Applying recurrence quantification and spectral analysis to represent nasalization in speech signals

Yousif A. El-Imam & Ahmad S. Elwakil

Both authors with the:
Dept. of Electrical and Computer Engineering,
College of Engineering,
The University of Sharjah,
Sharjah B. O. Box 27272,
United Arab Emirates.
Tel. 971 6 5050964
Fax 971 6 5585191

Abstract

Acoustic signals of vowels preceding nasal sounds were examined for nasalization by using nonlinear time-series analysis to represent the speech signal in a multidimensional embedding space. Recurrence quantification analysis (RQA), originally developed to study the behavior of nonlinear dynamical systems, were used on the speech signals to indicate their deterministic chaotic behavior and to show the differences between nasalized and oral vowels. The results obtained from RQA of the speech data are compared to the results obtained by using traditional spectral analysis of the same data. Characterizing of speech sounds, such as nasalized sounds, by acoustic or RQA parameters have potential applications in speech recognition because such parameters can be utilized in detection algorithms as acoustic clues for building robust speech recognition systems and in speech synthesis to build high quality synthesizers.

Keywords:
Speech processing, nonlinear time-series analysis, deterministic chaotic systems, speech synthesis and recognition, Arabic speech.

1-Introduction

The human speech signal is complex and highly nonstationary. This is due to the complex dynamic of the articulations involved in the speech production process. Previous work on the analysis of the speech signal resolved the nonstationarity problem of speech by windowing the speech signal and using short-time spectral analysis [1]. Recent work on modeling the speech production process using nonlinear time-series analysis has shown that the stationary articulated human voice has a low-dimensional dynamical origin [2]. The dynamics of speech production inside a maintainable phoneme are close to a low-order deterministic chaotic system. Therefore, by using nonlinear time series analysis to reconstruct the speech signal in an embedding space, the issue of nonstationarity can be resolved inside a sustainable sound such as a vowel. The properties of the speech signal that mimic structures present in deterministic chaotic systems are used to effectively reduce contamination caused by additive noise to human voice [3]. In this article, we apply recurrence quantification analysis (RQA) [4] to study nasalization of vowels, which is an important universal, phonetic variation phenomenon of human speech with applications in speech synthesis and recognition, [5]. Nasalization was previously studied by spectral analysis based on windowing the speech signal and using short-term spectral analysis [6, 7]. Nonlinear time series analysis confirms the findings of short-term spectral analysis.

2-Nasalization of vowels

For most languages, a nasalized vowel is produced when an oral opening is coupled with velopharyngeal opening. This happens when a nasal consonant occurs next to a vowel. There are production alterations in the vocal tract configuration in moving from the vowel to the nasal consonant articulation. Acoustically, the alterations from an oral to a nasalized vowel are manifested as 1) A decrease in the amplitudes of the second and higher formants over the 30-40 ms before the beginning of the alveolar or velar nasal closure [7]. 2) A noticeable spectral spread in the region of the low formants (noticeably in F1 & F2) due to vowel nasalization [6] and 3) shifting in the vowel formant values and possible appearance of additional poles and zeros in the transfer function that represents the combined vocal tract. In the context of speech synthesis, a perceptual study of nasalization aiming to improve the quality of synthesized speech was introduced in [8].

3-Short-term spectral analysis of oral and nasal vowels

Nasalized short Arabic vowels (high front short vowel /i/, low central /a/ and high back rounded /u/) were digitized and recorded in two phonetic contexts followed by bilabial nasal /m/ and alveolar nasal /n/. To provide a comparison platform, the same three vowels were digitized and recorded in oral articulation contexts. Short-term spectral analyses were conducted on nasalized and oral vowels. The characteristic spectral phenomenal changes discussed in section 2 are noticed in all these vowels. For a nasalized high front vowel /i/, the average amplitude drops in F2 and F3 are around 18 dB and 14 dB and the average spectral spread increases in F1 and F2 are around 16 Hz and 9 Hz. Similar short-term spectral analysis were carried out on samples of the other two short vowels /a/ and /u/. The same spectral trends show on nasalized versions of these vowels. The spectral trends are also visible in Figure 1 (b) which is the spectrum of a nasalized vowel /a/ as opposed to Fig. 1 (a) which is the spectrum of an oral /a/. Figure 1(b) also shows the shifting in the formant values.

4-Recurrence Quantification analysis of oral and nasal vowels

In [4], Zbilut and Webber proposed a set of quantifying measures suitable for analysis of a time series and derived from recurrence plots. Among the important measures is the 'percentage determinism', which can be used to measure how chaotic a time series is. The embedding dimension and spatio-temporal entropy (STE) are also important characteristics of a time-series. Table 1 summarizes the average values of these parameters performed for the same oral and nazal vowels. It is clear form Table 1 that the percentage determinism is less for all nazal vowels, which is confirmed by the spreading in the frequency spectrum. It is well known that a chaotic signal has a broadband noise-like spectrum. What is interesting to note is that the embedding dimension of all data files was between four and five, as shown in Table 1. This means that such data sets can be fully described by a system of fourth- or fifth- order nonlinear differential equations, which exhibit chaos. By using the estimated embedding dimension along with the necessary delay calculated from the autocorrelation function [4], the chaotic attractor shown in Fig. 2(a) was constructed from the data of the nazalized high-front /i/. Surprising enough, this attractor resembles that reported in [9], and shown in Fig. 2(b) for clarity, despite the large difference in the number of data points. In particular, the main features of these two attractors include a fast transition along the diagonal of the attractor, associated with a very small time constant, along with slow wiggling-type trajectories alternating around the diagonal and associated with a large time constant. Indeed, the model given in [9] is fourth-order with a hysteresis-type nonlinearity.

5-Conclusions

Previous work on vowel nasalization has focused on short-term spectral analysis in the frequency domain. By using nonlinear time-series analysis it is possible to calculate some recurrence quantification analysis parameters. These parameters show that speech signals of sustainable sounds, such the vowels, are characterized by deterministic chaotic behavior and that the signal of a nasalized vowel is indeed different from that of an oral vowel. Because of the coupling of the oral and nasal tracts during the production of a nasal vowel, the acoustic signal looses some of its power. The airflow in the combined tract becomes more turbulent, the lower formants become wider and the combined vocal tract will have different dynamics than when air flows through an oral tract alone. These characteristic changes show in the recurrence quantification analysis measurements that indicate deterministic chaos and that the nasalized signal data is less deterministic than the oral signal data.

Vowel type	**Oral/nasal pair**	**Vowel Context**							
		Average RQA of vowel preceding nasal /m/				**Average RQA of vowel preceding nasal /n/**			
		Time delay	**Embedding Dimension**	**STE**	**% max. Determinism**	**Time delay**	**Embedding Dimension**	**STE**	**% max. Determinism**
High front /i/	Oral	2	5	65	71	2	5	65	75
	nasalized	7	5	63	60	6	4	48	65
Central /a/	Oral	4	4	50	65	2	4	63	70
	nasalized	5	5	48	60	7	3	44	63
Back rounded /u/	Oral	4	4	55	62	3	4	67	70
	nasalized	6	5	42	58	7	4	49	65

Table 1: Average recurrence quantification analysis measurements

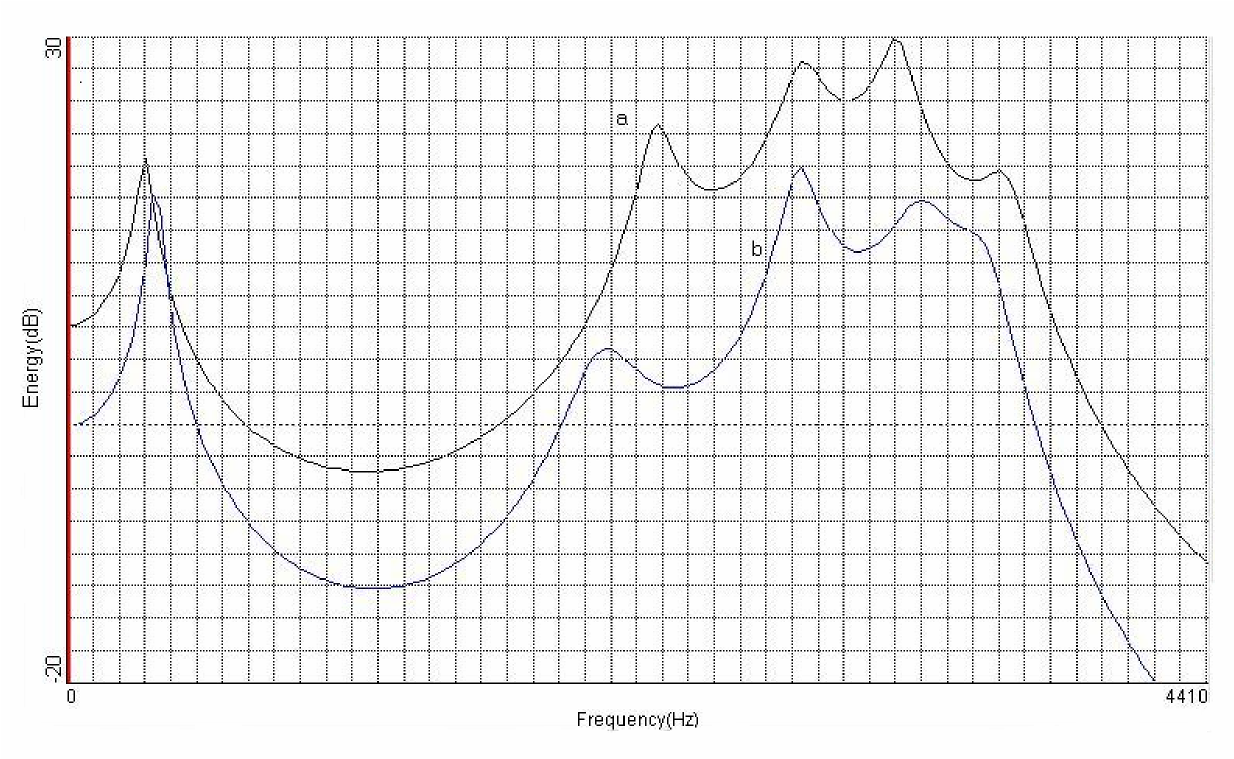

Figure 1: Spectrums of oral and nasalized vowel /a/: (a) oral and (b) nasalized

(a)

(b)

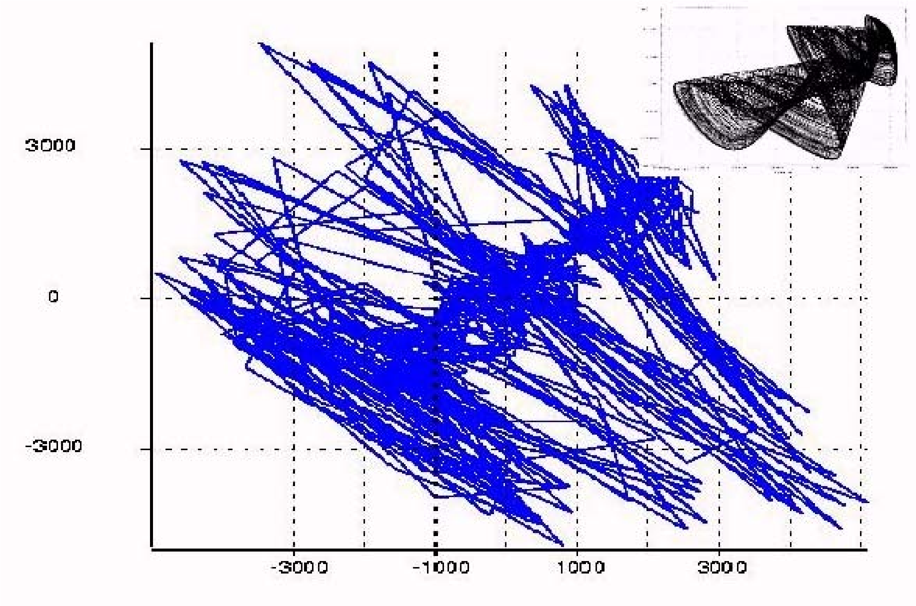

Figure 2
(a) Attractor for modeling of a vowel sound /i/ (b) Attractor given in [9] for a 4th order nonlinear circuit with hystresis

References

1. L. R. Rabiner and R. W. Schafer, Digital Signal Processing of Speech Signals, Prentice –Hall, New York, ISBN 0-13-213603-1, 1978.

2. H. Herzel and J. Wendler, “ Evidence of chaos in phonatory signals,” in Proceedings of EUROSPEECH (European Speech Communication Association, Genova, Italy), pp. 263-266 , 1991.

3. R. Hegger, H. Kantz and L. Matassini, “ Noise Reduction for Human Speech Signals by Local Projections in Embedding Spaces.” IEEE Trans. On Circuits and Systems, vol. 48, no. 12, 2001.

4. J. P. Zbilut and C. L. Webber, “Embedding and delays as derived from quantification of recurrence plots.” Physics Letters A, 1992, vol. 171, pp. 199-203.

5. Chen, M. Y., (2000), “Acoustic analysis of simple vowels preceding a nasal in Standard Chinese”, Journal of Phonetics, 28, pp. 43-67, Academic Press.

6. S. Maeda, “Acoustics of Vowel nasalization and Articulatory Shifts in French Nasal Vowels.” In Nasals, nasalization, and the velum: Phonetic and Phonology V (M. K. Huffman & R A. Krakow, editors), pp. 147-167. New York: Academic Press, 1993.

7. K. N. Stevens, “Acoustic Phonetics.” Boston: The MIT Press, 1998.

8. Y. A. El-Imam, “An unrestricted vocabulary Arabic speech synthesis system. ” IEEE Transaction on Acoustics, Speech and Signal Processing, Vol. 37, No. 12, 1989.

9. A. S. Elwakil and M. P. Kennedy, “A low voltage low power chaotic oscillator derived from a relaxation oscillator. ” Microelectronics journal, Vol. 31, 2000, pp 459-468.