THE BIFURCATION MODEL OF THE SPEECH RHYTHM AND STUTTERING

O.P. Skljarov
Research Institute of Otolaryngology and Speech Pathology
St.-Petersburg, Bronnitskaja, 9, Russia

Abstract

We are regarding the speech production as the result of the selforganization process arising, probably, on the control level. Let define the selforganization as the process of the change of the preceded stable state lost its stability by new stable state [1]. In dynamical aspect this stability is provided by bifurcation behavior of the system as the control parameter(s) is increasing. This behavior causes rhythmic organization (both temporal and spatial) of the speech production process. Impossibility to realize this rhythm we treat as stuttering.

Starting data and Hypothesis

We watched "bifurcation" behavior of the system in the comparative studies of the mean values $T$ for "sound" segments with durations $\{y_n\}$ and their standard deviation $\text{std}$. Patients were speaking the standard text from "Hunter's memoirs" by I.S. Turgenev, consisting of 120 syllables.

Results for explored group of healthy and stuttering subjects demonstrated such behavior, that points for healthy fall on the top (rings), and points for stuttering subjects (squares) separate on the two branches with larger and smaller durations and fall on these branches so far from top as std (severity of stutter) is larger (see fig. 1).

To explain these data we assume in the frame of the selforganization theory [1, 7, 16] that consequence $\{y_n\}$ is ruled by recurrent logistical equations of chaotic dynamics:
\[ \begin{align*}
Y_{n,k} &= F_{r_0}(Y_{n-1,k}) = r_0 Y_{n-1,k} (1 - Y_{n-1,k}) \\
Y_{n,k+1} &= F_{r_1}(Y_{n,k}) = r_1 Y_{n,k} (1 - Y_{n,k})
\end{align*} \tag{1} \]

Here, \( n, k \) are temporal and spatial discreet indexes, \( r_0 \) and \( r_1 \) are abstract feedback and forward coefficients. The changing of the preceded stable state, lost its stability, by new stable state and arising rhythmic structure is shown on the fig. 2a (normal speech) and on the fig. 2b ("stuttering" speech).

**Normal speech.** At the low exciting degree \((r_1<1)\) the first equation (1) has the single stability ("gomeostatic") fixed point. It's the situation before speech process ("silent" system). With a rise of the exciting degree (potential of initialization) or with a rise of backpropagation (fig. 2a, lower, \( r_0=3.6, r_1=3.7 \)) this gomeostatic point loses its stability at the one-step iterations, but two stability points are appearing from this point to its left and right at the two-step iterations. As this takes place, the one-step iterations generate steady transitions between the two-step stability states. It's our opinion that this limit cycle of the second order is the rhythm of the idealistic speech process. The system starts "speaking". Attractor of the type "coiling" on the limit cycle is occurred when \( r_0 \neq r_1 \). It's the situation of the real normal speech.

**"Stuttering" speech.** The absence or disappearance (treated as pathology) temporal \((n\rightarrow n+1)\) or/and spatial \((k\rightarrow k+1)\) mapping leads to rebuilding in the temporal-spatial organization of the process in such way that homeostatic situation with two stable points is set at the one-step iterations, and rhythm driver isn't realized (fig. 2b, upper). We have situation, which looked as superposition of the two independent mechanisms, shown on the fig. 2a. To start-up process of "speaking" for this system, it's necessary to destroy stability these two fixed points (analogously as in the case of fig 2a). That violence must bring into splitting these stable points, and arising two independent rhythms, according with two experimental branches on the fig. 1.

**Physiological and neurodynamics model of the rhythm (the voice source model)**

We shall show that recurrent eqs. (1) have real physiological and neurodynamical base (at least, for voice source model, alike the Fujisaki model [8]).

Our tool for research of voice source activity would be intonation contour. We determine the fundamental frequency \( F_0(t) \) by calculating autocorrelation shift \( T_0(t) = 1/F_0(t) \) for sound pressure \( p(t) \). We define intonation contour as the function of time \( F_0(t) \).

Fig. 3 show intonation contour (upper) and sound pressure (lower) for Russian phrase "Vot i nastupil vecher. Zalja zapylala..." ("Here the evening fell. The sunset was blazing as a fire...") from standard text, mentioned above. The syllables "Vot"
and "rom" are omitted. Perceptual identification of sound segments according the intonation segments is shown in the upper.

![Fig. 3. Intonation contour for standard phrase.](image)

From our researching we conclude that: a) the voice source is working in discrete regime and forming sequence of pairs \{\tilde{y}_n, F_0(t)\}, where \tilde{y}_n is the duration of \(n\)-th unbroken segment \(F_0(t)\); b) there are both "soft" (asymptotically infinite breaks) and "sharp" (finite breaks) regimes of on- and off-switching for voice source; c) the mean square of fundamental frequency for the \(n\)-th segments, as a rule, is proportional to \(y_n\) which in own turn (if to neglect the pure turbulent sounds) coincides with duration \(y_n\) of "sound" segment. The correlation coefficient is near 0.95 for "sound" \{\(y_n\)\} and "phonal" \{\(\tilde{y}_n\)\} sequences for our phrase.

By virtue of this high degree of correlation it's probably that the formation law for sequences \{\(y_n\)\} and \{\(\tilde{y}_n\)\} is the same. In our model the glottal pressure \(\Delta P\) is external parameter. Effects of this condition might be seen, e.g., on the stopping bursts as the "p".

As Beckman et al shown [9], the \(F_0\) variation strongly corresponds to the EMG level retrieved from cricothyroid when the folds are in phonal regime. In the other hand, the EMG level from muscle is defined by two factors: 1) the innervation frequency by motoneuron (MN) and 2) the number of recruited MNs [10]. As the effect of recruiting connected with synchronization of working of these recruited MNs, we limited our view by frequency of spikes in the axon of "generalized" MN only. This MN should be regarded, probably, as temporally formed coherent neural network. The innervation frequency (in limited diapason) is proportional to membrane potential \(\Phi(t)\) of MN. It may be suggested that phonal duration \(\tilde{y}_n\) (or \(y_n\)) is defined by duration of spike's batch \(\tilde{y}_n\) generated in the axon of MN controlling cricothyroid, that is, by the time when membrane potential \(\Phi(t)\) is more equilibrium potential.

Likely there is neurodynamical mechanism, controlling spike's generation in the axon of the "generalized" MN with spike's batch durations \(y_n\) described by eqs. (1). Assume that the control of the activity of cricothyroid is carried out by algorithm of McClelland-Rumelhart neural network [11] where role of parameter \(r_1\) plays input in the system and the role of feedback parameter \(r_o\) plays learning parameter \(\eta\). If to exclude the "fast" variables and to trace the dynamics of summary synaptic weights \(\bar{W}\) ("slow" variables) as a function of parameter \(r_o\) following van-der-Maas et al [12] then we get the recurrent equations, formally coincide with the eqs. (1). Apart this, if to allow that time-summation in \(\bar{W}\) includes the complete history of the network's output activity \(\bar{y}_n\) [13], then we get just equations for innervation durations \(\tilde{y}_n\), as that assumed by our hypothesis (1). It seems likely that neural structures with the spatial topology of the ring are used in the motor control of speech tract organ's movements (see, e.g. [14]). So, as it's shown in [15] there are recurrent mapping equations for spatial index \(k\rightarrow k+I\), having inherent bifurcation nature, when the parameters of "exciting" and "inhibition" \(r_1\) and \(r_o\) satisfy a certain condition, namely, the lack of "lateral inhibition" which, may be, is fulfilled for efferent control tract of speech production.

So, the base of the rhythmic structure of speech lies in the chaotic dynamics (1), describing the control processes using only two parameters \(r_1\) and \(r_o\). These control parameters may be matching by the degree of corre-
lation between the experimental and theoretical rhythm drawings, shown on the fig. 4.

\[ Y (\text{sec}) \]

\begin{align*}
\text{tnas} & \quad \text{tu} & \quad \text{pil} & \quad \text{zarja} & \quad \text{pil} & \quad \text{zar} & \quad \text{po} & \quad \text{zar} & \quad \text{rom} \\
0.5 & \quad 0.4 & \quad 0.3 & \quad 0.2 & \quad 0.1 & \quad 0.0 & \quad 0.0 & \quad 0.0 & \quad 0.0
\end{align*}

\[ X (\text{sec}) \]

\begin{align*}
\text{tnas} & \quad \text{tu} & \quad \text{pil} & \quad \text{zarja} & \quad \text{pil} & \quad \text{zar} & \quad \text{po} & \quad \text{zar} & \quad \text{rom} \\
0.5 & \quad 0.4 & \quad 0.3 & \quad 0.2 & \quad 0.1 & \quad 0.0 & \quad 0.0 & \quad 0.0 & \quad 0.0
\end{align*}

\textit{Fig. 4.} Empirical and theoretical rhythms.

\section*{Simulation}

The prediction following hypothesis (1) was made for parameters \( r_0 = 3.5 \) and \( r_1 = 2.5 \). The comparative results of experimental segmentation and the theoretical prediction are show on the fig. 4a. The experimental data are shown as the bold broken line with the black squares, marking the breakpoints. The vertical axis is durations of "sound" segments in sec. The perception identifications of these segments are shown on the horizontal axis. The results are shown as the bold broken line with the empty squares. It's seen that starting from the only one experimental segment "Vot" the recurrent prediction procedure (1) gives the good correlation with experimental data on the deep on 8-9 segments (correlation coefficient \( k \) for entire utterance is equal 0.61, and for the second part of utterance, namely, "Zarja zapylala pozarom..." is 0.69). From fig. 4a it's seen that after some transition region the rhythmic drawing of change of relative large and small durations is established in the system. The well correlation between experimental and theoretical data is observed if any experimental sequent segment (with the exception the starting segment of another sentence: "Zarja zapylala pozarom...") in our utterance is used as the start point, but in this case it's sufficiently only backpropagation equation (the first in the eqs. (1)). In the case of exception made above both eqs. (1) are necessary again.

When there is disbalance of the forward- and feedback relations, inherent to the specifically speech production process, then the speech rhythm disappears and the gomeostatic state is established. This situation is shown on the fig. 4b for \( r_1 = r_0 = 2.0 \).