Chaos tool implementation for non-singer and singer voice comparison. (Preliminary study).

**ME Dajer, JC Pereira, CD Maciel**
Department of Electric Engineering. School of Engineering of São Carlos, University of São Paulo, São Carlos, Brazil.
Av. Trabalhador São-Carlesnse, 400. CEP 13566-590. São Carlos. SP. Brazil.
E-mail: medajer@sc.usp.br, pereira@sel.eesc.usp.br

**Abstract.** Voice waveform is linked to the stretch, shorten, widen or constrict vocal tract. The articulation effects of the singer's vocal tract modify the voice acoustical characteristics and differ from the non-singer voices. In the last decades, Chaos Theory has shown the possibility to explore the dynamic nature of voice signals from a different point of view. The purpose of this paper is to apply the chaos technique of phase space reconstruction to analyze non-singers and singer voices in order to explore the signal nonlinear dynamic, and correlate them with traditional acoustic parameters. Eight voice samples of sustained vowel /i/ from non-singers and eight from singers were analyzed with “ANL” software. The samples were also acoustically analyzed with “Análise de Voz 5.0” in order to extract acoustic perturbation measures jitter and shimmer, and the coefficient of excess – (EX). The results showed different visual patterns for the two groups correlated with different jitter, shimmer, and coefficient of excess values. We conclude that these results clearly indicate the potential of phase space reconstruction technique for analysis and comparison of non-singers and singer voices. They also show a promising tool for training voices application.

**Introduction**
Human voice is one of the principal means of communication. This complex mechanism allows us to produce from primitive sounds like crying, screaming and laughing to more evolved and sophisticated communication sounds as talking and singing. All those different voice manifestations are acoustic signals with significant information about some individual characteristics, and have historically been matter of science interest.

Vowels are one kind of voiced signal; they are produced by vocal folds vibration (glottal excitation) and the vocal tract filtering. The vocal tract shape modifies the glottal excitation. If the vocal tract straitens or even temporarily closes, the airflow produces the consonant sound.

Different factors such as position, length and shape of the vocal tract are essential for vowel sounds production.[1]. The length of the vocal tract (distance from the glottis to the lips) can be modified by raising or lowering the larynx. The articulators, lips, tongue, teeth, etc. also affect directly the vocal emission. For example, lips posture can cause fluctuations to the vocal tract length, protruding the lips for lengthen and smiling for reduce it. This vocal tract transfer function can be characterized by formants, which are resonant peaks in the spectrum. The adjustment of any articulators generally affects the frequencies of formants. According to Sundberg the first three formants define the vowel
type and the fourth and the fifth formant are relevant for perceiving the voice timbre, which is the personal component of a voice [1].

Although normal voice and singing voice signals have the same production physiology, there are important differences between them.

The main six acoustic differences between speech and singing voice are: 1) the signal voiced and unvoiced ratio -voiced sounds are much increased in singing voice; 2) the “vibrato”- periodic modulation of the phonation frequency that occurs only in singing voice; 3) the voice dynamic range and loudness average -that is greater in singing than in speech; 4) the Singer's formant -one peak at approximately 2-3 kHz with a great magnitude value in the voice spectrum; 5) modification of vowels - in singing voice, in order to gain musical expression and loudness vowels, articulation are slightly modified; and 6) fundamental frequency -for talking, the frequency range is very small compared to singing; the fundamental frequency variations.

Traditionally, voice signal has been modeled as a linear process and acoustic analysis tools are based in linear system theory. The acoustic parameters evaluate perturbation or noise contents in the voice signal.

The classical perturbation parameters evaluate jitter (fundamental frequency variation), and shimmer (amplitude variation). Two parameters used to determine the voice signal noise quantity are the deterministic Harmonic to Noise Ratio (HNR) and the Coefficient of Excess (EX) that evaluate the noise from a statistical point of view [13].

Although, these linear model tools have been performed over the years, they are based on the assumption that voice is a linear phenomenon. But, voice production is a complex mechanism that involves different variables and shows several nonlinearities.

In “Chaos in voice research”, Hawkshaw, Sataloff and Bhatia have stated that the application of chaos to the voice analysis has already proven to be an exciting and promising approach. They summarized relevant studies about existence of chaos in human voice production and nonlinear dynamic analysis of voice signals [2].

Recently, several papers have shown the application of different nonlinear analysis tools to evaluate phonation with non-periodic segments or pathologic voice signals [3],[4] and [5].

Based on the nonlinearities, human voice can be described by a number of observable output states, and from that point, be used in the construction of a state space description of the system behaviour. Voice signal, as a time series data, makes possible the study of a system underlying dynamics, and provide the necessary information to obtain a reconstruction of the state space behaviour of the system[6].

The purpose of this paper is to apply the chaos technique of phase space reconstruction to analyze non-singers and singer voices in order to explore the signal nonlinear dynamic, and correlate them to traditional acoustic parameters.

Materials and Methods

1.1. Data Base:
Eight voice signal samples of sustained vowel /i/ from non-singers and eight from singers of Brazilian Portuguese, from bioengineering voice data base, were used for this work. Voice signals were recorded at 22,050 Hz sampling rate and processed in a personal computer Microsoft Windows XP Professional- Version 2002. Service Pack 2. AMD Athlon ™ XP 2800+2.08 GHz, 512 MB RAM.

1.2. Methods:

2.2.1 ANL. Phase space reconstruction technique.
In order to describe the nonlinear dynamic characteristics of voice signals, a sustained vowel data set was analyzed with “ANL” software, (“Análise Não Linear”). [7] “ANL”, was developed from Tisean
Package [8], and was run on Mathlab 7.0. The ANL is based on the phase-space reconstruction technique and it represents the vocal folds vibration as an orbit trajectory in phase-space with time evolution.

Procedure:
The ANL presents a voice sample in the traditional acoustic representation, in time domain and in frequency domain in order to choose any stationary part of the signal (figure 1). Subsequently, for a time series: x(ti), ti= t0+ iΔt, (i= 1, 2, …, N), sampled at the time interval Δt=1/fs, a phase space can be reconstructed with the time delay vector X(ti)= {x(ti), x(ti-τ), …, x(ti-(m-1)τ)}, where τ is the time delay and m is the embedding dimension [8]. Figure 2 shows the trajectory in the reconstructed (x(t); x(t+τ)) phase space of a voice signal of sustained vowel /e/, and the time delay τ was estimated as 7Δt using the mutual information method [9].

![Figure 1](image1.png)
**Figure 1.** Voice signal in time domain, and in frequency domain. Small stationary part of the signal.

![Figure 2](image2.png)
**Figure 2.** Curve of mutual information versus time delay τ.

![Figure 3](image3.png)
**Figure 3.** Reconstructed phase space for a time series x(ti) with the time delay technique of sustained vowel.

It is important for time delay that delayed versions of the time series have as little information redundancy as possible. Because of each signal particular dynamics, Fraser and Swinner [9] proposed that an effective criterion for choosing the proper time delay τ, that ensures that the variables be generally independent, is the first minimum value of the curve of mutual information versus time delay τ, represented in figure 3.

2.2.2 Traditional acoustic analysis.
The purpose of the traditional acoustic analysis is to extract some information from the voice signal. It is based on the principle that a voice signal contains fluctuations both in frequency and amplitude. Traditional acoustic analysis was performed by “Análise de Voz 5.0” [10].

Jitter refers to a short-term (cycle-to-cycle) perturbation in the fundamental frequency of the voice. Some of the early investigators [11] displayed speech waveforms oscillographically and concluded that two periods were not exactly alike. Shimmer was then proposed as a companion word for “amplitude-jitter”; a short-term (cycle-to-cycle) perturbation in amplitude [12].

The amplitude distribution of the residue signal would be useful for a statistical measure of the signal-to-noise ratio. The shape of this distribution may be quantified by a statistical measure called the coefficient of excess, (EX)[13].

3 Results

3.1. ANL Phase space reconstruction technique.
Singer and non-singer voice samples analyzed by means of phase space reconstruction technique with “ANL” showed differential visual patterns for each group.

To determinate the visual pattern characteristics three kind of orbits dynamic behavior were observed: number of loops; attractor course regularity and attractor trajectories distribution (divergence and convergence of attractor orbits trajectories).

For non-singer voice signals, phase space reconstruction for sustained vowel /i/ presented a visual pattern of a roughly single loop.

A high tongue position during production of vowel /i/ contributes to amplify a specific range region, that is, a dominant higher harmonic component, that usually covers the others harmonic frequencies, and appears in the phase space reconstruction pattern as a single loop orbit. As a function of the proportional relationship among the signal harmonic components, the orbit presented irregular trajectories with small loops in the reconstructed phase space.

Figure 4 shows four examples of phase space reconstruction for vowel /i/ with different time delay $\tau$ for each voice signal, according to Fraser and Swinner criterion. The time delay $\tau$ was estimated as $17\Delta t$, $17\Delta t$, $27\Delta t$ and $20\Delta t$, respectively. The four visual patterns showed a single loop, irregular trajectories and few small loops.

For the attractor course regularity and trajectories distribution, the patterns showed irregular characteristic and rough areas of the trajectories correlated with Ex effect, as figure 5A shows.

For voice signals samples with different shimmer values, patterns showed a divergent area of orbit trajectories, as well as the disperse characteristic of the attractor orbits remarked with a circle in figure 5B. Visual patterns that presented more orbit dispersion belonged to voices with higher values of shimmer.

Voice samples with curling trajectories (similar to the helix shape) presented higher values of jitter. The curling behavior when plotted in a two dimension pattern appears as convergent points in the trajectories.
In the “ANL” analysis, attractor trajectories converge into some specific region or, also, in different regions of it. Therefore, the lower the value for jitter, the smaller the number of curls. Consequently, the two dimension pattern presents a small number of convergent points, shown as small circles in figure 5C.

For singer voice samples, phase space reconstruction for sustained vowel /i/ displayed different visual pattern variability, from a single regular loop to an irregular and complex trajectory pattern.

The two visual patterns showed in figure 6 are related to singer voice signals with higher frequencies, (345 Hz and 350 Hz, respectively) and perceptually clean.

Although the harmonic components are present in the glottal pulse, the vocal tract equalization reinforces mainly the fundamental frequency, producing a single regular trajectory loop.

Figure 7 presents the visual patterns of four singer voice samples. In those signals, the vocal tract equalization established a medium proportional ratio gain among the component frequencies, producing visual patterns that look closer to the non-singer vowel /i/ patterns.
Figure 7. Reconstructed phase space for a time series $x(t_i)$ with the time delay technique of four sustained vowel /i/.

The most complex visual patterns for /i/ vowel are shown in figure 8. In those cases, the vocal tract equalization furnishes a high proportional ratio gain among the component frequencies and, consequently, a more complex visual pattern with several superposed and different sized loops.

Figure 8. Reconstructed phase space for a time series $x(t_i)$ with the time-delay technique of sustained vowel.

Dynamical orbits behavior, such as attractor course regularity and attractor trajectories distribution (divergence and convergence of attractor orbits trajectories) remained with the same characteristics indicated for as the non-singer voices.

3.2. Traditional acoustic analysis
Voice signals traditional perturbation analysis with “Análise de Voz 5.0” shown for non-singer voices varied from 0.25% to 4.97% for Jitter. For shimmer values, they ranged between 2.09% and 7.74%. And for (Ex) Coefficient of Excess, the values ranged from 17.09 to 37.5.

For singer voices values varied from 0.36% to 2.38% for Jitter. For shimmer values were from 0.65% to 7.29%; and for (Ex) Coefficient of Excess, the values ranged from 6.7 to 35.5.
4. Conclusions

In this paper we try to look into voice as a dynamical signal and consequently we explore a new processing technique for non-singer and singer voice analysis. We also try to present the practical application and advantages of dynamical analysis in combination with traditional methods. So we believe that chaos tools, as a phase space reconstruction technique, may help us review many of the voice signal dynamic properties in visual patterns.

The ANL and the phase space reconstruction have shown a potential value to describe non-singer and singer voice signals. The phase space depicts the vowel pattern in a dynamical way. Classical acoustic parameters have their counterpart in the patterns, also in a dynamical manner.

In this paper, the most complex characteristics of the singer voice signals have been outlined in comparison to non-singer voices. This technique allows us to visualize the differential dynamic between speech voice and singing voice.

But as a preliminary study, the relationship of this complexity with musical parameters remains to be analyzed and seems to be an exciting and promising field to explore.

Acknowledgements

The authors acknowledge the Program of Students - Post-graduation Agreement (PEC-PG) and the Department of Electric Engineering EESC-USP for the support and scholarship.

References

[1] Sundberg J 1987. The Science of Singing Voice. Illinois University Press.
[2] Bhati R, Hawkshaw MJ and Sataloff RT 1996. Chaos in Voice and Other Biomechanical Research. In: Professional Voice. The Science and Art of Clinical Care. Third Edition. Plural Publishing
[3] Zhang Y, McGilligan C, Zhou L, Vig M and Jiang J 2004. Nonlinear dynamic analysis of voices before and after surgical excision vocal polyps. J Acoust Soc Am.;115:2270–2277.
[4] Zhang Y and Jiang JJ 2003. Nonlinear dynamic analysis in signal typing of pathological human voices. Electron Lett.; 39:1021–1023.
[5] Douglas A, Rahn III, Maggie Chou, Jack J. Jiang, and Yu Zhang 2007. Phonatory Impairment in Parkinson’s Disease: Evidence from Nonlinear Dynamic Analysis and Perturbation Analysis. Journal of Voice, Vol. 21, No. 1, pp. 64–71.
[6] Dajer ME, Pereira JC and Maciel CD 2005. Nonlinear Dynamical Analysis of Normal Voices In: IEEE International Symposium on Multimedia (ISM2005),Irvine, California, USA.
[7] Dajer ME 2006. Padrões visuais de sinais de voz através de técnica de análise não linear. São Carlos. Dissertação de Mestrado Universidade de São Paulo – Campus de São Carlos.
[8] Hegger R., Kantz H and Schreiber T 1999. Practical implementation of nonlinear time series methods: The TISEAN package. Chaos; 9(2), 413—435.
[9] Fraser, AM and Swinney HL 1986. Independent coordinates for strange attractors from mutual information. Phys. Rev. Lett. vol. 33, pp. 1.
[10] Montagnoli NA 1998. Análise Residual do sinal de voz. São Carlos. Dissertação de Mestrado Universidade de São Paulo – Campus de São Carlos.
[11] Lieberman P 1961. Perturbations in vocal pitch. Journal of the Acoustical Society of America. Vol. 33, pp. 597-602.
[12] Wendahl RW 1966. Laryngeal analog synthesis of jitter and shimmer auditory parameters of harshness. Folia Phoniatrica; vol. 18, pp. 99-108.
[13] Davis, SB 1979. Acoustic characteristics of normal and pathological voices. Speech and Language: Advances in Basic Research and Practice..Vol 1.