Expressing Anger, Approval and Irony in Modern English," in "Materialy Vses. simpoziuma 'Rech' i emotsii'," Leningrad, 1975.

10. Vitt, N. V., "Expression of Emotional States in Speech Intonation," VOPROSY PSIKHOLOGII, No 3, 1965.

11. Luk'yanov, A. N., and Frolov, M. V., "Signaly sostoyaniya cheloveka-operatora" [Signals of Human Operator State], Moscow, Nauka, 1969.

12. Bondarko, L. V. et al., "Phonetic Correlates of Different Degrees of Speech Expressiveness and Emotionality," in "Materialy Vses. simpoziuma 'Rech' i emotsii'," Leningrad, 1975.

13. Nushikyan, E. A., "The Role of Tonal Characteristics in Transmission of the Speaker's Emotional State," in "Materialy Vses. simpoziuma 'Rech' i emotsii'," 1975.

14. Sobolev, V. N., and Baronin, S. P., "Investigation of the Displacement Method of Isolating the Fundamental Tone of Speech," ELEKTROSVYAZ', No 12, 1968, pp 30-36.

ANALYSIS OF ADAPTIVE MECHANISMS OF THE ARTICULATORY ORGAN CONTROL SYSTEM

A. A. Fedorov, G. P. Fisenko, A. O. Yurgenson

Being a communication source, the main objective of the speaker is to generate acoustic signals bearing certain semantic information.

In the ideal case all acoustic signals intended to communicate a fixed meaning must be identical to one another, irrespective of who the speaker is and of when he utters them. However, what we observe in fact is significant variation in spoken signals, albeit within certain limits imposed by the need for retaining semantic information. In this case we can isolate two forms of causes of such variation. Some of them are associated with different individual features of the speech forming systems of different people, while others are associated with certain differences in the way a certain individual's speech forming system functions at different moments in time due to changes in the state of the speaker or the environment. Individual differences may be produced by the most diverse causes: differences in the anatomy of the articulatory tract, in the structure of the control processes associated with speech forming organs, in the system responsible for monitoring speech formation and so on.

The speech forming mechanism is flexible, as can be seen from the fact that different articulatory configurations produce approximately the same formant frequencies. This phenomenon is known in phonetics as compensatory articulation. We know, of course, that each individual possesses sufficiently stable articulatory habits, and that they are basically the same for all other subjects of the same language group. When we say that articulation is compensatory, we mean that the similarity between two sounds, one formed normally and the other by compensatory articulation, is only approximate.

Discussing compensatory forms of articulation, we will distinguish between two types of variations in articulatory structures. The first type will include variations in the structure of different articulatory elements of different speakers. What we encounter here are differences in the anatomical structure of elements of the speech tract: the shape of the palate, the mass and shape of the tongue and the jaw structure of different people; such differences make individual "tuning" of the articulatory tract for formation of the required formant structure necessary. The second type of compensation in speech formation is associated with change in the articulatory structure of a particular speaker. If we ignore change in structure as the individual develops from child to adult, which would be a special task on its own, arisal of compensatory forms of articulation is usually associated with pathological changes in the shape of articulatory elements due to injury, surgery and so on, or due to changes in their control mechanisms--reduction of the mobility of articulatory organs, a decrease in the range of their movement and so on, which would require the subject to modify the structure if he is to produce the necessary phonemic sounds.

The accumulated experimental phonetic and logopedic material providing anatomical and physiological descriptions of articulation does not provide accurate data on compensatory forms of articulation exhibited by different speakers. What we find for the most part are averaged data on articulatory characteristics (roentgenograms and palatograms) of a number of Russian phonemes. Structural variability has not been considered, owing significantly to the methods used by the authors: Different elements of articulation characteristic of different speakers were studied separately, and then the results were combined into an overall picture of phonemic structure.

Particular information on variability of articulatory structure associated with consonants in Russian language may be obtained by comparatively analyzing the palatograms of several speakers. Though such data are significant to our understanding of the mechanisms of consonant formation, they cannot answer the question as to which elements of the articulatory tract produce the differences seen in palatograms, and how this difference influences the mutual arrangement of other elements of the tract. The shape of the hard palate is important, since together with the dental and maxillary systems it provides the foundation for articulatory movements of the tongue in speech.

It should be noted that investigations of the mechanisms of speech formation, to include compensatory forms, have rarely examined the acoustic characteristics of the spoken signal. This significantly reduces the value of such investigations, since they do not provide any more information on the acoustic signal other than a qualitative description of the phonetic sounds.

Our objective was to study the adaptive processes associated with formation of Russian consonant sounds by a speaker with a hard palate altered dramatically by artificial means. Such alteration disturbs the usual habits of articulation and causes arisal of compensatory articulatory dynamics associated with arisal of new myotactic sensations and with the consequent need for acoustic tuning.

The research was conducted on 10 speakers having no pronounced speech defect. The subjects were undergoing fitting for complete removable upper dentures.

Tape recordings were made of the speakers reading standard syllable tables.

They read the text before being fitted for the dentures, immediately after their manufacture and fitting, and following a week of adaptation to the dentures.

We also recorded texts read by speakers wearing dentures with dramatically altered shape: Adhesive wax pads were applied either to the front and side teeth or to the center of the palate (to the flat roof of the palate).

The texts read by the speakers were played back to five listeners, who wrote them down. Analysis of perception errors required plotting of matrices showing transitions of certain consonants into others. The consonant transition matrices were averaged in relation to all listeners, which made it possible to randomize the perception errors of different listeners.

In all, we obtained five consonant transition matrices for each speaker--normal, using a denture without adaptation, following adaptation to the denture and, correspondingly, alteration of the habits of articulation, and two matrices associated with disturbance of articulatory habits due to sharp distortion of the shape of the palate.

Our analysis of the transition matrices led to the important conclusion that the nature of transitions is common to all matrices, the only difference being the number of transitions, which is greater without adaptation and in the case of speech produced in the presence of the adhesive wax pads.

The largest number of transitions is associated with disturbances in the voiced-soft attribute, which may be explained by the fact that soft consonants are distinguished in Russian by a common articulatory attribute--mandatory elevation of the central ridge of the tongue toward the hard palate.

The transition from voiced to surd consonants differing only in the absence of a voice source may be explained by redistribution of the energy of the noise and voice components of the sound source due to impairment of articulation.

It should be emphasized that these are tentative data, and that dynamic electropalatography should provide a conclusive answer to the question as to what articulatory movements are responsible for compensation.

AN ALGORITHM FOR RECOGNIZING EMOTIONAL STATES OF SPEAKERS ON THE BASIS OF STRESSED VOWEL SOUNDS

M. V. Frolov, V. L. Taubkin, N. A. Luzhbin

Depending on the initial requirements, the task of monitoring a speaker's emotional state on the basis of his speech can be divided into four stages. They entail recognition of emotions on the basis of a given fragment of the speech of 1) one speaker and 2) a group of speakers, and evaluation of emotional states on the basis of any fragment of the spoken signal of 3) one speaker and 4) a group of speakers. The acoustic characteristics used in stages 1 and 2 permit partial or complete neutralization of the negative influences of the lexical-grammatical composition of speech and individual features of the speaker on the quality of recognition. Such neutralization is achieved by normalizing the parameters, by selecting invariant indicators and standards and by appropriately grouping the speakers. In distinction from the first two stages, effective completion of stages 3 and 4 would require, in addition to use of the acoustic-phonetic characteristics of the signal, consideration of the lexical-grammatical level of the communication, and possibly analysis of its semantic content. Variants of stages 1 and 2 were examined in (1-3). These works also provide the results of recognizing emotional states of speakers in real and laboratory conditions, both in the presence and in the absence of external and internal noises.

Stages 3 and 4 cannot be completed in their entirety by the methods available today. However, certain approximations may be made. They require, in particular, determination of formal characteristics of emotions that are invariant in relation to lexical-grammatical composition of speech and in relation to the individual characteristics of the speakers. Characteristics which satisfy these requirements to one degree or another were obtained from utterances of stressed vowel sounds. These sounds bear significant information on the speaker's emotional state (1), as is confirmed by studies on perception of emotional speech (4-6). In comparison with the influence of emotions, the dependence of the set of characteristics determined in these works upon the phonemic category of the vowels, the acoustic environment of the sound under analysis, the lexical-grammatical characteristics of the statement and so on was found to be low. By accounting for this fact and by introducing special acoustic standards, we were able to formulate and solve a variant of the problem of recognizing the emotional state of any speaker on the basis of an arbitrary speech fragment.

Stressed vowel sounds were extracted* from the speech of an actor simulating positive and negative emotions of different intensities, and from speech uttered in a state of

*Algorithms for automatic isolation of stressed vowel sounds are described in (8-9), and they were not analyzed in the present work.

rest. In the latter case the intensity and tempo of speech were intentionally distorted. These experiments are described in detail in (1). Material from six speakers was first subjected to listener analysis in order to permit its classification in terms of different states. Utterances which could not be classified uniquely were excluded from further analysis. The recognition algorithm was run in a YeS-1020 digital computer.

The information space consisted of eight parameters, which were determined by processing the recordings of the stressed vowel sounds. They included: 1) sound duration--T; average values (within time T), 2) fundamental tone frequency--F_{OT}, 3) power M, 4) number of times the spoken signal crossed the zero value--f_0, and 5) the dispersion of the fundamental tone frequency--$D(F_{OT})$. The i-th intervals of the fundamental tone T^i_{OT}, were isolated by the displacement method (7). In addition to the characteristics listed above (the characteristics are not presented here in the order of their importance), we used 6) asymmetry of the smoothed curve of the fundamental tone, $A(F_{OT})$:

$$A(F_{OT}) = \frac{(t_1 - t_0) - (t_2 - t_1)}{t_2 - t_0};$$

Smoothing was performed in a window of values for five samples for which $F_{OT} = (T_{OT})^{-1}$. t_0 and t_2 denote the beginning and end of $F_{OT}(t)$ on the time axis respectively, and the extreme point t_1 is such that $F_{OT}(t>t_1)<F_{OT}(t_1)$ for $t_1<t\leq t_2$, and such that the first derivative of function $F_{OT}(t_1) = 0$; 7) the average of the first derivative $F'_{1,OT}$ of function $F_{OT}(t)$ in the time (t_1-t_0):

$$\bar{F}'_{1,OT} = \frac{1}{t_1 - t_0}\int_{t_0}^{t_1} F'_{OT}(t)\,dt, \quad \text{where } \overline{F'_{OT}} = \begin{cases} \bar{F}_{1OT} & \text{at } \bar{F}'_{1OT} > 0 \\ 0 & \text{at } \bar{F}'_{1OT} \leq 0 \end{cases}$$

8) the average of the first derivative $\bar{F}_{2OT}$ of function $F_{OT}(t)$ in time (t_2-t_1):

$$\bar{F}'_{2OT} = \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} F'_{OT}(t)\,dt,$$

In this case $\bar{F}'_{2,OT} = 0$ if there is no descending segment in the smoothed curve of the fundamental tone. The parameters isolated for each speaker (except $A(F_{OT})$) were normalized in relation to their average values, obtained from material pronounced in normal style. Although some of the selected indicators (for example F_{OT}, f_0, M) exhibit certain correlations among each other, thus making it possible to reduce the parameter space, a certain amount of redundancy does raise the reliability of recognition on the background of different levels of noise, which have different effects on isolation of these indicators. In particular, it is with this purpose in mind that we suggest supplementing the parameter space by the trajectory of the first formant.

Preliminary research showed that for practical purposes the classes of states we analyzed do not intersect. Hyperplanes approximating characteristics separating the surface into an eight-dimensional space are sought in two successive stages: 1) we divide the entire set of utterances into two classes corresponding to the state of rest (accounting for distortions in speech tempo and intensity) and emotional tension, and 2) we find the hyperplane dividing the class of emotions into subclasses of positive and negative emotional states. The recognition process was organized as a hierarchical system in order to simplify the structure of the algorithm applied to our set of working characterstics. Only those solution rules that are found in each stage by means of the same procedures can be used.

With this purpose we developed the exploratory algorithm PATTERN RECOGNIZER, which can be used to determine the location and configuration of patterns and to seek possible points of their intersection in an n-dimensional space of continuous numerical characteristics. The algorithm makes use of a learning flowchart similar to that described in (10). It entails finding a dividing hyperplane W passing through the center of the segment connecting the closest points of convex envelopes stretched over points in the pattern of the learning sample. Our learning flowchart differed mainly in that it did not require complete sorting of points in the learning sample; instead, the results of intermediate iterations were subjected to analysis (points which would knowingly not participate in further learning were excluded). The parts of the algorithm and the order of their function are discussed below. The initial data--the learning sample together with the results of listener analysis and the test sample--were fed into the data bank (subroutine BANCD), in which they were subjected to primary processing and normalization. At each hierarchical level, possible intersections between patterns were verified, the area of pattern intersection was analyzed if it was present, and a decision was made on whether or not to exclude points within this area from further learning (subroutine CROSSING). The main working part of the algorithm consisted of three functional blocks: 1) the learning subroutine LEARN, which creates, on the basis of the learning sample and expert evaluations, hyperplane W separating two patterns A and B; 2) a testing subroutine TEST, which tests the quality of pattern division during learning, using the found equations for the hyperplane and the expert evaluations; 3) a recognition subroutine RECOGN, which uses the found solution rules on the test sample--that is, it correlates new material with pattern A or B. Learning and recognition were performed both in relation to each speaker taken alone and in relation to the entire group of speakers.

By applying the algorithm described above we obtained a rule for recognizing the following in relation to each speaker: a) the classes "rest" and "emotion" in 99 percent of all cases, b) the classes "positive and negative emotion" in 100 percent of the cases; and in relation to the whole group of speakers: a) the classes "rest" and "emotion" in 97 percent of all cases; b) the classes "positive and negative emotion" in 96 percent of the situations. In comparison with the first stage, in which all eight characteristics were used, sufficiently effective division of patterns was found to be possible in the second stage within a three-dimensional space of normalized (see above) parameters $A(F_{OT}), \bar{F}'_{1OT}, \bar{F}'_{2OT}$. A certain decline in the quality of recognition in relation to the group of speakers can be explained by individual characteristics of emotional expression, which complicates the configuration of the patterns and increases the number of intersection points. Selecting speakers appropriately within the group would obviously help to eliminate this shortcoming. It should be noted that the algorithm can be used to continue hierarchical division of previously obtained patterns; for example it can be used to divide the class of positive emotions into two subclasses depending on the degree of emotional tension. To permit analysis of this and other possibilities for improving the structure of the recognition system, we suggest supplementing the working space of characteristics with a number of spectral parameters (timbre characteristics) and with some characteristics of speech flow (average frequencies of the fundamental tone, intensity, tempo and so on).

It should be pointed out in conclusion that this algorithm can be supplemented by a procedure for approximately evaluating the level of emotional tension (within a given class of emotions), in proportion to the size of the vector extending in the two-dimensional space of normalized characteristics $\overline{F}_{OT}, \overline{f}_0$ from the coordinate origin to the point representing realization of the vowel sound. The degree to which the emotions examined in this work are expressed can be measured in precisely the same fashion as in the preliminary research, though in this case the coordinates would be $\overline{F}'_{1OT}, \overline{F}'_{2OT}, A(F_{OT})$.

BIBLIOGRAPHY

1. Luk'yanov, A. N., and Frolov, M. V., "Signaly sostoyaniya cheloveka operatora" [Signals of Human Operator State], Moscow, Nauka, 1969.

2. Frolov, M. V., and Taubkin, V. L., "Recognition of Emotional and Physical Stresses in a Group of Speakers by Means of Speech Parameters," in "Elektronnaya apparatura i metodicheskiye voprosy neyrofiziologicheskogo eksperimenta" [Electronic Apparatus and Procedural Questions Associated With Neurophysiological Experimentation], Moscow, Nauka, 1975.

3. Simonov, P. V., Frolov, M. V., and Taubkin, V. L., "An Invariant Method of Recognizing the Emotional State of a Group of Speakers on the Basis of Their Speech," ZH. VND, Vol 26, No 1, 1976.

4. Bondarko, L. V., et al., "Phonetic Correlates of Different Degrees of Speech Expressiveness and Emotionality," in "Materialy Vses. simpoziuma 'Rech' i emotsii'" [Proceedings of the All-Union Symposium "Speech and Emotions"], Leningrad, 1975.

5. Kolymba, S. N., Nushikyan, E. A., and Pirogova, A. A., "Acoustic Correlates of Emotionally Colored Phrases Expressing Anger, Approval and Irony in Modern English," in "Materialy Vses. simpoziuma 'Rech' i emotsii'," Leningrad, 1975.

6. Nushikyan, E. A., "The Role of Tonal Characteristics in the Transmission of a Speaker's Emotional State," in "Materialy Vses. simpoziuma 'Rech' i emotsii'," Leningrad, 1975.

7. Sobolev, V. N., and Baronin, S. P., "Investigation of the Displacement Method of Isolating the Fundamental Tone of Speech," ELEKTROSVYAZ', No 12, 1968, pp 30-36.

8. Khayretdinova, A. G., "A Method of Automatic Isolation of Accent in a Flow of Speech," in "Vychislitel'nyye sistemy" [Computer Systems], Novosibirsk, Nauka, 1971.

9. Andreyev, Yu. M., "Automatic Isolation of Accents in Oral Speech," in "Rechevoye obshcheniye v avtomatizirovannykh sistemakh" [Spoken Communication in Automated Systems], Moscow, Nauka, 1975.

10. Kozinets, B. N., "A Recurrent Algorithm for Dividing the Convex Envelopes of Two Sets," in "Algoritmy obucheniya raspoznavaniyu obrazov" [Pattern Recognition Learning Algorithms], Moscow, Sov. radio, 1973.

11004
CSO: 8144/1642

- END -