Recurrence Quantification Analysis: A Promising Method for Data Evaluation in Medicine

Jakub Schlenker[1], **Tomáš Nedělka**[1,2], **Lucie Riedlbauchová**[3], **Vladimír Socha**[1], **Karel Hána**[1], **Patrik Kutílek**[1]

[1] Czech Technical University in Prague, Faculty of Biomedical Engineering, Kladno, Czech Republic

[2] Department of Neurology, University Hospital Motol and 2nd Faculty of Medicine of the Charles University, Prague, Czech Republic

[3] Clinic of Cardiology, University Hospital Motol and 2nd Faculty of Medicine of the Charles University, Prague, Czech Republic

Abstract

Introduction: This paper describes principles of a promising method for medical data analysis called recurrence analysis, which is based on the chaos theory that better describes processes in the living organisms.
Methods: Phase space reconstruction and recurrence analysis are explained in brief. The main part of the work focuses on recurrence plot, which is a basic tool for recurrence analysis.

Conclusion: Possible clinical applications of recurrence analysis in the field of medicine are discussed and a short overview of pilot projects performed by our team on this topic is given.

Keywords

Nonlinear analysis, recurrence analysis, recurrence plot, phase space, heart rate variability, autonomic dysfunction.

Correspondence to:

Jakub Schlenker
Joint Department of Biomedical Engineering CTU
and Charles University in Prague, Czech Republic
Address: Studničkova 7/2028, 128 00 Prague 2, Czech Republic
E–mail: jakub.schlenker@gmail.com

EJBI 2014; 10(1):en35–en40
received: October 9, 2014
accepted: November 19, 2014
published: December 15, 2014

1 Introduction

Usage of nonlinear methods for data analysis is becoming increasingly popular in medicine due to the fact that they seem to be able to describe selected processes occurring in living organism more effectively than it is these days [1].

To increase the probability of full recovery or to minimize the health damages, it is important to detect diseases in their early or even in their subclinical stages. Since specific methods of nonlinear analysis seem to be sensitive enough to uncover these early phases of the disease development, their application in the data analysis may improve health care and help the physicians to understand better the physiological and pathophysiological processes occurring in the human body.

One such nonlinear method recently applied in medicine is the recurrence analysis. Method of so called recurrence analysis is derived from the chaos theory which describes the basic dynamics of a system with chaotic behaviour that can be found in every biological system [2]. Recurrence analysis has been successfully used in pilot projects in cardiology [4, 5] and neurology [7, 6, 9, 8, 10] where it was mainly used to describe dynamics of the heart rate and blood pressure regulation. These physiologic variables are under permanent control of the autonomic nervous system which may be viewed as an example of nonlinear deterministic system since the autonomic nervous system instantaneously changes its tone based on the actual demands and needs of the organism [11, 2]. Impaired function of the autonomic nervous system is therefore often associated with reduced variability of functions that this system controls, i.e. reduced heart rate variability. Due to this fact, the system (heart rate) tends to recur to a similar state and exhibit only limited changes in response to outer inputs when its control through the autonomic nervous system is harmed.

The effort of our team has recently focused on the evaluation of possible role of the recurrence analysis in the diagnostic of various diseases in their early phases (diseases origin of which is associated with the autonomic dysregulation) especially in the field of neurology [9, 10] and cardiology.

2 Methods

Recurrence analysis is one of the non-linear analysis of data derived from chaos theory. Recurrent graphs, basic tool recurrent analysis allow to visualize repetitive behaviour of dynamic systems. The method is also suitable for non-linear analysis of short-term and non-stationary data [12].

- The first step of analysis, as with most nonlinear techniques [12], is the reconstruction of the phase space.
- The second step is the formation of recurrent plot using the threshold distance.
- The last step in the analysis is to calculate the measures of recurrent plot.

2.1 Phase Space Reconstruction

The n state variables in time form vector (trajectory) in n-dimensional space (phase space). Trajectory in phase space represents all possible states of a system. Each state of the system corresponds to a specific point in the space phase. For N state variables N-dimensional phase space is created. However, very often it is not possible to observe more than one state variable of a system in the field of clinical medicine, because they are not known or it is difficult to measure them [12]. Nevertheless we can reconstruct the phase space trajectory from a single observation by using the Takens' theorem [12]. Takens' theorem is one of the most commonly used method for reconstruction of the phase space trajectory [12, 13]:

$$x_i = (y_i, y_{i+\tau}, \dots, y_{i+(m-1)\tau}),$$

where m is the embedding dimension, τ is the time delay and y_i is a single observation.

An optimal set of embedding dimension and time delay is important for reconstruction of phase space that fully describes the system dynamics [14]. One of the frequently used approaches for choosing time delay is the first minimum of mutual information [15]. The mutual information measures mutual dependence of two variables A and B and can be defined using entropy as [15]:

$$I(A, B) = H(A) + H(B) - H(A, B),$$

where $H(A)$ and $H(B)$ are the entropies and $H(A, B)$ is the joint entropy of A and B.

Entropy is generally, expressed in terms of a discrete set of probabilities s_i as [15]:

$$H(S) = -\sum_{i=1}^{n} p_s(s_i) \log p_s(s_i),$$

where $H(S)$ is entropy of system S, which consists of a set of possible messages $(s_1, s_2, ..., s_n)$ and the associated probabilities $p_s(s_1), p_s(s_2), ..., p_s(s_n)$.

False nearest neighbour method (FNN) is often used for optimal embedding dimension setting. Embedding dimension is a parameter which tells us how many dimensional phase space will be reconstructed. FNN method is based on the fact that choosing of low embedding dimension causes crossing of phase space trajectory[14]. We use modified FNN method using ratio of Euclidian distances of two neighbour's states in m and $m + 1$ dimensional space by Cao [16]:

$$\begin{aligned} a(i, m) &= \frac{\| y_i(m+1) - y_{n(i,m)}(m+1) \|}{\| y_i(m) - y_{n(i,m)}(m) \|}, \\ y &= 1, 2, \dots, N - m \cdot \tau, \end{aligned}$$

where $\| \ . \ \|$ is Euclidian distance, $y_i(m)$ is i-th reconstructed vector with embedding dimension m, $y_{n(i,m)}(m)$ is the nearest neighbor of $y_i(m)$.

The mean value of all $a(i, m)$ is introduced by Cao [16]:

$$E(m) = \frac{1}{N - m\tau} \cdot \sum_{i=1}^{N-m\tau} a(i, m).$$

The average $E(m)$ is dependent only on dimension m and delay τ. Finally, variable $E_1(m)$ is used to investigate variations from dimension m to dimension $m + 1$ [16]:

$$E_1(m) = \frac{E(m+1)}{E(m)}.$$

Value of $E_1(m)$ stops changing when dimension m is greater than the dimension of attractor m_0.

2.2 Recurrence Analysis

Recurrences in the nature were observed a long time ago [12], but the method of recurrence analysis was introduced just after the development of computer science in 1980s [19]. The basic tool of recurrence analysis is recurrence plot (RP) see Fig. 1 which can visualize recurrences in system dynamics using two-dimensional graph [19]. RP shows all moments in times when the phase space trajectory of the dynamical system visited roughly the same area in the phase space [17]. It is possible to visualise multidimensional phase space using two-dimensional graph (RP). We can demonstrate recurrences in a dynamical system, find interrelations between several systems or detect transitions between different states with RPs [6, 12, 17]. Mathematical expression of RP see equation 1.

$$R_{i,j} = \Theta(\epsilon - \|x_i - X_j\|), \text{ for } i, j = 1, 2, ..., N, \qquad (1)$$

where N is the number of states, ϵ is a threshold distance, $\|.\|$ a norm and $\Theta(.)$ the Heaviside function.

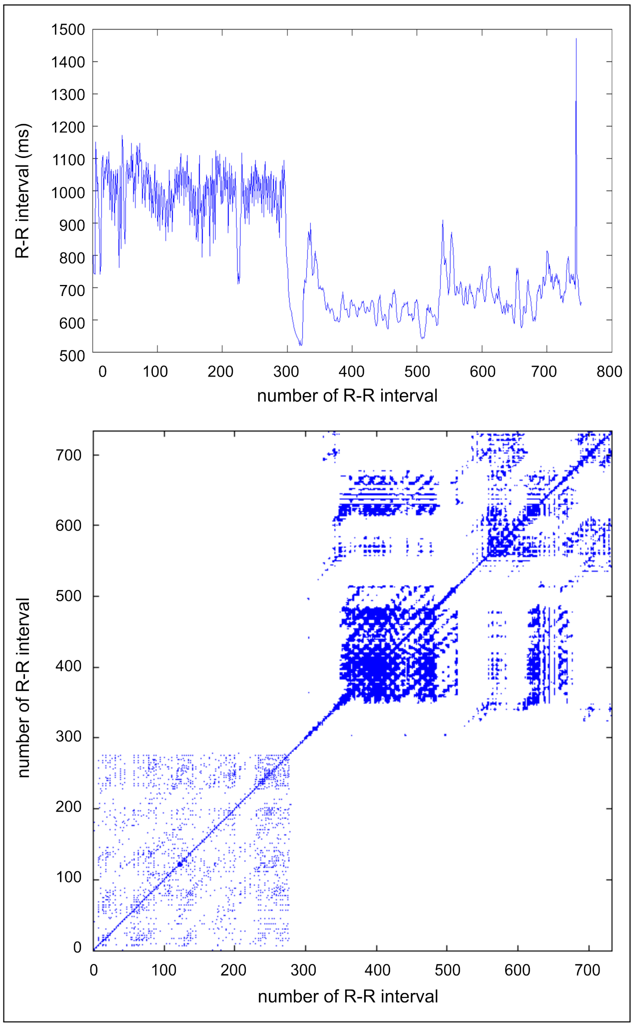

Figure 1: On the top figure, there is the input signal for recurrence analysis. The input signal is represented by the length of R-R intervals. On the bottom figure, there is the recurrence plot. On both figures we can distinguish two phases of the orthostatic test, resting in supine position and stand-up position, each for 5 minutes. In the second phase, after the verticalisation of the patient, we can recognise lower heart rate variability.

The most important parameter of recurrence analysis is the threshold distance. If the distance between two states on the phase space trajectory is smaller than a given threshold, the recurrence point in RP arises. Value of the point in recurrence matrix is one; otherwise it is zero [12]. This technique is called pair test and it is a pairwise test of all states (for N states we compute N^2 tests) [12].

With closer looking at the RP we can see single recurrence points, diagonal lines, vertical lines and horizontal lines. The structures formed by these elements are basis for recurrence quantification analysis (RQA). This analysis was introduced by Zbilut and Weber [19, 20, 21] to evaluate RPs quantitatively. RQA is a set of measures derived from diagonal and vertical structures of RP [17].

Diagonal line occurs when the trajectory visits the same area of the phase space at different times, it is when the system returns to the same or similar states at the different time points. Horizontal and vertical lines indicate how long the system stays in given state which does not change at all or just very slowly [17]. Single isolated recurrence point can represent a rare state [17].

The **Recurrence rate** is density of recurrence points in recurrence plot [17]:

$$RR = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j}.$$

This value corresponds to probability that a specific state will recur.

Determinism is a measure which represents the percentage of recurrence points which form diagonal lines [17]:

$$DET = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{i,j}^{N} R_{i,j}},$$

where $P(l)$ is the histogram of the lengths l of the diagonal lines.

Laminarity is the amount of recurrence points which form vertical lines. This variable is related with amount of laminar states [17]:

$$LAM = \frac{\sum_{v=v_{min}}^{N} v\ P(v)}{\sum_{v=1}^{N} v\ P(v)},$$

where $P(v)$ is the histogram of the lengths v of the vertical lines.

The average length of vertical lines is called **Trapping time** TT and it is related with laminarity time. This value contains information about frequency and length of laminar states [17]:

$$TT = \frac{\sum_{v=v_{min}}^{N} v\ P(v)}{\sum_{v=v_min}^{N} P(v)}.$$

The low value of LAM and TT indicate high complexity of a system [6].

The **Longest diagonal line** L_{MAX} is length of the longest diagonal line [17]:

$$L_{MAX} = max(l_i; i = 1 \ldots N_l).$$

Divergence DIV is the inverse of L_{MAX} and it is related with Kolmogorov-Sinai entropy [17].

$$DIV = \frac{1}{L_{MAX}}.$$

Other measures of RQA are average length of diagonal lines ($AVDL$), ratio ($RATIO$) between DET and RR, Shannon entropy ($ENTR$) and maximal length of vertical line (V_{MAX}) can be found in [17].

The advantages of RP include the ability to capture the chaotic properties of a system without a need of a long data series and the fact that it is relatively immune to noise and nonstationarity [17]. RQA is a sensitive tool for detecting any dynamic changes; however, it can be easily affected by inappropriate settings especially setting of the threshold distance where even a small change can dramatically affect the results of RQA [22, 23]. There are several methods for threshold distance setup [17, 22, 23]. One of the widely used method is setting of the threshold distance as a percentage of the maximum distance in phase space. Furthermore there is setting where values should not exceed 10% or 15% of the average or maximum distance in phase space [24, 18]. Quite frequently used method is called a fixed percentage of recurrent points. This means that we set a threshold distance value that guarantees the exact percentage of recurrent points [6]. Often is this value 1% [22, 24], but we can also find other values, such as 5% [6, 9, 25]. Selecting the optimal threshold distance appears to be a task for data mining [26].

3 Results

Our team currently evaluates possible applicability of the recurrence analysis in the field of medicine, specifically in neurology and cardiology. Our pilot study was focused on the early detection of the autonomic dysfunction in subclinical stages of diabetic cardiovascular neuropathy [9]. We found a significant increase in measures DET, LAM, $LMAX$ and TT in diabetic patients compared to control group. Based on the promising results of this pilot study, a bigger project on this topic has started, the aim of which is to confirm high sensitivity of the recurrence analysis in the early detection of diabetic autonomic disfunction and to identify typical values of RQA measures both in patients with diabetes and manifest autonomic disfunction and in the healthy control group. One of the goals is also the identification of optimal value of threshold distance for diagnosis of autonomic neuropathy. For now we reached promising results using a fixed percentage of recurrent points and our experimental setup using standard deviation:

$$\epsilon = \frac{0.02}{od},$$

where ϵ is threshold distance, od is a standard deviation of input signal.

This study is a joint project of Faculty of Biomedical Engineering CTU and University hospital Motol.

Next application of recurrence analysis in the field of neurology is the assesment of vasovagal syncopes. Results of a pilot project on this topic were recently presented by our team at the international conference in Slovakia [10]. We found reduced complexity of the heart rate control in young patients with vasovagal syncope using the recurrence analysis. RQA measurement showed significant differences in DET, $LMAX$, DIV and $ENTR$ see Fig. 2 However, future studies are needed in this particular field because of the limitations of our study - no gender match and small number of subject in groups.

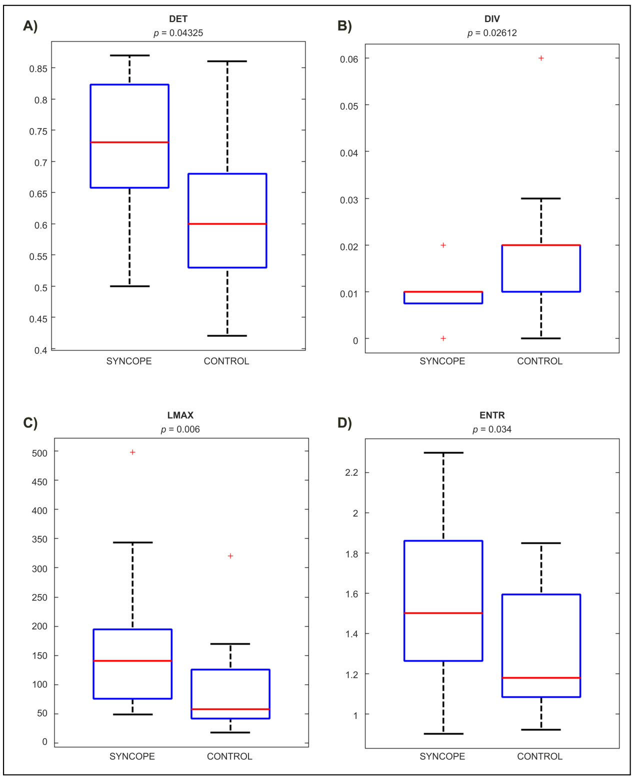

Figure 2: Boxplots[1] illustrating the comparison between patients with syncope and the control group. A – Determinism ($p = 0.04225$). B – Divergence ($p = 0.02612$). C – The length of the longest diagonal line ($p = 0.006$). D – The Shannon entropy ($p = 0.034$) [10].

Another cardiologic application of recurrence analysis that we are working on is the prediction of onset and termination of paroxysmal atrial fibrillation. It is a similar project to that of Mohhebi and Ghassemian [4] with the difference that we use real patient's Holter monitor data recorded at the Clinic of cardiology in the University hospital Motol. We hope, that our project will confirm capability of recurrence analysis in this field.

Currently a perspective applications of recurrence analysis focused on evaluation of psycho-physiological condition of pilots [27, 28]. When elevating pilots condition in stress situations, a change in physiological functions takes place, such as the change in heart rate, respiratory rate, blood pressure etc. The preliminary results show high potential of RQA for data evaluation of these type of measurement.

[1]Boxplots shows maximum, minimum, median, first and thirdquartile, cross points are outliers.

4 Conclusion

Detecting diseases in their early or even subclinical stages is important for increasing the probability of full recovery. Nonlinear methods seem to be useful for describing selected processes occurring in the living organism effectively. One of such methods is the recurrent analysis. Our pilot studies showed high potential of its use in medicine. However, there is a question about optimal setup of threshold distance and normative data are needed and results of pilot studies need to be confirmed in larger trials yet.

Acknowledgements

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant no. SGS14/169/OHK4/2T/17 and no. SGS14/170/OHK4/2T/17.

References

[1] Silipo R, Deco G, Vergassola R, Bartsch H. Dynamics Extraction in Multivariate Biomedical Time Series. Biol Cybern. 1998 Jul;79(1):15–27.

[2] Sharma V. Deterministic Chaos and Fractal Complexity in the Dynamics of Cardiovascular Behavior: Perspectives on a New Frontier. Open Cardiovasc Med J. 2009;3:110–123.

[3] Marwan N, Carmenromano M, Thiel M, Kurths J. Recurrence Plots for the Analysis of Complex Systems. Physics Reports. 2007;438(5-6):237–329. doi:10.1016/j.physrep.2006.11.001.

[4] Mohebbi M, Ghassemian H. Prediction of Paroxysmal Atrial Fibrillation Using Recurrence Plot-Based Features of the RR-interval Signal. Physiol. Meas. 2011;32:1147–1162.

[5] Naschitz JE, Rosner I, Shaviv N, Khorshidi I, Sundick S, Isseroff H et al. Assessment of Cardiovascular Reactivity by Fractal and Recurrence Quantification Analysis of Heart Rate and Pulse Transit Time. J Hum Hypertens. 2003 Feb;17(2):111–8.

[6] Javorka M, Trunkvalterova Z, Tonhajzerova I, Lazarova Z, Javorkova J, Javorka K. Recurrences in Heart Rate Dynamics are Changed in Patients with Diabetes Mellitus. Clinical Physiology and Functional Imaging. 2008;28(5):326-331. doi:10.1111/j.1475-097X.2008.00813.x.

[7] Acharya UR, Sree SV, Chattopadhyay S, Yu W, Ang PC. Application of Recurrence Quantification Analysis for the Automated Identification of Epileptic EEG Signals. Int J Neural Syst. 2011 Jun;21(3):199–211.

[8] Priano L, Saccomandi F, Mauro A, Guiot C. Non-linear Recurrence Analysis of NREM Human Sleep Microstructure Discloses Deterministic Oscillation Patterns Related to Sleep Stage Transitions and Sleep Maintenance. Conf Proc IEEE Eng Med Biol Soc. 2010;2010:4934–7.

[9] Nedělka T, Schlenker J, Riedlbauchová L, Mazanec R. Recurrence Quantification Analysis of Heart Rate Variability in Early Diagnosis of Diabetic Autonomic Neuropathy (in Czech). Cesk Slov Neurol N. 2012;75/108(6):721–728.

[10] Schlenker J, Nedělka T. Recurrence Analysis in Patients with Vasovagal Syncopes In: Measurement 2013 - 9th International Conference on Measurement (Proceedings). Bratislava: Institute of Measurement Science, 2013;145–148. ISBN 978-80-969-672-5-4.

[11] Rilley R, Balasubramaniam R, Turvey MT. Recurrence Quantification Analysis of Postural Fluctuations. Gait & Posture. 1999;9(1):65–78.

[12] Traut MH. MATLAB® Recipes for Earth Sciences. Potsdam: Springer, 2007, 288 s. ISBN 978-3-540-72748-4.

[13] Marwan N, Wessel N, Meyerfeldt U, Schirdewan A, Kurths J. Recurrence-Plot-Based Measures of Complexity and Their Application to Heart-Rate-Variability Data. Physical Review E. 2002;66(2), doi:10.1103/PhysRevE.66.026702.

[14] Kodera J, Tran VQ. Vizuální Nelineární Rekurentní Analýza a Její Aplikace na Český Akciový Trh. Politická ekonomie. Praha: VŠE, 2009;57(3). ISSN 0032-3233.

[15] Fraser A, Swinney H. Independent Coordinates for Strange Attractors from Mutual Information. Physical Review A. 1986;33(2):1134-1140. doi:10.1103/PhysRevA.33.1134.

[16] Cao L. Practical Method for Determining the Minimum Embedding Dimension of a Scalar Time Series. Physica D: Nonlinear Phenomena. 1997;110(1-2):43–50. doi:10.1016/S0167-2789(97)00118-8.

[17] Marwan N. Encounters with Neighbours: Current Developments of Concepts Based on Recurrence Plots and Their Applications [PhD thesis]. Potsdam: University of Potsdam; 2003. Available from: http://www.recurrence-plot.tk /furtherreading.php.

[18] Sun R, Wang Y. Predicting Termination of Atrial Fibrillation Based on the Structure and Quantification of the Recurrence Plot. Medical Engineering. 2008;30(9):1105–1111. doi:10.1016/j.medengphy.2008.01.008

[19] Marwan N. A Historical Review of Recurrence Plots. The European Physical Journal Special Topics. 2008;64(1):3–12. doi:10.1140/epjst/e2008-00829-1

[20] Zbilut JP, Webber CL. Embeddings and Delays as Derived from Quantification of Recurrence Plots. Physics Letters A. 1992;171(3-4):199–203.

[21] Webbber CL, Zbilut JP. Dynamical Assessment of Physiological Systems and States Using Recurrence Plot Strategies. Journal of Applied Physiology. 1994;76(2):965–973.

[22] Ding H, Crozier S, Wilson S. Optimization of Euclidean Distance Threshold in the Application of Recurrence Quantification Analysis to Heart Rate Variability Studies. Chaos, Solitons & Fractals. 2008;38(5):1457–1467.

[23] Garcia-Gonzales MA, Fernandez-Chimeno M, Ramos-Castro J. Errors in the Estimation of Approximate Entropy and Other Recurrence-Plot-Derived Indices Due to the Finite Resolution of RR Time Series. IEEE Transactions on Biomedical Engineering. 2009;56(2):345–351. doi:10.1109/TBME.2008.2005951.

[24] Schinkel S, Dimigen O, Marwan N. Selection of Recurrence Threshold for Signal Detection. The European Physical Journal Special Topics: Special Topics. 2008;164(1):45–53. Available from: http://www.springerlink.com/index/10.1140 /epjst/e2008-00833-5

[25] Javorka M, Turianikova Z, Tonhajzerova I, Javorka K, Baumert M. The Effect of Orthostasis on Recurrence Quantification Analysis of Heart Rate and Blood Pressure Dynamics. Physiological Measurement. 2009;30(1):29–41.

[26] Kalina J, D Tebbens J, Schlenker A. Robustness of High-Dimensional Data Mining. ITAT 2014, Part II, pp. 53—60.

[27] Kaľavský P., Socha L.. Výskum tréningových metód pilotov s využitím leteckých. In: FLYIN: aviation industry exhibition : FLYIN letecká konference VŠB-TUO 2013 : sborník příspěvků. 1st. ed. Ostrava: Vysoká škola báňská - Technická univerzita Ostrava, 2013. ISBN 978-80-248-3091-9.

[28] Socha V., Szabo S., Socha L., Kutílek P. 2014, "Evaluation of the variability of respiratory rate as a marker of stress changes", Proceedings of 18th International Conference. Transport Means, KAUNAS UNIVERSITY OF TECHNOLOGY, 23-24 October 2014 (in press)

