Role of perception of rhythmically organized speech in consolidation process of long-term memory traces (LTM-traces) and in speech production controlling

Skljarov O.P.
Research Institute of Otolaryngology and Speech Pathology
198013, Bronnitskaja, 9, St.-Petersburg, RUSSIA
e-mail: vigarb@thewall.ioffe.rssi.ru
telephone: +110 10 10

ABSTRACT

In this paper, proceeding from established during comparative research of speech of the normal persons and stutterers equations of a logistical type [1,2,3], hypothesis about discrete character of perception of a speech signal, occurrence of homeostatic state of speech reproduction system (speech memory), and also about ways revealing of such memory are offered. The results prove to be true by being available experimental data.

INTRODUCTION

In the last work [1,2,3,] we were convinced of gravity of the dual recurrent equations for segments, received from an acoustic speech signal, for an explanation of results average on sample of a comparative experimental research of stutterer's speech and persons without a speech pathology. The first stage of a procedure of segmentation is determination of a threshold. As the level of noise varies depending on conditions of registration of a signal, and also from peculiarities of a pronunciation examined, the program provides automatic installation of a threshold of detection of a signal. With this purpose a minimum level of a signal is found which allocation of given number of segments on an extent of a special fragment of initial record provides. At the second stage there is the detection of a signal on an extent of all record with the help of a threshold, found at the first stage, in such a manner that the whole signal is broken on "sounded" and "unsounded" segments. Equations for durations of these segments are written down as follows:

$$y_{n+1}=r_0 y_n(1-y_n)$$

$$y_{n+2}=r_1 y_{n+1}(1-y_{n+1}), \qquad (1)$$

here odd numbers n concern to sounded segments, and even - to "nonsounded" segments or gaps. r_0 is the coefficient of "inhibition", and r_1 - coefficient of "arousal" in the system. In our terminology the class of "sounded" segments is association of classes (in De Mori's terminology [4]) of vowels (V) and tracts of one or more voiced consonants (VC), while the class of "unsounded" segments is association of classes unvoiced tracts (UT) and silence (SL). These equations concern to a class of the logistical equations and describe so-called chaotic dynamics. Dynamics of these equations, along with the other surprising properties, have attractors of the second order, or supercycles of the second order.

However in an obvious kind such equations we nowhere met in the literature on speech production, as for speech actionable periphery, and for its central departments. Moreover, Tanaka [5], the considering process of self-organizing in cortex structures of a head brain, denies existence of attractors of the second order in such systems. Opposite, in paper of Cohen [6] an opportunity of presence of oscillatory phenomena in self-organizing systems, in a basis of which limiting cycles lay, is affirmed.

Whereas, for a long time the fact is known, what even in one dimensional, but discrete structures existence of the bifurcation diagram is possible, and, thus, existence of rhythm is supplied. So, the Schuster [7] brought an example how rotator's dynamics, described by the differential equation of a movement, under influence of periodically applied external force is reduced to dynamics, described by the difference equations in discrete time, and gets features, characteristic of chaotic dynamics, including and limiting cycles of the second order.

Therefore we have become interested, how within the framework of uniform paradigm of self-organizing there is the temporary organization, which is described by recurrent discrete equations as (1)? How in general in the differential equation which are describing activity of a brain, there is the disctrete behaviour?

We proceeded from: 1) classical, continuous in time equations of Grossberg, constructed within the framework of paradigm of self-organizing for the description of speech production [8]; 2) of the fact reliably established by us [1,2,3], that the speech signal, which is subjected to segmentation, describes behaviour of average values, and gives characteristic bifurcation behavior at stuttering; 3) and also from the that fact,

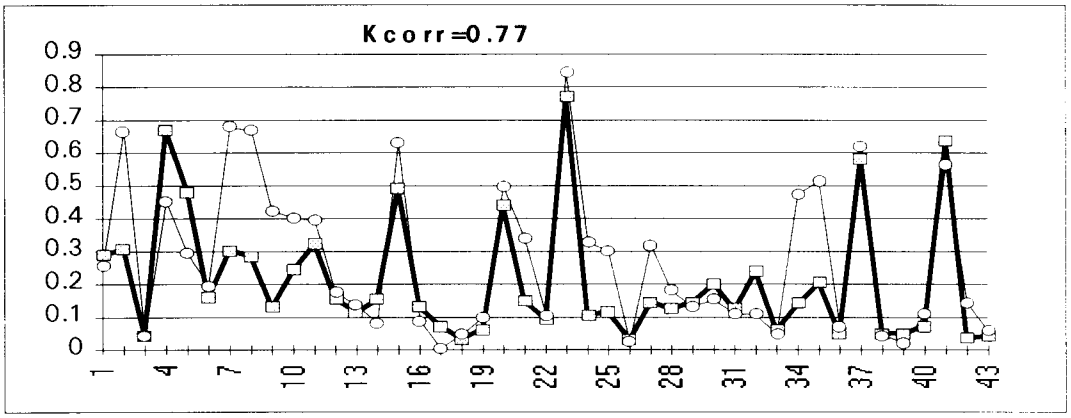

Fig. 1. On a vertical - duration of a sounded segment in seconds; on a horizontal - number of "sounded" segment. A greasy line - experiment; a thin line - forecast under the formula (1).

that the process of perception has essentially discrete character [9]. The last circumstance we observed experimentally. The speech signal, made by normal person and by stutterer, after segmentation and samples replacement by zero in "unvoiced" segments (that is signal on an input, and, hence, its intonation contour were discontinuous, with gaps), was perceived by the listeners completely without distortions. Moreover the use of system (1) to particular realizations (truth, with introduction of modulation in factor r_1), we achieved correlation coefficient up to 77 % between the forecast and real realizations of length up to 40-50 syllables (see fig.1).

All this has forced us to believe that system (1) is correct for a segmented speech signal. All this has forced us to believe that nature of speech perception is discrete also. That fact which was confusing us, namely, was not observed parabolic attractor $y_{n+2}=f(y_n)$ (as this attractor was specified by Titze and other in the work on bifurcations of voice fold's fluctuations [10]) on segments of a real speech signal, is disappeared now. Really Deco et al had shown [11], that two and more network layers with feedback destroy an attractor nature of internal variables. These networks complete decorrelation of external variables, keeping information entropy however. Jordan's network pertains to a class of these decorrelation networks, to all appearances. This network widely is used in speech production for care of management on peripheral level of articulators [12].

Perception and speech memory.

At the description of perception we proceeded from the self-organizational equations of Grossberg [5] for membrane potentials, or traces of short-term memory (SMT) and appropriate synaptic weights , or traces of long-term memory (LMT). The forcing force is in the right part of these equations. This force is arising owing to sensory afferentation. Using the fact established by us earlier, namely, chunking continuity of this afferentation, and also using connection between duration of a "sounded" segment and total synaptic weight [1], it was established, that in conditions such afferentation total synaptic weight, or LMT, submits to the same equations (1), but with new factor of excitation r'_1, which receives exponentially small multiplier (because of coherent attenuation of synaptic weights). By virtue of it, as it was shown in [1], the system has fall into a "silence zone", keeping only possibility of functioning, that is reproduction of a chain which was heard earlier. This possibility is realize when external or internal reasons will result in growth r'_1. This phenomenon we are inclined to interpret as speech memory. The speech specification in this case consists in the logistical recurrent law, to which entrance afferentation flow submits.

Such interpretation of memory differs from representations of Behtereva et al [13] about memory, as about some compressed patterns of words.

Our result, in any case, does not contradict theoretical and experimental results Alvares and Squire [14], which connect memory of recent events with large meaning r'_1, inherent in the medial temporal lobe, and on the contrary, memory of old events - with small meaning r'_1, inherent in connection between various cortex areas.

Generation of speech.

We noted that by virtue of appearance of exponentially small factor in new factor of excitation in result of perception of segmented speech, such situation was arising, that rhytmo-chaotical discrete process (that is the process of pronunciation) was not initialized, and the system of speech production is kept in a kind potentially ready to start.

But such situation is kept only so long as in system will not arise some global lateral inhibition. Imagining, that Grossberg's equations describe any set of coherently inside itself functioning subsets, we come to the representation about memory, namely, speech memory is a set of potentially frozen rhythmic

cues (or coherent cues). In physics such situation is rather standard, and it is known, that presence of global feedback in system (see, e.g. [15]) is necessary for generation potentially possible rhythmical (or coherental) dynamics.

Really, the reduction of parameter of excitation r'_1 appearing at perception can be compensated because of global lateral inhibition connections in Grossberg's equations [8]. There is growth of parameter of excitation up to some value R_1, sufficient the system has become rather "arousal" and has passed from the conditions of "silence", when exists only one (zero) steady point, in the conditions, when the system is capable to start process of "internal generation" speeches in result of realization of rhythm.

The stated circumstance allows to explain experimental facts on lateral inhibition. These facts were received at correlation processing of experimental data on positron topography of brain activity during realization of free verbal flow by the person [16]. Careful carrying out of measuring of changes of zones of activity in a head brain during realization of such verbal flow have allowed to the authors to assume, that interior temporal area of cortex are a place of storage of representation of words on one hand. On another hand the inhibition modulation of activity of these areas, which is realized by the left prefrontal cortex, is a basis of internal speech generation. Literary experimental data cited by the authors on anatomy of connections of various cortex and subcortex formations of a brain have allowed them to put forward a hypothesis that specified lateral inhibition is realized by deeply penetrating associations of fibres, outgoing from prefrontal cortex.

However, remaining within the framework of our model, it is possible only to describe process of imitation. Much more interesting to understand how the new phrase springs up. If to assume, that the new phrase is built from chunks of phrases which the person heard earlier, two stages are necessary at least:

I) It is necessary to highlight fragments of speech chains which the person heard earlier;

II) It is necessary to give these fragments unique semantic meaning, that is to subordinate this set of fragments by syntactic (or in oral speech - intonational) restrictions.

The logistical dependencies as (1) in some sense are similar Markov ergodic random processes usual for Boltzmann dynamics. The analogy consists in the appropriate Ljapunov function existence [7] (otherwise, Kolmogorov, or information entropy). This fact is ensuring relaxation to a global steady state. But the statistical chaos is provided not by thermal fluctuations, but by recurrent equations (1) - so-called "determined chaos". Thus, dynamics of a kind (1) is capable to generate information structures.

Here we pass to the last item of interpretation, which owes to clear situation, and also to justify discrete character of the equations (1). From the point of view of neuronal networks the association of the first and second equation in system (1) looks as follows. As to the first equation of system (1), it is known, that if Hebb's phenomenology about plasticity of synaptic connections to formalize with the help of discrete in time Hopfield's equations for neuron's ensemble, integrative dynamics for synaptic weights submits to the logistical equation as the first equation in (1), where r_0 - parameter of training [17]. If to consider, that motoneurons of speech production tract are loaded on this neuronal ensemble, and these motoneurons responsible for creation of speech segments, then this total synaptic weight can be connected to duration of the specified segments [1]. There is open the question, where there is this ensemble and why its dynamics is discrete? The question on a finding of the specified ensemble, coordinated with a question about "highlighting" of sites of speech, which the person heard earlier finds probable interpretation within the framework of a Crick's hypothesis about function thalamic reticular complex, as about some "searchlight", highlighting this or that sites of cortex [18]. The concentration of attention, by Crick, is supervised by processes, occurring on thalamic reticular complex. This reticular complex represents a thin layer of neurons, which is located between thalamus and neocortex and is penetrated by thalamocortical loops.

As intralaminar nuclei of the thalamus have projections to the striatum, also send axons through reticular complex, and the same is possible to tell about axons, leaving from pallidus, have a return projection to the thalamus, then striopallidus system (crucial for regulation of movements) is adjusted by processes played on this similar to retina reticular complex. As has shown Taylor [19], on this reticular layer process of a competition between input polysensoric activities and also between loops, connected as to memory and frontal zones of a brain is occurred. Taylor confirms existence of such mechanism of a competition by experiments on electrode stimulation of cortex, and describes mathematically this process with the help of the analysis of stability of the nonlinear bifurcational equations.

The described complex system of connections with regulation on the thalamic reticular complex allows us to assume, that if complex multifunctional striopallidus-thalamic-cortical loops have spatial topology as three-dimensional spatial topology of Lorentz's attractor [7] (figuratively speaking, representing butterfly, sitting on thalamus, with wings, concerning both hemispheres of neocortex), and reticular complex will form natural anatomical realization of Poincare's section [7], the consecutive passages of neuron signal in fibres of one wing through this section, as is known, are connected by a logistical difference equation, which at us is described by the "inhibition" equation of system (1), responsible (through training) for adaptation to environment, and second wing - for "arousal", excitation of system.

We needed to be answered a question, how on chaotic structure, produced by the equations (1), are imposed syntactic (intonational) restriction? As by experience on electroshock [20] was rather convincingly shown, that intonation information, on all

probability, is provided by the right hemisphere, and it, on all probability, appears inhibition influence on left (speech) hemisphere, resulting, as we saw, to actualization of traces of long-term memory, probably to assume the following mechanism. In result of these actions of "double searchlight" in the left hemisphere are found those sites of memory which, due to synchronous syntactic inhibition at the expense of the right hemisphere, are realized in semantically the high-grade message. Any disagreement in system (1) (or for the account "internal" deficit, that is distortions, in comparison with norm, r_1, or at the expense of unacceptable external conditions, that is r_0) result in infringement of process of speech generation. If in norm any optimum parity between parameters r_0 and r_1 is carried out (I admit their equality), then a rejection r_0 results in one branch, and deviation r_1 (generically caused by deficit of excitation, which can disappear in maturation process of nervous system, namely, maturation of loops of global inhibition connection, as, probably, it occurs at the children) - to other.

CONCLUSION

Thus, it is possible to conclude, that offered us earlier [1,2,3] the sight on speech production, as on process of self-organizing of complex structures, allows a treatment of speech memory as of temporarily "frozen" rhythmical structures. The starting-up of such rhythmical structures is possible at realization of global feedback in self-organizing system. The conclusions, received with such point of view, do not contradict the experimental neurophysiological facts, observable at activation of free verbal flow.

REFERENCES

1. *Skljarov O.P.* The bifurcation model of the rhythm and stuttering. In: Proc. Of the 1st ESCA Tutorial and Research Workshop on Speech Production Modeling: From Control Strategies to Acoustics; 4th Speech Production Seminar: Models and Data, Autrans, France. Grenoble, 1996.p. 89.

2. *Skljarov O.P.* The perception of the own delay speech as a tool for start-upping of the rhythm at the stuttering. In: Proc. Of the Workshop on the Auditory Basis of Speech Perception / Eds. W. Ainsworth et al, Keele University, UK. Keele, 1996. P. 279.

3. *Skljarov O.P.* The selforganization nature of speech rhythm and stuttering. Journal of Fluency Disorders, 1997, N2 (May), p. 177.

4. *De Mori R., Laface P., Piccolo E.* Automatic Detection and Description of Syllabic Features in Continuous Speech. IEEE Trans. on Acoust., Speech and Signal Proc., 1976, Vol. ASSP-24, N5, pp. 365-379.

5. *Tanaka S.* Theory of self-organization of cortical maps: mathematical framework. Neural Networks. -1990. -Vol.3. - P.625-640.

6. *Cohen M.A.* The construction of arbitrary stable dynamics in nonlinear neural networks. Neural Networks. -1992. -Vol. 5. - P. 83-92.

7. *Schuster H.G.* Deterministic chaos. Physik-Verlag, Weinheim, 1984.

8. *Grossberg S.* The adaptive self-organization of serial order in behavior: speech, language, and motor control. In: Pattern recognition by humans and machins. V.1. Speech Perception. Eds. E.C.Schwab et al. N.Y.: Acad. Press. Inc., 1986. P. 187.

9. *Ladefoged P.* The Perception of Speech. Proc. on Mechanization of Thought Processes, Nat. Rhys. Lab. Teddington, England Nov. 24-27, 1958

10. *Titze I.R., Baken R.J., Herzel H.* Evidence of chaos in vocal fold vibration. In: Vocal Fold Physiology. Frontiers in Basic Science. Ed. I.R.Titze, 1990, Singular Publ. Group, INC, San Diego, Calif. pp. 143-188.

11. *Deco G., Brauer W.,* Nonlinear Higher-Order Statistical Decorrelation by Volyme-Conserving Neural Architectures. Neural Networks, 1995, Vol. 8, N. 4. pp. 525-535.

12. *Saltzman E.L.* Intergestural timing in speech production: data and modeling. Proc. of XIII Int. Con. Phonetic Sci. (ICPhS 95), Stockholm, vol.2, pp. 84.

13. *Bechtereva N.P., Bundzen P.V.* Neurophysiological mechanisms and code oh higher brain functions. In: Speech Communication Seminar. Stockholm, 1974. pp. 311-320.

14. *Alvares P., Squire L.R.* Memory consolidation and the medial temporal lobe: A simple network model. Proc. Natl. Acad. Sci. USA, 1994, Vol. 91, pp. 7041-7045.

15. *Glauber R.J.* Coherent and noncoherent states of emission field. Phys. Rev., 1963, Vol. 131, p.2766.

16. *Frith C.D., Friston K.J., Liddle P.F., Frackowiak R.S.J.* A PET study of word finding. Neuropsychol. -1991. V.29. N12 -P. 1137-1147.

17. *Van der Maas H.L.J., Verschure F.M.J., Molenaar C.M.* A note on chaotic behavior in simple neural networks. Neural Nets., 1990, V. 3. P. 119

18. *Crick F.* Function of the thalamus reticular complex: The searchlight hypothesis. Proc. Natl. Acad. Sci. USA, 1988, vol. 81, p. 4586.

19. *Taylor J.G.* Goals, Drives, and Consciousness. Neural Nets., 1994, vol. 7, Nos 6/7, p. 1181.

20. Balonov L.Ja., Deglin V.L. Hearing and speech with dominant and nondominant hemispheres. 1976, Leningrad, "Nauka", (in Russian).