Chaotic Modeling and Simulation (CMSIM) 3: 207-220, 2014

Nonlinear Self-organization Dynamics of a Metabolic Process of the Krebs Cycle

Valeriy Grytsay[1], Iryna Musatenko[2]

[1] Bogolyubov Institute for Theoretical Physics, 14b, Metrolohichna Str., Kyiv 03680
(E-mail: vgrytsay@bitp.kiev.ua)
[2] Taras Shevchenko National University of Kyiv, Faculty of Cybernetics, Department of Computational Mathematics, 64, Volodymyrska Str., Kyiv, Ukraine
(E-mail: ivmusatenko@gmail.com)

Abstract. The present work continues studies of the mathematical model of a metabolic process of the Krebs cycle. We study the dependence of its cyclicity on the cell respiration intensity determined by the formation level of carbon dioxide. We constructed the phase-parametric characteristic of the consumption of a substrate by a cell depending on the intensity of the metabolic process of formation of the final product of the oxidation. The scenarios of all possible oscillatory modes of the system are constructed and studied. The bifurcations with period doubling and with formation of chaotic modes are found. Their attractors are constructed. The full spectra of indices and divergencies for the obtained modes, the values of KS-entropies, horizons of predictability, and Lyapunov dimensions of strange attractors are calculated. Some conclusions about the structural-functional connections of the cycle of tricarboxylic acids and their influence on the stability of the metabolic process in a cell are presented.

.

Keywords: *Krebs cycle, metabolic process, self-organization, strange attractor, bifurcation, Feigenbaum scenario.*

1 Introduction

One of the possible problems of synergetics is the study of the internal dynamics of metabolic processes in cells. Its solution allows one to find the structural-functional connections defining the self-organization of these processes and to answer the question how the catalyzed enzymatic reactions create the internal space-time ordering of the cell life.

The most general metabolic process in cells is the cycle of tricarboxylic acids [1]. This is the key stage of the respiration of all cells. In its course, the di- and tricarbon compounds, which are formed as intermediate products in the transformation of carbohydrates, fats, and proteins, are transformed up to CO_2. In this case, the released hydrogen is oxidized further up to water, by taking the direct participation in the synthesis of ATP, being the universal energy source.

Studies of the functioning of the Krebs cycle were carried out both experimentally and theretically in [2-10].

In the study of the given process, we use the mathematical model of the growh of cells *Candida utilis* on ethanol, which was developed by Professor

Received: 9 July 2014 / Accepted: 30 July 2014

ISSN 2241-0503

V.P. Gachok [11, 12]. With the help of this model, the unstable modes in the cultivation of cells observed in experiments were considered. The kinetic curves of the chaotic dynamics obtained with the help of computational experiments were in agreement with experimental data [13].

Then the given model was modified and refined in [14] due to the account for the influence of the CO_2 level on the respiration intensity. With the help of the model, the structural-functional connections of the metabolic process in a cell, which cause the appearance of complicated oscillations in the metabolic process, were investigated. It was concluded that the given oscillations arise on the level of redox reactions of the Krebs cycle, reflect the cyclicity of the process, and characterize the self-organization in a cell. The fractality of the dynamics of oscillations of the Krebs cycle was studied as well.

The analogous oscillatory modes were observed in the processes of photosynthesis and glycolysis, variations of the calcium concentration in a cell, oscillations in heart muscle, and other biochemical processes [15-19].

2 Mathematical Model

The general scheme of the process is presented in Fig. 1. According to it with regard for the mass balance, we have constructed the mathematical model given by Eqs. (1) - (19).

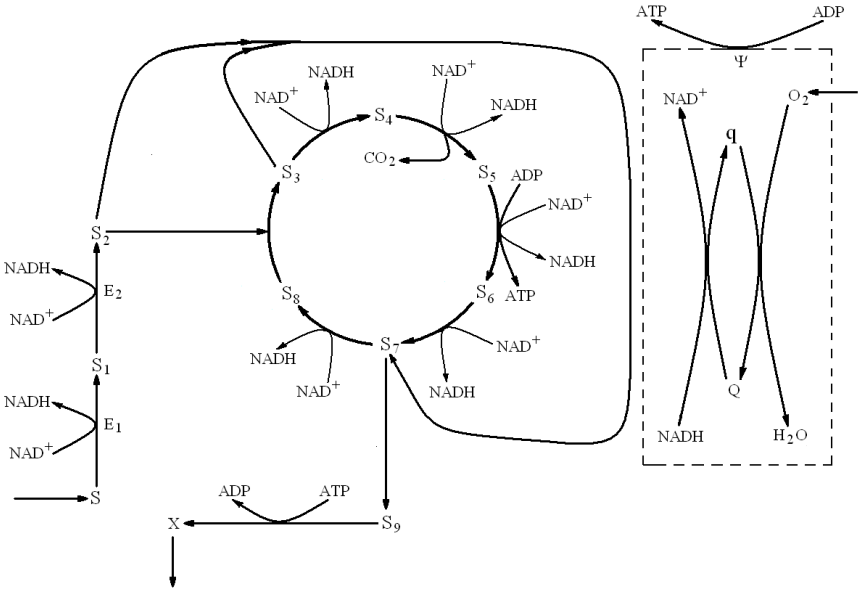

Fig. 1. General scheme of the metabolic process of growth of cells *Candida utilis* on ethanol.

$$\frac{dS}{dt} = S_0 \frac{K}{K + S + \gamma\psi} - k_1 V(E_1) \frac{N}{K_1 + N} V(S) - \alpha_1 S, \tag{1}$$

$$\frac{dS_1}{dt} = k_1 V(E_1)\frac{N}{K_1+N}V(S) - k_2 V(E_2)\frac{N}{K_1+N}V(S_1), \tag{2}$$

$$\frac{dS_2}{dt} = k_2 V(E_2)\frac{N}{K_1+N}V(S_1) - k_3 V(S_2^2)V(S_3) - k_4 V(S_2)V(S_8), \tag{3}$$

$$\frac{dS_3}{dt} = k_4 V(S_2)V(S_8) - k_5 V(N^2)V(S_3^2) - k_3 V(S_2^2)V(S_3), \tag{4}$$

$$\frac{dS_4}{dt} = k_5 V(N^2)V(S_3^2) - k_7 V(N)V(S_4) - k_8 V(N)V(S_4), \tag{5}$$

$$\frac{dS_5}{dt} = k_7 V(N)V(S_4) - 2k_9 V(L_1 - T)V(S_5), \tag{6}$$

$$\frac{dS_6}{dt} = 2k_9 V(L_1 - T)V(S_5) - k_{10} V(N)\frac{S_6^2}{S_6^2 + 1 + M_1 S_8}, \tag{7}$$

$$\begin{aligned}\frac{dS_7}{dt} = {} & k_{10} V(N)\frac{S_6^2}{S_6^2 + 1 + M_1 S_8} - k_{11} V(N)V(S_7) - \\ & - k_{12}\frac{S_7^2}{S_7^2 + 1 + M_2 S_9}V(\psi^2) + k_3 V(S_2^2)V(S_3),\end{aligned} \tag{8}$$

$$\frac{dS_8}{dt} = k_{11} V(N)V(S_7) - k_4 V(S_2)V(S_8) + k_6 V(T^2)\frac{S^2}{S^2 + \beta_1}\cdot\frac{N_1}{N_1 + (S_5 + S_7)^2}, \tag{9}$$

$$\frac{dS_9}{dt} = k_{12}\frac{S_7^2}{S_7^2 + 1 + M_2 S_9}V(\psi^2) - k_{14}\frac{XTS_9}{(\mu_1 + T)[(\mu_2 + S_9 + X + M_3(1 + \mu_3\psi)]S}, \tag{10}$$

$$\frac{dX}{dt} = k_{14}\frac{XTS_9}{(\mu_1 + T)[(\mu_2 + S_9 + X + M_3(1 + \mu_3\psi)]S} - \alpha_2 X, \tag{11}$$

$$\frac{dQ}{dt} = -k_{15} V(Q)V(L_2 - N) + 4k_{16} V(L_3 - Q)V(O_2)\frac{1}{1 + \gamma_1\psi^2}, \tag{12}$$

$$\frac{dO_2}{dt} = O_{2_0}\frac{K_2}{K_2 + O_2} - k_{16}(L_3 - Q)V(O_2)\frac{1}{1 + \gamma_1\psi} - k_8 V(N)V(S_4) - \alpha_3 O_2, \tag{13}$$

$$\begin{aligned}\frac{dN}{dt} = {} & -k_7 V(N)V(S_4) - k_{10} V(N)\frac{S_6^2}{S_6^2 + 1 + M_1 S_8} - k_{11} V(N)V(S_7) - \\ & - k_5 V(N^2)V(S_3^2) + k_{15} V(Q)V(L_2 - N) - \\ & - k_2 V(E_2)\frac{N}{K_1 + N}V(S_1) - k_1 V(E_1)\frac{N}{K_1 + N}V(S),\end{aligned} \tag{14}$$

$$\frac{dT}{dt} = k_{17}V(L_1 - T)V(\psi^2) + k_9V(L - T)V(S_3) - \alpha_4 T - $$
$$- k_{18}k_6V(T^2)\frac{S^2}{S^2 + \beta_1}\cdot\frac{N_1}{N_1 + (S_5 + S_7)^2} - \quad (15)$$
$$- k_{19}k_{14}\frac{XTS_9}{(\mu_1 + T)[\mu_2 + S_9 + X + M_3(1 + \mu_3\psi)S]},$$

$$\frac{d\psi}{dt} = 4k_{15}V(Q)V(L_2 - N) + 4k_{17}V((L_1 - T)V(\psi^2) - $$
$$- 2k_{12}\frac{S_7^2}{S_7^2 + 1 + M_2S_9}V(\psi^2) - \alpha\psi, \quad (16)$$

$$\frac{dE_1}{dt} = E_{1_0}\frac{S^2}{\beta_2 + S^2}\frac{N_2}{N_2 + S_1} - n_1V(E_1)\frac{N}{K_1 + N}V(S) - \alpha_5E_1, \quad (17)$$

$$\frac{dE_2}{dt} = E_{2_0}\frac{S_1^2}{\beta_3 + S_1^2}\frac{N_3}{N_3 + S_2} - n_2V(E_2)\frac{N}{K_1 + N}V(S_1) - \alpha_6E_2, \quad (18)$$

$$\frac{dC}{dt} = k_8V(N)V(S_4) - \alpha_7C. \quad (19)$$

where $V(X) = X/(1 + X)$ is the function that describes the adsorption of the enzyme in the region of a local coupling. The variables of the system are dimensionless [11, 12].

The internal parameters of the system are as follows:
$k_1 = 0.3$; $k_2 = 0.3$; $k_3 = 0.2$; $k_4 = 0.6$; $k_5 = 0.16$; $k_6 = 0.7$; $k_7 = 0.08$; $k_8 = 0.022$; $k_9 = 0.1$; $k_{10} = 0.08$; $k_{11} = 0.08$; $k_{12} = 0.1$; $k_{14} = 0.7$; $k_{15} = 0.27$; $k_{16} = 0.18$; $k_{17} = 0.14$; $k_{18} = 1$; $k_{19} = 10$; $n_1 = 0.07$; $n_2 = 0.07$; $L = 2$; $L_1 = 2$; $L_2 = 2.5$; $L_3 = 2$; $K = 2.5$; $K_1 = 0.35$; $K_2 = 2$; $M_1 = 1$; $M_2 = 0.35$; $M_3 = 1$; $N_1 = 0.6$; $N_2 = 0.03$; $N_3 = 0.01$; $\mu_1 = 1.37$; $\mu_2 = 0.3$; $\mu_3 = 0.01$; $\gamma = 0.7$; $\gamma_1 = 0.7$; $\beta_1 = 0.5$; $\beta_2 = 0.4$; $\beta_3 = 0.4$; $E_{1_0} = 2$; $E_{2_0} = 2$.

The external parameters determining the flow-type conditions are chosen as $S_0 = 0.05055$; $O_{2_0} = 0.06$; $\alpha = 0.002$; $\alpha_1 = 0.02$; $\alpha_2 = 0.004$; $\alpha_3 = 0.01$; $\alpha_4 = 0.01$; $\alpha_5 = 0.01$; $\alpha_6 = 0.01$; $\alpha_7 = 0.0001$.

The model covers the processes of substrate-enzymatic oxidation of ethanol to acetate, cycle involving tri- and dicarboxylic acids, glyoxylate cycle, and respiratory chain.

The incoming ethanol S is oxidized by the alcohol dehydrogenase enzyme E_1 to acetaldehyde S_1 (1) and then by the acetal dehydrogenase enzyme E_2 to acetate S_2 (2), (3). The formed acetate can participate in the cell metabolism and can be exchanged with the environment. The model accounts for this situation by the change of acetate by acetyl-CoA. On the first stage of the Krebs

cycle due to the citrate synthase reaction, acetyl-CoA jointly with oxalacetate S_8 formed in the Krebs cycle create citrate S_3 (4). Then substances S_4 - S_8 are created successively on stages (5)-(9). In the model, the Krebs cycle is represented by only those substrates that participate in the reduction of $NADH$ and the phosphorylation $ADT \to ATP$. Acetyl-CoA passes along the chain to malate represented in the model as intramitochondrial S_7 (8) and cytosolic S_9 (10) ones. Malate can be also synthesized in another way related to the activity of two enzymes: isocitrate lyase and malate synthetase. The former catalyzes the splitting of isocitrate to succinate, and the latter catalyzes the condensation of acetyl-CoA with glyoxylate and the formation of malate. This glyoxylate-linked way is shown in Fig. 1 as an enzymatic reaction with the consumption of S_2 and S_3 and the formation of S_7. The parameter k_3 controls the activity of the активность glyoxylate-linked way (3), (4), (8). The yield of S_7 into cytosol is controlled by its concentration, which can increase due to S_9, by causing the inhibition of its transport with the participation of protons of mitochondrial membrane.

The formed malate S_9 is used by a cell for its growth, namely for the biosynthesis of protein X (11). The energy consumption of the given process is supported by the process $ATP \to ADP$. The presence of ethanol in the external solution causes the "ageing" of external membranes of cells, which leads to the inhibition of this process. The inhibition of the process also happens due to the enhanced level of the kinetic membrane potential ψ. The parameter μ_0 is related to the lysis and the washout of cells.

In the model, the respiratory chain of a cell is represented in two forms: oxidized, Q, (12) and reduced, q, ones. They obey the integral of motion $Q(t)+q(t)=L_3$.

A change of the concentration of oxygen in the respiratory chain is determined by Eq. (13).

The activity of the respiratory chain is affected by the level of $NADH$ (14). Its high concentration leads to the enhanced endogenic respiration in the reducing process in the respiratory chain (parameter k_{15}). The accumulation of $NADH$ occurs as a result of the reduction of NAD^+ at the transformation of ethanol and in the Krebs cycle. These variables obey the integral of motion $NAD^+(t)+NADH(t)=L_2$.

In the respiratory chain and the Krebs cycle, the substrate-linked phosphorylation of ADP with the formation of ATP (15) is also realized. The energy consumption due to the process $ATP \to ADP$ induces the biosynthesis of components of the Krebs cycle (parameter k_{18}) and the growth of cells on the substrate (parameter k_{19}). For these variables, the integral of motion $ATP(t)+ADP(t)=L_1$ holds. Thus, the level of ATP produced in the redox

processes in the respiratory chain $ADP \to ATP$ determines the intensity of the Krebs cycle and the biosynthesis of protein.

In the respiratory chain, the kinetic membrane potential ψ (16) is created under the running of reducing processes $Q \to q$. It is consumed at the substrate-linked phosphorylation $ADP \to ATP$ in the respiratory chain and the Krebs cycle. Its enhanced level inhibits the biosynthesis of protein and process of reduction of the respiratory chain.

Equations (17) and (18) describe the activity of enzymes E_1 and E_2, respectively. We consider their biosynthesis (E_{1_0} and E_{2_0}), the inactivation in the course of the enzymatic reaction (n_1 and n_2), and all possible irreversible inactivations (α_5 and α_6).

Equation (19) is related to the formation of carbon dioxide. Its removal from the solution into the environment (α_7) is taken into account. Carbon dioxide is produced in the Krebs cycle (5). In addition, it squeezes out oxygen from the solution (13), by decreasing the activity of the respiratory chain.

The study of solutions of the given mathematical model (1)-(19) was performed with the help of the theory of nonlinear differential equations [20, 21] and the methods of mathematical modeling of biochemical systems applied and developed by the authors. in [22-38].

3 The results of Studies

For one cycle, there occurs the full oxidation of a molecule of acetyl-CoA up to malate and the formation of a new molecule of acetyl-CoA at the input. In such a way, the continuous process of functioning of the Krebs cycle is running. This process has the autooscillatory character.

The studies of the model with the help of computational experiments showed that if system's parameters vary, the appearance of autooscillations with various frequencies, as well as chaotic oscillations, becomes possible. Oscillations with the same frequency will occur in all components of the given metabolic process. In the present work, we will study the dependence of autooscillations of the system on the parameter k_8, which determines the level of formation of CO_2 in the cycle of tricarboxylic acid.

The different types of obtained autooscillatory modes are studied with the help of the construction of phase-parametric diagrams. The abscissa axis shows the values of parameter k_8, and the axis of ordinates gives the values of chosen variable $E_1(t)$, for example. Moreover, we used the method of cutting. In the phase space of trajectories of the system, we place the cutting plane $S_2 = 0.8$. Such a choice is explained by the symmetry of oscillations of acetate relative to this plane in a lot of earlier calculated modes. For every given value of k_8, we observe the intersection of this plane by the trajectory in a single direction, when it approaches the attractor. The value of $E_1(t)$ is put onto the phase-

parametric diagram. In the case where a multiple periodic limiting cycle arises, a number of points can be observed on the plane, and they will be the same in the period. If the deterministic chaos arises, the points of the intersection of the plane by the oscillating trajectory will be positioned chaotically.

In Fig. 2,a-d, we show the phase-parametric diagrams for the variable $E_1(t)$ versus the parameter k_8 changing in the appropriate intervals.

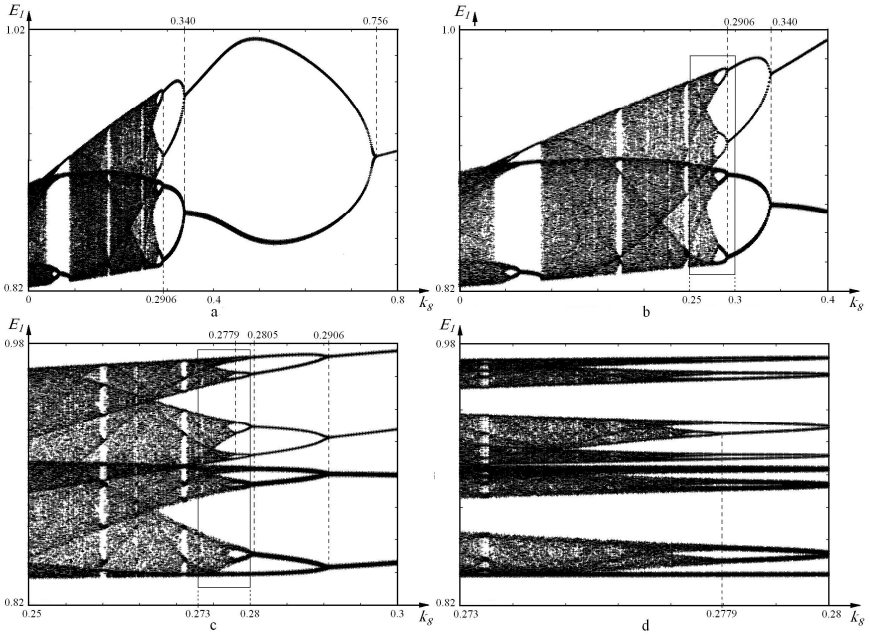

Fig. 2. Phase-parametric diagram for the variable $E_1(t)$: a - $k_8 \in (0.,0.8)$; b - $k_8 \in (0.,0.4)$; c - $k_8 \in (0.25,0.3)$; d - $k_8 \in (0.273,0.28)$.

As the parameter k_8 decreases, there occurs the subsequent doubling of the multilicity of the autoperiodic process. Such a sequence of the appearance of bifurcations creates a cascade of bifurcations, namely the Feigenbaum sequence [39]. After the multiple doubling of a period, the modes of aperiodic oscillations are eventually observed in the system. In other words, a chaos arises. As the parameter k_8 decreases further, we see the appearance of the windows of periodicity on the phase-parametric diagrams. The deterministic chaos is destroyed, and the periodic and quasiperiodic modes are established. The trajectory of a strange attractor in the chaotic mode is tightened to a regular attractor of the autoperiodic mode. We observe the self-organization in the system. Then the windows of periodicity are destroyed, and the chaotic modes arise again. Moreover, the transitions "order—chaos" and "chaos—order" happen. There occurs the adaptation of the metabolic process to varying conditions.

It is seen from the presented figures that, as the scale decreases, every subsequent phase-parametric diagram with doubling of a cycle and its windows

of periodicity are identicat to those of the previous diagram, as the scale decreases. The given sequence of bifurcations has a self-similar fractal structure.

In Figs. 3,e-f and 4, we present the examples of the projections of phase portraits for some values of parameter k_8, according to the phase-parametric diagram in Fig. 2.

In Fig. 5, we show the constructed kinetic curves for a strange attractor formed at $k_8 = 0.12$.

These figures indicate a variation of the dynamics of a metabolic process of the Krebs cycle, which depends on the intensity of formation of the final oxidation product, CO_2.

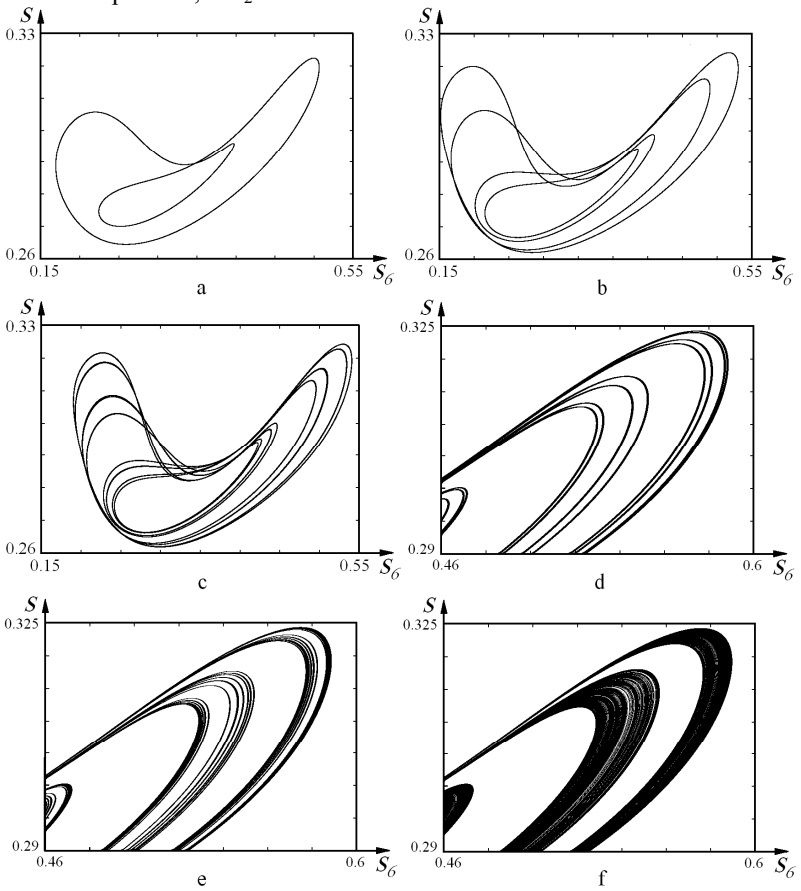

Fig. 3. Projections of system's phase portraits: a – regular attractor $2 \cdot 2^1$, $k_8 = 0.5$; b – regular attractor $2 \cdot 2^2$, $k_8 = 0.3$; c - regular attractor $2 \cdot 2^4$, $k_8 = 0.28$; d - regular attractor $2 \cdot 2^8$, $k_8 = 0.278$; e - regular attractor $2 \cdot 2^{16}$, $k_8 = 0.277$; f - strange attractor $2 \cdot 2^x$, $k_8 = 0.275$.

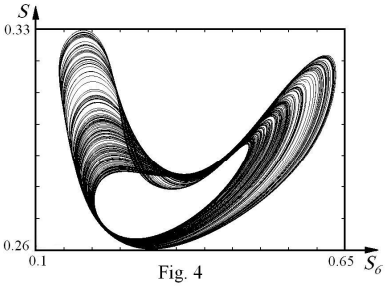

Fig. 4

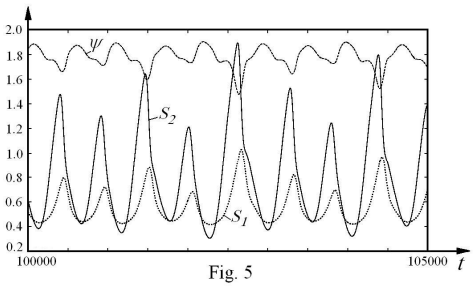

Fig. 5

Fig. 4. Projection of the phase portrait of the strange attractor $2 \cdot 2^x$ for $k_8 = 0.12$.

Fig. 5. Kinetic curve for the components S_1, S_2, and ψ of the Krebs cycle in the mode of the strange attractor $2 \cdot 2^x$ for $k_8 = 0.12$.

In order to uniquely identify the type of obrained attractors and to determine their stability, we calculated the full spectra of Lyapunov indices and their sum $\Lambda = \sum_{j=1}^{19} \lambda_j$ for the chosen points. The calculation was carried out by Benettin's algorithm with the orthogonalization of the vectors of perturbations by the Gram--Schmidt method [21].

The calculation of Lyapunov indices from this multidimensional system on a personal computer meets certain difficulties. The mathematical model of the given biochemical system contains many variables and parameters. The limitations in the solution of such problems arise due to the insufficient random-access memory of a computer in the processing of the $n \times n$ matrix of small perturbations. In addition, any inaccuracy on the stage of programming will essentially affect the redefinition of the vectors of perturbations, their orthogonalization, and, as a consequence, the result of calculations. Nevertheless, we solved the problem and obtained certain results. Below for the sake of comparison, we present the spectra of Lyapunov indices for some modes of the system. For brevity without any loss of information, we give the values of indices up to the fourth decimal point.

The ratios of the values of Lyapunov indices $\lambda_1 > \lambda_2 > \lambda_3 > ... > \lambda_{19}$ serve as the criterion of the validity of calculations. For a regular attractor, we have obligatorily $\lambda_1 \approx 0$. The remaining indices can be also ≈ 0 in some cases. In some other cases, they are negative. The zero value of the first Lyapunov index testifies to the presence of a stable limiting cycle.

For a strange attractor, at least one Lyapunov index must be positive. After it, the zero index follows. The next indices are negative. The presence of negative indices means the contraction of system's phase space in the corresponding directions, whereas the positive indices indicate the dispersion of trajectories. Therefore, there occurs the mixing of trajectories in narrow places

of the phase space of the system, i.e., there appears the deterministic chaos. The Lyapunov indices contain obligatorily the zero index, which means the conservation of the aperiodic trajectory of an attractor in some region of the phase space and the existence of a strange attractor.

For $k_8 = 0.01$, the strange attractor $2 \cdot 2^x$ arises. We have $\lambda_1 - \lambda_{19}$: .0007; .0000; -.0040; -.0125; -.0196; -.0200; -.0290; -.0299; -.0317; -.0416; -.0416; -.0416; -.0458; -.0816; -.0874; -.0874; -.1181; -.1539; -.2222; $\Lambda = -1.0672$.

For $k_8 = 0.075$ – regular attractor $3 \cdot 2^0$ (see the window of periodicity in Fig. 2,b). $\lambda_1 - \lambda_{19}$: .0000; -.0004; -.0040; -.0117; -.0194; -.0211; -.0285; -.0285; -.0326; -.0406; -.0406; -.0406; -.0451; -.0819; -.0883; -.0883; -.1182; -.1563; -.2241; $\Lambda = -1.0702$.

For $k_8 = 0.12$ – strange attractor $2 \cdot 2^x$; $\lambda_1 - \lambda_{19}$: .0008; .0000; -.0040; -.0125; -.0192; -.0210; -.0287; -.0300; -.0324; -.0406; -.0406; -.0406; -.0457; -.0822; -.0879; -.0879; -.1172; -.1542; -.2212; $\Lambda = -1.0653$.

For $k_8 = 0.27$ – strange attractor $2 \cdot 2^x$; $\lambda_1 - \lambda_{19}$: .0002; .0000; -.0040; -.0125; -.0192; -.0219; -.0281; -.0310; -.0320; -.0387; -.0387; -.0387; -.0436; -.0842; -.0889; -.0889; -.1185; -.1541; -.2209; $\Lambda = -1.0638$.

For $k_8 = 0.275$ – strange attractor $2 \cdot 2^x$; $\lambda_1 - \lambda_{19}$: .0001; .0000; -.0040; -.0125; -.0192; -.0219; -.0280; -.0312; -.0323; -.0383; -.0383; -.0383; -.0434; -.0842; -.0890; -.0890; -.1184; -.1540; -.2212; $\Lambda = -1.0631$.

For $k_8 = 0.278$ – regular attractor $2 \cdot 2^8$; $\lambda_1 - \lambda_{19}$: .0000; .0000; -.0041; -.0123; -.0193; -.0219; -.0283; -.0308; -.0320; -.0384; -.0384; -.0384; -.0435; -.0842; -.0890; -.0890; -.1187; -.1538; -.2212; $\Lambda = -1.0634$.

For $k_8 = 0.278$ – regular attractor $2 \cdot 2^8$; $\lambda_1 - \lambda_{19}$: .0000; .0000; -.0041; -.0123; -.0193; -.0219; -.0283; -.0308; -.0320; -.0384; -.0384; -.0384; -.0435; -.0842; -.0890; -.0890; -.1187; -.1538; -.2212; $\Lambda = -1.0634$.

For $k_8 = 0.28$ – regular attractor $2 \cdot 2^4$; $\lambda_1 - \lambda_{19}$: .0000; .0000; -.0041; -.0123; -.0193; -.0219; -.0282; -.0307; -.0322; -.0384; -.0384; -.0384; -.0436; -.0842; -.0890; -.0890; -.1186; -.1540; -.2212; $\Lambda = -1.0633$.

The presented results of calculations indicate that the sum Λ of all indices, which determine the flow divergencies and, hence, the evolution of the phase volume along the trajectory, is maximal for the regular attractor $3 \cdot 2^0$. It arises in the window of periodicity for $k_8 = 0.075$ ($\Lambda = -1.0702$). For the strange attractors on the left and on the right (for $k_8 = 0.01$ and $k_8 = 0.12$), the divergencies are, respectively, $\Lambda = -1.0672$ and $\Lambda = -1.0653$. This means that the phase volume element for the given attractor is contracted, on the whole, stronger along the trajectory. Here, we observe the self-organization of a stable cycle from chaotic modes. The Krebs cycle is adapted to the varying conditions.

By the given Lyapunov indices for strange attractors, we determine the KS-entropy (the Kolmogorov--Sinai entropy) [40]. By the Pesin theorem [41], the KS-entropy h corresponds the sum of all positive Lyapunov characteristic indices:

The KS-entropy allows us to judge about the rate, with which the information about the initial state of the system is lost. The positivity of the given entropy is a criterion of the chaos. This gives possibility to qualitatively estimate the properties of attractor's local stability.

We determine also the quantity inverse to the KS-entropy, $t_{\min}$. This is the time of a mixing in the system. It characterizes the rate, with which the initial conditions will be forgotten. For $t << t_{\min}$, the behavior of the system can be predicted with sufficient accuracy. For $t > t_{\min}$, only a probabilistic description is possible. The chaotic mode is not predictable due to the loss of the memory of initial conditions. The quantity $t_{\min}$ is called the Lyapunov index and characterizes the "predictability horizon" of a strange attractor.

In order to classify the geometric structure of strange attractors, we calculated the dimension of their fractality. The strange attractors are fractal sets and have the fractional Hausdorff-Besicovitch dimension. But its direct calculation is a very labor-consuming task possessing no standard algorithm. Therefore, as a quantitative measure of the fractality, we calculated the Lyapunov dimension of attractors by the Kaplan--Yorke formula [42, 43]:

$$D_{F_r} = m + \frac{\sum_{i=1}^{m} \lambda_i}{|\lambda_{m+1}|}, \tag{11}$$

where m is the number of the first Lyapunov indices ordered by their decreasing. Their sum $\sum_{i=1}^{m} \lambda_i \geq 0$, and $m+1$ is the number of the first Lyapunov index, whose value $\lambda_{m+1} < 0$.

For the above-considered strange attractors 2^{∞}, we obtained the following indices.

For $k_8 = 0.01$: $h = 0.0007$, $t_{\min} = 1428.6$, $D_{F_r} = 2.175$.

For $k_8 = 0.12$: $h = 0.0008$, $t_{\min} = 1250$, $D_{F_r} = 2.2$.

For $k_8 = 0.27$: $h = 0.0002$, $t_{\min} = 5000$, $D_{F_r} = 2.05$.

For $k_8 = 0.275$: $h = 0.0001$, $t_{\min} = 10000$, $D_{F_r} = 2.025$.

By these indices, we can judge about the difference of the given strange attractors.

Conclusions

With the help of the mathematical model of the Krebs cycle, we have studied the dependence of the cyclicity of the metabolic process on the amount of a final product of the oxidation, i.e., on the amount of the formed carbon dioxide. The multiplicity of the cycle is doubled by the Feigenbaum scenario, until the aperiodic modes of strange attractors arise. From them as a result of the self-organization, the stable periodic modes appear. This means that the system is adapted to the varying conditions. We have calculated the full spectra of Lyapunov indices and the divergencies for various modes. For the strange attractors, we have determined the KS-entropies, "predictability horizons," and Lyapunov dimensions of attractors. The results obtained allow us to study the structural-functional connections of the cycle of tricarboxylic acids, their influence on the cyclicity of metabolic oscillations in a cell, and the physical laws of self-organization in it.

The work is supported by the project No. 0113U001093 of the National Academy of Sciences of Ukraine.

References

1. H. A. Krebs and W. A. Johnson. The role of citric acid in intermediate metabolism in animal tissues. *Enzymologia*, 4, 148–156, 1937.
2. R. Bohnensack and E .E. Sel'kov. Stoichiometric regulation in the citric acid cycle. II. Non-linear interactions. *Studia biophysica*, 66, 47–63, 1977.
3. A. E. Lyubarev and B. I. Kurganov. Supermolecular organization of enzymes of the cycle of tricarboxylic acides. *Molek. Biol*., 21, 5, 1286–1296, 1987.
4. M. N. Kondrashova. Structural kinetic organization of Supermolecular organization of enzymes of the cycle of tricarboxylic acides under the active functioning of mitochondria. *Biofiz*., 34, 3, 450–457, 1989.
5. E. M. T. El-Mansi, G. C. Dawson and C. F. A. Bryce. Steady-state modelling of metabolic flux between the tricarboxylic acid cycle and the glyoxylate bypass in Escherichia coli. *Comput. Applic. Biosci*., 10, 3, 295–299, 1994.
6. R. Ramakrishna, J.S. Edwards, A. McCulloch and B. O. Palsson. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. *Am. J. Physiol. Regul. Integr. Comp. Physiol*., 280, 3, R695–R704, 2001.
7. S. Cortassa, M. A. Aon, E. Marban, R. L. Winslow and B. O'Rourke. An integrated model of cardiac mitochondrial energy metabolism and calcium dynamics. *Biophys. J*., 84, 2734–2755, 2003.
8. K. Yugi and M. Tomita. A general computational model of mitochondrial metabolism in a whole organelle scale. *Bioinformatics*, 20, 1795–1796, 2004.
9. V. K. Singh and I. Ghosh. Kinetic modeling of tricarboxylic acid cycle and glyoxylate bypass in Mycobacterium tuberculosis, and its application to assessment of drug targets. *Theor. Biol. and Med. Model.*, 3, 27, 2006.
10. E. Mogilevskaya, O. Demin and I. Goryanin. Kinetic model of mitochondrial Krebs cycle: unraveling the mchanism of salicylate hepatotoxic effects. *J. of Biol. Phys*., 32, 3-4, 245– 271, 2006.
11. V. P. Gachok. *Kinetics of Biochemical Processes*, Naukova Dumka, Kiev, 1988 (in Russian).

12. V. P. Gachok. *Strange Attractors in Biosystems*, Naukova Dumka, Kiev, 1989 (in Russian).
13. W. B. Armiger, A. R. Moreira, J. A. Phillips and A. E. Humphrey. Modeling cellulose digestion for single cell protein. *Utilization of Cellulose Materials in Inconventional Food Production*, Plenum Press, New York, 1979, 111–117.
14. V. I. Grytsay and I. V. Musatenko. Self-organization and fractality in a metabolic processes of the Krebs cycle. *Ukr. Biokh. Zh*., 85, 5, 191–199, 2013.
15. E. E. Selkov. Self-oscillations in glycolysis. *Europ. J. Biochem*., 4, 79–86, 1968.
16. B. Hess and A. Boiteux. Oscillatory phenomena in biochemistry. *Annu. Rev. Biochem*., 40, 237– 258, 1971.
17. A. Goldbeter and R. Lefer. Dissipative structures for an allosteric model. Application to glycolytic oscillations. *Biophys J*., 12, 1302–1315, 1972.
18. A. Godlbeter and R. Caplan. Oscillatory enzymes. *Annu. Rev. Biophys. Bioeng*., 5, 449–476, 1976.
19. *Chaos in Chemical and Biochemical Systems*, edited by R. Field, L. Györgyi, World Scientific, Singapore, 1993.
20. V. S. Anishchenko. *Complex Oscillations in Simple Systems*, Nauka, Moscow, 1990 (in Russian).
21. S. P. Kuznetsov. *Dynamical Chaos*, Nauka, Moscow, 2001 (in Russian).
22. V. P. Gachok and V. I. Grytsay. Kinetic model of macroporous granule with the regulation of biochemical processes. *Dokl. Akad. Nauk SSSR*, 282, 51–53, 1985.
23. V. P. Gachok, V. I. Grytsay, A. Yu. Arinbasarova, A. G. Medentsev, K. A. Koshcheyenko and V. K. Akimenko. Kinetic model of hydrocortisone 1-en dehydrogenation by Arthrobacter globiformis. *Biotechn. Bioengin.*, 33, 661–667, 1989.
24. V. P. Gachok, V. I. Grytsay, A. Yu. Arinbasarova, A. G. Medentsev, K. A. Koshcheyenko and V. K. Akimenko. A kinetic model for regulation of redox reactions in steroid transformation by Arthrobacter globiformis cells. *Biotechn. Bioengin.*, 33, 668–680, 1989.
25. V. I. Grytsay. Self-organization in the macroporous structure of the gel with immobilized cells. Kinetic model of bioselective membrane of biosensor. *Dopov. Nats. Akad. Nauk Ukr.*, 2, 175–179, 2000.
26. V. I. Grytsay. Self-organization in a reaction-diffusion porous media. *Dopov. Nats. Akad. Nauk Ukr*., 3, 201–206, 2000.
27. V. I. Grytsay. Ordered structure in a mathematical model biosensor. *Dopov. Nats. Akad. Nauk Ukr.*, 11, 112–116, 2000.
28. V.I. Grytsay. Self-organization of biochemical process of immobilized cells of bioselective membrane biosensor. *Ukr. J. Phys.*, 46, 1, 124–127, 2001.
29. V. V. Andreev and V. I. Grytsay. The modeling of nonactive zones in porous granules of a catalyst and in a biosensor. *Matem. Modelir.*, 17, 2, 57–64, 2005.
30. V. V. Andreev and V. I. Grytsay. Influence of heterogeneity of diffusion-reaction process for the formation of structures in the porous medium. *Matem. Modelir.*, 17, 6, 3–12, 2005.
31. V. I. Grytsay and V. V. Andreev. The role of diffusion in the active structures formation in porous reaction-diffusion media. *Matem. Modelir.*, 18, 12, 88–94, 2006.
32. V. I. Grytsay. Unsteady conditions in porous reaction-diffusion. *Medium. Romanian J. Biophys.*, 17, 1, 55–62, 2007.
33. V. I. Grytsay. The uncertainty in the evolution structure of reaction-diffusion medium bioreactor. *Biofiz. Visn.*, 2, 92–97, 2007
34. V. I. Grytsay. Formation and stability of morphogenetic fields of immobilized cell in bioreactor. *Biofiz. Visn.*, 2, 25–34, 2008.

35. V. I. Grytsay. Structural instability of a biochemical process. *Ukr. J. Phys.,* 55, 5, 599–606, 2010.
36. V. I. Grytsay and I. V. Musatenko. Self-oscillatory dynamics of the metabolic process in a cell. *Ukr. Biochem. J.*, 85, 2, 93–104, 2013.
37. V. I. Grytsay and I. V. Musatenko. The structure of a chaos of strange attractors within a mathematical model of the metabolism of a cell. *Ukr. J. of Phys.*, 58, 7, 677–686, 2013.
38. V. I. Grytsay and I. V. Musatenko. A mathematical model of the metabolism of a cell. Self- organisation and chaos. Chaotic modeling and simulation. *CMSIM*, 4, 539-552, 2013.
39. M. J. Feigenbaum. Quantative universality for a class of nonlinear transformations. *J. Stat. Phys*., 19, 1, 25–52, 1978.
40 A. N. Kolmogorov. On the entropy per unit time as a metric invariant of automorphisms. *DAN SSSR*, 154, 754–755, 1959.
41 Ya. B. Pesin. Characteristic Lyapunov indices and the ergodic theory. *Usp. Mat. Nauk*, 32, 4, 55–112, 1977.
42 J. L. Kaplan and J. A. Yorke. The onset of chaos in a fluid flow model of Lorenz. *Ann. N. Y. Acad. Sci.*, 316, 400–407, 1979.
43 J. L. Kaplan and J. A. Yorke. A chaotic behaviour of multidimensional differential equations. *Functional Differential Equations of Fixed Points*, edited by H. O. Peitgen, H. O. Walther, Springer, Berlin, 1979, 204-227.

Chaotic Modeling and Simulation (CMSIM) 3: 221-231, 2014

Key agreement protocol based on extended chaotic maps with anonymous authentication

Ping Zhen[1], Geng Zhao[2], Lequan Min[3] and Xiaodong Li[2]

[1] School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, 100083, China
(E-mail: zhenping1989@126.com)
[2] Beijing Electronic Science and Technology institute, Beijing, 100070, China
(E-mail: zg@besti.edu.cn, lxd@besti.edu.cn)
[3] School of Mathematics and Physics, University of Science and Technology Beijing, Beijing, 100083, China
(E-mail: minlequan@sina.com)

Abstract. Key agreement protocol is used to establish shared secret key for the network system, which is quite important to guarantee secure communication. This paper proposes a two-party key agreement protocol. In order to improve the efficiency and enhance the security, we utilize extended chaotic maps to generate the shared key, which can be used to encrypt and decrypt the transmitted messages in the subsequent communications. The proposed protocol can guarantee anonymity of user's identity and provide mutual authentication. In addition, it also can resist various attacks. The explicit analysis show that the protocol is secure, reliable and applicable in practice.
Keywords: Key agreement protocol, Chaotic maps, Anonymous authentication.

1 Introduction

Key agreement protocols are basic to modern cryptography, which are used to guarantee the security of secret keys which are exchanged over the insecure public network. The shared keys are used in the subsequent communication for encryption, authentication, access control, and so on. In 1976, Diffie and Hellman[1] introduced the first key agreement protocol. However, both of communication parties don't verity the identity of each other and it is vulnerable to man-in-the-middle attack. In order to solve the problem, an authenticated key agreement protocol[2] is proposed. The authenticated key agreement not only allow two parties to agree on a session key, but also ensure the authentication of the participant. Since then, many related key agreement protocols have been proposed[3-5].

Chaotic systems have complicated behaviors, which are sensitive to initial conditions and system parameters, and are not predictable in the long term. These properties, as required by several cryptographic primitives, render chaotic systems a potential candidate for constructing cryptosystem. The application of

Received: 14 July 2014 / Accepted: 30 July 2014

ISSN 2241-0503

chaotic maps in cryptography has been studied for more than twenty years. There are chaos-based symmetry key cryptosystem[6,7], public key cryptosystem[8,9], Hash functions [10,11], and so on.

In 2005, Xiao et al.[12] proposed a chaos-based key agreement protocol, which utilizes Chebyshev chaotic maps. Alvarez[13] demonstrated this protocol is vulnerable to man-in-the-middle attack. Xiao et al.[5] proposed an improved key agreement to enhance the security, but Han et al.[14] pointed out the improved protocol cannot resist the replay attack. Tseng et al.[15] proposed an anonymous key agreement protocol using smart cards. Niu et al.[16] demonstrated the protocol is vulnerable to the insider attacker and cannot protect user anonymity and then proposed a new key agreement protocol, which is also proved to have low computational efficiency problem by Yoon[17].

Recently, Tan[18] proposed a novel authenticated key agreement protocol with strong anonymity, which is based on smart cards. However, the expense of smart cards and readers will make the protocols costly in practical use. In Ref.[19], Gong et al. proposed a secure chaotic maps-based key agreement protocol without using smart cards and claimed that the protocol is secure. Wang et al.[20] pointed out that there are some problems existing in Gong et al.'s protocol, such as the stolen-verifier attack, forged message flood and key management problems. Then they proposed a new key agreement protocol. We have explicitly analyzed Wang et al.'s protocol. The protocol cannot provide the anonymity of users' identities. But in many insecure channels, especially in e-commerce applications, anonymity is also an very important issue. There also exits key distribution and management problems, which can be easily avoided. Lee et al.[21] proposed a three-party password-based authenticated key exchange protocol with user anonymity. However, the introduced trusted third party not only adds extra overhead, but also becomes another security and performance bottleneck, which will bring potential threats to the system. Motivated by this, this paper proposed a two-party key agreement protocol with anonymous authentication. an anonymous authenticated key agreement protocol based on extended chaotic maps to solve these problems. It doesn't need smart cards and at the same time preserves user anonymity. Besides, "two-party" will decrease the computation and communication cost and at the same time make the protocol secure and efficient. Explicit security analysis and performance analysis of the proposed protocol are also given in this paper.

This paper is organized as follows. Section 2 introduces the preliminaries about extend Chebyshev chaotic maps. Then the proposed two-party key agreement protocol is described in section 3. Security and performance analysis are given in section 4 and section 5 separately. The last section presents the conclusions.

2 Preliminaries

Definition 1. Let $n \in Z^+$ and $x \in [-1,1]$, then a Chebyshev polynomial 0 of order n, $T_n(x):[-1,1] \to [-1,1]$ is defined as:

$$T_n(x) = \cos(n \cdot \arccos(x))$$

It is recursively defined using the following recurrent relation:

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x), n \geq 2$$

where $T_0(x) = 1$ and $T_1(x) = x$.

The first few Chebyshev polynomials are

$$T_2(x) = 2x^2 - 1$$

$$T_3(x) = 4x^3 - 3x$$

$$T_4(x) = 8x^4 - 8x^2 + 1$$

L

The Chebyshev polynomials exhibit the following important properties: the semigroup property and the chaotic property.

(1) The semi-group property:

$$\begin{aligned} T_r(T_s(x)) &= \cos(r\cos^{-1}(\cos(s\cos^{-1}(x)))) \\ &= \cos(rs\cos^{-1}(x)) \\ &= T_{sr(x)} \\ &= T_s(T_r(x)) \end{aligned}$$

r and s are positive integer numbers and $x \in [-1,1]$.

(2) The chaotic property

When the degree $n > 1$, the Chebyshev polynomial map $T_n(x):[-1,1] \to [-1,1]$ of degree n is a chaotic map with its invariant density $f^*(x) = 1/(\pi\sqrt{1-x^2})$, and positive Lyapunov exponent $\lambda = \ln n > 0$.

To improve security, Zhang[22] proved that the semi-group property holds for extend Chebyshev polynomials defined on $(-\infty, +\infty)$, which can enhance the property, as follows:

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x) \bmod P$$

where $n \geq 2$ and P is a large prime. We can also obtain:

$$T_r(T_s(x)) \equiv T_{sr}(x) \equiv T_s(T_r(x)) \bmod P$$

Definition 2 The discrete logarithm problem (DLP) is explained by the following: Given an element y, the task of DLP is to find the integer s, such that $T_s(x) = y$.

Definition 3 The Diffie-Hellman problem (DHP) is explained by the following: Given the elements $T_r(x)$ and $T_s(x)$, the task of DHP is to compute $T_{rs}(x)$.

It is generally believed that there is no polynomial time algorithm to solve the DLP and DHP problems with non-negligible probability.

Table 1. The notations in the protocol

Notations	Descriptions
ID_i	Identity of client U_i
ID_S	Identity of server S
$E_k(\cdot), D_k(\cdot)$	Secure symmetric encryption and decryption
$H(\cdot)$	Secure one-way hash function
$T_k(\cdot)$	Cheybeshev chaotic map
x	The seed of Chebyshev chaotic map
r, s, r_1, r_2	The degree of Chebyshev chaotic map
PW_i	Password of client U_i
K_S	The secret key of server S
T_1, T_2, T_3	Time stamps
$\Delta T_1, \Delta T_2$	The specified valid time period
sn	The session identifier
KA	The established shared session key

3 The proposed protocol

This section will present our proposed two-party key agreement protocol based on extended Chebyshev chaotic maps. It consists of four phases: (1) the parameter generation phase; (2) the registration phase; (3) the key agreement phase; (4) the password updation phase. For the easy understanding of subsequent content, the commonly used notations are listed in Table 1.

1. Parameter generation phase

In order to perform the protocol, the server S firstly needs to generate some parameters as follow:

(1) S selects a secure symmetric cryptosystem with encryption $E_k(\cdot)$ and decryption $D_k(\cdot)$, where k is the key of symmetric cryptosystem;

(2) S selects a secure one-way hash function $H(\cdot)$;

(3) S select a private key K_S ,which is specialized for client registration.

(4) Utilizes the public key cryptosystem based on Cheybshev chaotic maps, S chooses two random large integers x and s as the seed and degree of Chebyshev maps respectively and computes $T_s(x)$. Then publish $(x, T_s(x))$ as the public parameters and keep s private.

2. Registration phase

The Client U_i with the identity ID_i registers with server S by the following two steps:

(1) U_i selects a password PW_i ,and sends the ID_i and PW_i to S through a secure channel.

(2) After receiving ID_i and PW_i , S use its private key K_S to computes $M_{reg} = H(ID_i, PW_i, K_S)$ and store M_{reg} as the register message securely.

3. Key agreement phase

The client and server need to perform the following four steps to realize mutual authentication and establish a common session key to complete the protocol. The simplified description of the phase is shown in Fig.1. The details are described in the following steps:

(1) $U_i \rightarrow S: M_1 = \{T_{r_1}(x), C_1 = E_{SK}(sn, ID_i, ID_S, PW_i, T_{r_1}(x), T_1)\}$.

U_i selects a random large integer r_1 , and computes $T_{r_1}(x)$ and $SK = T_{r_1}(T_s(x))$. SK is used as the temporary key of symmetric cryptosystem to compute $C_1 = E_{SK}(sn, ID_i, ID_S, PW_i, T_{r_1}(x), T_1)$, where sn is a session identifier and T_1 is a timestamp. Then U_i sends the message $M_1 = \{T_{r_1}(x), C_1\}$ to the server.

(2) $S \rightarrow U_i: M_2 = \{sn, C_2 = E_{SK}(sn, T_{r_2}(x), H_1 = H(KA, ID_S), T_1)\}$.

After receiving the message M_1 , S first compute $SK = T_s(T_{r_1}(x))$ and use it to decrypt C_1 . Then S checks whether $|T_2 - T_1| \le \Delta T_1$,where T_2 is the current timestamp and ΔT_1 is the specified valid time period. S continues to compute $M_{reg}' = H(ID_i, PW_i, K_S)$ and validates whether $M_{reg}' = M_{reg}$. If so, S can authenticate the identity of client U_i , otherwise, the process will be terminated immediately. S selects a random large integer r_2 , and computes $T_{r_2}(x)$, $KA = T_{r_2}(T_{r_1}(x))$, $H_1 = H(KA, ID_S)$ and $C_2 = E_{SK}(sn, T_{r_2}(x), H(KA, ID_S), T_1)$. S sends the message $M_2 = \{sn, C_2\}$ to the client.

(3) $U_i \rightarrow S: M_3 = \{sn, H_2 = H(sn, ID_i, KA)\}$.

Upon receiving the message M_2 from S , U_i first decrypts C_2 with the secret key SK . Then U_i checks whether $|T_3 - T_1| \le \Delta T_2$,where T_3 is the current timestamp. U_i computes $KA = T_{r_1}(T_{r_2}(x))$ and $H_1' = H(KA, ID_S)$, and validates whether $H_1' = H_1$. If so , U_i will authenticate the identity of S . Any fail will lead to the termination of the protocol. U_i continues to compute $H_2 = H(sn, ID_i, KA)$ and sends $M_3 = \{sn, H_2\}$ to the server.

(4) Having received the message M_3 from the client U_i, S will compute $H_2' = H(sn, ID_i, KA)$ and check whether $H_2' = H_2$. If so, the server S can affirm that U_i has received KA and KA will be the common session key used in the subsequent communications.

4. Password updation phase

If the client U_i want to update the password, U_i and S need to perform the following steps:

(1) U_i selects a random large integer r, and computes $T_r(x)$ and $K_{PW} = T_r(T_s(x))$. Similar with the first step in key agreement phase, K_{PW} will be used as the temporary key of symmetric cryptosystem. Then U_i encrypts $C_{PW} = E_{K_{PW}}(ID_i, PW_i, PW_i', T_r(x))$ and sends and $M_{PW} = \{T_r(x), C_{PW}\}$ to the server, where PW_i' is the updated password.

(2) Having received the message M_{PW} from U_i, S firstly computes $K_{PW} = T_s(T_r(x))$ and decrypts M_{PW}. Then S checks the validity of ID_i and PW_i. If so, then S continues to computes $M_{reg}' = H(ID_i, PW_i', K_S)$ and store M_{reg}' as the updated register message securely.

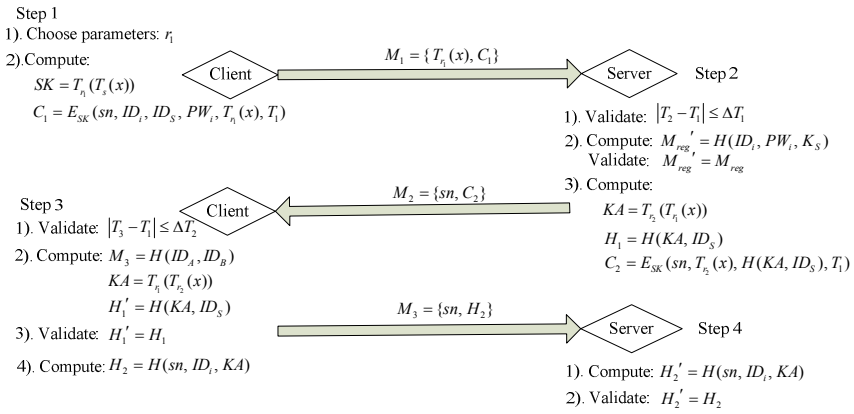

Fig. 1. The key agreement phase of the proposed protocol

4 Security analysis

In this section, we will analyze the security of the proposed protocol and show it can resist various attacks. Here, we claim that our protocol satisfy the following security properties:

(1) **Identity anonymity** With the popularization of internet application, identity privacy has become an important requirement. Identity anonymity means that in the key agreement phase, the attacker cannot find the information about user's ID by intercepting the communication messages. The attacker may eavesdrop the communication channel and try to find some sensitive information to trace the real identity. In the proposed protocol, the identity of Client and Server are encrypted by secure symmetric cryptosystem $C_1 = E_{SK}(sn, ID_i, ID_S, PW_i, T_{r_1}(x), T_1)$. In order to decrypt , the attack need the temporary secret key , which involve the DHP difficult problem mentioned in section 2. Only the server can decrypt the message and get the identity information. Thus, anonymity can be achieved during the key agreement phase.

(2) **Mutual authentication** The goal of mutual authentication is to confirm both the identities of the client and server and establish a common shared session key between them. In step 2 of the key agreement phase, only the server can decrypt the message $C_1 = E_{SK}(sn, ID_i, ID_S, PW_i, T_{r_1}(x), T_1)$ and authenticate the identity of the client by comparing the ID_i and PW_i with registered message M_{reg}. Client can authenticate the identity of server by the session identifier sn and comparing hash value $H_1' = H(KA, ID_S)$. The illegal attacker may modify the communication messages being transmitted over an insecure network. It is extremely difficult for the attacker to fabricate the false authentication information and any message modification during transmission will be detected by the protocol participant. So the proposed protocol can achieve the mutual authentication.

(3) **Resistance to tamper attacks** A tamper attack is an attempt by an adversary to modify information in an unauthorized manner. This is an attack against the integrity of the information. We have stressed the problem in the analysis above and will explain how our protocol can resist this attack in this part. In the key agreement phase, the session identifier sn and $T_{r_1}(x)$ are transmitted in the plaintext form and ciphertext form, respectively, which is used to validate whether the plaintext or cipherctext is being tampered. What is more, hash function is also utilized to further realize message integrity. If the adversary forges the message, the receiver can detect it by checking Hash value immediately. This leads to the termination of the protocol. According to the analysis, our protocol can resist the tamper attacks.

(4) **Fairness in the key agreement** The property fairness in the key agreement is also called the contributory property, which means that the session key is determined cooperationally by both the communicating parties. In 0, the author has given a strictly formal definition. The fairness in key agreement means that any communicating party cannot decide a shared session key in advance. In this protocol, we can see client and server choose random integers r_1 and r_2 separately. Through the commutative property of extended Chebyshev chaotic map, they can compute the shared session

key $KA = T_{r_1}(T_{r_2}(x)) = T_{r_2}(T_{r_1}(x))$. Therefore, the protocol can ensure the fairness in the key agreement.

(5) **Resistance to man-in-the-middle attack** Man-in-the-middle means that an active attacker intercepts the communication messages between communication participants and adopts some special means to successfully masquerade as the both parties communicate with each other. From previous analysis, the attack even doesn't know the identities of communicating parties since they are kept anonymous and any modification to the transmitted message will be detected. So the attacker cannot impersonate one participant to another during key agreement process. Therefore, the proposed protocol can withstand man-in-the-middle attack.

(6) **Resistance to replay attack** A replay attack is an offensive action in which an adversary impersonates or deceives another legitimate participant through the reuse of information obtained in a protocol. The proposed protocol can resist the replay attacks, which is realized by using the session identifier sn and time stamps (T_1, T_2, T_3). Time stamp is attached to verify freshness of every transmitted message. Furthermore, it cannot be modified because it is encrypted during transmission process. Thus, it is impossible for the replayed message to pass the verification with incorrect session identifier and timestamp. Therefore, our protocol can resist replay attack.

(7) **Resistance to password-based attacks** Dictionary attack is always used to crack the password in the protocol. There are three kinds of dictionary attack[21]: Off-line dictionary attack, undetectable on-line dictionary attack and detectable on-line dictionary attack. Both off-line and undetectable on-line dictionary attack can cause serious consequences among them. In the key agreement phase, the attacker needs to decrypt the message $C_1 = E_{SK}(sn, ID_i, ID_S, PW_i, T_{r_1}(x), T_1)$ to steal the password PW_i. To obtain the secret key SK, the attack faces the DHP difficult problem. So the attacker cannot launch any of these attacks. Therefore, our protocol is quite effective to resist password-based attacks.

(8) **Resistance to stolen-verifier attack** Then stolen-verifier attack means that an adversary who steals the password verification information from the server can use it directly to masquerade as a legitimate user in authentication phase[16]. In the protocol, we assume the registered message $M_{reg} = H(ID_i, PW_i, K_S)$ is safely stored by the server and cannot be accessed by the attacker. Even if it is stolen, the attacker still cannot carry out the stolen-verifier attack to get the client's password PW_i without the server's secret key K_S. So the secret key K_S can strength the security of password and resist the stolen-verifier attack.

(9) **High efficiency in key distribution and management** It need Server S to publish its public parameters $(x, T_s(x))$ and store the registered value $M_{reg} = H(ID_i, PW_i, K_S)$. Each entity only needs to keep his own password PW_i. This will improve the performance of the key distribution.

What's more, the symmetric secret keys SK are established temporarily utilizing the Chebyshev semigroup property and will be altered in each session according to the selected random numbers r_1 .So the communication entity does not need to store SK and it can decrease the key management cost and strengthen the security.

5 Performance analysis

In this section, we will compare the performance and security of our protocol with Tseng et al.'s protocol[15] and Wang et al.'s protocol[20]. For the convenience of evaluating the computational complexity, let T_X, T_S, T_C and T_H be the computation cost of one XOR operation, one symmetric encryption/decryption operation, one Chebyshev polynomial computation and one Hash operation, respectively. From table 2, we can see that our key agreement protocol need $(T_S + T_C)$ more computation cost for the client and $(T_S + T_C + T_H)$ more for the server than Wang et al.'s. In practical use, symmetric encryption/decryption and hash function can be quite efficient. As for the Chebyshev operation, the authors in[5,24,25] gave some implementation methods to decrease the computational cost. Our protocol provides user anonymity and can be more efficient in key distribution and management compared to Wang et al.'s protocol. What's more, our two-party protocol can decrease the communication cost. Our protocol only needs 3 times message transmission, which the number is 4 in Wang et al.'s protocol.

Table 2: Performance analysis and comparisons

	Tseng et al.'s	Wang et al.'s	Our protocol
User anonymity	No	No	Yes
Mutual authenticity	No	Yes	Yes
Fairness	Yes	Yes	Yes
Man-in-the-middle attack	No	No	No
Replay attack	No	No	No
Password-based attack	No	No	No
Stolen-verifier attack	No	No	No
Cost of Client	$2T_X + 2T_S + 2T_C + 5T_H$	$T_s + 2T_C + 2T_H$	$2T_s + 3T_C + 2T_H$
Cost of Server	$T_X + 2T_S + 2T_C + 3T_H$	$T_s + 2T_C + 2T_H$	$2T_s + 3T_C + 3T_H$

Conclusions

In this paper, we propose a two-party key agreement protocol based on extended chaotic maps. It securely establishes a shared session key, and provides identity anonymity and mutual authentication at the same time. It is demonstrated that

the protocol can resist various attacks, such as man-in-the-middle attack, replay attack, stolen-verifier attack, and so on. The protocol is also very efficient in key distribution and management. Compared with some previously proposed protocols, our protocol has shown its advantage in security and efficiency, which can be applicable in practical use. However, the two-party party protocol may not be suitable in large peer-to-peer network situations, which still needs further research.

Acknowledgements

The authors would like to thank the anonymous reviewers for helpful comments and suggestions. This research is supported by the National Natural Science Foundation of China (No. 61170037) and the specialized Research Fund for Doctoral Program of Higher Education of China (No. 06198016).

References

1. W. Diffe, Hellman, and M. E., New directions in cryptography, IEEE Trans. Inf. Theory, vol.22, no.6, pp. 644-654, 1976, doi:10.1109/TIT.1976.1055638.
2. M. Bellare, D. Pointcheval and P. Rogaway, Authenticated key agreement secure against dictionary attacks, Advances in Cryptography, Eurocrypt'00, Bruges, Belgium. LNCS, 2000, vol. 1807, pp. 139-155.
3. T. Y. Change, W. P. Yang and M. S. Hwang, Simple authenticated key agreement and protected password change protocol, Comput. Math. Appl., vol.49, no.5-6, pp.703-714, April-May 2005, doi: 10.1016/j.camwa.2004. 11.007.
4. T. Y. Change, M. S. Hwang, and W. P. Yang, A communication efficient three-party password authenticated key exchange protocol, Inf. Sci., vol.181, no.1, pp.217-226, January 2011, doi: 10.1016/j.ins.2010. 08.032.
5. D. Xiao, X. F. Liao and S. J. Deng, A novel key agreement protocol based on chaotic maps, Inf. Sci., vol.177, no.4,15, pp. 1136-1142, February 2007, doi: 10.1016/j.ins.2006.07.026.
6. H. Liu and X. Wang, Color image encryption based on one-time keys and robust chaotic maps, Computers & Mathematics with Applications, vol.59, no.10, pp.3320-3327. May 2010, doi: 10.1016/j.camwa.2010.03.017.
7. W. Zhen, H. Xia, L. Ning and S. X. Na, Image encryption based on a delayed fractional-order chaotic logistic system, Chin. Phys. B, , Vol. 21, No.5, 2012, doi: 10.1088/1674-1056/21/5/050506.
8. B. Ranjan, Novel public key encryption technique based on multiple chaotic systems, Phys. Rev. Lett, vol.95, no.9, pp. 098702, September, 2005, doi:10.1103/Phys RevLett. 95.09870.
9. L. Kocarev and Z. Tasev, Public-key encryption based on Chebyshev maps, In: Proceedings of the IEEE international symposium on circuits system, 25-28 May 2003, vol.3, pp.28-31, doi: 10.1109/ISCAS.2003.1204947.
10. Y. Wang, X. F. Liao, D. Xiao D and K. W. Wong, One-way Hash function construction based on 2D coupled map lattices, Inf. Sci., vol.178, no.5, pp.1391–1406, March 2008, doi: 10.1016/j.ins.2007.10.008.

11. M. Amin, O. S. Faragallah, A. A. A. El-latif, Chaos-based Hash function (CBHF) for cryptographic applications, Chaos Solitons & Fractals, vol.42, no.2,30, pp.767–772, October 2009, doi: 10.1016/j.chaos.2009. 02.001.
12. D. Xiao, X. F. Liao and K. W. Wong, An efficient entire chaos-based scheme for denial authentication, Chaos Solitons & Fractals, vol.23, no.4, pp.1327-1331,February 2005, doi: 10.1016/j.chaos.2009.02.001.
13. G. Alvarez, Security problems with a chaos-based denial authentication scheme, Chaos Solitons & Fractals, vol.26, no.1, pp.7-11, October 2005, doi: 10.1016/j.chaos. 2004.12.023.
14. S. Han and E. Chang, Chaotic map based key agreement with/out clock synchronization, Chaos Solitions & Fractals, vol.39, no.3, pp.1283-1289, February 2009, doi: 10.1016/j.chaos.2007.06.030.
15. H. Tseng, R. Jan, and W. Yang, A chaotic maps-based key agreement protocol that preserves user anonymity, In: IEEE Inter. Conf. Commun., ICC'09, Dresden, Germany, 14-18 June 2009, pp. 1-6, doi: 10.1109/ICC.2009.5198 581.
16. Y. Niu and X. Y. Wang, An anonymous key agreement protocol based on chaotic maps, Commun. Nonlinear Sci. Number. Simul., vol.16, no.4, pp.1986-1992, April 2011, doi: 10.1016/j.cnsns.2010.08.015.
17. E. J. Yoon, Efficiency and security problems of anonymous key agreement protocol based on chaotic maps, Commun. Nonlinear Sci. Number. Simul., vol.17, no.717, pp.2735-2740, July 2012, doi: 10.1016/j.cnsns. 2011.11.010.
18. Z. Tan. "A chaotic maps-based authenticated key agreement protocol with strong anonymity," Nonlinear Dyn., vol.72, no.1-2, pp. 311-320, April 2013, doi: 10.1007/s11071-012-0715-5.
19. P. Gong, P. Li and W. Shi, A secure chaotic maps-based key agreement protocol without using smart cards, Nonlinear Dyn., vol.70, pp. 2401-2406, 2012, doi: 10.1007/s11071-012-0628-3.
20. X. Y. Wang and D. P. Luan, A secure key agreement protocol based on chaotic maps, Chin. Phys. B., vol.22, no.11, 2013, doi: 10.1088/1674-1056/22/11/110503.
21. C. C. Lee, C. T. Li and C. W. Hsu, A three-party password-based authenticated key exchange protocol with user anonymity using extened chaotic maps, Nonlinear Dyn., Vol.73, no.1-2, pp.125-132, July 2013, doi: 10.1007/s11071-013-0772-4.
22. L. Zhang, Cryptanalysis of public key encryption based on multiple chaotic systems, Chaos Solitons & Fractals, vol.37, no.3, pp.669-674, August 2008. Doi: 10.1016/j. chaos.2006.09.047.
23. A. C. Yao and Y. Zhao, Computationally-Fair group and identity-based key-exchange, Proceedings of the 9th international conference on Theory and Application of Models of Computation, TAMC'12, Beijing, China, May 16-21, 2012, pp.237-247, doi: 10.1007/978-3-642-29952-0_26.
24. X. Y. Wang and J. F. Zhao, An improved key agreement protocol based on chaos, Commun. Nonlinear Sci. Number. Simul., vol.15, no.12, pp.4052-4057, December 2010, doi: 10.1016/j.cnsns.2010.02.014.
25. Z. H. Li, Y. D. Cui and H. M. Xu, Fast algorithms of public key cryptosytem based on Chebyshev polynomials over finite field, The Journal of China Universities of Posts and Telecommunications, vol.18, no.2, pp.86-93, April 2011, doi: 10.1016/S1005-8885(10)60049-0.

Chaotic Modeling and Simulation (CMSIM) 3: 233–241, 2014

Chaos in Pendulum Systems with Limited Excitation in the Presence of Delay

A.Yu. Shvets and O.M. Makasyeyev

National Technical University of Ukraine Kyiv Polytechnic Institute, Kyiv, Ukraine
(E-mail: alex.shvets@bigmir.net, makaseyev@ukr.net)

Abstract. Dynamic system "pendulum - source of limited excitation" with taking into account the various factors of delay is considered. Different approaches to write a mathematical model of this system using three- or fifteen-dimensional systems of differential equations without delay is suggested. It is established that for small values of the delay it is sufficient to use three-dimensional mathematical model, whereas for relatively large values of the delay the fifteen-dimensional mathematical model should be used.

Genesis of deterministic chaos is studied in detail. Maps of dynamic regimes, phase portraits of attractors of systems, phase-parametric characteristics, Poincare sections and maps are constructed and analyzed. The scenarios of transition from steady-state regular regimes to chaotic ones are identified. It is shown, that in some cases the delay is the main reason of origination of chaos in the system "pendulum - source of limited excitation".

Keywords: pendulum system, limited excitation, delay, deterministic chaos..

1 Introduction

In mathematical modeling of oscillatory processes a mathematical model of a relatively simple dynamical system is often used to study the dynamics of much more complex systems. A typical example of this approach is the extensive use of pendulum models to study the dynamics of systems of an entirely different nature. Pendulum mathematical models are widely used to describe the dynamics of various technical constructions, machines and mechanisms, in the study of cardiovascular system, financial markets, etc. Such widespread use of pendulum models makes it relevant to study directly the dynamics of pendulum systems.

The study of the non-ideal by Zommerfeld–Kononenko [1] dynamical system "pendulum–electric motor" in the absence of any delay factors was initiated in [2], [3]. In this system the existence of deterministic chaos was identified and studied. It was proved that limited excitation is the main cause of chaos in this system.

Received: 4 April 2014 / Accepted: 16 July 2014

ISSN 2241-0503

In this paper the oscillations of "pendulum–electric motor" system with taking into account various factors of delay are considered. The delay factors are always present in rather extended systems due to the limitations of signal transmission speed: waves of compression, stretching, bending, current strength, etc. The aim of this work is to study the influence of various factors of delay on steady-state regimes of this system.

2 Delay factors in "Pendulum–electric motor" system

In the absence of any delay factors the equations of motion of the system "pendulum–electric motor" were obtained in [2]:

$$\begin{cases} \dfrac{dy_1}{d\tau} = Cy_1 - y_2y_3 - \dfrac{1}{8}(y_1^2y_2 + y_2^3); \\ \dfrac{dy_2}{d\tau} = Cy_2 + y_1y_3 + \dfrac{1}{8}(y_1^3 + y_1y_2^2) + 1; \\ \dfrac{dy_3}{d\tau} = Dy_2 + Ey_3 + F; \end{cases} \tag{1}$$

where phase variables y_1, y_2 describe the pendulum deviation from the vertical and phase variable y_3 is proportional to the rotation speed of the motor shaft. The system parameters are defined by

$$C = -\delta_1\varepsilon^{-2/3}\omega_0^{-1}, D = -\frac{2ml^2}{I}, F = 2\varepsilon^{-2/3}(\frac{N_0}{\omega_0} + E) \tag{2}$$

where m - the pendulum mass, l - the reduced pendulum length, ω_0 - natural frequency of the pendulum, a - the length of the electric motor crank, $\varepsilon = \dfrac{a}{l}$, δ_1 - damping coefficient of the medium resistance force, I - the electric motor moment of inertia, E, N_0 - constants of the electric motor static characteristics.

Let us consider the following system of equations [4]:

$$\begin{cases} \dfrac{dy_1(\tau)}{d\tau} = Cy_1(\tau-\delta) - y_2(\tau)y_3(\tau-\gamma) - \dfrac{1}{8}(y_1^2(\tau)y_2(\tau) + y_2^3(\tau)); \\ \dfrac{dy_2(\tau)}{d\tau} = Cy_2(\tau-\delta) + y_1(\tau)y_3(\tau-\gamma) + \dfrac{1}{8}(y_1^3(\tau) + y_1(\tau)y_2^2(\tau)) + 1; \\ \dfrac{dy_3(\tau)}{d\tau} = Dy_2(\tau-\gamma) + Ey_3(\tau) + F. \end{cases} \tag{3}$$

Positive constant parameter γ was introduced to account the delay effects of electric motor impulse on the pendulum. We assume that the delay of the electric motor response to the impact of the pendulum inertia force is also equal to γ. Taking into account the delay γ conditioned by the fact that the wave velocity perturbations on the elements of the construction has a finite value that depends on the properties of external fields, for instance, the temperature field. In turn, the constant positive parameter δ characterizes the delay of the

medium reaction on the dynamical state of the pendulum. This delay is due to the limited sound velocity in that medium.

Assuming a small delay, we can write

$$y_i(\tau-\gamma) = y_i(\tau) - \frac{y_1(\tau)}{d\tau}\gamma + ..., \quad i = 2,3$$

$$y_i(\tau-\delta) = y_i(\tau) - \frac{y_1(\tau)}{d\tau}\delta + ..., \quad i = 1,2$$

Then, if $C\delta \neq -1$, we get the following system of equations [4]:

$$\begin{cases} \dot{y_1} = \dfrac{1}{1+C\delta}\left(Cy_1 - y_2\left[y_3 - \gamma\left(Dy_2 + Ey_3 + F\right)\right] - \dfrac{1}{8}(y_1^2y_2 + y_2^3)\right); \\ \dot{y_2} = \dfrac{1}{1+C\delta}\left(Cy_2 + y_1y_3 - y_1\gamma(Dy_2 + Ey_3 + F) + \dfrac{1}{8}(y_1^3 + y_1y_2^2) + 1\right); \\ \dot{y_3} = (1-C\gamma)Dy_2 - \dfrac{D\gamma}{8}(y_1^3 + y_1y_2^2 + 8y_1y_3 + 8) + Ey_3 + F. \end{cases} \tag{4}$$

The obtained system of equations is already a system of ordinary differential equations. Delays are included in this system as additional parameters.

In order to approximate the system (3) another, more precise, method can be used [5], [6]. Let us divide each of the segments $[-\gamma;\ 0]$ and $[-\delta;\ 0]$ into m equal parts. We introduce the following notation

$$y_1(\tau - \frac{i\delta}{m}) = y_{1i}(\tau),\ y_2(\tau - \frac{i\gamma}{m}) = y_{2i}(\tau),\ y_2(\tau - \frac{i\delta}{m}) = \tilde{y}_{2i}(\tau),$$
$$y_3(\tau - \frac{i\gamma}{m}) = y_{3i}(\tau), i = \overline{0,m}.$$

Then, using difference approximation of derivative [5], [6] the system of equations with delay (3) can be reduced to the following system of equations without delay:

$$\begin{cases} \dfrac{dy_{10}(\tau)}{d\tau} = Cy_{1m}(\tau) - y_{20}(\tau)y_{3m}(\tau) - \dfrac{1}{8}(y_{10}^2(\tau)y_{20}(\tau) + y_{20}^3(\tau)); \\ \dfrac{dy_{20}(\tau)}{d\tau} = C\tilde{y}_{2m}(\tau) + y_{10}(\tau)y_{3m}(\tau) + \dfrac{1}{8}(y_{10}^3(\tau) + y_{10}(\tau)y_{20}^2(\tau)) + 1; \\ \dfrac{dy_{30}(\tau)}{d\tau} = Dy_{2m}(\tau) + Ey_{30}(\tau) + F; \\ \dfrac{dy_{1i}(\tau)}{d\tau} = \dfrac{m}{\delta}(y_{1\ i-1}(\tau) - y_{1i}(\tau)), \quad i = \overline{1,m}; \\ \dfrac{dy_{2i}(\tau)}{d\tau} = \dfrac{m}{\gamma}(y_{2\ i-1}(\tau) - y_{2i}(\tau)), \quad i = \overline{1,m}; \\ \dfrac{d\tilde{y}_{2i}(\tau)}{d\tau} = \dfrac{m}{\delta}(\tilde{y}_{2\ i-1}(\tau) - \tilde{y}_{2i}(\tau)), \quad i = \overline{1,m}; \\ \dfrac{dy_{3i}(\tau)}{d\tau} = \dfrac{m}{\gamma}(y_{3\ i-1}(\tau) - y_{3i}(\tau)), \quad i = \overline{1,m}. \end{cases} \tag{5}$$

Should be noted that the main variables in this system are only y_{10}, y_{20}, y_{30}. In other words the solutions y_1, y_2, y_3 of the system (3) are described by the functions y_{10}, y_{20}, y_{30} of the system (5).

The system (5) is a system of ordinary differential equations of $(4m+3)$-th order. Choosing a sufficiently large m in the system (5), the system (3) will be very well approximated by the system (5) [5]. In this paper the system of equation (5) was considered at $m = 3$. In this case, the system (5) has 15 equations. The calculations of cases $m > 3$, with a significant increase the number of equations, were also carried out. It was established, that increasing the number of equations has practically no effect on identification and description of steady-state regimes of "pendulum–electric motor" system. But it significantly increases the complexity of constructing characteristics, which are necessary for study the steady-state regimes of oscillations. Therefore, the use of mathematical model (5) at $m = 3$ is the most optimal for studying the influence of delay on regular and chaotic dynamics of "pendulum–electric motor" system.

3 Maps of dynamic regimes

Therefore, we obtained three-dimensional (4) and fifteen-dimensional (5) models each describing the system of equations with delay (3). These models are the systems of non-linear differential equations, so in general the study of steady-state regimes can be carried out only by using numerical methods and algorithms. The methodology of such studies is described in detail in [2].

In the study of dynamical systems the information about the type of steady-state regime of the the system depending on its parameters is crucial. This information can provide a map of dynamic regimes. It is a diagram on the plane, where two parameters are plotted on axes and the boundaries of different

dynamic regimes areas are shown. The construction of dynamic regimes maps is based on analysis and processing of spectrum of Lyapunov characteristic exponents [2,7]. Where necessary, for more accurate determination of steady-state regime of the system, we study other characteristics of attractors: phase portraits, Poincare sections and maps, Fourier spectrums and distributions of the invariant measure.

Let us consider the behavior of the systems (4) and (5) when the parameters are $C = -0.1$, $D = -0.6$, $E = -0.44$, $F = 0.3$. In fig. 1 the maps of dynamic regimes are shown. The map in fig. 1a was built for three-dimensional model (4) and the map in fig. 1b was built for fifteen-dimensional model (5). These figures illustrate the effect of delays γ and δ on changing the type of steady-state regime of the systems. The dark-grey areas of the maps correspond to equilibrium positions of the system. The light-grey areas of the maps correspond to limit cycles of the system. And finally, the black areas of the maps correspond to chaotic attractors.

We can notice a certain similarity the maps in fig.1a, b. In delay absence in these systems, the steady-state regime is stable equilibrium position. With an increase of the delay of the medium δ the type of steady-state regime of the systems (4) and (5) does not change. It still remains an equilibrium position (dark-grey areas in the figures). However, with an increase of the delay of interaction between pendulum and electric motor γ, the equilibrium position is replaced by the area of limit cycles with "mounted" area of chaos. With further increase of the delay γ, the attractor of both systems is again equilibrium position.

Let us study the dynamics of the system (4) and (5) at other values of the parameters. At $C = -0.1$, $D = -0.58$, $E = -0.6$, $F = 0.19$ the steady–state regime of both systems is limit cycle. In fig. 2a the map of dynamic regimes of three-dimensional system (4) and in fig. 2b the map of dynamic regimes of fifteen-dimensional system (5) are shown. At small values of the delays the steady-state regime of both systems does not change, it is periodic.

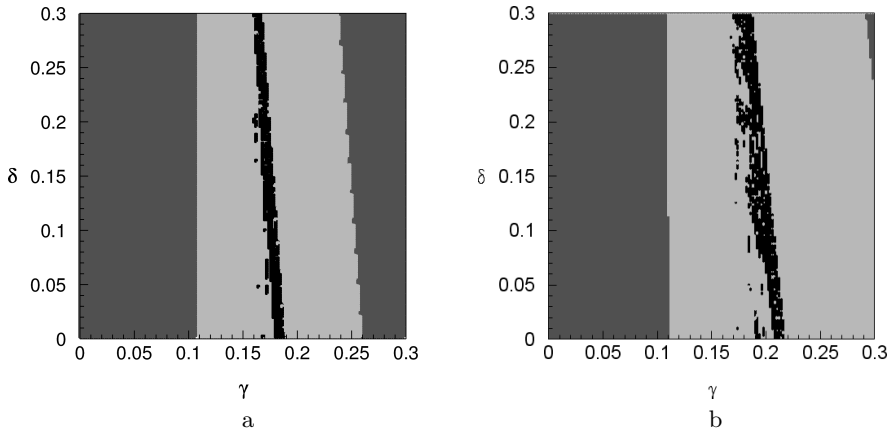

Fig. 1. Maps of dynamic regimes

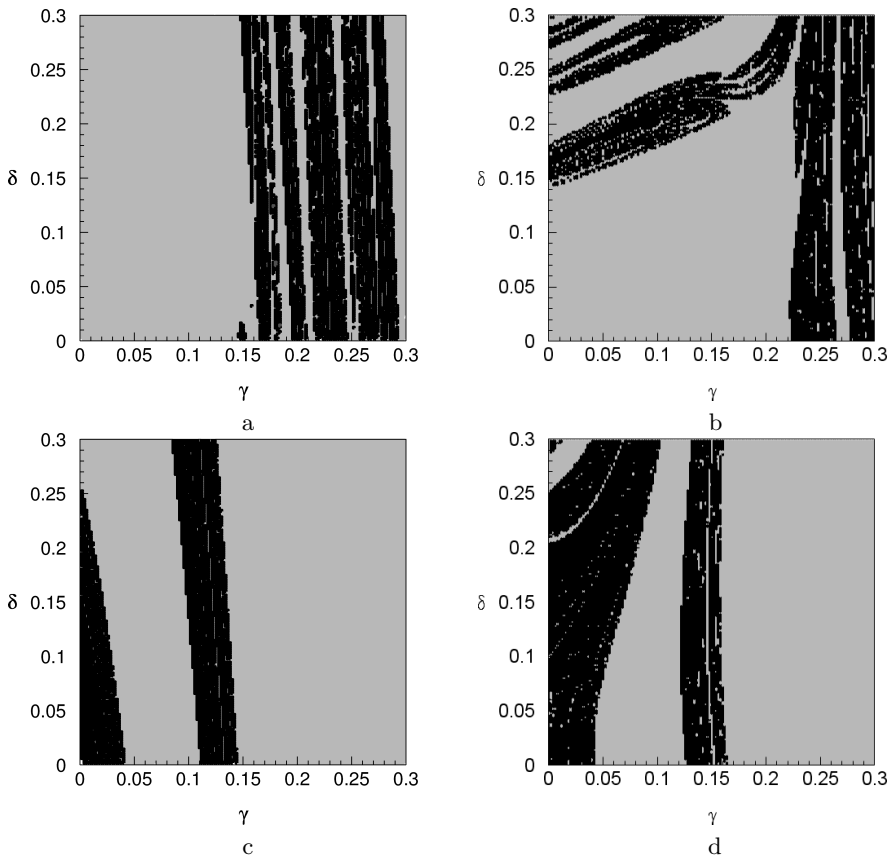

Fig. 2. Maps of dynamic regimes

The attractors are limit cycles (light-grey areas in the figures). With a further increase of the delay values the maps in fig. 2a, b are certainly different. At small values of the delay γ and with increase of the delay δ the type of steady–state regime of the system (5) is replaced by chaotic regimes, whereas the type of steady–state regime of the system (4) does not change, it remains periodic. Further in both figures there are a rather wide area of chaos in which fairly narrow strips of periodic regimes are built in.

In fig. 2c, d the maps of dynamic regimes of respectively the system (4) and the system (5) at $C = -0.1$, $D = -0.53$, $E = -0.6$, $F = 0.19$ are constructed. In delay absence and at small values of the delays both systems have chaotic attractors (black areas in the figures). With an increase of the delay values the region of chaos is replaced by the region of periodic regimes. Then again chaos arises in the system. Further this area is replaced by the area of limit cycles.

As seen from the constructed maps of dynamic regimes, the dynamics of the system (4) and (5) is the same only for small values of the delay γ and δ. With an increase of the delays the differences of the dynamics of these systems is very significant.

4 Regular and chaotic dynamics

Let us study the types of regular and chaotic attractors that exist in the systems (4) and (5). We implement a horizontal section of the maps of dynamic regimes in fig.2c, d along the delay γ at $\delta = 0.15$. In other words, let us consider the behavior of the systems (4) and (5) when parameters are $C = -0.1$, $D = -0.53$, $E = -0.6$, $F = 0.19$ and the delays $\delta = 0.15$ and $0 \leq \gamma \leq 0.3$.

In fig. 3a,b the dependence of maximum non-zero Lyapunov's characteristic exponent and phase-parametric characteristic of three-dimensional system (4) are shown respectively. These figures illustrate the influence of the delay of interaction between pendulum and electric motor γ on chaotization of the system (4).

Let us construct the same characteristics at the same values of the parameters for fifteen-dimensional system (5). In fig. 4a,b respectively the dependence of maximum non-zero Lyapunov's characteristic exponent and phase-parametric characteristic are shown.

In fig.3a, 4a we can clearly see the presence of intervals γ in which maximum Lyapunov exponent of the systems is positive. In these intervals the systems have chaotic attractors. The area of existence of chaos is clearly seen in phase-parametric characteristics of the systems. The areas of chaos in the bifurcation trees are densely filled with points. A careful examination of the obtained images allows not only to identify the origin of chaos in the systems, but also to describe the scenario of transition to chaos. So with a decrease of γ there are the transitions to chaos by Feigenbaum scenario (infinite cascade of period-doubling bifurcations of a limit cycle). Bifurcation points for the delay γ are clearly visible in each figures. These points are the points of approaches of the Lyapunov's exponent graph to the zero line (fig.3a, 4a) and the points of splitting the branches of the bifurcation tree (fig.3b, 4b). In turn, the transition to chaos with an increase of the delay happens under the scenario of Pomeau-Manneville, in a single bifurcation, through intermittency.

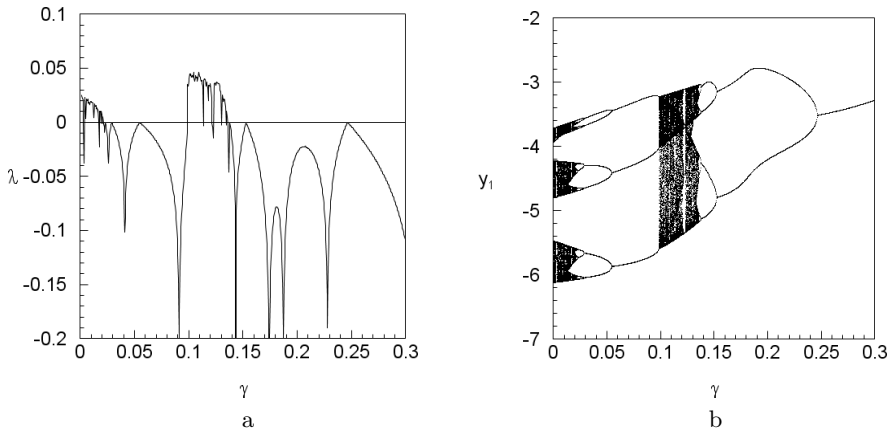

Fig. 3. The dependence of maximal non-zero Lyapunov's characteristic exponent (a), phase-parametric characteristic (b) of three-dimensional system (4)

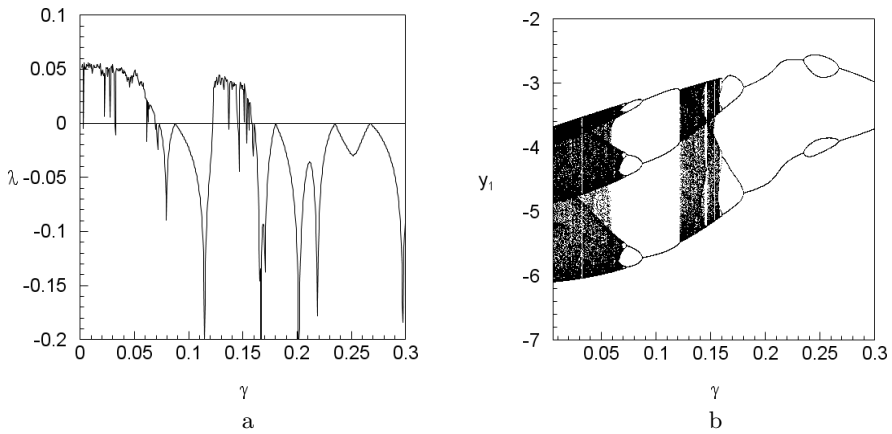

Fig. 4. The dependence of maximal non-zero Lyapunov's characteristic exponent (a), phase-parametric characteristic (b) of fifteen-dimensional system (5)

A careful analysis of these figures allows to see qualitative similarity of the respective characteristics of the systems (4) and (5). However, with increasing the delay the differences in the dynamics of these systems become very significant. So for instance at $\gamma = 0.05$ the steady–state regime of the system (4) is limit cycle. While at this value of the delay the attractor of the system (5) is chaotic attractor. Conversely, for example at $\gamma = 0.11$ the system (4) has steady–state chaotic regime. While at this value of the delay the system (5) has periodic regime of oscillations.

This suggests that three-dimensional system of equations (4) should be used to study the system (3) only at very small values of the delay. With increasing values of the delay to study regular and chaotic oscillations of "pendulum–electric motor" system, fifteen-dimensional system of equations (5) should be used.

5 Conclusion

Various factors of delay have significant influence on the dynamics of "pendulum–electric motor" system. The presence of delay in such systems can affect the type of steady-state regime change. It is shown that for small values of the delay it is sufficient to use three-dimensional mathematical model, whereas for relatively high values of the delay the fifteen-dimensional mathematical model should be used.

In future research is planned to construct and research mathematical models of "pendulum–electric motor" system in the presence of variable in time delay factors.

References

1.V. O. Kononenko, *Vibrating System with a Limited Power-supply*, Iliffe, London, 1969.

2.T. S. Krasnopolskaya, A. Yu. Shvets. *Regular and chaotical dynamics of systems with limited excitation.* R&C Dynamics, Moscow, 2008.
3.A. Yu. Shvets, A. M. Makaseyev. Chaotic Oscillations of Nonideal Plane Pendulum Systems. *Chaotic Modeling and Simulation (CMSIM)*, 1:195–204, 2012.
4.A. Yu. Shvets, A. M. Makaseyev. Delay Factors and Chaotization of Non-ideal Pendulum Systems. *Chaotic Modeling and Simulation (CMSIM)*, 4:633–642, 2012.
5.N. A. Magnizkiy, S. V. Sidorov. *New methods of chaotic dynamics.* Editorial URSS, 2004.
6.A. A. Samarskiy, A. V. Gulin. *Numerical methods.* Nauka, 1989.
7.S. P. Kouznetsov, *Dynamic chaos.* Physmatlit, Moscow, 2001.

Chaotic Modeling and Simulation (CMSIM) 3: 233–254, 2014

Chaotic Solutions in non Linear Economic - Financial models

Beatrice Venturi

Department of Economics and Business, University of Cagliari, Italy.
(E-mail: venturi@unica.it)

Abstract. Following Mulligan and Sala-i-Martin (1993) we study a general class of endogenous growth models formalized as a non linear autonomous three-dimensional differential system. We consider the abstract model. By using the Shilnikov Theorem statements, we determine the parameters space in which the condition for the existence of a homoclinic Shilnikov orbit and Smale horseshoe chaos are true.

The Lucas model (1998) can be considered as an application of the general result. The series expression of the homoclinic orbit is derived by the undetermined coefficient method. We show the optimality for the solutions path based on the Shilnikov Theorem. Some economic implications of this analysis are discussed.
Keywords: homoclinic Shilnikov bifurcation, Smale horseshoe chaos..

1 Introduction

We consider a class of endogenous growth two sector models as formulated by Mulligan B. and X. Sala-i-Martin (1993). Some examples of this class are well known and deeply study in recent literature (see inter al. D. Fiaschi, S. Sordi, 2002). A lot of research has been done in indeterminacy results and in the conditions of existence and stability of cycles, in a special way Hopf bifurcations (see G. Benhabib and R. Perli, 1994, p.124; P. Mattana and B. Venturi, 1999; M. Boldrin, K. Nishimura, T, Shigoka, M. Yano, 2000; D. Fiaschi, S. Sordi, 2002; P. Mattana, 2004; S. Slobodyan, 2005; K. Nishimura , T.Shigoka, 2008; A. Antoci, M. Galeotti, and P. Russu, 2011; G.Bella, P. Mattana and B.Venturi, 2013).

In particular, our analysis focuses on the context in which the application system, the Lucas model, admits only one steady state which corresponding, after a change of variables, in standard way, to an equilibrium point of a non linear three-dimensional autonomous system.

As described by Guckenheimer J. and Holmes P. (1983), and Wiggins S. (1990) usually a chaotic attractor has two or more fixed points: one determines the location and the structure of the attractor, and another is used to build a suspended flow which forms the spine of the attractor. However, as reported in recently papers one equilibrium point is still possible to form a chaotic attractor.

Received: 9 February 2014 / Accepted: 12 May 2014

ISSN 2241-0503

(In order to study the long-run properties of the equilibrium) we treat this class as a general dynamical system. We give the conditions under which the Shilnikov chaos occurs in a appropriated parameter set. Using Cardano formula and series solution of the differential equations, the eigenvalues problem and the rigorous proof of the existence of the homoclinic orbit are pursued and applied to the Lucas model.

The work develops as follows. The second Section introduces the considered class of generalized two sector models of endogenous growth, as a dynamical system. We refer to the original paper of B. Mulligan and X. Sala-i-Martin,1993 and R. Lucas 1988 for an appropriate economic description of the system and its application. The third Section is devoted to characterize the parameter set in which the Shilnikov Theorem statements hold. We give a rigorous proof of the emergence of a homoclinic Shilnikov orbit. In view of its evaluation , in the first we found the set in which the system has a saddle-focus (of index 2) and in the second, we determined the coefficients of the series expression of the stable and unstable manifolds of such equilibrium point (the saddle-focus). As application of these results we consider the Lucas model. At the end we show the optimality for the solutions path based on the Shilnikov Theorem. Numerical simulation demonstrate that there is a route to chaos. Some economic implications of this analysis are discussed.

2 The Generalized Class of Two Sector Models of Endogenous Growth

We review the generalized class of two sector models of endogenous growth, with externalities, as formulated by B. Mulligan and X. Sala-i-Martin (1993).The model deal with the maximization of a standard utility function:

$$\int_0^\infty \frac{c^{1-\sigma}-1}{1-\sigma} e^{-\rho t} dt \tag{2.1}$$

where c is per-capita consumption, ρ is a positive discount factor and σ is the inverse of the intertemporal elasticity of substitution. The constraints to the growth process are represented by the following equations

$$\dot{k} = A((h(t)^{\alpha_h} u(t)^{\alpha_u})(\nu(t)^{\alpha_\nu} k(t)^{\alpha_k})\hat{h}(t)^{\alpha_{\hat{h}}} k(t)^{\alpha_{\hat{k}}} - \tau_k k(t) - c(t) \tag{2.2}$$

$$\dot{h} = B((h(t)^{\beta_h}(1-u(t)^{\beta_u}))((1-\nu(t)^{\beta_\nu} k(t)^{\beta_k})\hat{h}(t)^{\beta_{\hat{h}}} k(t)^{\beta_{\hat{k}}} - \tau_h h(t)$$

where k is physical capital, h is human capital, $\alpha_k$and $\alpha_h$being the private share of physical and the human capital in the output sector, β_k and β_h being the corresponding shares share in the education sector, u and v are the fraction of aggregate human and physical capital used in the final output sector at instant t (and conversely, $(1-u)$ and $(1-v)$ are the fractions used in the education sector), A and B are the level of the technology in each sector, τ is a discount factor, $\alpha_{\hat{k}}$ is a positive externality parameter in the production of physical capital, $\alpha_{\hat{h}}$is a positive externality parameter in the production

of human capital. The equalities $\alpha_k + \alpha_h = 1$ and $\beta_k + \beta_h = 1$ ensure that there are constant returns to scale at the private level. At the social level, however, there may be increasing, constant or decreasing returns depending on the signs of the externality parameters. All other parameters $\omega = (\alpha_k, \underset{k}{\alpha_{\wedge}}, \alpha_h, \underset{h}{\alpha_{\wedge}}, \beta_k, \underset{k}{\beta_{\wedge}}, \beta_h, \underset{h}{\beta_{\wedge}}, \sigma, \gamma, \delta, \rho)$ live inside the following set $\Omega \subset (0,\,1)\times(0,\,1)\times(0,\,1)\times(0,\,1)\times(0,\,1)\times(0,\,1)\times(0,\,1)\times \mathrm{R}^4_+$

The representative agent's problem (1.1)-(1.2) is solved by defining the current value Hamiltonian.

$$H = \frac{c^{1-\sigma}-1}{1-\sigma} + \lambda_1(A((h(t)^{\alpha_h}u(t)^{\alpha_u})(\nu(t)^{\alpha_\nu}k(t)^{\alpha_k})\hat{h}(t)^{\alpha_{\hat{h}}}k(t)^{\alpha_{\hat{k}}} - \tau_k k(t) - c(t)) +$$
$$+\lambda_2(B((h(t)^{\beta_h}(1-u(t)^{\beta_u}))((1-\nu(t)^{\beta_\nu}k(t)^{\beta_k})\hat{h}(t)^{\beta_{\hat{h}}}k(t)^{\beta_{\hat{k}}} - \tau_h h(t)) \quad (2.3)$$

where λ_1 and λ_2 are co-state variables which can be interpreted as shadow prices of the accumulation. The solution candidate comes from the first-order necessary conditions (for an interior solution) obtained from the Maximum Principle, with the usual transversality condition

$$\lim_{t\to\infty}\left[e^{-\rho t}\left(\lambda_1 k + \lambda_2 h\right)\right] = 0 \quad (2.4)$$

consider only the competitive equilibrium solution. After eliminating $v(t)$ the rest of the first order conditions and accumulation constraints entail four first order non linear differential equations in four variables: two controls (c and u) and two states (k and h). By using new variables, since h, k and c grow at a constant rate and u is a constant, Mulligan B.-Sala-i-Martin X.(1993) have transformed a system of ordinary differential equations for c, u, k and h, , into a system of three first order ordinary differential equations.

Setting $A = B = 1$ and

$$x_1 = h^{\frac{a_{\hat{h}}}{(a_{\hat{h}}-1)}}k\ ; \quad x_2 = u; \quad x_3 = \frac{c}{k} \quad (2.5)$$

we get:

$$\begin{aligned}
\dot{x}_1 &= \phi_1(x_1, x_2, x_3, \alpha_k, \alpha_{\hat{k}}, \alpha_h, a_{\hat{h}}, \beta_k, \beta_{\hat{k}}, \beta_h, \beta_{\hat{h}}, \sigma, \gamma, \delta, \rho)\\
\dot{x}_2 &= \phi_2(x_1, x_2, x_3, \alpha_k, \alpha_{\hat{k}}, \alpha_h, \alpha_{\hat{h}}, \beta_k, \beta_{\hat{k}}, \beta_h, \beta_{\hat{h}}, \sigma, \gamma, \delta, \rho)\\
\dot{x}_3 &= \phi_3(x_1, x_2, x_3, \alpha_k, \alpha_{\hat{k}}, \alpha_h, \alpha_{\hat{h}}, \beta_k, \beta_{\hat{k}}, \beta_h, \beta_{\hat{h}}, \sigma, \gamma, \delta, \rho)
\end{aligned} \quad (2.6)$$

where the ϕ_i with $i = 1, 2, 3$ are complicated nonlinear functions ; which depend of the parameters $(x_1, x_2, x_3, \alpha_k, \underset{k}{\alpha_{\wedge}}, \alpha_h, \underset{h}{\alpha_{\wedge}}, \beta_k, \underset{k}{\beta_{\wedge}}, \beta_h, \underset{h}{\beta_{\wedge}}, \sigma, \gamma, \delta, \rho)$ of the model.

3 Shilnikov Theorem and The Emergence of a Homoclinic Orbit.

In order to verify that our system satisfies the Shilnikov Theorem statements, we follow strictly D. Shang M.Han, 2005. In the first we determine the parameter

space in which our system has a homoclinic orbit. We remember that a homoclinic orbit is a transversal intersection between the stable manifold with the unstable manifold of a hyperbolic equilibrium point (connects a saddle to itself). Under regularity conditions (continuity since the second order) the model (2.6), has at least one stationary point $P^*(x_1^*, x_2^*, x_3^*)$.

Lemma 1. *In Ω exists a parameters subset $\hat{\Omega}$ such that the equilibrium point $P^*(0,0,0)$ is a saddle focus of index 2.*

Proof. By using Cardano's formula, we determine a parameters space in which the solutions (roots) r_i, $i = 1,2,3$ of the polynomial characteristic of the Jacobian matrix J, evaluated in the stationary point $J^* = J(P^*)$ satisfies the following conditions

$$r_1 = -\frac{\hat{a}}{3} + u + v \tag{3.1}$$

$$r_{2,3} = -\frac{\hat{a}}{3} - \frac{u+v}{2} \pm \sqrt{3}\frac{u-v}{2}i$$

where $i = \sqrt{-1}$ is the imaginary root, $u = \sqrt[3]{-\frac{m}{2} + \sqrt{\Delta}}$ and $v = \sqrt[3]{-\frac{m}{2} - \sqrt{\Delta}}$, with, $l = \frac{3\hat{b}-\hat{a}^2}{3}$ and $m = \hat{c} + \frac{2\hat{a}^3}{27} - \frac{\hat{a}\hat{b}}{3}$, $\hat{a} = -\mathrm{Tr}(\mathbf{J}^*)$, $\hat{b} = \mathrm{B}(\mathbf{J}^*)$, and $\hat{c} = -\mathrm{Det}(\mathbf{J}^*)$, whereas $\Delta = \left(\frac{l}{3}\right)^3 + \left(\frac{m}{2}\right)^2$ is the discriminant. For the scope of our paper, a saddle-focus (of index 2) emerges when

$$\Delta > 0 \tag{3.2}$$

$$\sqrt[3]{-\frac{m}{2} + \sqrt{\Delta}} + \sqrt[3]{-\frac{m}{2} - \sqrt{\Delta}} < -\frac{2\hat{a}}{3}$$

that is explicitly

$$\left(\frac{\hat{c}}{2} + \frac{\hat{a}^3}{27} - \frac{\hat{a}\hat{b}}{6}\right)^2 > \left(\frac{\hat{a}^2 - 3\hat{b}}{9}\right)^3 \tag{3.3}$$

Thus (3.2) holds the characteristic equation has one real root and a conjugate pair of complex , and the real root is positive (negative) since $\mathrm{Det}(\mathbf{J}^*) > 0 (< 0)$.

To ensure that the real part of the complex conjugate roots is positive (negative) and that the equilibrium point is a saddle focus (of index 2) it is further required that :

$$\sqrt[3]{-\left(\frac{\hat{c}}{2} + \frac{\hat{a}^3}{27} - \frac{\hat{a}\hat{b}}{6}\right) + \sqrt{\Delta}} + \sqrt[3]{-\left(\frac{\hat{c}}{2} + \frac{\hat{a}^3}{27} - \frac{\hat{a}\hat{b}}{6}\right) - \sqrt{\Delta}} < -\frac{2\hat{a}}{3} \tag{3.4}$$

In other word when (3.1) (3.2) (3.3) are satisfied the characteristic equations of the Jacobian J^*in $\hat{\Omega}$ has one positive real and two complex conjugate eigenvalues whose real parts is negative: then the equilibrium point P^*in $\hat{\Omega}$ is a saddle focus and the real eigenvalue is bigger than the absolute value of the real part of the complex conjugate eigenvalues.

Lemma 2. *In* $\overset{\wedge}{\Omega} \subset \Omega$ *the system (2.6) has an homoclinic Shilnikov orbit.*

Proof. We compute the stable and unstable manifolds of the saddle focus equilibrium point to construct the Shilnikov type homoclinic orbit in an analytic style.

Theorem 1. *The system (2.6) exhibits a Smale horseshoe type of chaos. In other words (2.6) has at least a finite number of Smale horseshoes in the discrete dynamics of the Shilnikov map defined near the homoclinic orbit.*

Proof. Theorem 1 is a direct application of the Shilnikov Theorem (see Guckenheim-Holmes1983, pp.151-152). We only have to verify that the assumptions of Shilnikov theorem are satisfied.

4 Application: The Lucas Model

The general model just presented collapses to Lucas's model (1988) that is analyzed by Benhabib and Perli (1994), Mattana and Venturi (1999) and Mattana (2004) when depreciation is neglected and the following restrictions are imposed

$$\alpha_\nu = \alpha_{\hat{k}} = 0; \beta_{\hat{k}} = \beta_{\hat{h}} = \beta_\nu = \beta_k = 0; \alpha_\nu = \alpha_h = 1 - \alpha_k; \beta_u = \beta_h \tag{4.1}$$

The equations of the Lucas's model can be formalized in R^3 in the following form

$$\begin{aligned} \dot{x}_1 &= x_1^\beta x_2^{\beta-1} - x_1 x_3 + \psi \tfrac{(\beta-1)}{\beta}(1 - x_2) \\ \dot{x}_2 &= \eta x_2^2 + \psi \tfrac{(\beta-1)}{\beta} x_2 + x_1 x_3 \\ \dot{x}_3 &= \phi x_2^{1-\beta} x_1^{\beta-1} x_3 - \tfrac{\rho}{\sigma} x_3 + x_3^2 \end{aligned} \tag{4.2}$$

as a system of three first order differential equations where

$$\phi = \frac{\beta - \sigma}{\sigma} \quad \eta = \frac{\delta(\beta - 1)}{\beta} \quad \psi = \frac{\delta(1 - \beta + \gamma)}{\beta - 1} \qquad \xi = \frac{\rho}{\sigma} \tag{4.3}$$

A stationary (equilibrium) point $P^*$of the system is any solution of

$$\begin{aligned} x_1^{*1-\beta} x_2^{*\,\beta-1} - x_1^* x_3^* + \psi(1 - x_2) x_1^* &= 0 \\ \eta x_2^2 + \psi \tfrac{(\beta-1)}{\beta} x_2 - x_2^*\, x_3^* &= 0 \\ \phi x_2^{*1-\beta} x_1^{*\beta-1} x_3^* - \ \xi x_3^* + x_3^{*2} &= 0 \end{aligned} \tag{4.4}$$

Then, we solved the system in (4.4) and we get the following steady state values

$$x_1^* = x_2^* \left[\frac{\beta\rho - \delta\sigma\,(1 - u^*) + \delta(\beta - \gamma)}{\beta\,(\beta - \sigma)} \right]^{1/(\beta-1)} \tag{4.5a}$$

$$x_2^* = \frac{(1 - \beta)\ (\rho - \delta)}{\delta - [\gamma - \sigma(1 - \beta + \gamma)]} \tag{4.5b}$$

$$x_3^* = \eta x_2^* + \delta \frac{(1-\beta+\gamma)}{\beta} \tag{4.5c}$$

where $\phi = \frac{\beta-\sigma}{\sigma}$ simplifies the notation.

The system (4.2) possesses an interior steady-state characterized by the stationary values in (4.5.*a*), (4.5.*b*) *and* (4.5.*c*) for x_1^*, x_2^* and x_3^*. It is well-known that many theoretical results relating to the system depend upon the eigenvalues of the Jacobian matrix evaluated at the stationary point.

Let J be the Jacobian matrix and $P^*(x_1^*,\ x_2^*,\ x_3^*)$ the stationary point ($J(P^*) = J^*$, see appendix A). The "feasible" restrictions in the parameters are satisfied if and only if the parameters lie in one of the following subsets

Remark 1. i) if $\omega \in \Omega^1,\ J^*$ has one negative eigenvalue and two eigenvalues with positive real parts. (This means that the competitive equilibrium path is locally unique). ii) $\omega \in \Omega^2,\ J^*$ has one positive eigenvalue and two eigenvalues with negative real parts. iii) $\omega \in \Omega^3$ there exist two subsets Ω_3^A *and* Ω_3^B, and such that:
if $\omega \in \Omega_3^A$ J^**has one eigenvalue with a positive real part and two eigenvalues with negative real parts.* $\Omega_3^A = \left\{\rho \in (\delta, -\psi), \sigma \in (0.1, \rho/\psi), \gamma \in (\frac{(1-\beta)(\rho-\delta)}{\delta}, \overset{\rightsquigarrow}{\gamma})\right\}$ where $\overset{\rightsquigarrow}{\gamma}$ is the Hopf bifurcation value found in Mattana P. and Venturi B.(1999);
if $\omega \in \Omega_3^B$ J^**has three eigenvalues with positive real parts:*
$\Omega_3^B = \left\{\rho \in (\delta, -\psi), \sigma \in (0.1, \rho/\psi), \gamma \in (\overset{\rightsquigarrow}{\gamma}, \beta)\right\}$. *So, there is either a continuum of equilibria converging towards the steady-state or no stable transitional paths at all.*

We focuses our attention in the set Ω_3^A and we rigorously prove that our system in this subset satisfies all conditions stated in the Shilnikov Theorem.

In the first we translate the unique equilibrium point P^* in the origin W^*,we get

$$\frac{dw_i}{dt} = f_i(w_1, w_2, w_3) \ with \quad i = 1, 2, 3 \tag{4.6}$$

and we make use of the normal form (see Appendix B and Mattana and Venturi, 1999).

Lemma 3. *If* $\omega \in \Omega_3^A$ *the equilibrium point* $W^*(0,0,0)$ *is a saddle focus.*

The Jacobian J^*in Ω_3^A has one positive real and two complex conjugate eigenvalues whose real parts is negative: then the equilibrium point W^*in Ω_3^A is a saddle focus and the real eigenvalue is bigger than the absolute value of the real part of the complex conjugate eigenvalues. By using Cardano Formula we have verified analytically, and numerically the statement.

Lemma 4. *In* Ω_3^A *the system (4.6) has an homoclinic Shilnikov orbit* Γ.

Proof. We show that the **equilibrium point** $W^*(0,0,0)$ of system (4.6) is **doubly asymptotic** with respect to time t along the solution manifold. See Appendix B for details.

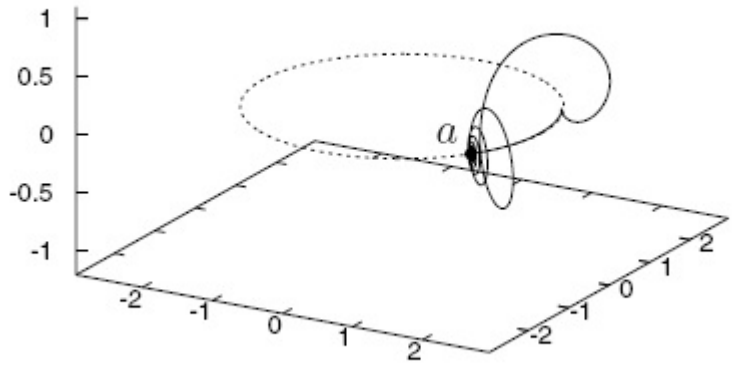

Figure 1: The homoclinic orbit

In Figure 1. we have the graph of the Homoclinic Shilnikov Orbit Γ .

Remark 2. In the set Ω_3^B the Jacobian J^* has one positive real and two complex conjugate eigenvalues whose real parts is positive. In this situation the model is expanding and thus it cannot have homoclinic orbits .

Theorem 2. *The system (4.6) exhibits a Smale horseshoe type of chaos. In other words (4.6) has at least a finite number of Smale horseshoes in the discrete dynamics of the Shilnikov map defined near the homoclinic orbit.*

Proof. By lemma1 the equilibrium point $W^*$is a saddle focus in Ω_3^A and the real eigenvalue $r_1 \in \mathbb{R}$ is bigger than the real part of the complex conjugate eigenvalues $r_{2/3} = -p \pm iq$ and $r_1 p > 0$. with a further constraint $|r_1| > |p|$. By lemma 2 the system has a homoclinic Shilnikov orbit Γ in Ω_3^A. It follows directly from the Shilnikov Theorem that if the third-order autonomous system (4.6) has a saddle-focus (of index 2) in the unique equilibrium points, W^* with eigenvalues associated to J^* given by $r_1 \in \mathbb{R}$ and $r_{2/3} = -p + iq \in \mathbb{C}$, such that $r_1 p > 0$. with a further constraint $|r_1| > |p|$, and there exists a homoclinic orbit Γ connecting W^*, then the Shilnikov map, defined in a neighborhood of the homoclinic orbit of the system, possesses a countable number of Smale horseshoes in its discrete dynamics. and for any sufficiently small C^1-perturbation $\mathbf{g}$ of $\mathbf{f}$ the perturbed system $\frac{dw_i}{dt} = g_i(w_1, w_2, w_3)$ with $i = 1, 2, 3$ exhibits a Smale horseshoe type of chaos has at least a finite number of Smale horseshoes in the discrete dynamics of the Shilnikov map defined near the homoclinic orbit.

5 Transversality Conditions

Proposition 1. *The transversality conditions are satisfied on the homoclinic orbit Γ.*

As shown in BP the transversality conditions are satisfied on the balanced growth paths. Let W^* $(\beta^*, \delta^*, \rho^*, \sigma^*, \gamma^*)$ be the only steady state in Ω_3^A. Let U in R^3be a small open neighborhood of W^*.So for each $(\beta, \delta, \rho, \sigma, \gamma) \in \Omega_3^A$, if we choose U sufficiently small, each path inside, starting from a point in the homoclinic orbit Γ, satisfies the transversality conditions. It follows directly from continuity arguments (the theorem of the permanence of the sign for continuous functions).

Proposition 2. *The transversality conditions hold near the homoclinic orbit where the Shilnikov Theorem is true.*

Proof. Let g_i , $i = 1, 2, 3$, be a C^1perturbation of f_i, $i = 1, 2, 3$, where the Shilnikov Theorem is true near the homoclinic orbit. Then for each $(\bar{\beta}, \bar{\delta}, \bar{\rho}, \bar{\sigma}, \bar{\gamma}) \in \Gamma$ in Ω_3^A there exists a constant $\mathbf{L}$ such that

$$|\mathbf{f}(\bar{w}(t)) - \mathbf{g}(\mathbf{w}(t))| < \mathbf{L}\,|\bar{w}(t) - \mathbf{w}(t)| \tag{5.1}$$

i.e., in vectorial form the distance between a path starting in the homoclinic Shilnikov orbit Γ and a Smale horseshoe chaotic path of $\mathbf{g}$ can be arbitrary small. From proposition *1* the transversality conditions are satisfied on the homoclinic orbit Γ then their are satisfied also in the chaotic solutions. We can choose an arbitrary small open set U of f a path starting in the homoclinic orbits in which there is a path that exhibits a Smale horseshoe chaos. But the Shilnikov Theorem stated that for any sufficiently small C^1-perturbation g of f, the perturbed system exhibits a Smale horseshoe chaos. Then the transversality condition is satisfied.

6 Conclusions

This paper aims to give a contribution of research to conditions which determine a chaotic behavior in the long-run properties of an economic model . Investigations of this kind are important in economic theory since help mapping the regions of the parameters space in correspondence of which the capacity of the models to produce indications on future economic outcomes starting from given fundamentals is drastically impaired. The aim of the present paper is to point out some basic ideas that may be useful to prove the transition to bounded and complex behavior, and to explain how the presence of an Homoclinic Shilnikov orbit and chaos in a model of a general class of economic-financial models can be interesting from an economic and dynamic point of view.

References

1. A. Antoci, M. Galeotti, and P. Russu, Poverty trap and global indeterminacy in a growth model with open-access natural resources, Journal of Economic Theory 146 (2011) 569-591.
2. C. Azariadis, R. Guesnerie, Sunspot and cycles, Review of Economic Studies 53 (1986) 725-736.
3. G. Bella, P. Mattana, B. Venturi, The double scroll chaotic attractor in the dynamics of a fixed-price IS-LM model, Int. J. Mathematical Modelling and Numerical Optimisation, Vol. 4, No. 1, (2013), 1-13.
4. J. Benhabib, K. Nishimura, T. Shigoka, Bifurcation and sunspots in the continuous time equilibrium model with capacity utilization, International Journal of Economic Theory 4 (2008) 337-355.
5. J. Benhabib, R. Perli, Uniqueness and indeterminacy: on thedynamics of endogenous growth. Journal of Economic Theory 63, (1994), 113-142.
6. J. Benhabib, S. Schmitt-Grohé, M. Uribe, The perils of Taylor rules, Journal of Economic Theory 96 (2001) 40-69.
7. J.B. Bullard, A Two-Headed Dragon for Monetary Policy, Business Economics 44 (2009) 73-79.
8. J.B. Bullard, S.H. Russel, An Empirically Plausible Model of Low Real Interest Rates and Unbaked Government Debt, Journal of Monetary Economics 44 (1999) 477-508.
9. D. Fiaschi, S. Sordi, Real business cycle models, endogenous growth models and cyclical growth: A critical survey," Computing in Economics and FinanceReal business cycle models, 16, (2002), 306-329.
10. G. Gandolfo, Economic dynamics, Springer-Verlag, Berlin, 1997.
11. J.M. Grandmond, On endogenous business cycles, Econometrica 53 (1985) 995-1046.
12. N. Kopell, L.N. Howard, Bifurcations and trajectories joining critical points, Advances in Mathematics 18 (1975) 306-358.R.
13. E. Lucas, On the mechanics of economic development. Journal of Monetary Economics 22, (1998), 3-42.
14. P.Mattana , B.Venturi ,. Existence and Stability of Periodic Solutions in the Dynamics of Endogenous Growth. RISEC, International Review of Economics, vol. 46, fas. 2, (1999), 259-284.
15. C.B. Mulligan, X. Sala-i-Martin, Transitional dynamics in two-sector models of endogenous growth. Quarterly Journal of Economics 103, (1993), 739-773.
16. P. Mattana, K. Nishimura, T. Shigoka, Homoclinic bifurcation and global indeterminacy of equilibrium in a two-sector endogenous growth model, International Journal of Economic Theory 5 (2009) 1-23.
17. M. Neamtu, D. Opris, C. Chilarescu, Hopf bifurcation in a dynamic IS–LM model with time delay, Chaos, Solitons and Fractals 34 (2007) 519-530.
18. U. Neri, B. Venturi, Stability and Bifurcations in IS-LM economic models, International Review of Economics 54 (2007) 53-65.
19. D. Shang , M. Han., The existence of homoclinic orbits to saddle-focus, Applied Mathematics and Computation 163, (2005), 621–631.

7 Appendix A

As shown in the text, Luca/s model gives rise to the following system of first-order differential equations

$$\begin{aligned} \dot{x_1} &= x_1^{\beta} x_2^{1-\beta} - x_1 x_3 + \psi(1-x_2)x_1; \\ \dot{x_2} &= \eta x_2^2 + \psi\tfrac{(\beta-1)}{\beta} x_2 - x_2 x_3; \\ \dot{x_3} &= \phi x_2^{1-\beta} x_1^{\beta-1} x_3 - \ \xi x_3 + x_3^2; \end{aligned} \tag{A.1}$$

where $\phi = \frac{\beta-\sigma}{\sigma}$, $\eta = \frac{\delta(\beta-\gamma)}{\beta}$, $\psi = \frac{\delta(1-\beta+\gamma)}{\beta-1}$, $\xi = \frac{\rho}{\sigma}$.

The system has the single equilibrium point: $P^*(x_1^*, x_2^*, x_3^*)$

$$\begin{aligned} x_1^* &= x_2^* \left[\tfrac{\beta\xi - \delta(1-\beta+\gamma) + \delta(\beta-\gamma)x_2^*}{\beta\phi}\right]^{1/(\beta-1)} \\ x_2^* &= \tfrac{(1-\beta)(\rho-\delta)}{\delta-[\gamma-\sigma(1-\beta+\gamma)]} \\ x_3^* &= \eta x_2^* + \delta\tfrac{(1-\beta+\gamma)}{\beta} \end{aligned} \tag{A.2}$$

The Jacobian matrix J associated with the system (A.1) evaluated at the unique equilibrium point P^* is given by $J(P^*)$:

$$J(P^*) = \begin{bmatrix} J_{11}^* & \frac{x_1^*}{x_2^*}(J_{11} + \psi x_2^*) & -x_1^* \\ 0 & -\eta x_2^* & x_2^* \\ \frac{J_{11}\phi x_3^*}{x_1^*} & \frac{J_{11}\phi x_3^*}{x_2^*} & x_3^* \end{bmatrix} \tag{A.3}$$

where

$$J_{11}^* = \frac{(\beta-1)[\gamma\rho - \delta\sigma(1-\beta+\gamma)]}{\beta[\gamma - \sigma(1-\beta+\gamma)]} \tag{A.4}$$

and

$$Tr(J^*) = \frac{\delta(2\beta-\gamma)}{\beta} x_2^* \tag{A.5}$$

$$Det(J^*) = J_{11}^* x_2^* x_3^* \frac{\delta(\gamma - \sigma(1-\beta+\gamma))}{\sigma(\beta-1)} \tag{A.6}$$

$$B(J^*) = J_{11}^* x_3^* + \frac{\delta^2(\beta-\gamma))}{\beta} x_2^{*2} \tag{A.7}$$

8 Appendix B.

The Shilnikov type homoclinic orbit in an analytic style.

To apply the Shilnikov theorem to the system $(A.1)$, we have to prove that the system has a homoclinic Shilnikov orbit at the equilibrium point P^*. If the parameters lie in the following subsets:

$\Omega_3^A = \left\{\rho \in (\delta, -\psi), \sigma \in (0.1, \rho/\psi), \gamma \in (\frac{(1-\beta)(\rho-\delta)}{\delta}, \overset{\leftrightsquigarrow}{\gamma})\right\}$,

where $\overset{\leftrightsquigarrow}{\gamma}$ is the Hopf bifurcation value found in Mattana and Venturi (1999). By using Cardano Formula, and numerical evaluation we shown that the singular equilibrium point $P^* \in \Omega_3^A$ is a hyperbolic saddle focus of index 2 .

In other words, the eigenvalues of the Jacobian matrix of the system (A.1) evaluated in $P^*$are of the form $r = r_1$ and $r_{2/3} = -p \pm iq$,a saddle focus, with $r_1 > 0, p > 0,\ q \neq 0$ and $r_1 > p > 0$. We remember that a homoclinic orbit joining the **equilibrium point** $P^*$of system $(A.1)$ is **doubly asymptotic** with respect to time t along the solution manifold.

We translate the equilibrium point P^* in the origin $W^*(0,0,0)$ and we put the system (A.1) in normal form

Body Math

$$\begin{aligned}
\dot{w_1} &= r\ w_1 + F_{1a}w_1w_2 + F_{1b}w_1w_3 + F_{1c}w_2w_3 + F_{1d}w_1^2 + F_1ew_2^2 + F_{_1f}w_3^2;\\
\dot{w_2} &= pw_2 - qw_3 + F_{2a}w_1w_2 + F_{2b}w_1w_3 + F_{2c}w_2w_3 + F_{2d}w_1^2 + F_{2e}w_2^2 + F_{2f}\ w_3^2\\
\dot{w_3} &= qw_2 + pw_3 + F_{3a}w_1w_2 + F_{3b}w_1w_3 + F_{3c}w_2w_3 + F_{3d}w_1^2 + F_{3e}w_2^2 + F_{3f}\ w_3^2
\end{aligned} \tag{B.1}$$

We compute the stable and unstable manifolds of the saddle focus equilibrium point W^* to construct the Shilnikov type homoclinic orbit in an analytic style. Parameter values are set as $\beta = 0.76$, $\rho = 0.055$, $\delta = 0.05499$, $\sigma = 0.1$ and $\gamma = 0.\ 042$. So let's begin with the **analytic expression** of the **one-dimensional unstable manifold** associated with **the real eigenvalue** r_1 where $a_m.b_m.c_{\mathrm{m}}$.are undetermined coefficients.

So for t¡0 the trajectory will tend to zero (to steady state) along the unstable manifolds.

$$w_1(t) = a_0 + \sum_{k=1}^{\infty} a_k e^{krt};$$
$$w_2(t) = b_0 + \sum_{k=1}^{\infty} b_k e^{krt};$$
$$w_3(t) = c_0 + \sum_{k=1}^{\infty} c_k e^{krt}$$

When $k = 4$ we get:

$$w_1(t) = \xi e^{0,053071\,21t} - \xi^2 0,10856 e^{0,106\,14t} - \xi^4 0,00064 e^{0,159\,214t} - \xi^8 1,5E-08\ e^{0,212\,285t}$$

$$w_2(t) = -0,275059245\xi^2\ e^{0,106\,142t} + 0,5911103\ \xi^4 e^{0,159\,214t}\ - \xi^8 83,1E-07\ e^{0,212\,285t}$$

$$w_3(t) = -0,3075101\xi^2\ e^{0,106\,142t} - 0,008405\xi^4\ e^{0,159\,214t} + \xi^8 80,0003196 e^{0,212\,285t}$$

We choose $\xi \leq 1$.

As t$\to \infty$, the trajectory will tend to zero along the stable manifold. We choose $r_2 = -p + iq$ the complex eigenvalue

$$w_1(t) = a_0 + \sum_{k=1}^{\infty} a_k(\varsigma,\eta) e^{k(-p+iq)t} = \sum_{k=2}^{\infty} [a_k^1(\varsigma,\eta) + i a_k^2(\varsigma,\eta)] e^{k(-p+iq)t}$$

$$w_2(t) = b_0 + \sum_{k=1}^{\infty} b_k e^{k(-p+iq)t} = \sum_{k=1}^{\infty} [b_k^1(\varsigma,\eta) + i b_k^2(\varsigma,\eta)] e^{k(-p+iq)t}$$

$$w_3(t) = c_0 + \sum_{k=1}^{\infty} c_k e^{k(-p+iq)t} = \sum_{k=1}^{\infty} [c_k^1(\varsigma,\eta) + i c_k^2(\varsigma,\eta)] e^{k(-p+iq)t}$$

$$w_1(t) = e^{-2pt}\ [a_2^1(\varsigma,\eta)\cos(2q) + i a_2^1(\varsigma,\eta) sin(2q)] + i[a_2^2(\varsigma,\eta)\cos(2q) + \grave{\imath} a_2^2(\varsigma,\eta)\sin(2q)] +$$

...

$w_2(t) = e^{-pt}[-\varsigma\cos(q) - i\varsigma\sin(q)] + i[\eta\cos(q) + \grave{\imath}\eta\sin(q)]+$
$+e^{-2pt}\ [b_2^1(\varsigma,\eta)\cos(2q)+ib_2^1(\varsigma,\eta)\sin((2q)]+i[b_2^2(\varsigma,\eta)\cos(2q)+\grave{\imath}b_2^2(\varsigma,\eta)\sin(2q)]....$
$w_3(t) = e^{-pt}[-\eta\cos(q) - i\eta\sin(q)] + i[\varsigma\cos(q) + \grave{\imath}\varsigma\sin(q)]$
$+e^{-2pt}\ [c_2^1(\varsigma,\eta)\cos(2q)+ib_2^1(\varsigma,\eta)\sin((2q)]+i[c_2^2(\varsigma,\eta)\cos(2q)+\grave{\imath}c_2^2(\varsigma,\eta)\sin(2q)]....$

a_{12}=$-3,7E-06$;b_{12}=$-0,004825$;c_{12}=$-0,00046$
a_{13}=$5,49E-10$;b_{13}=$4,03E-06$;c_{13}=$4,66E-07$
a_{22}=-1,1E-09;b_{22}=7,148E-07;c_{22}=$-1E-07$
a_{23}=-1,1E-23;b_{23}=-1,2E-20;c_{23}=-4,1E-20
Body Math When $k = 3$ we get:

$w_1(t) = e^{-0,00302t}\ \varsigma^2([(3,7E{-}06)\cos(0,11053t){-}(-1,1E{-}09)\sin(0,11053t)]+$
$+\grave{\imath}[(-1,1E-09)\cos(0,11053t) + (3,7E-06)sin(0,11053t))+$
$+e^{-0,0045t}\ \varsigma^4([(5,49E-10)\cos(0,016579t) - (1,1E-23)\sin(0,016579t) +$
$i[(1,1E-23)\cos(0,016579t) + (5,49E-10)\sin(0,016579t)])...$

$w_2(t) = e^{-0,00151t}\varsigma([-\cos(0,005523t) - \sin(0,005523t)] + i[\cos(0,005523\ t) -$
$\sin(0,005523\ )])+$
$+e^{-0,00302t}\ \varsigma^2(([(-0,004825)\cos(0,11053t) - (7,148E-07)\sin(0,11053t)] +$
$i[(7,148E-07)\cos((0,11053t)) - 0,004825\sin(0,11053t]+$
$+.e^{-0,0045t}\ \varsigma^4([(4,03E-06)\cos(0,016579t) - (1,2E-20)\sin(0,016579t) +$
$i((-1,2E-20))\cos(0,016579t) + (4,03E-06)\sin(0,016579t))]$

$w_3(t) = e^{-0,00151t}\varsigma([-\cos(0,005523t) - \sin(0,005523t)] + i[\cos(0,005523\ t) -$
$\sin(0,005523\ )])+$
$+e^{-0,00302t}\ \varsigma^2(([(-0,00046)\cos(0,11053t) - (7,148E-07)\sin(0,11053t)] +$
$i[(7,148E-07)\cos((0,11053t)) - (0,00046)\sin(0,11053t]+$
$+.e^{-0,0045t}\ \varsigma^4([(4,66E-07)\cos(0,016579t) + (4,1E-20)\sin(0,016579t) +$
$i((-4,1E-20)\cos(0,016579t) + (4,66E-07)\sin(0,016579t))]$

We assume $\varsigma = \eta \leq 1$.

Chaotic Modeling and Simulation (CMSIM) 3: 255-259, 2014

Comparison of non-relativistic and relativistic Lyapunov exponents for a low-speed system

Shiuan-Ni Liang and Boon Leong Lan*

School of Engineering, Monash University, 46150 Bandar Sunway, Selangor, Malaysia
*(E-mail: lan.boon.leong@monash.edu)

Abstract. The Newtonian and special-relativistic Lyapunov exponents are compared for a low speed system – the periodically-delta-kicked particle. We show that although the agreement between the Newtonian and special-relativistic transient Lyapunov exponents rapidly breaks down initially, they converge to values which are very close to each other.
Keywords: kicked particle, Lyapunov exponent, special relativity, Newtonian approximation

1 Introduction

It is conventionally believed [1-3] that if the speed v of a dynamical system is *low* compared to the speed of light c, that is, $v << c$, then the special-relativistic dynamical predictions for the system will be well-approximated by the Newtonian predictions. However, it was shown in recent numerical studies [4-9] that, contrary to the conventional belief, the agreement between the Newtonian and special-relativistic dynamical predictions for a single trajectory [4-7] and for an ensemble of trajectories [8,9] can break down completely although the speed of the system is low. Here, we extend the previous studies [4-9] to a comparison of the Newtonian and special-relativistic predictions for the Lyapunov exponent of a prototypical chaotic Hamiltonian system – the periodically-delta-kicked particle – at low speed. Details of the system and calculations will be given next, followed by the results and discussion.

2 Method

In the Newtonian framework, the equations of motion for the periodically-delta-kicked particle are reducible to an exact mapping, which is called the standard map [10,11]:

$$P_n = P_{n-1} - \frac{K}{2\pi}\sin\left(2\pi X_{n-1}\right) \quad (1)$$

$$X_n = \left(X_{n-1} + P_n\right) \bmod 1 \quad (2)$$

Received: 17 March 2014 / Accepted: 12 July 2014

ISSN 2241-0503

where X_n and P_n are, respectively, the dimensionless scaled position and momentum of the particle just before the nth kick (n = 1, 2, …), and K is a dimensionless positive parameter.

In the special-relativistic framework, the equations of motion for the periodically-delta-kicked particle are also reducible to a mapping, which is called the relativistic standard map [12,13]:

$$P_n = P_{n-1} - \frac{K}{2\pi}\sin(2\pi X_{n-1}) \tag{3}$$

$$X_n = \left(X_{n-1} + \frac{P_n}{\sqrt{1+\beta^2 P_n^2}} \right) \bmod 1 \tag{4}$$

where β, like K, is also a dimensionless positive parameter.

The transient Lyapunov exponent for a map is generally defined [14] as

$$\lambda_n = \frac{1}{n}\ln[\mathrm{abs}(\mathrm{trace}\, M_n)] \tag{5}$$

where $M_n = J_n\, J_{n-1} \ldots J_2\, J_1$ and J_n is the Jacobi matrix. In the limit $n \to \infty$, λ_n yields [14] the largest Lyapunov exponent. A hallmark of chaos is the existence of a positive Lyapunov exponent. For the standard map in Eqs. (1) and (2), the Jacobi matrix is

$$J_n = \begin{bmatrix} 1 & -K\cos(2\pi X_n) \\ 1 & 1-K\cos(2\pi X_n) \end{bmatrix}. \tag{6}$$

For the relativistic standard map in Eqs. (3) and (4), the Jacobi matrix is

$$J_n = \begin{bmatrix} 1 & -K\cos(2\pi X_n) \\ \left(1+\beta^2 P_{n+1}^2\right)^{-3/2} & 1-\left(1+\beta^2 P_{n+1}^2\right)^{-3/2}[K\cos(2\pi X_n)] \end{bmatrix}. \tag{7}$$

In each theory, the transient Lyapunov exponent [Eq. (5)] is calculated twice to determine its accuracy. The calculation for the transient Lyapunov exponent is first performed in 32-significant-figure precision and then repeated in quadruple (35 significant figures) precision. The accuracy of the transient Lyapunov exponent is determined by the common digits of the 32-significant-figure-precision and quadruple-precision calculations. For example, if the former calculation yields 1.234… and the latter calculation yields 1.235…, the transient Lyapunov exponent is accurate to 1.23.

3 Results and discussion

Here we will present an example to illustrate the typical result. In this example, $X_0 = 0.5$, $P_0 = 99.9$, $K = 7.0$ and $\beta = 10^{-7}$. For these initial conditions and parameters, both the Newtonian and special-relativistic trajectories are

chaotic. In this case, the speed of the particle is low, about $10^{-5}c$, up to 8800 kicks.

Fig. 1, which plots the Newtonian and special-relativistic transient

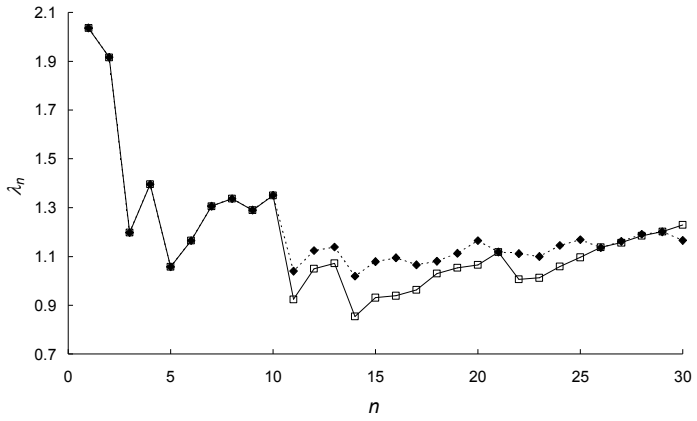

Fig. 1. Newtonian (squares) and special-relativistic (diamonds) transient Lyapunov exponents versus kick.

Lyapunov exponents for the first 30 kicks, shows that the two transient Lyapunov exponents agree with each other for the first 10 kicks but the agreement breaks down from kick 11 onwards. The agreement between the Newtonian and special-relativistic transient Lyapunov exponents breaks down rapidly because the difference between the two grows, on average, exponentially – see Fig. 2. The exponential growth constant of the difference

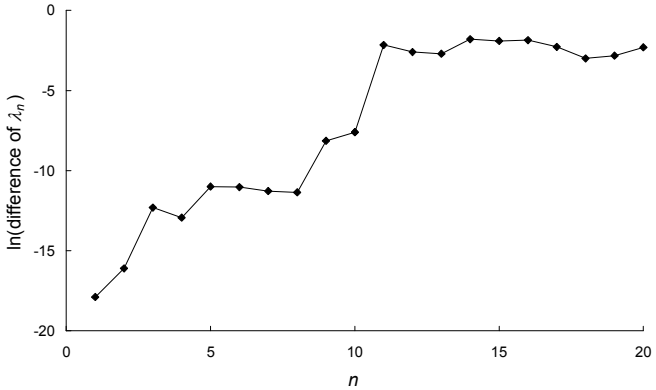

Fig. 2. Difference between the Newtonian and special-relativistic transient Lyapunov exponents versus kick.

between the two transient Lyapunov exponents, measured from kick 1 to kick 10, is 0.96.

However, asymptotically, the Newtonian and special-relativistic transient Lyapunov exponents converge to values which are very close to one another. In particular, at kick 8800, the Newtonian and special-relativistic transient Lyapunov exponents are both accurate to 1.27, which is quite close to the analytical estimate [10] of the asymptotic Newtonian Lyapunov exponent given by $\ln(K/2) = 1.253$. This result is surprising since the chaotic trajectories predicted by the two theories agree only for the first 16 kicks, which suggests that the two asymptotic Lyapunov exponents should not agree.

Conclusions

We have shown that although the agreement between the Newtonian and special-relativistic transient Lyapunov exponents rapidly breaks down initially, the asymptotic special-relativistic Lyapunov exponent is well-approximated by the asymptotic Newtonian value. The same result should hold for other low-speed chaotic Hamiltonian systems since the periodically-delta-kicked particle is a prototype.

Acknowledgment

This work was supported by a Fundamental Research Grant Scheme, FRGS/1/2013/ST02/MUSM/02/1.

References

1. A. P. French. Special relativity, Thomas Nelson & Sons, Cambridge, 1968. p. 167.
2. W. D. McComb. Dynamics and relativity, Oxford University Press, Oxford, 1999. preface.
3. J. B. Hartle. Gravity: an introduction to Einstein's general relativity, Addison-Wesley, San Francisco, 2003. p. 88.
4. B. L. Lan. Comparison of the Newtonian and relativistic predicted trajectories for a low-speed periodically delta-kicked particle, Chaos, 16, 033107, 2006.
5. B. L. Lan and H. Y. Cheng. Newtonian and special-relativistic predictions for the trajectory of a slow-moving dissipative dynamical system, Commun. Nonlinear Sci. Numer. Simulat., 15, 2497-2503, 2010.
6. B. L. Lan and F. Borondo. Newtonian and special-relativistic predictions for the trajectories of a low-speed scattering system, Phys. Rev. E, 83, 036201, 2011.
7. B. L. Lan. Implications of the disagreement between Newtonian and relativistic low-speed trajectories, Chaos, Solitons and Fractals, 42, 534-537, 2009.
8. S. N. Liang, F. Borondo and B. L. Lan. Newtonian versus special-relativistic statistical predictions for low-speed scattering, PLoS ONE, 7, e48447, 2012.

9. S. N. Liang and B. L. Lan, Statistical predictions for the dynamics of a low-speed system: Newtonian versus special-relativistic mechanics, PLoS ONE, 7, e36430, 2012.
10. B. V. Chirikov. A universal instability of many-dimensional oscillator systems, Phys. Rep., 52, 263-379, 1979.
11. G. Casati, B. V. Chirikov, F. M. Izrailev, and J. Ford. Stochastic behavior of a quantum pendulum under a periodic perturbation, in Stochastic behavior in classical and quantum Hamiltonian systems, edited by G. Casati and J. Ford, Springer, Berlin, 1979. pp. 334–351.
12. A. A. Chernikov, T. Tél, G. Vattay, and G. M. Zaslavsky. Chaos in the relativistic generalization of the standard map, Phys. Rev. A, 40, 4072-4076, 1989.
13. Y. Nomura, Y. H. Ichikawa, and W. Horton. Nonlinear dynamics of the relativistic standard map, Phys. Rev. A, 45, 1103-1115, 1992.
14. R. F. Fox and T. C. Elston. Chaos and the quantum-classical correspondence in the kicked pendulum, Phys. Rev. E, 49, 3683-3696, 1994.

Chaotic Modeling and Simulation (CMSIM) 3: 261-269, 2014

Spatial and temporal imaging of a plasma jet plume

V J Law[1*], L J Cox[2], W Adress[2], W G Graham[2] and D P Dowling[1]

[1]School of Mechanical and Materials Engineering, University College Dublin, Belfield, Dublin 4, Ireland.
[2]Centre for Plasma Physics, Queen's University of Belfast, Belfast, BT7 INN, UK
*(e-mail: viclaw66@gmail.com)

Abstract: Cold atmospheric pressure plasma jets have been shown to exhibit considerable potential for use in plasma medicine applications such as in wound treatment. New pulsed atmospheric pressure plasma jets are being developed that have inherent plasma stability and low gas temperatures. This study examines a new digital enhancement technique to characterise the far field plasma plume and effluent region of the plasma. The digital technique provides spatial information that identifies possible gas treatment zones for medical applications. Using images from a fast a capture (10 μm second) ICCD camera the study shows the luminous plume extends up to 7 mm from the reactor exit nozzle and has a kinked, or wrinkled, appearance but nonluminous perturbation of the gas is detected up to 3 cm away to the front and either side of the visible plasma plume.

Keywords: Imaging, Atmospheric plasma jet; Diagnostics.

1. Introduction

The development of the cold temperature atmospheric pressure plasma jets in recent years has led to the promising new science of plasma medicine. Treatments are generally applied using a hand-held atmospheric plasma sources that utilise a wide range of electric drive frequencies and reactor geometries. Examples of cell treatment leading to apoptosis using these plasma jets have been reported by a number of authors [1, 2]. One of the first clinically proven hand-held plasma jets is the kINPen med® developed by the Leibniz Institute for Plasma Science and Technology (INP), Greifswald, Germany in cooperation with neoplas GmbH, Greifswald, Germany is now undergoing in-vivo clinical investigation of plasma antiseptic properties on human skin [3], chronic venous leg ulcers [4] and cosmetic surgery [5]. These clinical trials require the relatively small 1.6 mm diameter plasma to treat large areas of thermally sensitive living tissue and microorganisms. Earlier studies using the kINPen 0.9 versions [6-8] of the plasma jet on microorganism have shown that cells are killed outside the visible plasma plume immediate treatment area, indicating

Received: 12 December 2013 / Accepted: 25 July 2014

ISSN 2241-0503

what has been termed a 'spillover' occurs [9]. Further to this, kINPen med® plasma induced activation studies on poly(ethylene-terephthalate) PET at a nozzle-to-surface distance of 5-15 mm have shown that a similar immediate activation (1 day) post treatment 'spillover' can be induced up to 20 mm in diameter on the polymer surface [10].

This work reports on the spatial and temporal visual imaging of the kINPen med® plasma plume fluid structure using a photodiode (PD) to trigger a gated ICCD camera, with the addition of a new digital image processing technique of the ICCD camera images. This post image processing technique is used to enhance the immediate area (up to a distance of approximately 3 cm) around the luminous plasma plume to reveal the fluid structure emanating from the gas flow. This digital image enhancement approach differs from the shadowgraph and Schlieren imaging technique previously for air/hydrogen jet [11], air discharge [12] and helium jets [13, 14] all of which probe the use of back lighting to probe the refractive index changes by density gradients in the fluid distortion Here no back lighting is required. This approach differs from the high temporal resolution flame-front visualization technique [15], and also differs from large time scale (10s) flow imaging of complex vortex mixing in DBDs [16]. In this work the widely available National Instrument LabVIEW software packages is used as an example.

2. Experiment apparatus and methods

Figure 1a shows a photograph of the plasma jet used in this study. The plasma reactor is a cylindrical dielectric barrier discharge made from a glass ceramic with an internal diameter of D = 1.6 mm. The inner metal electrode has a diameter of ~0.3 mm. The outer body is grounded to produce a cross-field jet configuration i.e. an electric field perpendicular to the gas flow. Here a gas flow rate of 5 SLM of 99.99% pure argon is used, equating to a gas velocity through the reactor tube of v = 36.7 $m.s^{-1}$. Since the plasma region is 20 mm long there is a gas residence time of about 0.5 ms.

Figure 1: Photograph of the kINPen Med® plasma interacting with a fingertip.

The inner electrode is powered by a 1 MHz electrical drive frequency that is

pulse modulated with 2.5 kHz square wave (50 % duty cycle) signal [10]. In this work the plasma plume, and effluent expands unobstructed into atmospheric pressure air, and is investigated with a PD, fast imaging camera and post image capture enhancement technique to reveal the fluid structure around the plasma plume (0-30 mm) in front and to the side of the plume.

The PD used was a Hamamatsu MPPC with a rise time of 10 ns and spectral range between 320 and 900 nm [17]. The light was collected at right angles to the plume, 1 mm downstream of the nozzle exit, via a fibre optic with a collimating lens, the combination producing a focal area of 1 mm in diameter at a length of 6 mm from the lens: Thus making the interrogation area smaller than to the diameter of the jet discharge (~1.6 mm). The rising edge of the 2.5 kHz modulated plasma light is used to trigger the ICCD Camera.

The Andor iStar 334T ICCD camera is used to capture the plasma images. A 14 cm focal length glass lens focused the region from between 2 mm upstream to 20 mm downstream of the exit nozzle. Using this combination the overall optical chain (between camera and plasma-plume) is of the order of 2 m and the camera spectral range is restricted to 300 to 850 nm by the glass lens. The camera was triggered, via a delay generator, from the rising edge of the PD signal. Within the camera the images are processed using a false-colour scale from blue (low intensity) to yellow (high intensity) for maximum visual differentiation the gain was set to 2817 out of a maximum of 4095, where the final digital images are formatted as a 24-bit red-green-blue (RGB) JPEG (Joint Photographic Experts Group) with a N x N pixel array, where N = 1024. Through an initial survey of the pulse-on and pulse-off periods of plasma the ICCD was synchronised to the respective time periods.

The gas behaviour beyond the luminous plasma region is explored by using LabVIEW based software [18]. This software essentially extracts the lowest intensity colour plane (blue plane) from the original RGB image and then uses pixel resolution enhancement through digital filtering and a thresholding algorithm. Care is taken at each step to ensure that the morphology in the recorded data is not distorted by reference at each step to experimentally available information, and the goals of the operation and limitations of the algorithms. The final images were achieved using four standard sequential steps.

1. The 8-bit “blue” plane is selected from the original 24-bit RGB image.
2. A fast Fourier Transform (FFT) is then applied to this plane to convert the spatial information into its frequency domain. A low-pass filter is used to smooth the noise with a truncation process to remove any remaining high frequency component above the user defined cut-off point.
3. An inverse FFT is then applied to bring the frequency domain data back into the spatial domain.
4. A local Nibalck thresholding segmentation algorithm is then used to produce a binary image. In this operation the background particles are set to I = 0 (black) while setting fluid structure to a pixel value of I = 1 (white). The result of this process produces a black-and-white binary image that represents the fluid structure within the original blue image.

3. Results.

3.1 Visible plasma imaging

Figures 2 provide examples of 15 individual ICCD images sampled from 31 images obtained for the argon plasma. The images span from the beginning of the pulse at t = 0 μs to the end stages of the pulse at, t = 185 μs. With the gain fixed at 2817 each image has the same intensity scale and therefore their intensities may be compared directly. To add comparison a scale bar is displayed at the top of the figure. The figure shows a linear increase in the length of the plume between 0 μs to 40 μs and rapid decrease in length beyond 185 μs when the plume is almost completely gone. Apart from the earliest and latest times the plumes vary in visible length and exhibit a kinked or wrinkled structure along the length of each plume.

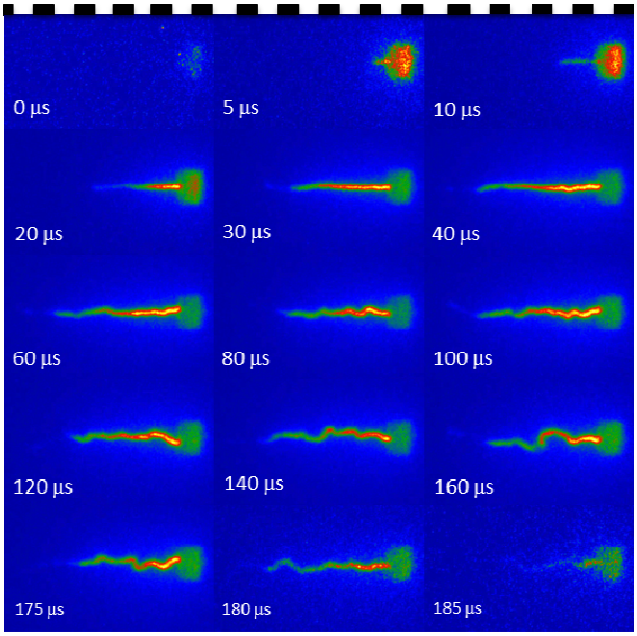

Figure 2: A selection (a total 15) of space and time resolved images of the nozzle and argon plasma.

Using all the 31 ICCD images, the distal length of each discharge plume have been calculated but are not shown here. The calculations reveal that the plasma expands from the nozzle and reaches, and maintains, a maximum length of about 4.5 or 6 mm until the voltage pulse is terminated. The initial velocity of the visible plume front is about 200 $m.s^{-1}$. However at about 4.25 mm the argon the front rapidly accelerates to about 300 $m.s^{-1}$ before reaching its maximum length with a periodic cycling ranging from 6.5 to 5 mm: with each cycle period taking 40 to 45 microseconds, which equates to a frequency of 20 to 22 kHz.

3.2 Spatial enhancement of non-visible plasma region

Figure 3 show a screen shot image of the LabVIEW colour plane extraction and line profile front panel for the pulse-on period. In this figure the left-hand images is the original 32-bit image with interactive line intensity profile (LIP) cursor; the second column of images are the three extracted blue, green and red planes (presented here in grayscale); the third column of graphs depict the selected LIP for each plane; and the final column is basic descriptive analysis of the LIP for each plane. The information presented on this front panel reveals that majority of the plasma information (white to grey colours) is aligned along the flow of the plume in the red and green planes. In contrast the far-field low intensity fluid structure information is captured within the blue plane as speckled noise surrounding the plume with an outer white ring at a typical distance of 2-4 plume diameters either side of the plume.

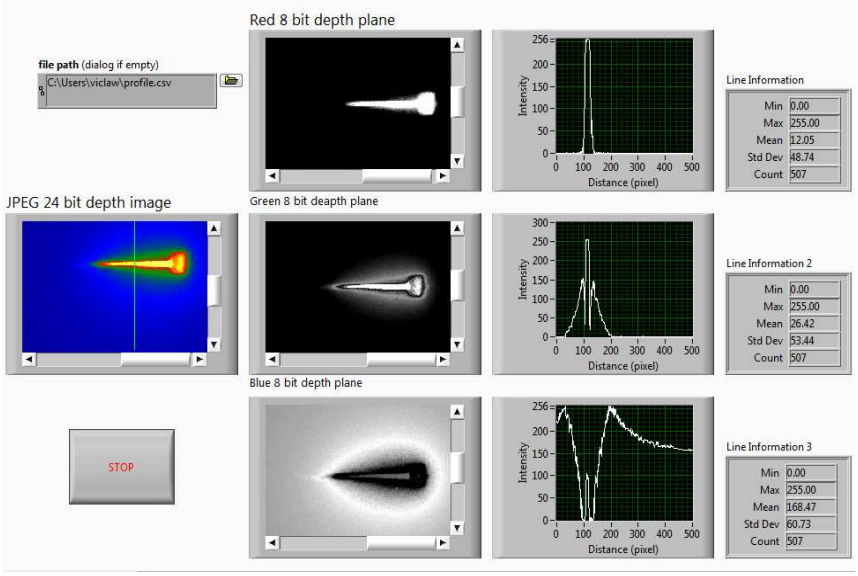

Figure 3: LabVIEW RGB colour plane and line profile.

We now turn to the digital filtering and threshold processing of the blue image. Figure 4 shows the processing of the duration of the pulse-on period and the duration of pulse-off period. It is interesting the structure observed on short time scale (figures 2) is absent in the long exposure image. It is also apparent from figure 4b that there is afterglow. In figure 4c we are imaging the structure of the background gas. This shows a distinct ripple-like feature centred in the proximity of the maximum light emission from the plume. Figure 4f show that this is absent when there is no discharge present.

To understand these fluid structure images we consider the dimensionless Reynolds number (R_e) as defined in equation 1 when interpreting figures 4c and 4f, as it provides a measure of the ratio of inertial forces to viscous forces and quantifies the relative importance of these two types of forces.

$$R_e = \frac{QD}{v_k A} \tag{1}$$

Where Q is the neutral argon flow rate (8.35 x 10-5 $m^3.s^{-1}$), D is the diameter of the nozzle (0.0016 m), v_k is the gas kinematic viscosity (0.000014 $m^2.s^{-1}$) and A is the cross sectional area of the nozzle (2 x 10^{-6} m^2). For argon gas flows of 5 SLM, R_e equates to 4465 which implies the inertial forces are expected to be more dominant than viscous forces and large-scale fluid motion would be un-damped in the pulse-off period. However when the plasma is turn-on, the neutral gas flow rate will increase due to the associated gas heating.

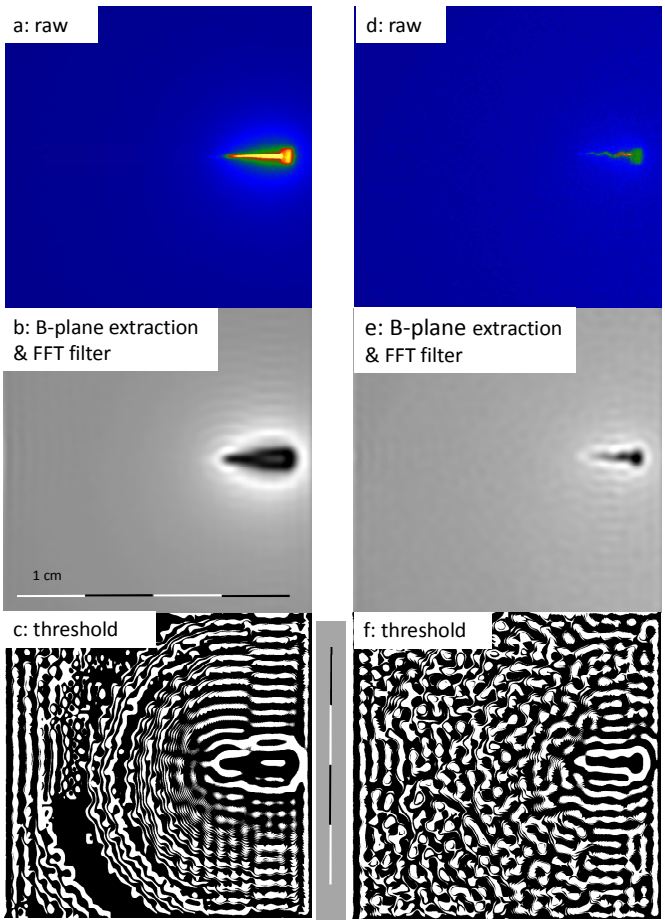

Figure 4: 0.2 μ second exposure ICCD images of plasma in pulse-on (a) and pulse-off period (d); images (b) and (c) depict the image enhancement of the pulse-on period; and images (e) and (f) depict the image enhancement for the pulse-off period.

Considering the processed image of the pulse-on period (figure 4c) a ripple structure is observed to radiate from a point along the axis of the plasma plume and extends with a complex structure in the direction of effluent flow up to 4 cm from the nozzle. This repeating far-field wave-like structure with a white peak distances separation of typically 1-2 mm is within an order of magnitude of the expected travel distance of the neutral gas within the capture time-frame of the camera image. In addition the ripple pattern is found to be asymmetric with respect to the effluent flow axis, producing a complex broken structures to the top beyond which the discontinuities the structures extend into the ambient air. The distance disturbance occurs at around 0.5 cm from the plume distal point.
In the case of the pulse-off period (figure 4f) the wave-like structures has collapsed to form irregular and small-scale chaotic structures with scale lengths of the order of the nozzle diameter. These observations are consistent with the Reynolds number dimensionless analysis and the loss of driving force to heat the plasma gas when the electrical drive power is switched-off. Under these conditions the heated gas is expected to begin to equilibrate with the surrounding ambient air. This imaging technique is also supported by the work of Roberts et al who have used the Schlieren technique to look at a pulsed helium jet. In their work they found. They found rapid changes in the stability at the start and end of the pulse period.

Using the work of Ghasemi *et al* [14], we can obtain, to a first approximation, the increase in gas velocity when the plasma is turn-on by using the continuing mass flow equation (2). In this equation: $\rho_{1,2}$ are the densities of the argon gas at room temperature (1.62 kg.m^3) and plasma temperature 1.47kg.m^3 (3*30 K* [1, 2]). A is the cross-sectional area of the jet nozzle and $V_{1,2}$ are the argon gas velocities at room temperature and plasma temperature, respectively.

$$\rho_1 A V_1 = \rho_2 A V_2 \tag{2}$$

Assuming the argon gas mass flow rate is the same in both case (5 SLM), the argon gas velocity increase from 41 m.s^{-1} at room temperature (~300 *K*) to 45 m.s^{-1} at the expected plasma temperature of ~33.0 *K*.

4. Conclusion

The spatial and temporal visual imaging of an argon-based pulsed plasma jet designed for medical use has been studied using photodiode and ICCD camera imaging, plus post exposure enhancement of the camera images. This combined measurement and diagnostic approach provides a spatial and temporal picture of the plasma plume and its effluent. The PD measurements show that the plasma is modulated by a fast rising and falling 2.5 kHz square wave time-base profile. Microsecond time scale imaging of the discharge using the ICCD camera reveals that the argon plasma plume is continuous through the 0.2 ms pulse-on period of the discharge. However the plume morphology takes on a kinked or wrinkled appearance. In addition the plume rapidly decays at the end of the

voltage pulse suggesting micro-turbulence is the driving force in the production of the kinks within the plasma jet.

To gain access to the effluent gas being expelled from the plasma plume the technique of image plane extraction has been developed and demonstrated. Here the blue plane of the ICCD digital images has revealed pulsed plasma induced fluid structures extending up to 2-3 cm form the visible plume. This far-field fluid structure information may be used in the understanding 'spillover' effect when plasma treating thermally sensitive polymers and their biomaterial counter parts.

Acknowledgement

L. J. Cox is supported by a grant from the Department of Education and Learning Northern Ireland, and W. Adress is supported by the Iraqi Ministry of Higher Education and Scientific Research. This work is part supported by Science Foundation Ireland Grant 08/SRC114; COST ACTION programme and; the 7th CHAOS2014 International Conference, Lisbon, Portugal organizing committee.

References

1. Georgescu. N, and Lupu. A R. Tumoral and normal cells treatment with high-voltage pulsed cold atmospheric plasma jets. IEEE. Trans. Plasma Sci. 38(8), 1945-1955, 2010.
2. Ahn. H. J, Kim. K Il, Moon. Kim. G, Yang. S. S, Lee. J-S. Atmospheric-pressure plasma jet induces apoptosis involving mitochondria via generation of free radicals. PluS One. Plasma Sci. 6(11), e28154 (7pp), 2011.
3. Lademann. J, *et al*. B. Risk assessment of the application of tissue-tolerable plasma on human skin. Clinical Plasma Medicine 1, 5-10, 2013.
4. Isbary. G, *et al*. Non-thermal plasma-More than five years of clinical experience. Clinical Plasma Medicine 1, 19-23, 2013.
5. Metelmann. H-R, *et al*. Experimental recovery of CO_2-laser skin lesions by plasma stimulation. American Journal of Cosmetic Surgery 29(1), 52-56, 2012.
6. Weltmann. K-D, *et al*. Atmospheric pressure plasma jet for medical therapy: plasma parameters and risk estimation, Contrib. Plasma Phys. 49(9) 631-64, 2009.
7. Daeschlein. G, *et al*. In vitro killing of clinical fungal strains by low temperature atmospheric pressure plasma jet, IEEE Trans. Plasma Sci. 39(2), 815-821, 2011.
8. Daeschlein. G, *et al*. Skin decontamination by low temperature atmospheric pressure plasma jet and dielectric barrier discharge plasma, J. Hosp. Infect. 81 177-183, 2012
9. Kolb. J. F, *et al*. Cold dc-operated air plasma jet for the inactivation of infectious microorganisms. IEEE Trans, Plasma Sci. 40(11), 3007-3026, 2012.
10. Law. V. J, *et al*. Resonances and patterns within the kINPen-MED

atmospheric pressure plasma jet. CMSIM Journal. 1, 3-10, 2014.
11. Carlsson. T. E, *et al*. Combination of schlieren and pulsed TV holography in the study of a high-speed flame jet. Optic and Lasers in Engineering. 44(6), 535–554, 2006.
12. Sutton. Y, Moore. J, Sharp. D, and Braithwaite. N. St. Looking into a plasma loudspeaker. IEEE Trans, Plasma Sci. 39(11), 2146-21487, 2011.
13. Ghasemi. M, Olszewski. P, Bradley. J. W, and Walsh. J. L. Interaction of multiple plasma plumes in an atmospheric pressure plasma jet array. J. Phys. D: Appl. Phys, 46, 052001 (6pp), 2013.
14. Roberts. E, Sarron. V, Darny. T, Ries. D, Dozias. S, Fontane J, Joly. L, and Pousvesle J-M. Rare gas flow structuration in plasma jet experiments. PSST. 22 012003 (5pp), 2014.
15. Wisman. D, and Ganguly. B. Flame-front visualization using argon–oxygen plasma discharge. IEEE Trans, Plasma Sci. 39(11), 2066-2067, 2011.
16. Dahai Yu, *et al*. Influence of a rotating electrode on the uniformity of an atmospheric pressure air filamentary barrier discharge. Plasma Process. Polym. 2013.
17. O'Neill. F. T, Twomey. B, Law. V. J, Milosavljević. V, Kong. M. G, Anghel. S. D, and Dowling. D. P. Generation of active species in a large atmospheric pressure plasma jet. IEEE Trans. Plasma Sci. 40(11) 2994-3004, 2012.
18. Law. V. J, O'Neill. F. T, and Dowling. D. P. Automatic computation of crossing points within orthogonal interpolation line graphs. *How nature works: Emergence, Complexity and Computation*. 195-216, Eds. I Zelinka, A. Sanayei, H. Zenil, and O. E. Rössler. (Springer International Publishing Switzerland 2013).

Chaotic Modeling and Simulation (CMSIM) 3: 271-279, 2014

Reconstruction of Evaporation Dynamics from Time Series

Özlem BAYDAROĞLU and Kasım KOÇAK

Istanbul Technical University, Department of Meteorology, 34469, Istanbul, Turkey

E-mails: baydaroglu@itu.edu.tr, kkocak@itu.edu.tr

Abstract: The maximum amount of water loses from the reservoirs take place through evaporation. Thus it is important to know the dynamical system that governs the evaporation process. In this study, the Trajectory Method has been applied in order to obtain the differential equation from reconstructed phase space using evaporation time series. The trajectory method has been successfully applied in order to obtain the dynamical system that represents the periodic behavior of evaporation process.

Keywords: Dynamical system, Trajectory Method, Ordinary Differential Equations, Water Losses, Evaporation

1. Introduction

Water is the most vital substance for sustainability of life on planet earth. Unfortunately its distribution on earth both in time and in space is not uniform. This means that the water problem existed in the past, exists today and will exist in the future. On the other hand, especially in recent years water problem has gained much importance due to climate change. The state of the art climate models have shown that water related problems will be experienced more frequently in the future. This worsens the water related problems to a great extent. Thus it is mandatory to make intensive researches on the water resources and managements. In this context, water loses from all kind of water reservoirs are very important to be brought to a minimum level. As known well, the maximum amount of water loses from the reservoirs take place through evaporation. Thus it is important to know the dynamical system that governs

Received: 17 July 2013 / Accepted: 12 July 2014

ISSN 2241-0503

the evaporation process. In this study, the Trajectory Method has been applied to reconstruction of differential equation that governs the behavior of evaporation process. The brief history of the trajectory method used in this study is as follows. Crutchfield and McNamara (1986) have made some important attempt to reconstruct the differential equation from time series. These two researchers have suggested two approximations about the issue. The first of them is the determination of local dynamic that considers the short-term behavior of the system while the second approach deals with the dynamic of the whole attractor that consider the long-term behavior of the system. Almost at the same time with the aforementioned studies, Cremers and Hübler (1986) have developed the flow method that considers the sort-term behavior of the system. The flow method is applied to all points on the attractor. Thus it does not consider the long-term behavior of the system dynamic. Then Breeden and Hübler have developed this approach to include all of the system variables that could not be observed. In the end, Eisenhammer et al. (1991) have combine both short and long-term behavior of the system and they called their approach "trajectory method". In this study, the trajectory method has been successfully applied in order to obtain the dynamical system of evaporation process.

2. Trajectory Method

Trajectory method is based on the reconstruction of differential equations which produce the trajectory resembling the original trajectory. In other word, the reconstructed model is the best possible model reflecting the original model (Perona et al., 2000).

A set first order ordinary differential equations can be given as

$$\dot{x} = f(x,t) \tag{1}$$

where x and t represent the variable vector and time, respectively. To reconstruct the equation of motion it is necessary to obtain the differential equations of model trajectory as close as possible to the original trajectory. On the other hand, mathematical form of the model should be determined ab initio.

According to theory of dynamical system, time evolution of a system can be given by its trajectories in a phase space. Coordinates of this space are formed by state variables which are necessary to reflect the time evolution of the system under study. Every trajectory in this space represents the different time evolution of the system that corresponds to different initial conditions. Phase portraits have distinct patterns that attract all trajectories. This type of a pattern is called attractor. All initial conditions of which trajectories captured from the attractor defines a domain of attraction. Systems that show deterministic evolution have low dimensional attractors like point, limit cycle and torus. These kinds of attractors can be characterized by an integer dimension. An important property of these kinds of attractors is that trajectories that converge onto them remain in a fixed distance from each other. This property ensures the

system to be predictable for a long period of time (Koçak, 1996).

It is possible to reconstruct the phase space from a time series of one state variable sampled at regular time intervals Δt. For this to be done, some information and topological properties (e.g. dimension) of the attractor should be first estimated from the time series. Dimension of an attractor is the number of variable necessary to define the dynamics of the underlying system.

Packard et al., (1980) have suggested the reconstruction of phase space in order to obtain some invariant measures from an observed turbulent or chaotic flow. This can be achieved via transformation of the dynamical process to a higher dimensional space (embedding) by adding an extra independent dimension until no further information gain is impossible. One of these coordinates is formed by the time series itself and the remaining independent coordinates are formed by derivatives of the time series up to $(m\text{-}1)^{th}$ order. As a result, phase portrait of time evolution of a dynamical system can be represented in a new m-dimensional space spanned by a single state variable and its successive derivatives.

In this study, phase space is reconstructed from univariate or single time series (evaporation). Thus it is necessary to mention briefly from phase space reconstruction. Let's take a time series given as

$$x_i \in R, \qquad i = 1, 2, \ldots, N. \tag{2}$$

Then the reconstruction procedure is given as

$$\begin{aligned} &X_i = (x_i, x_{i-\tau}, \ldots, x_{i-(m-1)\tau}) \in R^m \\ &i = 1+(m-1)\tau, 2+(m-1)\tau, \ldots, N-1, N \end{aligned} \tag{3}$$

where X_i is an m-dimensional vector.

This pseudo-phase space preserves the structure of the attractor embedded in the original phase space, (Takens, 1981). In Eq (3) τ is called time delay and should be calculated from time series by using autocorrelation function or mutual information function. Differential equation used in the trajectory method is assumed in the following form:

$$\dot{x}_i = \sum_{k=1}^{K} c_{i,k} F_{i,k}(x_1, x_2, \ldots, x_D) \qquad i = 1,2,\ldots,D \tag{4}$$

where $c_{i,k}$s are coefficients of differential equation and $F_{i,k}(x_1, x_2, \ldots, x_D)$s are approximating functions. On the other hand K and D represent the number of approximating function and state variable, respectively. If $F_{i,k}$ is chosen as

the 3rd degree polynomial then Eq (4) can given as

$$\dot{x}_i = \left\{ \begin{array}{l} c_{i,1} + c_{i,2}x_1 + c_{i,3}x_2 + c_{i,4}x_3 + c_{i,5}x_1x_2 + c_{i,6}x_1x_3 \\ + c_{i,7}x_2x_3 + c_{i,8}x_1^2 + c_{i,9}x_2^2 + c_{i,10}x_3^2 + c_{i,11}x_1x_2x_3 \\ + c_{i,12}x_1^2x_2 + c_{i,13}x_1^2x_3 + c_{i,14}x_1x_2^2 + c_{i,15}x_2^2x_3 \\ + c_{i,16}x_1x_3^2 + c_{i,17}x_2x_3^2 + c_{i,18}x_1^3 + c_{i,19}x_2^3 + c_{i,20}x_3^3 \end{array} \right\} \quad (5)$$

The trajectory method is very effective way of representing both short and long term behavior of dynamical system in the space of *K* functions.

Figure 1 outlines the trajectory method. As shown in this figure, model (Eq (4)) is run with the initial conditions (j=1,2,…,j_{max}) chosen along the original trajectory ($x_r(t_n)$, n=1,2,…,N).

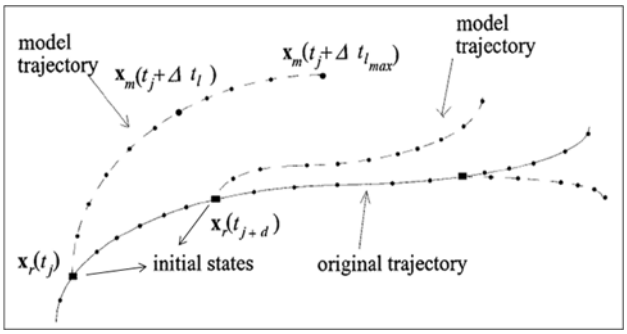

Figure 1. Schematic presentation of trajectory method in a phase space (after Perona et al., 2000).

The model equation is used to predict the state variable at the instants ($t_j + \Delta t_l$). A quality function Q is obtained by repeating this approach for different initial conditions.

$$Q = \sum_{j=1}^{j_{max}} \sum_{l=1}^{l_{max}} \left\| x_m(t_j + \Delta t_l) - x_r(t_j + \Delta t_l) \right\| \quad (6)$$

where the notation $\left\| \cdot \right\|$ shows the Euclidean norm. $x_r(t_j)$ and $x_m(t_j)$ in Eq (6) are the initial conditions on the original trajectory and trajectory produced by the model, respectively. At the beginning $x_r(t_j)$ and $x_m(t_j)$ are the

same data point. On the other, hand l_{max} determines how many steps the model will be run in order to catch both sort and long-term behavior of the system. In other words, l_{max} is the number of points used for comparison between the single reconstructed trajectory and the original trajectory, starting from the initial state set on the latter. Δt_l in Eq (6) is the time interval between the integration steps of the model equation. This quantity can be calculated as

$$\Delta t_l = h(2^{l-1}) \tag{7}$$

where h is the interval between the observations or integration step in case of numerical integration. The optimum value of $c_{i,k}$ are obtained by minimization the quality function Q.

$$Q_{\min} = \min_{C_{i,k}} Q \qquad (i = 1,2,\ldots,D;\ k = 1,2,\ldots K) \tag{8}$$

Eq (6) can be stated as given below

$$Q = \sum_{j=1}^{j_{\max}} \sum_{l=1}^{l_{\max}} \sqrt{\sum_{i=1}^{D} \left[\left(\int_{t_j}^{t_j+\Delta t_l} \dot{x}_{m_i}(\tau) d\tau \right) + x_{m_i}(t_j) - x_{r_i}(t_j+\Delta t_l) \right]^2} \tag{9}$$

The integral given in Eq (9) represents the change of $x_{m_i}(t)$ between the time interval $[t_j, t_j+\Delta t_l]$ and can be stated as

$$\int_{t_j}^{t_j+\Delta t_l} \dot{x}_{m_i}(\tau) d\tau = x_{m_i}(t_j + \Delta t_l) - x_{m_i}(t_j) = c_{i,1} \int_{t_j}^{t_j+\Delta t_l} F_{i,1}(\tau) d\tau + \ldots + c_{i,k} \int_{t_j}^{t_j+\Delta t_l} F_{i,k}(\tau) d\tau \tag{10}$$

The integrals in Eq (10) should be calculated numerically because the functions $F_{i,k}$ are all unknown functions. If the partial derivative of Q with respect to unknown coefficients $c_{i,k}$ is set to zero, then the following set of linear equation is obtained:

$$\frac{\partial Q}{\partial c_{i,k}} = \left(\sum_{z=1}^{K} c_{i,z} A_{k,z}^{(i)} \right) - B_k^{(i)} = 0 \qquad z,k = 1,\ldots,K \tag{11}$$

The matrix $A_{k,z}^{(i)}$ and the vector $B_k^{(i)}$ are as given in Eqs (12) and (13), respectively.

$$A_{k,z}^{(i)} = \sum_{j=1}^{j_{max}} \sum_{l=1}^{l_{max}} \left[\left(\int_{t_j}^{t_j+\Delta t_l} F_{i,k}(\tau) d\tau \right) \left(\int_{t_j}^{t_j+\Delta t_l} F_{i,z}(\tau) d\tau \right) \right] \tag{12}$$

$$B_k^{(i)} = \sum_{j=1}^{j_{max}} \sum_{l=1}^{l_{max}} \left[\left(x_{r_i}(t_j + \Delta t_l) - x_{m_i}(t_j) \right) \left(\int_{t_j}^{t_j+\Delta t_l} F_{i,k}(\tau) d\tau \right) \right] \tag{13}$$

The matrix *A* is reversible. By solving Eq (11) a new set of coefficients $c_{i,k}$ are obtained then these coefficients are used in the next optimization cycle. This process continues until the optimum values of coefficients are obtained (Perona et al., 2000).

3. Application to Evaporation Data

Daily evaporation totals used in this study are observed in the Ercan Meteorology Station located in North Cyprus. Observation period covers 2001-2010; total number of data points is 3652. In this study, before the application of the trajectory method, the original time series smoothed out by using loess method (Cleveland, 1979). Figure 2 shows the original and the smoothed out time series together.

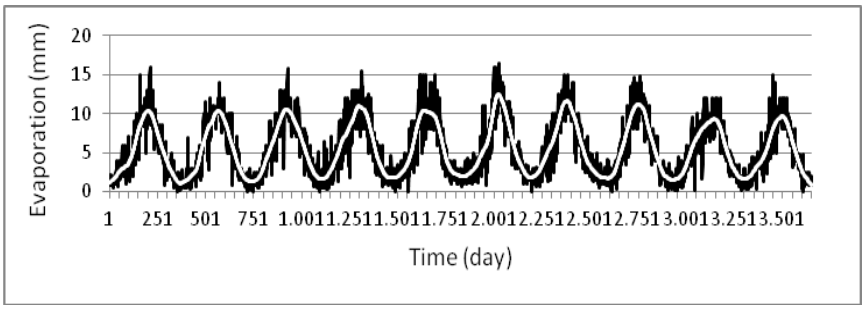

Figure 2. Evaporation time series (black) and smoothed out series (white).

By using smoothed time series phase space is reconstructed. As mentioned before, for phase space reconstruction two parameters namely time delay and embedding dimension are necessary. The time delay is determined by using Mutual Information Function (MIF) approach (Fraser, 1986). The first minimum value is taken as the optimum time delay (see Figure 3). As seen from Figure 3,

the first minimum of the MIF is τ=112. On the other hand embedding dimension is assumed m=3.

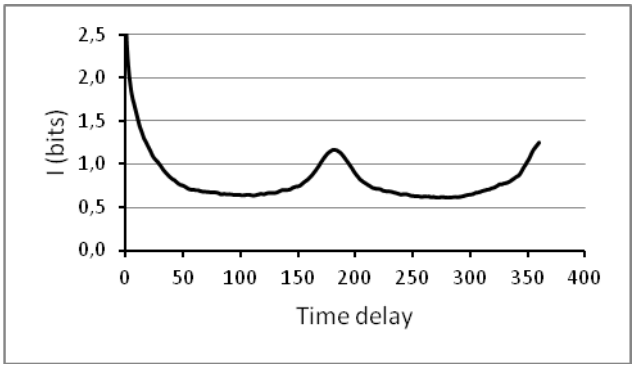

Figure 3. Mutual information of smoothed evaporation time series.

The phase space of evaporation process is reconstructed by taking time delay 112 and embedding dimension 3. Projection of the resulting attractor onto 2-dimension is given in Figure 4. As depicted in this figure smoothed attractor shows almost quasi-periodic behavior. Put another way, the behavior of this attractor in phase space is neither periodic nor aperiodic. This result shows that it will be reasonable to model the periodic structure or limit cycle of this attractor. The trajectory model has been applied to smoothed evaporation time series. The resulting limit cycle is given in Figure 5. As shown from this figure starting from an initial condition, the trajectory eventually converge the stable periodic orbit.

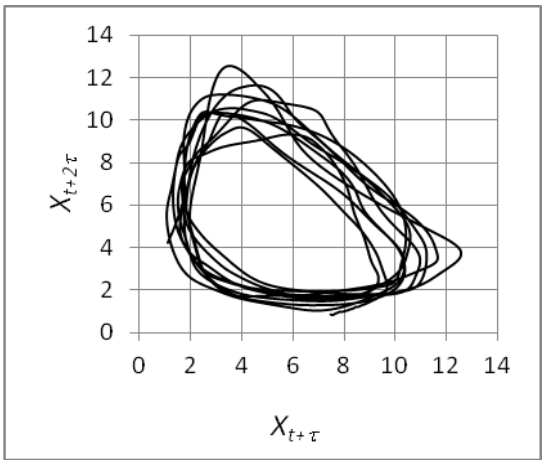

Figure 4. Projection of the attractor onto plane.

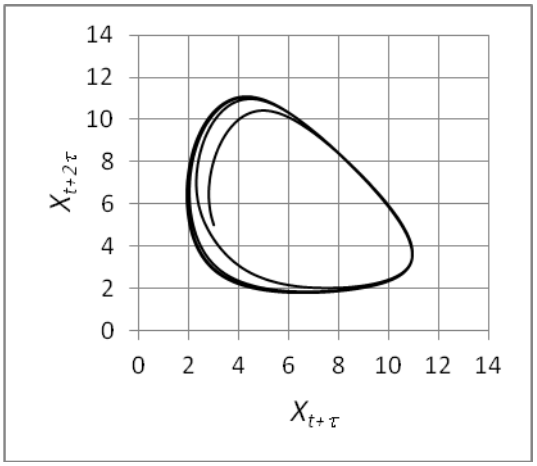

Figure 5. Periodic attractor of evaporation process.

4. Results and Discussion

Water reservoirs are very important in producing hydraulic energy, irrigation, flood control, drinking water, recreational purposes, etc. On the other hand there are some water loses from water reservoirs. The most important water loses take place by evaporation process. Thus, it is important to know the main dynamic of the evaporation.

In this study the trajectory method, the state art of the inverse problem solving method, is applied to evaporation process. Other variables that affect the evaporation such as temperature, wind speed, relative humidity, solar radiation, etc. are not considered in this application. In other words phase space reconstruction from univariate time series is used instead of multivariate approach. After the reconstruction process, the trajectory method is applied to smoothed evaporation data. The limit cycle or periodic behavior of the evaporation has been successfully reconstructed in the form of a set of differential equation which has three state variables.

References

Breden, J.L., and Hübler, A., 1990. Reconstructing equation of motion from experimental data with unobserved variables, *Physical Review A*, 42(10), 5817-5826.

Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368), 829-836.

Cremers, J., and Hübler, A., 1986. Construction of differential equations from experimental data, *Zeitschrift für Naturforshung A*, 42, 797-802.

Crutchfield, J.P., and McNamara, B.S., 1986. Equations of motion from a data series, *Complex Systems*, 1, 417-452.

Eisenhammer, T., Hübler, A., Packard, N., and Kelso, J.A.S., 1991. Modeling experimental time series with ordinary differential equations, *Biological Cybernetics*, 65, 107-112.

Fraser A.M., Swinney H.L, 1986. Independent coordinates for strange attractors from mutual information. *Physical review A*, 33(2), 1134-1140.

Koçak, K., 1996: Fractal Dimension Variation as a Chaotic Behaviour Criterion and its Application to Dynamical Systems. PhD Thesis, ITU Institute of Science (in Turkish).

Packard, N.H., Crutfield, J.P., Farmer, J.D., Shaw, R.S., 1980. Geometry from a time series, Physical Review Lett., 45(9), 712-716.

Perona, P., Porporato, A., and Ridolfi, L**.,** 2000. On The trajectory Method for the Reconstruction of Diffrerential Equations from Time Series, *Nonlinear Dynamics*, 23, 13-33.

Perona, P., D'Odorico, P., Porporato, A., and Ridolfi, L.**,** 2001. Reconstructing the Temporal Dynamics of Snow Cover from Observations, *Geophysical Researh Letters*, 28, 2975-2978.

Takens, F. (1981) "Detecting strange attractors in turbulence", in Lectures Notes in Mathematics, edited by D. A. Rand and L. S. Young, 898, 366-381, Springer-Verlag, New York.

Chaotic Modeling and Simulation (CMSIM) 3: 281-286, 2014

Influence of geomagnetic activity on recurrence quantification indicators of human electroencephalogram

I. E. Kanunikov, B. V. Kiselev

Saint-Petersburg State University, Russia
E-mail: igorkan@mail.ru

Abstract: The investigation deals with the revealing of influence of a geomagnetic field on human electroencephalogram by means of recurrence quantification analysis (RQA). The EEG base of 10 subjects was processed. The database included electroencephalogram records carried out from 16 points under three background conditions. Each subject took part in 15–50 experiments. EEG was registered from frontal, temporal, central, parietal and occipital areas of the left and right hemispheres. For every subject for each of 16 points of EEG registration 9 recurrent measures of EEG were calculated (RR, DET, L, DIV, ENTR, RATIO, LAM, TT, CLEAN). Then the factor of correlation of these measures with a planetary index of geomagnetic activity of Ap and local daily K-index in a day of carrying out experiment was calculated. As a result of this research the following conclusions were received.
1. Significant influence of intensity of a geomagnetic field on recurrent EEG dynamics indicators is shown. Thus the relationship between recurrent EEG measures and indexes of local intensity of a geomagnetic field appeared higher than with planetary indexes.
2. Existence of significantly bigger number of relations between geomagnetic activity and recurrent measures of the left hemisphere EEG is shown.
3. The conclusion suggests that the geomagnetic field makes the main impact on a chaotic component of EEG.
Keywords: Nonlinear methods, Recurrance quntification analysis, Electroencephalogram, Geomagnetic field, Magnitobiology.

1. Introduction

The investigation deals with the revealing of influence of a geomagnetic field on human electroencephalogram by means of recurrence quantification analysis (RQA). In contrast with chaos method, an important advantage of RQA is that it can deal with a noisy and short time series.

2. Methods and experiments

Recurrence Plots are introduced by Eckmann et. al. (1987) as a tool for visualization of recurrence of states Xi in phase space. This approach enables us to investigate the m-dimensional phase space through a two-dimensional representation of its recurrences.

Zbilut and Webber (1992, 1994) developed RQA for definition of numerical indicators. They offered the measures using density of recurrent

Received: 7 August 2013 / Accepted: 15 April 2014

ISSN 2241-0503

points and diagonal structures of the diagram: indicator of similarity (RR), determinism (DET), maximum length of diagonal lines (L), the maximal length of diagonal structures or its inversion — the divergence (DIV), entropy (ENTR), the ratio between DET and RR (RATIO). Slightly after Marwan et.al. (2004, 2007) offered the measures based on horizontal (vertical) structures of recurrent diagrams: laminarity (LAM) and indicator of a delay (TT). V.B.Kiselev (20007) suggests the indicator CLEAN which shows influence of a stochastic component of process, thus prevalence of the stochastic component leads to increase of CLEAN value.

Expressions for RQA measures are shown below.

The simplest measure of the RQA is the recurrence rate (RR) or percent recurrences which is a measure of the density of recurrence points in the recurrent points. Note that it corresponds to the definition of the correlation sum.

The ratio of recurrence points that form diagonal structures (of at least length lmin) to all recurrence points is introduced as a measure for determinism (DET*)* (or predictability) of the system. The threshold lmin excludes the diagonal lines which are formed by the tangential motion of the phase space trajectory.

L is the average time that two segments of the trajectory are close to each other. This measure can be interpreted as the mean prediction time.

Another RQA measure considers the length Lmax of the longest diagonal line found in the recurrent points, or its inverse, the divergence, DIV=1/Lmax. These measures are related to the exponential divergence of the phase space trajectory. The faster the trajectory segments diverge, the shorter are the diagonal lines and the higher is the measure DIV.

ENTR refers to the Shannon entropy of the frequency distribution of the diagonal lines lengths. This measure reflects the complexity of the deterministic structure in the system.

RATIO is the ratio between DET and RR. This measure is useful to discover transitions when RR decrease and DET does not change at the same time.

LAM is analogous to the definition of determinism. This measure is the ratio between the recurrence points forming the horizontal structures and the entire set of recurrence points. The computation of LAM is realized for horizontal line length that exceeds a minimal length V_{min}.

TT shows average length of laminar states in the system.

In periodical systems fluctuations and noise influence leads in separate points and very short diagonals. The measure *cleanness* (CLEAN) is the ratio between recurrence points in diagonals with lengths less than l_{min} and recurrence points in diagonal lines with lengths equal or more than l_{min}. The measure quantifies influence of noise and fluctuations on system trajectory and should be used if studied system shows periodic behavior.

In this work the EEG base of ten clinically normal subjects (six males and four females in the age range 20– 65 years) was processed. The database included records of electroenchephalogram, carried out from 16 sites under

three background conditions: two with open eyes and one with close eyes. During background condition with open eyes subject has to look passively at a picture or thumb through the book. During close eyes subject has to consider drops which were modelled by phonostimulator. In our opinion such simple activity more will balance subjects with each other in comparison with a standard background condition at which it is impossible to check internal state of the subject.

Each subject took part in 20-50 experiments which are carried out to the period of time from half a year till two years. Registration of EEG was carried out in the international system 10/20 in frontal (Fp1, Fp2, F3, F4, F7, F8), temporal (T3, T4, T5, T6), central (C3,C4), parietal (P3, P4) and occipital (O1, O2) sites of the left and right hemispheres. The length of record EEG was about 1 minutes for each of three backgrounds, EEG was quantized with frequency of 250 times a second. The constant of time was 0.3 seconds, and the top frequency of a cut equaled 30 Hz.

3. Results

Before data processing all records were filtrated to escape EEG from different artifacts. For every subject for each of 16 sites and the 3rd background conditions 9 recurrent measures of EEG were calculated (RR, DET, L, DIV, ENTR, RATIO, LAM, TT, CLEAN). Then the coefficient of correlation of these measures with an index of geomagnetic activity was calculated. The coefficient of correlation was calculated on two rows: one row corresponded to defined EEG indicator, and the second – represented values of an index of geomagnetic activity in day of carrying out experience.

As a result of carrying out one experiment about 500 values of recurrent measures (9x16x3) turned out. Two geomagnetic indexes were thus used: planetary Ap and local daily K-index which undertook from a site of the Finnish observatory (Sudancula). At calculation of coefficients of correlation with an index of geomagnetic activity value of correlation were averaged on three background conditions. Tests were significant at $P < 0.05$.

At the first analysis stage significant correlations of 9 recurrent measures of EEG were compared with indexes of planetary and local geomagnetic activity. It appeared that all measures significantly correlated with geomagnetic activity. Total number of significant interrelations for all 10 subjects made in relation to a planetary index was 271, and in relation to a local indicator - 347. Considering that fact that the local index of geomagnetic activity was more sensitive to recurrent EEG measures in comparison with a planetary index, in further calculations it was used only. Thus the maximum quantity of correlations made 44 (for an indicator of DIV), and the minimum number equaled 32 (for a TT indicator). Statistically significant distinctions between quantity of correlations for each of measures it was revealed not. On this basis in the subsequent analysis data on all measures were averaged.

In table 1 are submitted data by number of statistically significant coefficients of correlation between recurrent measures of EEG and local K-

indexes of geomagnetic activity. First, the fact of individual differences in number of correlations which are in range from 14 to 57 attracts attention.

Table 1. Quantity of significant correlations of recurrent measures of EEG with local K-index

Subjects	RR	DET	L	DIV	ENTR	RATIO	LAM	TT	CLEAN	Summa
1	5	4	5	7	5	7	4	5	4	46
2	2	1	4	4	3	1	1	3	0	19
3	11	4	2	8	7	7	9	1	8	57
4	2	2	1	2	1	1	3	0	2	14
5	1	4	3	3	3	4	2	4	4	28
6	4	5	7	9	8	3	3	5	3	47
7	1	5	2	5	3	4	8	3	5	36
8	4	8	5	1	5	2	6	5	7	43
9	1	5	5	4	6	7	4	5	5	42
10	3	1	1	1	1	3	1	1	2	14
Summa	34	39	35	44	42	39	41	32	40	346

The second interesting result consisted that all recurrent measures were characterized by a large amount of correlations for EEG of the left hemisphere in comparison with right. However statistically significant differences took place only for DET measure ($P < 0.02$). As a whole, when averaging all 9 recurrent EEG measures differences between the left and right hemisphere were statistically high-significant ($P < 0.001$).

At the following analysis stage interhemisphere differences of coefficients of correlation for each pair of sites (tab. 2) were considered. Except for pair of sites of C3 and C4 where in the right hemisphere the quantity of correlations was higher, than in left, and in T5, T6 sites where it was equal, in all other pairs of EEG sites the number of correlations at the left was higher than in right. However statistically significant difference was observed only between temporal sites T3 and T4.

Table 2. Quantity of significant correlations of 9 recurrent measures of EEG in different sites with local K-index of geomagnetic activity (data were avaraged on 10 subjects)

C3	C4	F3	F4	F7	F8	Fp1	Fp2
24	34	16	14	22	12	37	19

O1	O2	P3	P4	T3	T4	T5	T6
32	21	10	6	35	12	26	26

Research of changes of classical rhythms EEG (α, β, θ) in reply to changes of a geomagnetic field hasn't revealed significant interrelations with K

index. On the other hand primary not filtered signal EEG has revealed such relationship.

4. Discussion

The fact of existence of a large number of correlations between various recurrent EEG measures and index of geomagnetic activity appeared the most important. It testifies that the nonlinear component of EEG for which analysis the RQA method was used, is very sensitive to changes of a geomagnetic field. Carruba et.al. (2007) show that magnetosensory evoked potentials weren't detected when the EEGs were analyzed by time averaging, indicating that the evoked potentials were nonlinear in origin. Obviously, the geomagnetic field influences electric activity of a brain in a nonlinear way. This fact can cause failures in search of reflections in EEG of influences from a geomagnetic field.

That fact that a local index was more closely connected with recurrent EEG measures in comparison with a planetary index is explained by that a local index more precisely, in comparison with planetary, reflects a condition of a geomagnetic field in St. Petersburg being on close longitude.

The fact of very high individual differences found in work concerning quantity of correlations of various recurrent EEG measures with geomagnetic activity, was explained obviously, existence of individual differences concerning sensitivity of subjects to influence on the central nervous system of changes of a geomagnetic field. It should be noted that subjects differed concerning that what by sites EEG significantly correlated with indicators of geomagnetic activity. At the 4th of 10 subjects correlated mainly frontal and temporal sites, at 4 subjects significant correlations were observed practically for all sites, at 2 subjects correlated either frontal, or temporal sites. Similar individual differences were observed in the work of Carruba et.al. (2007). They show that magnetosensory evoked potentials so strongly differ at various subjects that when the results obtained within subjects were averaged across subjects, evoked potentials couldn't be detected.

The most interesting fact concerns high-significant differences concerning number of correlations with recurrent EEG measures of the left and right hemispheres. This result based on a tendency to excess of number of correlations with every recurrent measures of the left hemisphere in comparison with right, and on the high-significant difference received at averaging of all recurrent measures of EEG. The question of why the bigger number of EEG sites of the left hemisphere correlates with changes of a geomagnetic field, remains open. We know that the right hemisphere is closely connected with adaptation processes. So, for example, V.P. Leutin and E.I.Nikolayeva (1988) on the basis of numerous experimental studies drew a conclusion that right brain hemisphere activation is decisive factor, providing adaptation to extreme climate conditions. In our experiments devoted to studying of influence of a geomagnetic field on an indicator of spatial synchronization of EEG, it was

shown that in reply to changes of a geomagnetic field activation of the right hemisphere authentically increases. We connected this result with the stress reaction caused by changes of a geomagnetic field.

In the real experiments more sensitive in relation to variations of a geomagnetic field there was a nonlinear component of EEG of the left hemisphere. The understanding of this result will require further researches.

5. Conclusions

As a result of this research the following conclusions were received.

1. Significant influence of intensity of a geomagnetic field on recurrent EEG dynamics indicators is shown. Thus the relationship between recurrent EEG measures and indexes of local intensity of a geomagnetic field appeared higher than with planetary indexes.
2. Existence of significantly bigger number of relations between geomagnetic activity and recurrence measures of the left hemisphere EEG is shown.
3. The conclusion suggests that the geomagnetic field makes the main impact on a chaotic component of EEG.

References

1. Carrubba S., Frilot C., Chesson A.L., MarinoA.A. Evidence of a nonlinear human magnetic sense //Neuroscience 144, 2007, 356–367.
2. Eckmann J.P., Kamphorst S. O., Ruelle D. Recurrence Plots of Dynamical Systems // Europhysics Letters 5, 1987, pp. 973–977.
3. Kanunikov I.E. et.al. Vlijanie geomagnitnoi aktivnosti na electroencephalogrammu cheloveka (Влияние геомагнитной активности на электроэнцефалограмму человека) // Ekologia cheloveka. 2010. №6. P.6-11.
4. Kiselev V.B. Opredelenie stabilnosti traectorii prozessa v phazovom prostransntve pri pomochi recurrentnogo analisa (Определение стабильности траектории процесса в фазовом пространстве при помощи рекуррентного анализа)// Nauchno-tecnicheski vestnik SPbUITMO. SPb.: SPbUITMO, 2007, vip.20. P.121-130.
5. Leutin V.P., Nikolaeva E.I. Psychofisiologicheskie mechanismi adaptazii i funkzionalnaja asimmetria mozga (Психофизиологические механизмы адаптации и функциональная асимметрия мозга) Novosibirsk, Nauka SO. - 1988. - 192 p.
6. Marwan N., Romano M.C., Thiel M., Kurths J. Recurrence plots for the analysis of complex systems// Physics Reports 438, 2007, 237–329.
7. Marwan N., Meinke A. Extended Recurrence Plot Analysis and its Application to ERP Data // International Journal of Bifurcation and Chaos “Cognition and Complex Brain Dynamics” 14 (2), 2004.
8. Webber Jr. C. L., Zbilut J. P. Dynamical assessment of physiological systems and states using recurrence plot strategies // Journal of Applied Physiology 76 (2), 1994, pp. 965–973.
9. Zbilut J.P., Webber Jr.C.L. Embeddings and delays as derived from quantification of recurrence plots // Physics Letters A 171 (3–4). 1992. pp. 199–203.

Chaotic Modeling and Simulation (CMSIM) 3: 287–294, 2014

Mode locking, chaos and bifurcations in Hodgkin-Huxley neuron forced by sinusoidal current

Himanshu Gangal[1] and Gaurav Dar[2]

[1] BITS Pilani K K Birla Goa Campus, Goa, India
(E-mail: himanshugangal@gmail.com)
[2] BITS Pilani K K Birla Goa Campus, Goa, India
(E-mail: gdar@goa.bits-pilani.ac.in)

Abstract. The action potentials in a sinusoidaly forced Hodgkin-Huxley neuron are known to possess mode locked or chaotic oscillations depending on the values of forcing parameters. We have numericaly studied the spiking dynamics of the sinusoidaly forced Hodgkin-Huxley neuron by making fine variations in the amplitude while keeping the frequency fixed. We find that the dynamics of the neuron is far richer than previously known. Increasing the resolution of forcing amplitude (I_0) uncovers $1/m$ mode locked oscillations with increasingly larger values of m. Moreover, a mode locked oscillation of type $1/m$ can exist over multiple disconnected intervals of forcing amplitude. Chaotic oscillations are found interspersed with mode locked oscillations. By varying I_0 we have further explored the transition between qualitatively different types of oscillations. On increasing I_0, every $1/m$ mode locked oscillation is found to go through a sequence of period doubling bifurcations giving rise to $1/2m, 1/4m, ...$ mode locked oscillations and finally chaos. Chaotic oscillations further undergo a transition to a $1/m'$ mode locked oscillation through a tangent bifurcation. The observed spiking patterns in mode-locked oscillations are unusual and encode the stimulus strength.

Keywords: Hodgkin-Huxley model, Neurons, Bifurcation.

1 Introduction

Hodgkin Huxley model serves as a paradigm for axonal membranes of spiking neurons. The model arose from the electrophysiological experiments of Hodgkin-Huxley with squid giant axons. Consequently, a lot of experimental work with squid axons and theoretical work with the Hodgkin-Huxley (HH) model has been carried out.

A nerve membrane is an excitable system. An appropriate stimulus evokes a strong response (action potential) resulting in a train of spikes in the membrane potential. For forcing by a steady current, a subcritical Hopf bifurcation causes the rest state of the neuron to become unstable giving rise to a periodic train of spikes (a limit cycle) (Xie *et. al.*[1]). Periodically varying stimuli evoke a

Received: 5 April 2014 / Accepted: 27 July 2014

ISSN 2241-0503

rich variety of response. Mode-locked (periodic), chaotic, and quasiperiodic oscillations of membrane voltage have been found in experiments with squid giant axons (Kaplan *et. al.* [2], Matsumoto *et. al.*[3], Aihara and Matsumoto[4], Guttman *et. al.*[5]) and in numerical simulations of the HH model (Lee[6], Borkowski[8], Borkowski[7], Parmananda *et. al.*[9]).

A periodicaly stimulated neuron does not fire action potentials unless the forcing amplitude is above a threshold value. The threshold amplitude depends on the forcing frequency. The firing threshold curve (in forcing parameter space) of a HH neuron under sinusoidal forcing has been explored extensively. Firing onset occurs through a variety of bifurcation mechanisms(Lee[6]). The firing region in parameter space is dominated by mode locked oscillations of the type 1/1, 1/2, and 1/3 while there is a smaller region that exhibits chaotic oscillations. Bifurcations mechanisms that bring about a change in the mode-locking ratio of the periodic oscillations have not been explored so far. Our work explores this question.

In their simulations Lee[6] carried out a characterization of the HH neuron's firing response in the forcing amplitude-frequency parameter space. However, there exists a strip in parameter space lying between the 1/1 and 1/2 mode locked regions that has not been explored adequately. In order to uncover the bifurcations between various mode-locked oscillations, it is imperative to carry out an exhaustive investigation of this strip. We have found that a complex structure of interwoven periodic and chaotic dynamics connected by period doubling and tangent bifurcations exist in this strip.

2 Hodgkin-Huxley Model

The Hodgkin-Huxley model of an axon describes the dynamics of its membrane voltage (V), activation variable (m) and the inactivation variable (h) of its sodium channels, and the activation variable (n) of its potassium channels. The model consists of the following set of four coupled differential equations

$$C\frac{dV}{dt} = -\bar{G}_{Na}m^3h(V - V_{Na}) - \bar{G}_K n^4(V - V_K) - \bar{G}_L(V - V_L) + I_{ext}, \quad (1)$$

$$\frac{dm}{dt} = \alpha_m(1 - m) - \beta_m m, \quad (2)$$

$$\frac{dh}{dt} = \alpha_h(1 - h) - \beta_h h, \quad (3)$$

$$\frac{dn}{dt} = \alpha_n(1 - n) - \beta_n n, \quad (4)$$

where,

$$\alpha_m = \frac{0.1(25 - V)}{exp\left[(25 - V)/10\right] - 1}, \qquad \beta_m = 4exp\left[-V/18\right], \quad (5)$$

$$\alpha_h = 0.07exp\left[-V/20\right], \qquad \beta_h = \frac{1}{exp\left[(30 - V)/10\right] + 1}, \quad (6)$$

$$\alpha_n = \frac{0.01(10-V)}{exp\left[(10-V)/10)\right]-1}, \qquad \beta_n = 0.125exp\left[-V/80\right]. \tag{7}$$

Capacitence of the axonal membrane $C = 1\mu F/cm^2$. The reversal potentials of sodium, potassium, and leakage channels are $V_{Na} = 115mV$, $V_K = -12mV$, and $V_L = 10.5995mV$ respectively. The maximal conductances of the membrane for sodium, potassium, and leakage currents are $\bar{G}_{Na} = 120mS/cm^2$, $\bar{G}_K = 36mS/cm^2$, $\bar{G}_L = 0.3mS/cm^2$ respectively. In our work we stimulate the neuron with a sinusoidal current $I_{ext} = I_0 sin(2\pi\nu_f t)$, where I_0 is the forcing amplitude and ν_f is its frequency.

In our work we choose the frequency $\nu_f = 50Hz$ and the amplitude I_0 is varied in the range $1.6\mu A/cm^2 - 2.0\mu A/cm^2$. At the lower and upper end of this range, the neuron exhibits 1/1 and 1/2 mode locked spiking oscillations (Lee [6]). By carrying out fine variations in the amplitude over this range, we have uncovered a complex dynamical structure between these two periodic spiking oscillations.

We carry out numerical simulations of the Hodgkin-Huxley equations (Eq. 1- 4) using the fourth order Rungke-Kutta method. We choose the time step dt in our simulations as $dt = T_f/1000$.

3 Results

Dynamics of forced nonlinear systems are often studied by sampling their phase space trajectory stroboscopically. Following this approach, we sample the phase space trajectory of the HH model once every time period of the sinusoidaly varying external current. Doing so, yields a sequence of voltage values $V_0, V_1, V_2, ..., V_i, ...$ For periodic oscillations, a repetitive sequence will be present. Let T be the the time taken for the neuron's phase space trajectory to complete one full cycle. For periodic oscillations $T_f/T = 1/m$. We will characterize periodic oscillations by this ratio and refer to these as $1/m$ mode locked oscillations. The repetitive sequence of voltage values for a $1/m$ oscillation will contain m distinct values.

We have plotted the stroboscopically generated voltage sequences against the forcing amplitude as a bifurcation parameter. The resulting bifurcation plot is shown in Figure 1 over the amplitude range $I_0 = (1.6 - 2.0)\mu A/cm^2$. Two ends of the plot display the 1/2 and 1/1 mode locked oscillations, known from Lee's work (Lee [6]). This interval is believed to contain a rich dynamical structure (Lee [6], Parmananda *et. al.* [9]) but very few details are known.

Figure 1 shows that the amplitude interval between the known 1/1 and 1/2 oscillations contain many more periodic oscillations. Infact, the interval is dominated by periodic oscillations. On increasing the forcing amplitude from $I_0 = 1.6\mu A/cm^2$ onwards we observe $1/2, 1/3, 1/4, 1/5, ..., 1/m, 1/(m+1), ...$ mode locked oscillations. A $1/m$ oscillation contains m branches in the bifurcation diagram. A new branch gets added to the bifurcation diagram on crossing over from a $1/m$ to a $1/(m+1)$ oscillation. Figure 1 suggests the presence of a $1/m$ mode locked oscillation for every positive integer m.

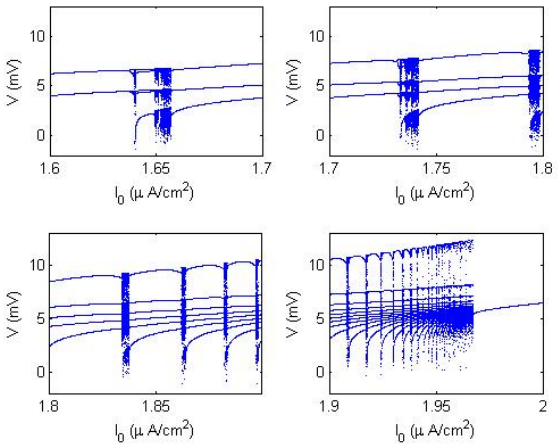

Fig. 1. Bifurcation diagram with variation in forcing amplitude I_0 with forcing frequency fixed to $\nu = 50Hz$.

The amplitude interval lying between $1/m$ and $1/(m+1)$ oscillations described above contains a rich dynamical structure not discernible in Fig. 1. We see an instance of this richness on magnifying the amplitude interval lying between 1/3 and 1/4 mode locked oscillations (see Fig. 2(a)). This interval contains a myriad of periodic and chaotic oscillations. The region between every $1/m$ and $1/(m+1)$ oscillations of Fig. 1 contain such periodic and chaotic oscillations.

In Fig. 2(a)-(b) we observe that the 1/3 mode locked oscillations (on the extreme left of the figure) undergo a cascade of period doubling bifurcations giving rise to a sequence of 1/6, 1/12,... oscillations finally converging to chaos. Similar period doubling bifurcations are present in other periodic windows in Fig. 2(a). In general, starting from a $1/n$ periodic window, period doublings will result in $1/(2n)$, $1/(4n)$,... oscillations. Each successive periodic oscillation obtained through period doubling takes double the time to go around its phase space trajectory once. Each cascade of period doublings finally converges to chaos.

Periodic windows in Fig. 2(a) emerge from chaotic oscillations through a tangent bifurcation. The bifurcation is identified by plotting a return map between V_i and V_{i+n} if a $1/n$ mode locked oscillation results from the bifurcation. Close to tangent bifurcation, the return map has n curve segments tangent to a 45^0 line. After the tangent bifurcation occurs the return map crosses the 45^0 line at $2n$ points. Half of these points lie on a stable trajectory and the other half like on an unstable trajectory. All periodic windows arise in the same manner. Once a periodic oscillation is created through a tangent bifurcation, the subsequent changes in the qualitiative dynamics of the membrane voltage arise from period doubling bifurcation.

Typical spike sequences generated due to sinusodial forcing are depicted in Figs. 3 and 4. Figure 3(a) shows a 1/3 mode locked oscillation obtained

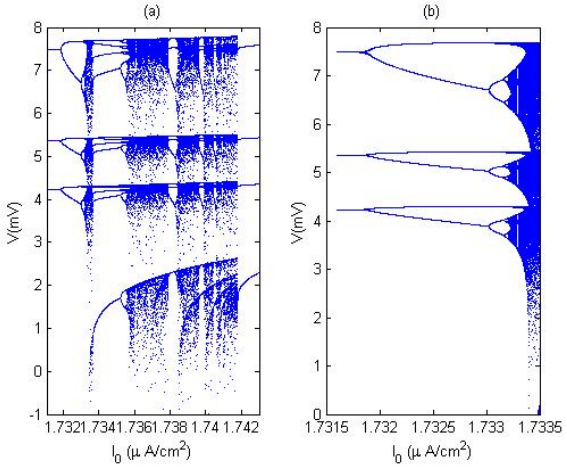

Fig. 2. (a) Bifurcation diagram for the amplitude interval lying between the 1/3 and 1/4 mode locked oscillation of Fig. 1. (b) Panel (a) figure is further magnified over its initial amplitude interval to depict period doubling.

with forcing parameters chosen from the large region of 1/3 oscillations in Lee's paper. Here a spike occurs once every three cycles of forcing. In our work we have found a novel set of spike patterns.

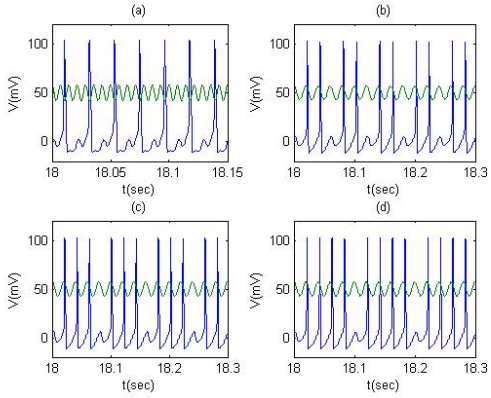

Fig. 3. Some typical spike sequences for periodic oscillations. Each figure gives the result for a different I_0 (in units of $\mu A/cm^2$) and $\nu = 50Hz$. The sinusoidal curve in each figure indicates the profile of this current (a)$I_0 = 4$, (b) $I_0 = 1.7$, (c) $I_0 = 1.78$, (d) $I_0 = 1.82$. Repeating units of spike sequences in (a)-(d) are of form {1..} , {2.}, {3.}, and {4.} respectively.

The spike sequences in Fig. 3(b)-(d) are representative of the $1/m$ periodic oscillations that dominate the amplitude interval in Fig. 1. In each of these periodic oscillations, we find that a spike occurs consecutively over $(m-1)$

forcing cycles, following which there is no spike in the m^{th} forcing cycle. We will represent this spike pattern by $\{(m-1).\}$, with $(m-1)$ representing the group of consecutive $(m-1)$ spikes and the dot $'.'$ representing the missing spike in the m^{th} forcing cycle. The $\{(m-1).\}$ pattern repeates itself every m forcing cycles and thus we will regard it as a repeating unit. The 1/3 and 1/4 mode locked oscillations in Figs. 3(b) and (c) have $\{2.\}$ and $\{3.\}$ as their repeating units respectively. In contrast, the 1/3 oscillation [Fig. 3(a)] from Lee's work has a repeating unit of the form $\{1..\}$.

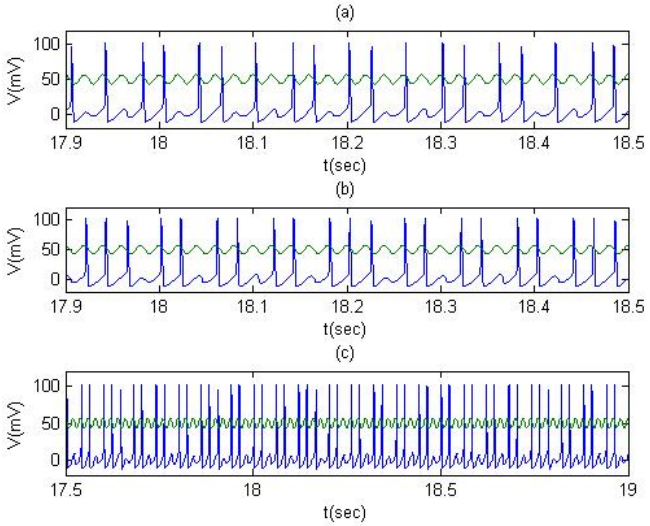

Fig. 4. Each figure gives the result for a different I_0(in units of $\mu A/cm^2$) and $\nu = 50Hz$. (a) and (b) show some typical spike sequences of the fundamental oscillation of a periodic window. Here $I_0 = 1.6518$ in (a) and $I_0 = 1.7345$ in (b). In (c) we see an intermittent spike sequence for $I_0 = 1.73365$. A few sequences $\{3.2.2.2.\}$ appear intermittently here.

Figure 4 shows spike patterns of oscillations in periodic windows. Here the repeating units have a form different from the ones in Fig. 3. A typical repeating unit is of the form $\{m_1.m_2.m_3.\}$, with multiple groups of spikes, whereas oscillations in Fig. 3 contain only one group of spikes. Here we have shown three groups of spikes containing m_1, m_2, and m_3 spikes each separated by a missing spike. However, the number of groups can be more or less (but not less than two) than represented by $\{m_1.m_2.m_3.\}$. Figure 4(a) and (b) shows a 1/8 and 1/13 oscillations with repeating units $\{2.2.1.\}$, and $\{3.2.2.2.\}$ respectively.

Figure 5 shows the typical changes in $V(t)$ that accompany period doubling bifurcations. As an illustrative example we choose the period doubling cascade starting from the 1/3 mode locked oscillation in Fig. 2. The repeating unit is $\{2.\}$ here. We find that the number of spikes per group remain unchanged (equal to two) across all period doubling bifurcations starting from the 1/3

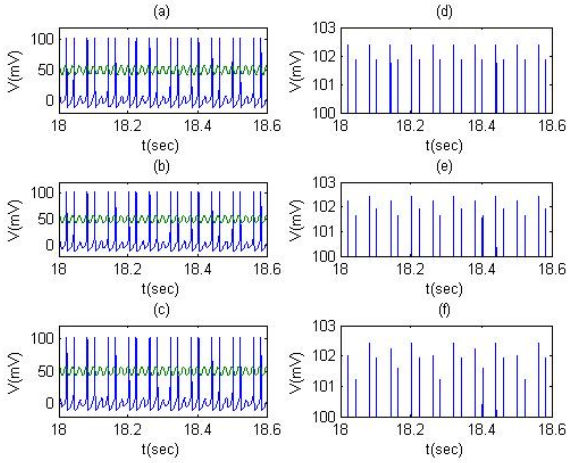

Fig. 5. Changes in the repeating unit of spike sequences across a period doubling bifurcation are shown here. All the figures are plotted for $\nu = 50Hz$ and I_0 is in units of $\mu A/cm^2$(a) $I_0 = 1.731$ gives 1/3 mode locking (b) $I_0 = 1.7324$ gives 1/6 mode locking, and (c) $I_0 = 1.73315$ gives 1/12 mode locking, (d)-(e) show the variation in spike amplitude for (a)-(c) respectively by plotting $V(t)$ on a smaller scale.

oscillation. However, the amplitudes of spikes undergoes a change. Hence, the repeating unit for 1/6, and 1/12 are $\{2.2.\}$ and $\{2.2.2.2.\}$ respectively. Likewise, in a period doubling of any other periodic oscillation with a repeating unit $\{m_1.m_2.m_3.\}$, every period doubling doubles the length of the repeating unit to $\{m_1.m_2.m_3.m_1.m_2.m_3.\}$.

A tangent bifurcation is known to be preceded by intermittency. We find that intermittency occurs through an interesting set of changes in spike patterns as we approach the bifurcation point on varying the forcing amplitude A. Far from the bifurcation point $V(t)$ is chaotic. The spike sequence is of the same form as that for peridic oscillations. However, there exists no repetitive sequence for chaotic oscillations. As the amplitude is brought closer to the bifurcation point the frequency of a specific spike sequence $\{m_1.m_2.m_3.\}$ within the chaotic sequence increases. Once the tangent bifurcation occurs $\{m_1.m_2.m_3.\}$ becomes the repeating unit. Every periodic window arises through a similar increase in the frequency of some unit.

4 Discussion

In the paper we presented a few results of stimulating a HH neuron by a sinusoidal current in the regime where it evokes action potentials. We found a complex structure of $1/m$ mode locked and chaotic oscillations between the 1/1 and 1/2 oscillations of Lee's work(Lee[6]). Chaotic oscillations arise through the period doubling route to chaos. Periodic windows emerge through a tangent bifurcation preceded by an intermittent spike sequence. Starting from 1/2 mode locked oscillations, period doubling cascade gives rise to 1/4, 1/8,.. mode

locked oscillations. Across a period doubling bifurcation, spike sequences do not undergo any change. However, the amplitudes of the spikes undergo an alteration. Intermittent spike sequences before a tangent bifurcation contain glimpses of the spike sequences that are finally realized in the periodic oscillation across the bifurcation. Infact, the neuron enters the 1/1 oscillation through a tangent bifurcation.

In our work the ratio $1/m$ for a periodic oscillation is the ratio of time taken for one forcing cycle to the time taken for the neuron to go once around its closed orbit in phase space. In going around the limit cycle once the neuron may fire several spikes. However, in literature $1/m$ usually implies that the neuron fires one spike in every m cycles of forcing.

The spike sequences presented in our paper are distinct from those obtained earlier. Spikes are organized in groups where each group may contain a different number of spikes. These sequences of spikes alone can carry information about the forcing amplitude. Implying that no knowledge of the rate of spiking or that of the interpsike interval is necessary to extract information about the forcing parameters.

Acknowledgments

This work was supported by computing facilities acquired from funds provided by the BITS Seed grant and DST-FIST grant

References

1.Yong Xie, Luonan Chen, Yan Mei Kang, and Kazuyuki Aihara. Controlling the onset of Hopf bifurcation in the Hodgkin-Huxley model. *Physica Review E*, 77, 061921, 2008.

2.Daniel T. Kaplan, John R. Clay, Timothy Manning, Leon Glass, Michael R. Guevara, and Alvin Shrier. Subthreshold Dynamics in Periodically Stimulated Squid Giant Axons. *Physical Review Letters*, 76, 4074, 1996.

3.Gen Matsumoto, Kazuyuki Aihara, Michinori Ichikawa, and Akira Tasaki . Periodic and nonperiodic responses of membrane potentials in squid giant axons during sinusoidal current stimulation. *J. Theoret. Neurobio.*, 3, 1-14, 1984.

4.Kazuyuki Aihara, Gen Matsumoto. Chaotic oscillations and bifurcations in squid giant axon. In A.V. Holden, editor, *Chaos*, pages 257-269, Princeton, 1986. Princeton University Press.

5.Guttman R, Feldman L, Jakobsson E. Frequency entrainment of squid axon membrane. *J. Membr. Bio.* 21;56(1):9-18, 1980.

6.Sang-Gui Lee and Seunghwan Kim . Bifurcation analysis of mode-locking structure in a Hodgkin-Huxley neuron under sinusoidal current. *Physical Review E*, 73, 041924, 2006.

7.L. S. Borkowski . Bistability and resonance in the periodically stimulated Hodgkin-Huxley model with noise. *Physical Review E*, 83, 051901, 2011.

8.L. S. Borkowski . Response of a Hodgkin-Huxley neuron to a high-frequency input . *Physical Review E*, 80, 051914, 2009.

9.P. Parmananda, Claudia H. Mena, and Gerold Baier. Resonant forcing of a silent Hodgkin-Huxley neuron. *Physical Review E*, 66, 047202, 2002

Chaotic Modeling and Simulation (CMSIM) 3: 295-313 , 2014

Phonological Word Proximity in Child Speech Development

Elena Babatsouli[1], David Ingram[2], Dimitrios A. Sotiropoulos[3]

[1] Division of Linguistics, University of Crete, Rethymnon 74100, Greece
(Correspondence author: ebabatsouli@yahoo.com)
[2] Department of Speech and Hearing Science, Arizona State University, Tempe, AZ 85287, USA
[3] Department of Sciences, Technical University of Crete, Chania 73100, Greece

Abstract: In order to establish a universal perspective on phonological word proximity in child speech development, the relationship between phonological word proximity (PWP) and the proportion of consonants correct (PCC) is derived analytically for a whole speech sample, in terms of the proportion of vowels (PV) and the proportion of phonemes deleted minus phonemes added (PPD). PWP depends linearly on the weighted averages of PCC and PPD and non-linearly on the weighted average of PV; the PV dependence is linearized quite accurately for a wide range of PV values. Upper and lower bounds on PWP are obtained for minimum and maximum PPD. Further, PWP changes are obtained relative to PCC and PPD changes, thus, determining which of these measurements better discriminates performance between speech samples. The method and analysis is applied to PWP, PCC and PPD computations from the data of a bilingual child's speech traced longitudinally from age 2;6 to age 3;9. The results reveal a growth pattern for PCC and PPD and, consequently, for PWP, which is associated with three stages of phonological development. The middle stage is nearly cyclic with a strange attractor and seven months long while the other two stages are progressive of the double-logistic type. The developmental PWP values lie chaotically inside a trapezoid within a triangle bounding general child phonological development in a (PCC, PWP) plane.
Keywords: Child Speech, Development, Phonological Word, Measurement.

1 Introduction

Quantifying progress in child speech has been of interest in the literature since the 1920s. Nice [1] introduced the average length of sentence (ALS) as such an attempt which, in her words, 'may well be the most important single criterion for judging a child's progress in the attainment of adult language'. McCarthy [2] set specific rules on how to count words in the sentence and re-named Nice's measure as mean length of response (MLR). Brown [3] introduced a similar measure, the mean length of utterance (MLU), counting however morphemes in the utterance which, in his words, is 'an excellent simple index of grammatical development'. In language sample analysis (LSA) which is widely used by speech-language pathologists (see, for example, Kemp & Klee [4]), the mean length of response (MLR) has found yet another name, mean length of utterance

Received: 12 July 2013 / Accepted: 19 April 2014

ISSN 2241-0503

in words (MLUw), to distinguish it from Brown's mean length of utterance in morphemes (MLUm). These two measures were compared by Parker & Brorson [5] for 40 language transcripts of 28 typically developing English speaking children between the ages of 3;0 and 3;10. The two measures were found to be perfectly correlated suggesting that, the simpler to calculate, MLUw may be used instead of MLUm. However, correctness of segments or words is totally ignored in all these measures as they are grammatical and not phonological in nature.

Consonant correctness and its measurement has been discussed just about as long (e.g. Wellman et al. [6], Poole [7]) but Shriberg's work with colleagues has refined the measure by addressing issues such as distortions (Shriberg & Kwiatkowski [8]) and speech profiles (Shriberg et al. [9]). Their proposed proportion of consonants correct (PCC) measures the number of consonants produced correctly in context in proportion to the targeted consonants in the speech sample. As for whole-word correctness, Schmitt et al. [10] suggested that the measure of whole-word accuracy (WWA) would favorably complement other measures such as the proportion of consonants correct (PCC). They based this result on data they collected from children between the ages 3;0 and 3;6. Whole words, however, do not only vary in their correctness but also in their complexity and intelligibility. Ingram & Ingram [11] and Ingram [12] proposed to measure phonological word complexity in child speech by the phonological mean length of utterance (PMLU). It is a similar measure to Brown's grammatical mean length of utterance in that it also measures length of utterance even though utterance in PMLU refers to word length while utterance in MLU refers to sentence length. PMLU measures individual segments (consonant or vowel sounds) in the utterance while MLU measures morphemes. But PMLU differs substantially from MLU in that it does not count all the measurable quantities equally, doubling the count of consonant segments produced correctly in the context of intended target, to emphasize the fact that children's errors more often occur in consonants (e.g. Ingram [13], Stemberger [14]) and do not vary nearly as much as vowel errors between transcribers (e.g. Powell [15]).

Further, Ingram & Ingram [11] and Ingram [12] introduced the phonological whole-word proximity (PWP), an indirect indication of word intelligibility, as a measurement of the phonological proximity between produced and targeted words in child speech. PWP was defined as the ratio of the produced phonological mean length of utterance, PMLU, to the targeted one in which all the consonants are by definition correct in context. For utterances consisting of more than one word, PMLU and PWP were defined as the arithmetic mean of their corresponding single word values.

Following the proposed whole word phonological measures, PMLU and PWP, several studies have used them to assess sample utterances of monolingual or bilingual, normal or phonologically impaired children across languages. Taelman et al. [16] discussed how to use CLAN (MacWhinney [17]) to compute PMLU from children's data. Other works on the subject published between 2005 and 2009 are described in Bunta et al. [18] and will not be repeated here.

Bunta et al. [18] compared 3-year old Spanish-English bilingual children to their monolingual peers to compute, among other quantities, PWP and the proportion of consonants correct, PCC. They found that while PWP and PCC differ in general, bilinguals only differ on PCC from their monolingual peers in Spanish. They further found that when comparing the Spanish and English of the bilingual participants, PCC was significantly different but PWP was almost the same. Burrows & Goldstein [19] compared PWP and PCC accuracy in Spanish-English bilinguals with speech sound disorders to age-matched monolingual peers. Macleod et al. [20] compared the change in PWP to that in PCC for two samples of twenty children each, both taken at the age of 18 months and at 36 months. One of the samples involved monolingual English children while the other involved bilingual French-English children. For each sample, their results showed that the change in PWP was larger than that in PCC. Saaristo-Helin [21] measured PMLU and PWP for both typically developing children and children with a specific language impairment acquiring Finnish and concluded that the phonologies of the impaired children largely resembled the ones of younger, typically developing children. Goldstein & Bunta [22] compared PWP and PCC for Spanish-English bilingual children, who have parent-reported language use and proficiency measures commensurate with those of their monolingual peers, to the PWP and PCC for their monolingual peers. Bilingual children did not differ from their monolingual peers in Spanish while they outperformed their monolingual English-speaking peers. Last, Freedman & Barlow [23] examined the effect of phonotactic probability and neighborhood density on PWP and compared it between five Spanish-English children and five age-matched monolingual peers. Phonotactic probability refers to the frequency with which sounds occur and co-occur in the language, while neighbourhood density is defined as the number of real words that can be created by adding, substituting, or deleting a phoneme in any word position. No differences were found between bilinguals and monolinguals in the respective languages or between languages, even though bilinguals evidenced greater phonological complexity in Spanish than English on words with low phonotactic probability and low neighborhood density.

Besides phonological whole-word proximity and its related phonological mean length of utterance, Ingram & Dubasik [24] proposed six other measures in a multidimensional assessment of phonological similarity (MAPS) for a complete comparison between children's utterance samples or within a child's sample. The number of preferred syllable shapes, the proportion of monosyllables, the phonetic inventory articulation score for onsets and codas and the relational articulation score onsets and codas in word initial position and word final position, respectively.

While most studies cited compare speech sample values for the two phonological measures, PWP and PCC, the comparison between them has not been examined in general; Flipsen et al. [25] has compared the two measures by looking into word intelligibility. As a consequence, results though interesting, are not placed in perspective in child speech development, which, if done, would make them more meaningful and useful for practical applications. The

present study achieves this by obtaining analytically the relationship between phonological word proximity (PWP) and proportion of consonants correct (PCC) in terms of the proportion of vowels (PV) and the proportion of phonemes deleted minus added (PPD). Cumulatively, for all the words in a speech sample, PWP is computed as the weighted average of single-word PWPs and not as their arithmetic mean as done in previous studies. This way, the relationship between PWP and PCC for the whole speech sample is expressed analytically and has the same form as for the single word, enabling us to obtain upper and lower PWP bounds in terms of PCC and PV for minimum and maximum PPD, respectively.
The analytical results for general child speech development are applied to a speech sample of a bilingual normal child traced longitudinally between the ages 2;6 and 3;9. The growth patterns of the child's computed PCC, PPD and PWP values are obtained and are placed in perspective in general child speech development.

2 Method

2.1 The PWP-PCC relationship

In order to place speech sample values of phonological word proximity (PWP) in a universal perspective in child speech development, we obtain the analytical relationship between PWP and PCC. Before considering the whole speech sample, we take a single word. Ingram & Ingram [11] and Ingram [12] defined PWP as the ratio of the produced phonological mean length of utterance (PMLU) to its targeted counterpart, that is, the ratio of the sum of in context correctly produced consonants in context (called from now on correct consonants) and all produced segments (consonants plus vowels) to weighted targeted segments (targeted consonants plus all targeted segments). We re-arrange this formula first by separating the two terms of the sum in the numerator resulting in the sum of two ratios: correct consonants to weighted targeted segments plus all phones to weighted targeted segments. In turn, each ratio is now expressed as the product of two ratios. The former ratio, call it ratio-1, is written as the correct consonants divided by the targeted consonants, called in the literature proportion of consonants correct (PCC) (Shriberg & Kwiatkowski [8], Shriberg et al. [9]), times the targeted consonants divided by the weighted targeted segments. In terms of the proportion of vowels (PV) to segments in the targeted word, this last ratio, call it p, becomes (1-PV)/(2-PV). Ratio-2 is expressed as the product of the produced segments to targeted segments times the ratio of targeted segments to the weighted targeted segments. This last ratio is clearly equal to 1-p. Since targeted segments may equivalently be written as the produced segments plus the deleted segments minus the epenthetic segments, the former ratio in the product may be expressed as 1-PPD, where PPD stands for the proportion of phonemes (segments) deleted minus phonemes added. Deleted and epenthetic segments are taken into account since it is known (Ingram, [26]; Stoel-Gammon & Dunn [27], Bernhardt &

Stemberger [28]) that during child phonological development it is not unusual to have targeted consonants deleted and vowels added. It is less usual to have epenthetic consonants, even adjacent to a targeted consonant, in normal as well as disordered children (e. g. Ingram [29], Stemberger [14], Babatsouli [30]).
We have, therefore, derived an expression for the phonological word proximity (PWP) of a single word in terms of three phonological parameters PCC, PV and PPD, as

$$PWP = pPCC + (1-p)(1-PPD) \tag{1a}$$

$$p = (1-PV)/(2-PV) \tag{1b}$$

Going now from a single word's PWP to the PWP of a speech sample consisting of N words, we propose taking the *weighted* average of the single word PWPs instead of their arithmetic mean so that the cumulative PWP becomes the ratio of the arithmetic mean of the produced single word PMLUs to the arithmetic mean of the targeted PMLUs, i.e.,

$$PWP = (\sum PMLU^{(p)}/N)/(\sum PMLU^{(t)}/N) \tag{2}$$

The choice of weighted average yields a cumulative PWP of exactly the same form as that of the single word given by Eq. (1a, b), with the three phonological parameters PCC, PV and PPD now computed as weighted averages directly from the whole sample. For example, PCC is now the ratio of the correct consonants produced in context to the targeted consonants in the whole sample. Moreover, the choice of weighted average makes it possible to obtain upper and lower bounds on the cumulative PWP by taking the minimum and maximum PPD, respectively.
In general, the three phonological parameters PCC, PV and PPD vary across speech samples within a child or between children. Eq. (1a, b) shows that the phonological word proximity (PWP) depends linearly on PCC and PPD and, nonlinearly, on PV. However, the PV value may be kept constant when comparing samples by appropriately selecting the words in them. At the early age of a few months, an infant's PCC is negligible giving PWP as (1-p)(1-PPD). At complete acquisition, attained by normal children usually at school age or later, PCC is almost one, PPD is negligible and, therefore, PWP becomes one independent of PV.

2.2 The weighting parameter p

The weighting parameter p which represents the proportion of targeted consonants to weighted targeted segments, weighs the contribution of PCC to the value of PWP as shown in Eq. (1a). Similarly, 1-p weighs the contribution of 1-PPD. It may be seen that p, as given by Eq. (1b), monotonically decreases with PV. The minimum PV is 0, yielding 0.5 as the maximum p. Thus, 1-p is

greater than p for all the values of PV. The maximum PV is 1 giving 0 as the minimum p. PV norms in adult speech are 0.45 for English and Dutch, 0.5 for Italian and Spanish and 0.55 for Japanese (Ramus et al. [31]). For example at PV=0.45, p=0.35 and 1-p=0.65. For children, the proportion of vowels (PV) in targeted word samples in English usually varies in practice between 0.25 and 0.5. When most words contain consonant clusters, the lower value is approached as will be seen in the samples below. For this range of PV values, p is proposed to approximately depend on PV in a linear fashion as

$$p \approx PV^*(1+PV^*-PV), \quad PV^* = (3-\sqrt{5})/2 \tag{3}$$

This is the one-term Taylor series expansion of p about PV*≈0.382, the only acceptable PV value less than 1 for which p=PV. For PV between 0.25 and 0.5, the root mean square error in approximating p by Eq. (3) is calculated to be 0.0018, while the maximum error is 0.0038 at PV=0.25. For the range of PV values in discussion, we compared the approximation given by Eq. (3) to that of a linear interpolation of p between 3/7 (PV=0.25) and 1/3 (PV=0.5). It turns out that the latter approximation's root means square error is 0.0026, larger than the former's. Therefore, it is more accurate to use the linear approximation given by Eq. (3).

A consequence of approximating p linearly on PV values between 0.25 and 0.5 is the linear approximation of PWP of Eq. (1a) on PV. In this range of PV values, the error in approximating PWP, denoted by Δ(PWP), is given in terms of the error in p, denoted by Δp, as

$$\Delta(PWP) = -(1-PCC-PPD)\Delta p \tag{4}$$

Since the ratio of the produced segments to targeted segments, 1-PPD, is larger than the proportion of consonants correct, as will be discussed in the subsection that follows, the factor multiplying Δp is greater than zero. It is also smaller than one since PCC and PPD are positive. As a result, the error in approximating PWP using, instead of the exact p, that of Eq. (3) is smaller than the error in approximating p, itself.

2.3 Upper and lower PWP bounds

For a given sample of targeted words, the proportion of vowels (PV) is known and, thus, the PWP value depends on the PCC and PPD values which are measured from the produced speech.

It is seen from Eq. (1a) that, for the same PCC, PWP is larger for a smaller PPD. Therefore, PPD's minimum and maximum values yield upper and lower PWP bounds. We take the smallest PPD to be equal to zero when there is no deletion or epenthesis. The largest PPD is equal to (1-PCC)(1-PV) when there is no epenthesis, no deleted vowels, and no substitutions for targeted consonants; only correct consonant productions and consonant deletions. Substituting these two

extreme values of PPD in Eq. (1a) we obtain, respectively, the upper and lower PWP bounds as:

$$\textit{Upper bound}: \qquad PWP_{max} = PCC + (1-p)(1-PCC) \tag{5}$$

$$\textit{Lower bound}: \qquad PWP_{min} = PCC + (1-2p)(1-PCC) \tag{6}$$

Clearly, PWP is larger than PCC, except at complete acquisition when PCC equals one and they become equal. For the same p and PCC values, subtracting the two bounds yields the largest possible spread in PWP values between any children as p(1-PCC). For $p_2<p_1$ and $PCC_2>PCC_1$, which is the case of more consonants produced correctly in context when the proportion of vowels is larger, subtracting the upper bound on PCC_2 from the lower bound on PCC_1 gives the largest spread of PWP as $(2p_1-p_2)(1-PCC_2)+2p_1\Delta(PCC)$. The PWP bounds may also be used, as follows, to determine sufficient conditions on the changes of p and PCC across samples in order to have an increasing PWP.

For the same targeted sample, PCC generally changes with a child's age. It also generally changes within a child and between children for two targeted samples of the same p but of distinctly different word constituency, for example, singleton and cluster words. Taking the lower PWP bound for the larger PCC, say PCC_2, and the upper PWP bound for the smaller PCC_1, we derive from Eqs. (5) and (6) that PWP_2 will be for sure larger than PWP_1 only when PCC_2 is larger than $(1+PCC_1)/2$, meaning that PCC_2 will necessarily have to be larger than 0.5. When PCC_2 is smaller than $(1+PCC_1)/2$, then PWP_2 is either larger or smaller than PWP_1 depending on the values of PPD_1 and PPD_2. This is investigated further in the subsection that follows.

With p also changing across samples, PWP_2 is for sure larger than PWP_1 only when PCC_2 is larger than $1-[p_1(1-PCC_1)/(2p_2)]$, meaning that PCC_2 will necessarily have to be larger than $1-[p_1/(2p_2)]$. However, since PCC_2 was taken larger than PCC_1 we must have p_2 larger than $p_1/2$, which from Eq. (1b) yields, PV_2 smaller than $2/(3-PV_1)$.

2.4 PWP and PCC changes

Generally, PWP changes between speech samples within a child and between children. It is of interest to determine the magnitude of PWP change in terms of changes in PCC, PV and PPD. In doing so, an answer will be given to the question: which is a better measurement PCC or PWP, in the sense of discriminating performance between speech samples?

Consider two targeted samples with their corresponding single productions or one targeted sample with two different productions. In either case, the parameters in the two sets are distinguished by the subscripts 1 and 2 and the change in their values (2 minus 1) is denoted by the Greek capital letter Δ. Then, subtracting Eq. (1a) with subscript 1 in the parameters from Eq. (1a) with

subscript 2 in the parameters, results in the following expression for the change of PWP, Δ(PWP),

$$\Delta(PWP) = -(1 - PCC_2 - PPD_2)\Delta p + p_1\Delta(PCC) - (1 - p_1)\Delta(PPD) \quad (7)$$

The change in p, Δp, in terms of the change in PV, Δ(PV), is obtained similarly from Eq. (1b) as

$$\Delta p = -(1 - p_1)(1 - p_2)\Delta(PV) \quad (8)$$

However, for PV values between 0.25 and 0.5, p may be approximated by Eq. (3), as discussed above, which yields a much simpler expression for Δp,

$$\Delta p \approx -(3 - \sqrt{5})/2\ \Delta(PV) \quad (9)$$

We see from Eq. (7) that PWP increases in proportion to increases in PV and PCC and a decrease in PPD. When the same sample is targeted or when two targeted samples have the same PV, it is concluded from Eq. (7) that PWP increases whenever the change in PCC is larger than the change in PPD divided by (1-PV), that is,

$$\Delta(PWP) \geq 0 : p\Delta(PCC) \geq (1 - p)\Delta(PPD) \quad (10)$$

We note that positive Δ(PCC) and negative Δ(PPD) automatically satisfy (10), resulting to a positive change in PWP.

Changes of PCC and PPD are, however, bounded since 1-PPD-PCC is bounded above and below by 1 and 0, respectively, as discussed above. This means that when changes in PCC and PPD values have opposite signs, their magnitudes may vary anywhere between 0 and 1. But when the PCC and PPD changes have the same sign, they are bounded above by the sum of their magnitudes not exceeding 1. That is, the bounds on Δ(PCC) and Δ(PPD) are given as

$$\begin{aligned} &Bounds\ on\ \Delta(PCC), \Delta(PPD): \\ &i)\ \Delta(PCC)\Delta(PPD) < 0,\ |\Delta(PCC)| + |\Delta(PPD)| \leq 1 \\ &ii)\ \Delta(PCC)\Delta(PPD) > 0,\ 0 \leq |\Delta(PCC)|, |\Delta(PPD)| \leq 1 \end{aligned} \quad (11)$$

A practical question that may arise is: which is a better measurement, PWP or PCC, with regard to discriminating performance in two productions? In other words, what is really the difference between the two measurements in practice? We will answer this, generally, by taking the same targeted sample and two different productions and compare the magnitude of the PWP change to that of PCC. Practitioners may want either a small or a large disparity between the two

values of the measurement they use, depending on whether they want to discriminate performance in the two productions. Taking the absolute value of Δ(PWP) of Eq. (7) with Δp=0 and comparing it with the absolute value of Δ(PCC), we find that it is larger only when Δ(PCC) Δ(PPD) satisfy the following conditions:

$$\begin{aligned} &|\Delta(PWP)| \geq |\Delta(PCC)|: \\ &i)\ \Delta(PCC)\Delta(PPD) < 0,\ |\Delta(PCC)| \leq |\Delta(PPD)| \\ &ii)\ \Delta(PCC)\Delta(PPD) > 0,\ (1+p)|\Delta(PCC)| \leq (1-p)|\Delta(PPD)| \end{aligned} \tag{12}$$

in which Δ(PCC) and Δ(PPD) are bounded according to (11). If conditions (12) are violated, the magnitude of Δ(PCC) is larger than the magnitude of Δ(PWP).

Considering (10) and (12) simultaneously, since (1-p)/p is larger than (1-p)/(1+p) we conclude that only when Δ(PPD) is negative, Δ(PWP) may be positive and, at the same time, larger than Δ(PCC). Even then, this will be true only when

$$|\Delta(PCC)| \leq |\Delta(PPD)|,\ \ (1+p)|\Delta(PCC)| \leq (1-p)|\Delta(PPD)| \tag{13}$$

with Δ(PCC) and Δ(PPD) bounded according to (11).

When comparing two productions between two stages in general, the largest possible Δ(PWP)/Δ(PCC) as obtained in the preceding subsection is 2p+p(1-PCC_2)/Δ(PCC), where Δ(PCC) is positive. This will approach 2p from above as stage 2 reaches complete acquisition, where PCC_2 is almost 1. Therefore, Δ(PWP)/Δ(PCC) will generally be larger than 1 when necessarily (1-2p)Δ(PCC)<p(1-PCC_2) in accordance with the first case of (12), as Δ(PPD) is -(1-PCC_1)(1-PV). Last, when two speech samples are compared in general, and Δ(PCC) is positive while Δ(PPD) is negative, we are in the first case of (12) and, therefore, the ratio Δ(PWP)/Δ(PCC) will be smaller than 1 only when the magnitude of Δ(PCC) is larger than the magnitude of Δ(PPD).

These observations have also implications on the age dependence of the PWP change relative to the PPC change between two speech samples. We take sample-2 to be much easier to produce correctly than sample-1. For example, take the words in sample-2 to contain only singletons and all the words in sample-1 to contain consonant clusters. Then, near complete acquisition, the first term of Eq. (7) is negligible, Δ(PPD) is expected to be negative approaching zero before Δ(PCC) which is expected to be positive and, thus, Δ(PWP)/Δ(PCC) becomes smaller than one, approaching p_1. As we get away from complete acquisition, the first term of Eq. (7) is positive since p_2 is smaller than p_1, and we expect that in general Δ(PWP)/Δ(PCC) will increase with decreasing age with its largest value, obtained in the preceding subsection, being equal to $2p_1+(2p_1-p_2)(1-PCC_2)$/Δ(PCC) as Δ(PPD) is -(1-PCC_1)(1-PV_1).

Therefore, Δ(PWP)/Δ(PCC) will generally be larger than 1 when necessarily (1-$2p_1$)Δ(PCC)<($2p_1$-p_2)(1-PCC_2). Since Δ(PPD) is negative and Δ(PCC) is positive for the case at hand, the necessary and sufficient condition for Δ(PWP) to be larger than Δ(PCC), on use of Eq. (7), becomes

$$\Delta(PCC) \le |\Delta(PPD)| + |\Delta p|(1 - PCC_2 - PPD_2)/(1 - p_1) \quad (14)$$

Comparing (14) with (12), we see that when the proportion of vowels is not the same between targeted samples, the range of Δ(PCC) values for which Δ(PWP) is greater than Δ(PCC) is larger. This means that, when Δ(PCC) is smaller than Δ(PPD), condition (14) is automatically satisfied and Δ(PWP) is for sure larger than Δ(PCC).

If one wants to measure only consonants in a speech production and ignores vowels altogether, the phonological word proximity (PWP) becomes what we will call 'phonological word consonants proximity' (PWCP). It is interesting to compare directly the two measures, PCC and PWCP, as the first measure counts consonants only when they are produced correctly in context while the second measure counts consonants when they are produced correctly, independent of context, even though the correct consonants in context are counted twice. In this case, the magnitude of Δ(PWCP) is larger than the magnitude of Δ(PCC) only when the changes of the proportion of consonants deleted minus added, Δ(PCD), and of Δ(PCC) satisfy the inequalities given by (12) with PWCP in place of PWP, PCD in place of PPD, and p=0.5 since PV=0.

3 The speech data

The data is taken from a Greek/English bilingual female child's speech in English from age 2 years and 6 months to age 3 years and 9 months. Her spontaneous speech in English during thirty-minute daily routine interactions with the first author was recorded and, subsequently, time aligned and phonetically transcribed by the first author in a CLAN (MacWhinney [17]) database, using the International Phonetic Association (IPA) symbols. The purpose here is to compute the child's phonological word proximity (PWP) and trace its monthly change with age together with its components PCC and PPD, placing them in perspective within general child speech development, as examined above. For this reason, the same sample of targeted words was considered at each age. The sample taken consists of 25 words which were selected in order to satisfy two main criteria: first, that the same 25 words could be found in the child's speech at least once a month between the ages of 2;6 and 3;9 and, second, that they are a mixture of different complexities in terms of consonant place and manner of articulation, consonant position in the word, singleton consonants and consonant clusters, and number of syllables. As expected, the child's natural utterances contained a varying number of words, so that the 25 word types in the sample were extracted from different utterances, on a different day or week of the month, in general. However, the first production

of each word in the month was included in the sample, so that the child's age increased by about a month between word productions. The targeted words in the child's speech sample in alphabetical order are: *again, also, and, another, any, bag, blanket, close, clothes, come, don't, English, finished, give, go, hold, inside, make, play, ready, the, together, took, why, wolf.*

It is expected that when the speech sample considered changes substantially, PWP will in general also change as its components PV, PCC and PPD vary between samples. In the present study, this is exemplified by selecting a second sample comprising of all the word types in the child's speech at the age of 3;0 that contain at least one consonant. In order to have a large sample of word types, we selected the words within ten days after the child's third birthday upon first production. As a result, the following 158 targeted word types are included in the sample: *accident, again, airplane, already, also, and, animals, another, any, back, bag, balance, beach, because, bed, birdie, bit, blanket, block, boots, box, bread, breakfast, bridge, bring, brush, bunnies, bunny, called, case, cat, chicken, chicken, chocolate, clean, clock, close, clothes, colors, come, cotton, counter, crunchy, cucumber, destroying, dirty, dog, dolphin, don't, donkeys, door, downhill, downstairs, dream, English, every, excellent, falling, farm, finished, fish, five, floor, food, found, full, garden, give, glasses, go, grab, grandpa's, hair, have, head, help, here, hide, hold, inside, juice, kettle, kiss, later, leave, left, lick, licking, look, loose, lost, make, meatballs, middle, milk, moon, more, morning, myself, nice, no, nose, now, once, open, outside, panty, pieces, piglet, plain, play, polite, potatoes, pull, pushing, put, puzzles, rain, ready, red, remember, restaurant, scatter, seeds, shopping, shoulder, shower, slide, small, someone, space, spaghetti, stopped, street, stroller, sunscreen, table, teacher, the, there, things, throw, toast, today, together, took, train, trash, trouble, umbrella, upset, washed, what, where, why, wolf, working, yes*. The changes in the phonological parameters PWP, PCC, and PPD between the two samples will be calculated and viewed in relation to the method and analysis presented above.

4 Numerical results

4.1 Child data in general

The method and analysis for general child speech development, as far as phonological word proximity (PWP) and its components are concerned, were examined above. Here, numerical results will be presented graphically. In terms of its components, PWP is given by Eq. (1a, b). In a three-dimensional (PCC, PV, PWP) rectangular coordinate system, all children's PWP values lie inside a body which is bounded above and below by the surfaces given by Eq. (5) and Eq. (6), respectively. These bounds on PWP are calculated for PCC values ranging from 0 to 1 and PV from 1/3 to 3/4 and are plotted in a (PCC, (2-PV)/(1-PV), PWP) space where they are easy to view. Note that (2-PV)/(1-PV)=1/p ranges from 2.5 to 5. The results are shown in Fig. 1.

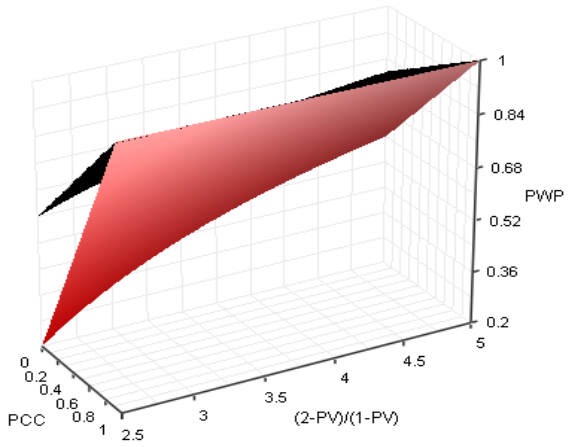

Fig. 1. Upper (black) and lower (red) bounds on PWP in child speech development.

The black surface in Fig.1 is the upper bound while the red surface is the lower bound. These surfaces are shaped as hyperbolic paraboloids and they form the wings of a phonological word proximity (PWP) glider. When we have the same sample of targeted words, PV does not change and children's PWP values lie inside the glider's section which, as seen in the figure, is triangular with its base equal to p at zero PCC. As PV increases, 1/p also increases and the base of the bounding triangle becomes smaller. The largest triangular section base in the figure is equal to 0.4 at 1/p=2.5 (the left end of the figure) and the smallest is equal to 0.2 at 1/p=5 (the right end of the figure). At complete acquisition, PCC is one, PPD is zero and PWP becomes one independent of PV. This defines the glider's ceiling shown in the figure along 1/p.

When PV is the same between speech samples, it was shown in the method and analysis above by conditions (10) - (13) that what matters in the change of PWP is the change of PCC relative to the change of PPD. This is shown schematically in Fig. 2. Regions of positive and negative Δ(PWP) are bounded by the blue-green and blue-red lines respectively in a Δ(PCC), Δ(PPD) plane. On the blue line which represents the equation in (10), PWP remains unchanged between the two speech samples. In the figure, the irregular hexagon bounding Δ(PCC) and Δ(PPD) values represents the equations in (11).

Discriminating measurements between two productions of the same targeted speech sample is of interest to practitioners. To this end, a comparison of the magnitude of Δ(PWP) to that of Δ(PCC) was made in the method above andwas given by (12). In Fig. 3, the regions where (12) is satisfied are drawn in dashed lines in a Δ(PCC), Δ(PPD) plane. That is, the magnitude of Δ(PWP) is

larger than that of Δ(PCC) in the dashed regions and smaller in the rest. As in Fig. 2, changes in PCC and PPD are bounded by the irregular hexagon shown also in this figure.

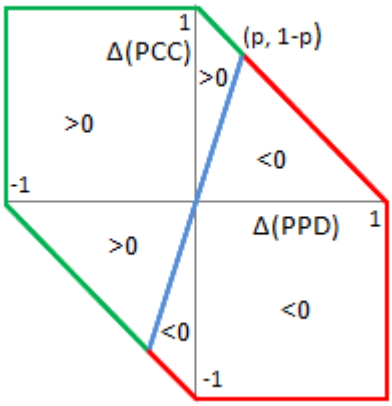

Fig. 2. PWP changes (>0, <0) relative to PPD, PCC changes (Δ) in child speech development.

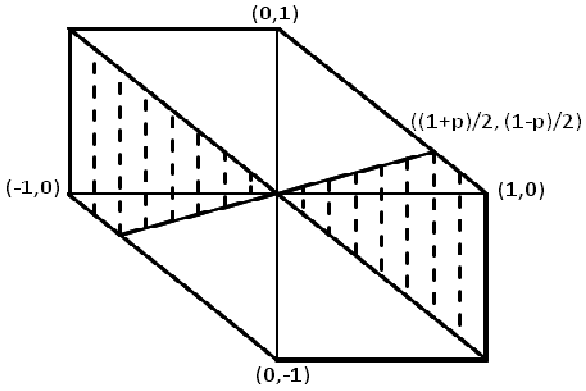

Fig. 3. The shaded region of the Δ(PPD), Δ(PCC) plane where the magnitude of Δ(PWP) is larger than that of Δ(PCC) in child speech development.

4.2 The phonological data of this study's child

For the bilingual child's speech sample described above, which was taken monthly between the ages 2;6 and 3;9, we calculate PCC, PPD and subsequently

PWP. Cumulatively for the 25 words in the speech sample, PCC and PPD are computed as two ratios, respectively. The first is the ratio of the number of produced correct consonants divided by 63, the total number of consonants in the speech sample, while the second ratio is the number of deleted consonants and vowels minus the added ones divided by the total number of segments in the speech sample, which is 109. Thus, the proportion of vowels, PV, in the targeted sample is 46/109 or 0.42. The developmental PCC and PPD values were subsequently computed monthly. In turn, PWP was computed using Eq. (1a, b).

The numerical results are depicted in solid lines in Fig. 4. We see that three distinct stages of phonological development may be identified, associated with the growth patterns of the phonological parameters PCC, PPD and PWP. In each stage their growth pattern may be fitted by a straight line shown by the dashed line in Fig. 4. As a result, the overall developmental pattern is tri-linear.

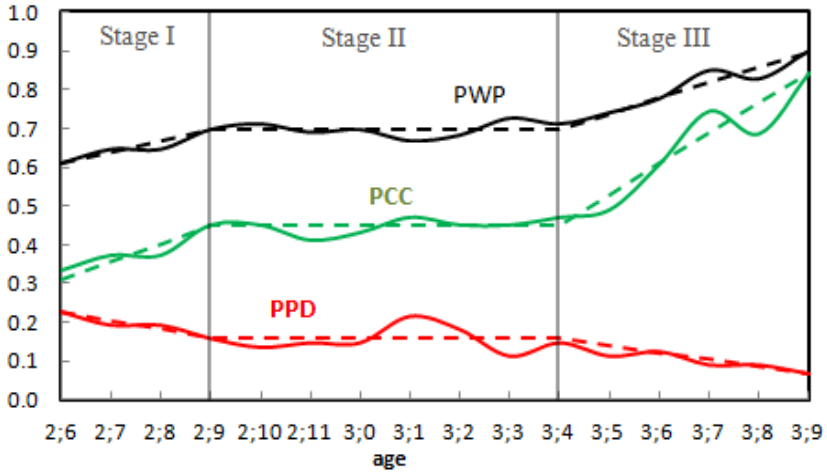

Fig. 4. The bilingual female child's phonological parameters PPD, PCC and PWP in speech development.

The first stage lasts for three months from age 2;6 to age 2;9 and is progressive in PWP as, according to (10), PCC increases and PPD decreases. The increase in PWP is from 0.32 to 0.70 and is of the double-logistic type. The second stage lasts for seven months and is nearly cyclic as PCC, PPD and also PWP fluctuate, even though non-uniformly, about the same level. In fact, at this stage, PCC has a strange attractor of the value 0.44 and PCC has a strange attractor of the value 0.16 resulting in a strange attractor for PWP of the value 0.69. The third stage is again progressive with PCC increasing, PPD decreasing and PWP increasing from 0.70 to 0.90 in a double-logistic fashion with a nearly cyclic regime separating the two logistic-like sub-stages within this stage.

The existence of a plateau stage during speech development has been reported in the literature on a qualitative basis (e.g. Ingram [13]). Moreover, the plateau is the well known middle stage of the U-shaped learning pattern in developmental psychology (e.g. Werker et al. [32]). Here, on a quantitative basis, we see that this stage exists and is, in fact, nearly cyclic.

Now, in view of the analysis and discussion above, it will be interesting to compare the ratio $\Delta(PWP)/\Delta(PCC)$ in the child's speech performance between different ages. Calculating this ratio between the first and last speech samples in stage-1 and in stage-3 we obtain 0.73 and 0.50 respectively. In both stages, the ratio is smaller than 1 since $\Delta(PCC)$ is 0.118 and 0.372 in the two stages respectively, while $\Delta(PPD)$ is negative and its magnitude is smaller at 0.068 and 0.148 and, therefore, the first of (12) is violated in both stages. The targeted sample is the same all along with $p=0.37$. Then, according to the method and analysis above, the ratio will approach 0.37 when two speech productions are compared near this child's or any child's complete phonological acquisition. In fact, the ratio between the last two months (3;8 and 3;9) in stage-3 is 0.46, that is, even closer to 0.37 than the average ratio 0.50 over the whole stage-3.

The bilingual child's developmental PWP and PCC values are placed in perspective in general child speech development by comparing them to the upper and lower bounds on phonological word proximity (PWP) given by Eqs. (5) and (6), respectively. This comparison is depicted graphically in Fig. 5 where PWP is plotted versus PCC.

The upper bound on PWP given by Eq. (5) for all children is represented by the black dashed line in Fig. 5, while the lower bound given by Eq. (6) is the red dashed line. These two lines meet at the point PCC=1, PWP =1 forming a triangular bounding region in which all children's values lie during speech development. The vertices of the triangle's base are given by the points (0,1-p=0.63) and (0, 1-2p=0.26). This triangular region is the section of the PWP glider of Fig. 1 at $1/p=2.7$. However in practice, as it is also the case here, a child's speech samples are taken following increased production of intelligible words. Thus, it is expected that the smallest computed PCC will be larger than zero and the largest computed PPD will be smaller than all children's maximum value which is given by (1-PV)(1-PCC), resulting in a smaller bounding region for PWP. In Fig. 5, this region for the bilingual child is defined inside the black solid lines because the child's minimum PCC is 0.33 and maximum PPD is 0.23 at this PCC. Inside this trapezoid, the child's actual developmental (PCC, PWP) values are shown by dots. Their correlation is rather chaotic in the nearly cyclic regime of their developmental paths where PCC has a strange attractor of the value 0.44 and PPD has a strange attractor of the value 0.16.

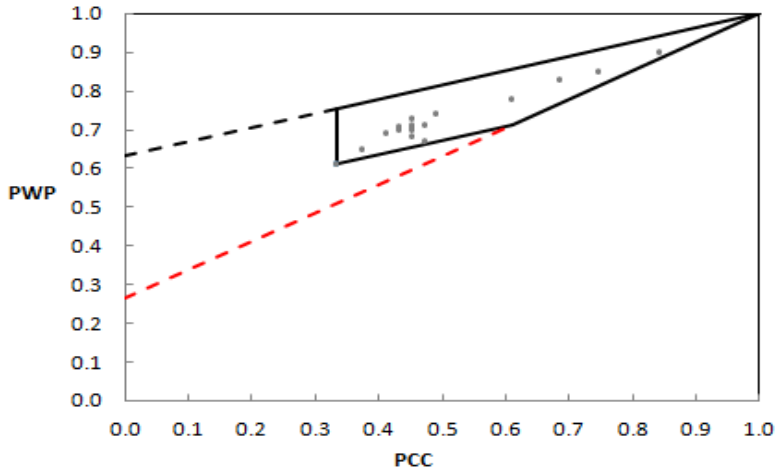

Fig. 5. The bilingual female child's developmental PWP (dots) and its bounds (solid lines) within the bounds of general child speech development (dashed lines).

Now, it will be of interest to compare the bilingual child's PWP, PCC and PPD values between productions of the targeted sample considered above and the larger targeted sample at age 3;0 described in the methodology. The 158 words in the larger sample have PV= 0.37 (p=0.39). The child's production resulted in PCC equal to 0.53 and PPD equal to 0.127. Then, on use of Eq. (1), we get 0.74 for PWP. The corresponding values for the smaller targeted sample at age 3;0 are PWP=0.713, PCC=0.47 and PPD=0.145. This shows that the smaller targeted sample traced along development is more difficult for the bilingual child to produce correctly than the larger sample. Therefore, the PCC and PWP growth patterns of Fig. 1 are conservative. Even though the two targeted samples differ significantly in size, we see that the disparity in their PWP, PCC and PPD values is relatively small with the largest disparity being that of PCC. Calling sample-2 the larger targeted sample, we have Δ(PCC)=0.06, Δ(PPD)=-0.018 and Δ(PWP)=0.027. The ratio Δ(PWP)/Δ(PCC) is 0.45 smaller than 1 since Δ(PPD) is negative and Δ(PCC), Δ(PPD) are such that (14) is violated.

It is also of interest to compare the child's performance between words that contain singleton consonants and words that contain consonant clusters (at least two consonants next to each other). We call here sample-1 the 89 cluster words included in the 158 words sample whose weighted proportion of vowels is PV_1=0.43. We call sample-2 the 69 singleton words with PV_2=0.34. The child's corresponding productions give (PCC_1, PPD_1, PWP_1) and (PCC_2, PPD_2, PWP_2), respectively, as: (0.52, 0.18, 0.70) and (0.66, 0.03, 0.86). We see that the child

produces singleton words better than cluster words. The disparity in the values of PCC, PPD and PWP is expressed in terms of the ratio $\Delta(PWP)/\Delta(PCC)$ which becomes 1.14. It is larger than 1 since $\Delta(PPD)$ is negative and its magnitude is larger than $\Delta(PCC)$, so that (14) is automatically satisfied. As discussed in the methodology following Eq (7), the ratio will overall decrease with age and it will approach $p_1=0.36$ near complete phonological acquisition, where PCC_2 is nearly 1 and PPD is negligible for both singleton and cluster words.

5 Conclusions

Measurements of phonological word proximity (PWP) and proportion of consonants correct (PCC) in child speech development are placed in perspective having obtained analytically their relationship as well as upper and lower bounds in terms of the proportion of phonemes deleted minus added (PPD) and the proportion of vowels (PV). Child data reveal the existence of a nearly cyclic stage with strange attractors for PCC and PPD and, consequently for PWP, before the final progressive double-logistic stage in phonological development, and the relative advantages of using PWP instead of PCC in discriminating performance between speech samples within a child and between children, of the same or different age.

References

1. M. M. Nice. Length of sentences as a criterion of a child's progress in speech. *Journal of Educational Psychology, 16,* 370-379, 1925.
2. D. McCarthy. *The language development of the pre-school child*. Child Welfare Monographs. Minneapolis: University of Minnesota Press, 1930.
3. R. Brown. *A first language: the early stages*. Cambridge, MA: Harvard Univ. Press, 1973.
4. K. Kemp and T. Klee. Clinical language sampling practices: Results of a survey of speech-language pathologist in the United States. *Child Language Teaching and Therapy, 13,* 161-176, 1997.
5. M. D. Parker and K. Brorson. A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language, 25,* 365-376, 2005.
6. B. L. Wellman, I. M. Case, I. G. Mengert and D. E. Bradbury. *Speech sounds of young children.* University of Iowa Studies in Child Welfare, volume 5, No 2. Iowa City: The University of Iowa, The Iowa Child Welfare research Station, 1931.
7. I. Poole. Genetic development of articulation of consonants sounds in speech. *Elementary English Review, 11,* 159-161, 1934.
8. L. Shriberg and J. Kwiatkowski. *Natural process analysis (NPA)*. New York: John Wiley, 1980.
9. L. Shriberg, D. Austin, B. Lewis, J. McSweeney and D. Wilson. The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech,Language, and Hearing Research, 3,* 708-722, 1997.

10. L. S. Schmitt, B. H. Howard and J. F. Schmitt. Conversational speech sampling in the assessment of articulation proficiency. *Language, Speech and Hearing, Service in Schools*, *14*, 210-214, 1983.
11. D. Ingram and K. Ingram. A whole-word approach to phonological analysis and intervention. *Language, Speech and Hearing Services in Schools, 32,* 271-283, 2001.
12. D. Ingram. The measurement of whole-word production. *Journal of Child Language*, *29,* 713-733, 2002.
13. D. Ingram. *First language acquisition: method, description and explanation*. Cambridge: Cambridge University Press, 1989.
14. J. P. Stemberger. Speech errors in early language production. *Journal of Memory and Language, 28,* 164-188, 1989.
15. T. W. Powell. Phonetic transcription of disordered speech. *Topics in Language Disorders, 21,* 52-72, 2001.
16. H. Taelman, G. Durieux and S. Gillis. Notes on Ingram's whole-word measures for phonological development. *Journal of Child Language, 32,* 391-400, 2005.
17. B. MacWhinney. *The Childes project: tools for analyzing talk*. 3rd edition. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
18. F. Bunta, L. Fabiano-Smith, B. A. Goldstein and D. Ingram. Phonological whole-word measures in three-year-old bilingual children and their age-matched monolingual peers. *Clinical Linguistics and Phonetics, 23,* 156-175, 2009.
19. L. Burrows and B. A. Goldstein. Whole word measures in bilingual children with speech sound disorders. *Clinical Linguistics and Phonetics, 24,* 357-368, 2010.
20. A. A. Macleod, K. Laukys and S. Rvachew. The impact of bilingual language learning on whole-word complexity and segmental accuracy among children aged 18 and 36 months. *International Journal of Speech and Language Pathology, 13,* 490-499, 2011.
21. K. Saaristo-Helin. *Phonological mean length of utterance as an indicator of typical and impaired phonological development in children acquiring Finnish*. Unpublished doctoral dissertation, University of Helsinki, 2011.
22. B. A. Goldstein and F. Bunta. Positive and negative transfer in the phonological systems of bilingual speakers. *International Journal of Bilingualism*, *16 (4)*, 388-401, 2012.
23. S. E. Freedman and J. A. Barlow. Using whole-word production measures to determine the influence of phonotactic probability and neighborhood density on bilingual speech production. *International Journal of Bilingualism, 16 (4)*, 369-387, 2012.
24. D. Ingram and V. L. Dubasik. Multidimensional assessment of phonological similarity within and between children. *Clinical Linguistics & Phonetics, 25,* 962-967, 2011.
25. P. Jr. Flipsen, J. B. Hammer and K. M. Yost. Measuring severity of involvement in speech delay: Segmental and whole-word measures. *American Journal of Speech-Language Pathology, 14,* 298-312, 2005.
26. D. Ingram. *Procedures for the phonological analysis of children's language*. Baltimore, MD: University Park Press, 1981.

27. C. Stoel Gammon and C. Dunn. *Normal and disordered phonology in children*. Baltimore, MD: University Park Press, 1985.
28. B. M. Bernhardt and J. P. Stemberger. *Handbook of phonological development from the perspective of constraint-based nonlinear phonology*. New York, NY: Academic Press, 1997.
29. D. Ingram. *Phonological disability in children*. London, England: Edward Arnold, 1976.
30. E. Babatsouli. Beyond the norm developmental cluster production. Poster presented at the *14th International Clinical Phonetics and Linguistics Association Conference*, Cork, Ireland, 2012.
31. F. Ramus, M. Nespor and J. Mehler. Correlates of linguistic rhythm in the speech signal. *Cognition, 73,* 265-292, 1999.
32. J. F. Werker, D. G. Hall and L. Fais. Reconstruing U-shaped functions. *Journal of Cognition and Development, 5,* 147-151, 2004.