Example of nonlinear dynamical parameters for machine learning-based speech pathology classifier

To cite this article: Aleksey Kharitonov et al 2019 J. Phys.: Conf. Ser. 1176 032037

View the article online for updates and enhancements.
Example of nonlinear dynamical parameters for machine learning-based speech pathology classifier

Aleksey Kharitonov1,*, Vladimir Antonets1,3, Kirill Aleshin2, Kirill Gromov1

1Department of social sciences, Lobachevsky State University of Nizhny Novgorod
2Faculty of Radiophysics, Lobachevsky State University of Nizhny Novgorod
3Federal Research Center Institute of Applied Physics of the Russian Academy of Sciences

*Corresponding author e-mail: 4alekseykharitonov@gmail.com

Abstract. In light of muscle coordinative structures might be defined as time-invariant dynamical systems that underlie an action’s form, we find out if the ability of dynamical system to grade the levels of complexity according to external conditions is a feature of speech without sound disorders. Results revealed that such metrics of speech signal as correlation dimension and sample entropy as indicators of complexity vary widely under environmental conditions (acoustic fatigue) for participants without speech sound disorders, unlike for participants, who admitted the ones (sibilant /s/ considered). Supposedly, mentioned parameters may be of interest to machine learning-based speech pathology classifiers.

1. Introduction

The issue of the physical space in which the motor task is planned becomes particularly complex, since the distal space can be defined either by articulatory positions, or by spectral properties of the speech signal [1-3]; each movement parameter defines a dimension, which one can think of as an axis in a potentially high-dimensional parameter space. From this point of view, fricative sounds as sounds with evident spectral properties are illustrative. The focus on /s/ in this work was motivated by following considerations: /s/ provides insight into laryngeal, oral, and combined laryngeal-oral activities; /s/ requires both laryngeal abduction for devoicing and a tongue-tip constriction. A useful feature of oral airflow data is that they permit noninvasive assessment of both articulatory actions: airflow rises with vocal fold abduction and decreases with tongue-tip constrictions. The fact the fricatives were most heavily affected demonstrates very high precisions for corresponding articulatory effort.

Here is a typical reaction of a person without /s/ sound disorder (fig.1). There’s mismatch between /s/ spectrum in case of acoustic fatigue (bottom) and without it (upper); the deformation lies in 3-7kHz frequency band, completely defining the corresponding sound. You may find detailed analysis of spectral properties of fricative consonants, including with regard to locations of the important poles and zeros in the spectra, in [4].
When a parameter specifying the configuration or state of an articulatory structure is manipulated through a range of values, some acoustic parameter describing the resulting sound often changes in a non-monotonic fashion. In particular, there appear to be ranges of the articulatory parameter for which there is very little change in the acoustic parameter and other ranges where the acoustic parameter is more sensitive to changes in articulation. The model articulators are controlled by transforming the tract-variable dynamical system into model articulator coordinates. This coordinate transformation creates a set of gesture-specific and articulatory posture-specific coupling functions among the articulators. These functions create a dynamical system at the articulatory level whose modal, cooperative behaviors allow them to flexibly and autonomously attain speech relevant goals; though, it’s not guaranteed for participants with speech disorders. In terms of /s/ spectrum deformation it means the spectrum in case of different conditions remains mostly unchanged.

2. Methods

2.1. Correlation dimension

The technique of nonlinear time series analysis is used for speaker identification [5], classification of emotional speech [6,7], sleepiness detection [8], studies of pathology [9] etc. In analyzing the time series, the main task is to reconstruct the model/features of underlying dynamical system. According to Takens theorem, [10], the acceptable description of dynamical system phase space can be obtained, constructing k–dimensional delay vectors from time series. In other words, if the embedding dimension is adequate, then a «natural» attractor of dynamical system and time series reconstructed-attractor (pseudoattractor) are topologically equivalent and have the same quantitative characteristics. Nowadays, the most popular algorithm to estimate the correlation dimension is Grassberger-Procaccia algorithm [11]. The correlation dimension gives an idea of the complexity of the dynamics and the attractor-more complex systems have a higher correlation dimension.

2.2. Sample Entropy

Recently, the application of entropy measures to investigate signal complexity and irregularity in human data has become quite popular [12]. Regularity and complexity statistics such ApEn and SampEn are measures without the shortcomings that correlation dimension and other metrics of nonlinear time series analysis techniques possess. ApEn and SampEn can effectively discriminate both stochastic processes and noisy deterministic data sets in instances where measures such as spectral and autocorrelation analyses exhibit minimal distinctions [13]. SampEn was introduced as an improvement of ApEn where self-matches are excluded, i.e., vectors are not compared to themselves.
3. Results
In statistics below, *Seria means the level of acoustic fatigue (from 1 to 4: no fatigue-slight fatigue-strong fatigue-no fatigue (to evaluate the inertia)).

In our first analysis we entered Group (3 levels: Male, Female, Focus), and Seria (4 levels: no.1 – no.4) into Factorial ANOVA. For each seria (no.1-no.4), correlation dimension value for each repetition was calculated. The main effect of Group (F(10,328)=1.95, p<0.05), Seria (F(15,453)=18.21, p<0.05) and interaction between Group and Seria (F(30,658)=4.5602, p<0.05) reached statistical significance. This result is displayed in fig.2.

![Figure 2. Statistics for Correlation Dimension](image)

Figure 2. Statistics for Correlation Dimension

In our next analysis for SamEn values we added the effect of Set_r (9 levels, 0.2-1.8 at intervals of 0.2). First of all, we were interested in full ((Group//Seria//Set_r ) interaction (F(48, 1511)=11.617, p<0.05). As can be seen from the fig.3, the value of SamEn for Focus group changed slightly (decreased) for all r, whereas the corresponding values for M-F controls significantly decreased, again. Taking into account the fact that a lower value of SampEn also indicates more self-similarity in the time series, it’s possible to link these results to spectral deformation, mentioned above (see fig.1).

![Figure 3. Statistics for Sample Entropy](image)

Figure 3. Statistics for Sample Entropy

4. Conclusion
However, despite the high degree of generality, the obtained values lead to the main conclusion of this investigation: the complexity of underlying motor control system in subjects without speech sound disorders varies depending on acoustic fatigue, unlike in subjects with ones; this can be interpreted in
a way that there’s no required degrees of freedom to enable the variability of articulatory settings. Apparently, the ability of dynamical system to grade the complexity of its inner structure according to external conditions, and provide the range of control regimes, is a feature of «healthy» systems.

Acknowledgments
Research reported in this publication was supported by IAP RAS fundamental research program (2014-2017), theme № 0035-2014-0008 «Acoustic and optical methods for dynamics of physiological processes in biological tissues»

References
[1] Stevens K. N. On the quantal nature of speech, Journal of Phonetics 17 (1989), pp. 3–46.
[2] Savariaux C., Perrier P., and Orliaguet J.-P. Compensation strategies for the perturbation of the rounded vowel [u] using a lip-tube: A study of the control space in speech production. J. Acoust. Soc. Am. 98 (1995), pp. 2428-2442.
[3] Guenther, F.H., Hampson, M., and Johnson, D. A theoretical investigation of reference frames for the planning of speech movements. Psychological Review, 105 (1998), pp. 611–633
[4] Hughes, G. W., Halle M. (1956). Spectral properties of fricative consonants. J. Acoust. Soc. Am. 28(2) (1956), pp. 303–310.
[5] Petry A., Barone D.A.C. Speaker identification using nonlinear dynamical features. Chaos, Solitons and Fractals 13 (2002), pp. 221-231.
[6] Henríquez P., Alonso J.B., Ferrer M.A., Travieso C.M., Orozco-Arroyave J.R. Application of Nonlinear Dynamics Characterization to Emotional Speech. In: Travieso-González CM, Alonso-Hernández JB, editors. Advances in Nonlinear Speech Processing. Springer Berlin Heidelberg (2011), pp. 127–136.
[7] Giannakopoulos T., Pikrakis A., Theodoridis S.A. Dimensional Approach to Emotion Recognition of Speech from Movies. In: IEEE Int.Conf. on Acoustic, Speech and Signal Proc. (2009), pp. 65-68
[8] Krajewski J., Schnieder S., Sommer D., Batliner A., Schuller B. Applying multiple classifiers and non-linear dynamics features for detecting sleepiness from speech Neurocomputing 84 (2012), pp. 65–75.
[9] Henríquez P., Alonso J.B., Ferrer M.A., Travieso C.M., Godino-Llorente J.I., Diaz-de-Maria F., Characterization of healthy and pathological voice through measures based on nonlinear dynamics, IEEETrans.AudioSpeechLang.Proc. 17(6) (2009), pp. 1186–1195.
[10] Takens F. Detecting strange attractors in turbulence. In D. A. Rand and L.-S. Young. Dynamical Systems and Turbulence, Lecture Notes in Mathematics, vol. 898. Springer-Verlag. 1981, pp. 366-381
[11] Grassberger P., Procaccia I. Characterization of Strange Attractors. Phys. Rev. Lett., vol.50, № 5 (1983), p. 346 – 349.
[12] Yentes J. M., Hunt N., Schmid K. K., Kaipust J. P., McGrath D., and Stergiou N. The appropriate use of approximate entropy and sample entropy with short data sets. Annals of Biomedical Engineering, Vol. 41, No. 2 (2013), pp. 349–365.
[13] Pincus S. M. (2001) Assessing serial irregularity and its implications for health. Annals New York Academy of Sciences, 954, pp. 245-267