CONTEXT-INDEPENDENT DURATION MODEL ON CATEGORIES OF VOICE AND UNVOICE SEGMENTS

O.P.Skljarov

Research Institute of Ear, Throat, Nose and Speech 198013, Bronnitskaja st., 9, St.-Petersburg, Russia
E-mail: skljarov@usa.net

ABSTRACT

Trying to understand the experimental data on segmentation of a speech signal by a principle "Voice/Unvoice" has led us to the hypothesis about a pair of logistical dependence between durations of these segments. The segmentation was carried out with the help of the computer program working in quasi real time. The hypothesis about logistic recurrent dependence for sequence of segments durations has allowed to make a conclusion about quasi rhythmical organization of this sequence. With the help of offered recurrent dependences it is possible to explain statistical peculiarities of speech behaviour of stutterers in comparison with normal speech behaviour. These logistic dependences were confirmed by direct experimental data. The assumption of origins of specified rhythm is made. These origins are hidden at the level of control of speech production and perception. Is shown, that the chaotic nature of offered dynamics of formation of large-scale temporary structure allows to enter concept of the information into consideration by a natural way.

1. INTRODUCTION

Concept of rhythm is one of fundamental concepts in biology. Rhythm is also basic characteristic of large-scale (about tens and hundreds msec.) temporary organization of speech production process. However till now almost nothing is known about the origin of speech rhythm.

We offer algorithm of rhythm, proceeding from idea of self-organizing of complex system of speech production. Trying to understand the experimental data on segmentation of a speech signal by a principle "Voice/Unvoice" has led us to the hypothesis about a pair of logistical dependences [3-9] between durations of these segments (segment duration- SD). Is shown, that the reproduction of rhythmic speech with concrete meanings of control parameters is possible only if the neuron ensemble perceived similar rhythmically organized sequences. This circumstance, in our view, demonstrates the speech specification of our approach.

Further, we shall show, how concept of the information arises naturally within the framework of our representations.

And, at last, we shall tell about clinical use of the program which realized our representations. Certificated by Ministry of Health of Russian Federation, this program is introduced into practice of correction of stutter, because this program is instrument of the diagnosis and planning of logo- and psyhotherapeutic treatment measures.

2. EXPERIMENT

Signals. We used an acoustic signals arising at reading by patients of the standard text, consisting from 120 syllables. On latest stages of a research we used the spontaneous speech of patients of the same size also. Each patient before the basic text made a test phrase: "papa, papa, papa".

Transduction. We used a dynamic microphone 82• -12 executed by Leningrad Optical-Mechanical Factory. Performances of microphone: frequencies range: 50-12000 Hz; sensitivity on frequency 1 kHz - 1,8 mV/Pa; directedness - cardioid). The signal from microphone was introduced (through the block of preliminary amplifiers of an input - output) into the computer on 12-digit converter AD-DA, which produced digitization of signal with frequency 10 kHz.

Segmentation. The input signals were normalized on all dynamic range of the computer. At first we determined threshold of segmentation for each individual patient with the help of handling of test phrase. The amplitude threshold of a gradually grows from 0 with a rather small step so long as for the first time will arise 6 (and only 6) Voice segments. We used a temporal parameter about several periods of basic tone also. If the following sample did not occur at t > this parameter, we considered Voice segment as completed segment and Unvoice segment begins. Then this threshold was used for segmentation of a basic signal. The comparison of results of our segmentation method with both results of hand-operated segmentation, and with results of segmentation by HMM-ANN method (as for an acoustic signal, and EGG signal) has given satisfactory coincidence.

Subjects. The patients were both men, and women in the age of from 17 till 50 years. Number of healthy patients was some tens; of stutterers (various degree of severity) - some hundreds. For each patient the average duration of segments T (on a joined set of Voice and Unvoice segments) and their standard deviation std (ratio of root-mean-square deviation to average) was calculated. If the point (T, std) corresponds to each patient, on a plane of indicated parameters there is characteristic fork diagram (see figure 1a). The patients with normal speech are grouped in top of the diagram (circles), and stutterers have place on branches of the diagram (rhombs), and the further from top, than more considerable severity of stuttering. The upper branch corresponds to patients with slow speech, lower - with accelerated speech. Shown on Figure 1a bold lines are examples of trajectories for typical evolution of patients during course of stuttering correction. Appropriate to patients points were labeled by letters • ., • . are A. These letters are put in the beginning of trajectories of an evolution, spent by bold lines and have on three measureable points: prior to the beginning the course of treatment, in the middle of the course and in the end. The vertical lattice labels a transitive zone from "norm" to the "mild" form of stuttering. Horizontal line in centre of drawing - conditional border, sharing "tachylalia" (from below) from "bradylalia" (from above). Such trajectories have allowed to make the diagnosis of stutter severity and to plan logo- and psychotherapeutic measures during of stutter correction. Not shown on a figure (to not block up a drawing) evolution trajectory of spontaneous speech pass in parallel specified bold lines.

This is statistical handling of signals. Besides we have carried out dynamic (step by step, without an average on number of segments) research of sets of segments for concrete patients. We shall now consider concrete realizations for normal utterance of phrase "Vot i nastupil vecher. Zarja zapylala pozarom" (Here an evening has come . Evening-glow has flared by a fire). The spline approximation of set of points with coordinates in 3D-space T_n, T_{n+1}, T_{n+2} $(1 \leq n \leq 21)$ has given a smooth surface for norm, shown on fig. 1b. It is visible, that the dependence T_{n+2} over T_n at fixed T_{n+1} represents characteristic "fork" curve. The polynomial approximation of dependences between segment durations in 2D-space for the first half of specified above phrase is submitted on fig. 1• for norm, and on fig. 1d for stutterer accordingly . On the fig. 1c the bold curve is dependence T_{n+1} vs. T_n; thin curve is T_{n+2} vs. T_n, $(1 \leq n \leq 8)$; on fig. 1d: bold curve is dependence T_{n+2} vs. T_n; thin curve is dependence T_{n+4} vs. Tn, $(1 \leq n \leq 15)$.

3. HYPOTHESIS AND ITS ORIGINS

For an explanation of dependences submitted on drawings from fig. 1a to fig. 1d we have assumed [3-9], that the discrete sequences of segment durations $\{T_n\}$ are generated according to two discrete logistical transformations (reflections):

$$\left.\begin{aligned} T_{n+1} &= F_{r_0}(T_n) = r_0 T_n (1 - T_n) \\ T_{n+2} &= F_{r_1}(T_{n+1}) = r_1 T_{n+1} (1 - T_{n+1}) \end{aligned}\right\} \qquad (1)$$

where $0 \leq T_n \leq 1$, $n=1,2,3,...$; the odd indexes concern to voice segments, and even - to unvoice; ; r_0 and r_1 are abstract coefficients of "inhibition" and "arousal".We shall consider system (1) in two cases:

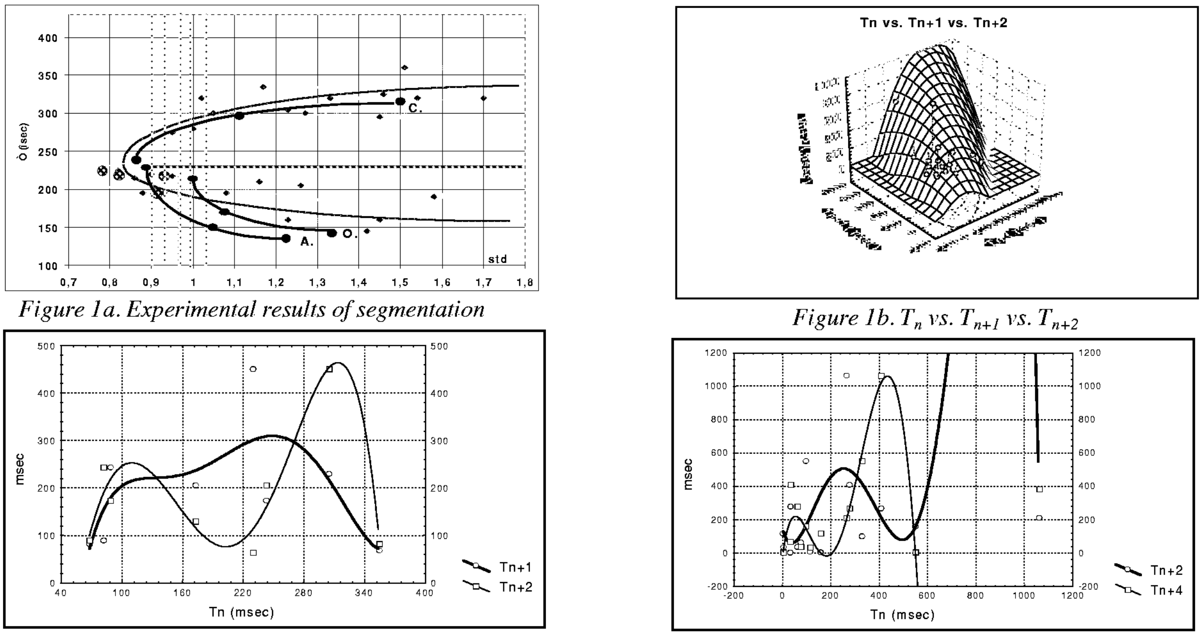

Figure 1a. Experimental results of segmentation

Figure 1b. T_n *vs.* T_{n+1} *vs.* T_{n+2}

***Figure** 1d.* T_{n+2} *vs.* T_n *(bold line);* T_{n+4} *vs.* T_n *(thin line)*

Figure 1c. T_{n+1} *vs.* T_n *(bold line);* T_{n+2} *vs.* T_n *(thin line)*

Figure 1.

Case •). $r_0=r_1=r$. In this case the system (1) is reduced to one reflection as:

$$T_{n+1} = F_r(T_n) = rT_n(1 - T_n) \qquad (1a)$$

Case b). $r_1 \neq r_0$. Absolutely other situation arises in this case. Excitation and inhibition are not counterbalanced in system ($r_1 \neq r_0$), and to determine duration of voice segments it is necessary already to proceed from bilogistical recurrent equation:

$$T_{n+2} = F_{r_1}[F_{r_0}(T_n)] = r_1 r_0 T_n (1 - T_n)[1 - r_0 T_n (1 - T_n)] \quad (1b)$$

The theoretical modeling of experimental curves (fig. 1c and fig. 1d) is submitted on fig. 2a) - case of normal speech and 2b) – case of stuttering. In the top part of fig. 2a) two parabolas are represented: the bottom parabola in general has not steady points (points of crossing with bisector), except zero; top parabola has one point T_n^*, with already lost stability by virtue of growth of parameter *r*. This point is splitted to two new steady points: T_{1n}^* and T_{2n}^*, represented on the figure as points of crossing of a curve in the bottom of drawing with bisector.

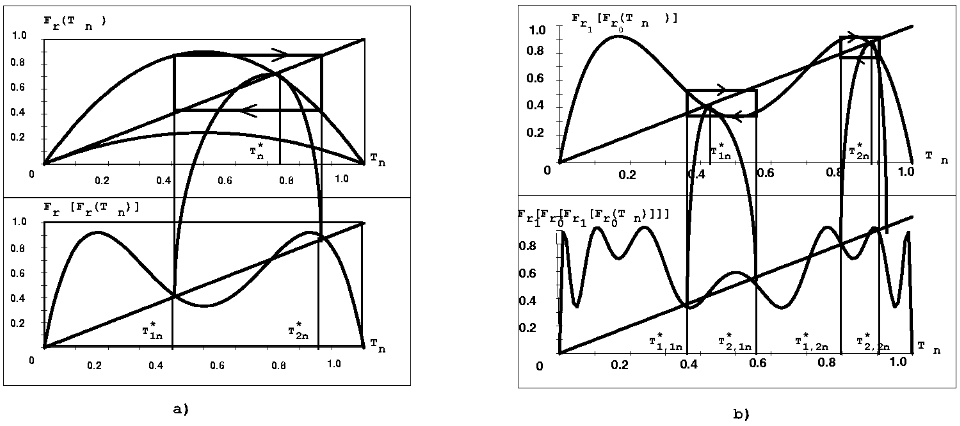

Figure 2.

In accordance with growth of parameter *r* these points form a geometrical place of points shaping bifurcation curve which extended on both top and bottom of fig. 2a). This is bifurcation curve which we observed as section of surface by coordinate plane $T_{n+1}=const$ on fig. 1b. Between these points the rhythmic transitions will be organized. These transitions schematically are represented as a rectangle with arrows and represent the elementary rhythm as change of durations of voice segments.

In a case of stuttering (fig. 2b), the rhythmic transitions already cannot be between steady points T_{1n}^* and T_{2n}^*, (see top of fig. 2b) because there is no previous state which lost its stability. It is necessary to split one of these points on new two steady points (that is, to repeat a situation of fig. 2a), for example, on $T_{1,1n}^*$ and $T_{2,1n}^*$. left). In this case there is the opportunity of appearance of both bifurcation curve and cycle (rectangle with arrows on fig. 2b - at the left). On an ensemble of patients realization both one rhythm (left), and other (right) are possible. These rhythms are on the different parties in respect of norm rhythm (see experimental data on fig. 1a).

3.1. Perception of speech and memory

The speech specification of made hypothesis will be more clear, if to consider perception of speech and memory [6, 7, 10, 11]. At the description of perception we proceeded from Grossberg's equations [1] for both membrane potentials of neurons in network and synaptic weights. In the right part of these equations there is the external forces, arising owing to sensory afferentation.

As this afferentation has structure of discrete set of continuous chunks (remember chunk structure of intonation contour [9]), it turns out that total synaptic weight submits to the same equations (1), but with new control parameter *r* which receives exponentially small multiplier [6, 7, 10, 11]. Due to linear dependence between duration of voice segment and total synaptic weight established by us in [9], it is fixed that the system falls into silence zone at the speech perception, keeping potency of reproduction of received temporal chain. This reproduction turns out when external or internal reasons will result in growth of *r*. Speech specification in this case consists in fact, that entrance afferentation flow is ruled by logistic recurrent transformations (1).

3.2. Generation of speech

But the conditions for such homeostatic state is kept only so long as global lateral inhibition will not arise in system. Really [6, 7, 10, 11], the account of global lateral feedback in the Grossberg equations [1] can compensate reduction of control parameter *r* occurring at perception. There is the growth of control parameter up to some size *R*, sufficient for system arousing. Then system makes transition from state of "silence" in the state, when the system is capable to start process of «internal generation" of speech in result of realization of rhythm.

This circumstance allows to explain the experimental facts on lateral inhibition, received at correlation processing of experimental data on positron tomography of brain activity during realization of verbal flow by the person [12]. With help of modeling of modulation's influence of global feedback as dependence: $r(n) = 2\{1 + sin[(n + 1)T_n]\}$, we achieved good correlation between theory and experiment not only in limits one or two syntagmas (or speech expirations), but on an extent of the statement, consisting of 85 segments.

4. INFORMATION IN MODEL

As is known [13], for discrete systems with Jakobian less than 1 (that is, for dissipative systems), the phase space is compressed in area, named as attractor of the given system. It is known also, that the system is capable to generate the information, if her attractor has fractional dimension. For one-dimensional case information entropy (Kolmogorov entropy) for systems with fractional dimension coincides with a Ljapunov index. It is possible to define the information of a speech flow (in syntactic sense) with help of dependence for Ljapunov index vs. control parameter *r*, because this dependence is known for logistical transformation [13]. Parameter *r* can be determined with help of using of modulation equation for *r(t)* by achieving the greatest correlation of results of the theory and experiment. Thus, the system, described by the equations (1), is capable to generate the new limited information at those meanings of *n*, when this system is in a zone of chaos (information is positive). This circumstance is necessary condition of speech process communicativeness. As is known, the dimension of strange attractor for logistical transformation is equal 0.5 at $r=r_\infty$. On the other hand it is known, that for experimental data is possible to determine the attractor dimension as arctangent of the binary logarithm of correlation integral. It is may to show, that the dimension of strange attractor is estimated as approximately 0.4 from our experimental data.

5. CLINICAL PRACTICE [14]

Numerous inspections of different groups of stutterers of different age and sex (the number of surveyed patients is a few hundreds) confirm [14] submitted diagnosis peculiarities of the segmentation program. This program got Certificate of Ministry of Health of Russian Federation N245.

6. CONCLUSIONS

Used earlier formal and non measurable concept of a segment in the theory of speech production is replaced with such concept of a temporary segment, which admits his fixing with help of hardware and has well determined biophysical sense of time of voice folds' vibration. Stochastic (in sense of chaotic dynamics) model of speech rhythm is offered, based on recurrent dependence between durations of segments. This model is agreed as with theoretical representations about occurrence of rhythm in complex systems, and with experimental data. On the basis of such recurrent dynamic model the bifurcational diagram is predicted. Is shown, that the speech specificity of model is reached at the simultaneous consideration both process of perception, and process of generation of rhythmically organized discrete temporal sequences. Information properties of offered model are shown.

7. REFERENCES

1. Grossberg, S. «The adaptive self-organization of serial order in behavior: speech, language, and motor control,» In: Pattern recognition by humans. V.1. Speech Perception, Acad. Press. Inc., N.Y., pp. 187-197, 1986.

2. Bailly, G.: «Sensory-motor control of speech movements,» In: Proc. of the 1st ESCA Tutorial and Research Workshop on Speech Production Modeling, Autrans, France, Grenoble, pp. 145-148, May 21-24, 1996.

3. Skljarov, O.P. «The bifurcation model of the speech rhythm and stuttering,» In: Proc. of the 1st ESCA Workshop on Speech Production Modeling, Autrans, France, pp. 89 – 92, May 21-24, 1996.

4. 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 ESCA Tutorial and Research Workshop « The Auditory Basis of Speech Perception.». Keele University, UK, pp. 279-282, 15-19 July, 1996.

5. Skljarov, O.P. «The selforganization nature of speech rhythm and stuttering,» Journal of Fluensy Disorders. 22: 139, 1997.

6. Skljarov, O.P. «Role of perception of rhythmically organized speech in consolidation process of long-term memory traces (LTM-traces) and in speech production controlling.» In: Proc. of the 5th European Conference on Speech Comm., Greece, 4: 2147-2150, 1997.

7. Skljarov, O.P. «The logistical generator of speech». In: Proc. of the International Conference on Speech Proc., Seoul National Univ. Seoul, Korea, August 26-28, 1997.

8. Skljarov, O.P. «The chaotical generator of speech», In: Informal Proc. of the NATO ASI Conference «Computational Models of Speech Pattern Processing». St. Helier, Jersey, UK, Vol. 1, July 1997.

9. Skljarov O.P. «Self-organizational nature of speech rhythm (model of voice source),» Biofizika 43: 152-158, 1998 (in Russian).

10. Skljarov, O.P. «The possible mechanism of being up of traces of LTM at perception of rhythmically organized speech,»News of otorhinolaryngol. and speech pathol. 1:36-45, 1997 (in Russian).

11. Skljarov, O.P. «A role of lateral inhibitions at actualization of long-term memory traces,» News of otorhinilaryngol. and speech pathol. 2: 27-30, 1997.

12. Frith, C.D. et.al. «A PET study of word finding,» Neuropsychol. 29: 1197-1206, 1991.

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

14. Skljarov, O.P. «The program of segmentation of speech signal as means of planning of correction measures at stuttering,» News of otorhinolaryngol. and speech pathol. 3: 61-65, 1998 (in Russian).