Vocal Dynamic Visual Pattern for voice characterization

M E Dajer¹, F A S Andrade ², A N Montagnoli ³, J C Pereira ², D H Tsuji ¹
¹Department of Otorrinolaryngology, School of Medicine, University of São Paulo-SP, Brazil
²Department of Electric Engineering, School of Engineering of São Carlos, University of São Paulo, São Carlos-SP, Brazil
³Department of Bioengineering, School of Engineering of São Carlos, University of São Paulo, São Carlos-SP, Brazil

E-mail: eugedajer@yahoo.com.ar

Abstract. Voice assessment requires simple and painless exams. Modern technologies provide the necessary resources for voice signal processing. Techniques based on nonlinear dynamics seem to assess the complexity of voice more accurately than other methods. Vocal dynamic visual pattern (VDVP) is based on nonlinear methods and provides qualitative and quantitative information. Here we characterize healthy and Reinke’s edema voices by means of perturbation measures and VDVP analysis. VDPD and jitter show different results for both groups, while amplitude perturbation has no difference. We suggest that VDPD analysis improve and complement the evaluation methods available for clinicians.

1. Introduction

The effect of voice pathologies causes many physiological modifications that result in an unhealthy pattern of vocal folds vibration. Consequently pathological voice assessment should search the different aspects of voice production. Reinke’s edema is a chronic inflammatory disease of the vocal folds and is considered a benign lesion. It is usually bilateral but asymmetric in volume and manifests different voice characteristics depending on the degree of inflammation and the severity of voice impairment. In general, patients with Reinke’s edema are women, heavy smokers and report vocal abuse [1].

Voice disorders as Reinke’s edema are usually evaluated by a perceptual analysis and perturbation measures such as jitter and shimmer. However, Titze [2] suggested that perturbation analysis might not be applicable to aperiodic signals, as patients’ voices with benign mass lesions of the vocal fold.

In the last years, techniques based on nonlinear dynamic models have been applied to investigate the dynamic behaviour of biomedical systems, including voice signals [3], [4], [5]. Here we characterize normal and Reinke’s edema voices using qualitative and quantitative analysis of Vocal Dynamic Visual Pattern (VDVP) and acoustic perturbation measures.

¹ E-mail eugedajer@yahoo.com.ar
2. Methods

2.1. Database
Forty eight female voices of sustained vowel /a/ were analyzed by means of perturbation measures, fundamental frequency and VDVP analysis. Voices were divided in two groups; one group includes 23 signals of female healthy volunteers ranging from 25 to 45 years old (mean age 33), with no voice complaints and no laryngeal pathology. The other group includes 25 voice signals from female patients with Reinke’s edema in different stages of disease evolution, ranging from 28 to 45 years old (mean age 37). All signals were from Signal Processing Laboratory database of the School of Engineering of São Carlos, University of São Paulo.

Procedures:
For recordings, subjects were asked to produce a sustained vowel /a/ at a comfortable pitch and loudness level for about 5 seconds. The microphone was placed at 45 degrees and 5 cm from the person’s mouth [2]. Consecutive trials were performed, selecting the signal with less voice variability. For data analysis we selected 500 milliseconds and the amplitude of the signal was normalized according to its absolute maximum value. Signal samples were quantized in amplitude with 16 bits and recorded in mono-channel .WAV format to preserve the fidelity of the signal. The sampling frequency was 44,100 Hz.

For statistical analysis we applied T-test on perturbation measures, fundamental frequency (F0) and on quantitative measures of VDVP. Statistical significance level was set at p = 0.05.

2.2. Acoustic analysis
Perturbation measures jitter and shimmer, and fundamental frequency were extracted from healthy and Reinke’s edema voices. Voice analysis was performed using “Analise de Voz 4.0” [6]. Frequency perturbation or Jitter and amplitude perturbation or shimmer were estimated using correlation technique.

2.3. Vocal Dynamic Visual Patterns
Vocal Dynamic Visual Pattern (VDVP) analysis was developed from phase space reconstruction technique with time delay technique [7] and Poincaré Section [8]. The analysis is run in Matlab 7.0 and provides qualitative and quantitative measures for healthy and pathological signals. According to Packard et al the time delay technique can be used to reconstruct the phase spaces of systems. A time series with length N is measured and recorded as x(t1),x(t2),x(t3),..., where x(ti) ϵ R, ti = t0 + iτ (i=1,2,...,N) at the discrete time interval τ [9].

Then, the time delay vector creates the reconstructed phase space as

\[ X(t) = \{ x(t), x(t - \tau), ..., x(t - (m-1)\tau) \} \tag{1} \]

The reconstructed phase space preserves the topological properties and geometrical invariance of the original attractor, so the system dynamics can be studied in the reconstructed phase space [10]. According to Titze it is important to choose a proper time delay for phase space reconstruction because the time series have a finite length and finite precision. If τ is too small, the time delay vector will be strongly correlated. But if τ is too long, the reconstructed attractor will produce a self Intersection [11]. Figure 1 shows a single cycle wave and VDPV, and 10 cycle waves and VDVP of the same sustained vowel /a/.
2.3.1. Qualitative measures
VDVP analysis describes the nonlinear dynamic characteristics of voice signal in two and three dimensional plots. Therefore qualitative and quantitative measures can be extracted from these VDVP. For qualitative analysis we observed three characteristics of the visual pattern: a) configuration or number of loops (L); b) the trajectories regularity degree (R) – how stable and smooth or rough and irregular are the trajectories – and c) convergence degree (C) – how close or disperse are the trajectories. These characteristics were scored on a five-point scale.

**Number of loops**: a decreasing scale form 4 to 0 describes the VDVP configurations. Patterns with more than three loops are scored with degree 4; three loops = degree 3; two loops = 2; one loop = 1, and 0 = non defined configuration. For vowel /a/ degree 4 and 3 are the expected degrees [5].

**Trajectory Regularity**: was classified from 4 to 0. Where 4 is the highest degree and corresponds to completely regular and smooth trajectories and the lowest degree is 0 that describes completely irregular or rough trajectories. Middle scores – 3, 2 and 1 – are different degrees for combine behaviour of regularity.

**Convergence**: For trajectory convergence scores are: 4= trajectories with high degree of trajectories convergence and 0 for disperse or no-convergence trajectories; scores 3, 2 and 1 are different degrees of convergence.

2.3.2. Quantitative measures
We performed an algorithm based on Pointcaré Section to calculate the dispersion of points in one or more sections in the VDVP. Each voice signal was cut in ten different Poincaré sections, and then a mean value was calculated for each of the forty eight voices, as shown in Figure 2.
To determine the Poincaré Section we first select a tangent line in a trajectory segment. Then, we select three points that determine the cut region and two other points for establishing the perpendicular line to the Poincaré plan (Figure 3).

The next step is to calculate the dot product between the perpendicular line to Poincaré plan and the lines formed by the midpoint of the cut region with the other points of the cut region. Dot product near zero guarantees that points belong to Poincaré Section.

Then, a 3 dimensional rotation of the points is performed using the perpendicular line to the plane. Next, all points will be parallel to the x axis, and all points will be projected in the x=0 plane. When the points are in the x=0 plane we extract the principal components, in order to avoid the correlation of the points and finally, the Poincaré section is achieved (Figure 4).
In order to quantify the dispersion of the points we applied traditional statistical methods:

• DSD: Dispersion Standard Deviation - deviation from the midpoint of the section.
• ASD: Axis standard deviation – related the coordinate axes x and y.

3. Results

3.1. Perturbation Measures
Results for perturbation measures Jitter and Shimmer, and fundamental frequency of healthy and Reinke’s edema female voices are on Table 1.
Statistical analysis with T-Test for jitter showed a significant difference between the groups (P = 0,005). Power of performed test with alpha = 0,050: 0,785.
For shimmer T-test showed no significant difference between the (P = 0,075). Power of performed test with alpha = 0,050: 0,303.
And for fundamental frequency there was a significant difference between the groups (P = <0,001). Power of performed test with alpha = 0,050: 1,000.

Table 1. Values of maximum, minimum, mean and standard deviation for jitter and shimmer percentages and fundamental frequency (F₀) in healthy and Reinke’s edema voices.

|            | Healthy       | Reinke’s edema |
|------------|---------------|----------------|
| Jitter (%) | 0,18          | 0,35           |
| Min        | 0,12          | 0,11           |
| Max        | 0,34          | 1,55           |
| SD         | 0,05          | 0,55           |
| Fo (Hz)    | 218,12        | 149,88         |
| Shimmer (%)| 1,01          | 1,24           |
| Min        | 0,51          | 0,46           |
| Max        | 249,33        | 227,64         |
| SD         | 1,55          | 2,26           |
| Fo (Hz)    | 191,34        | 105,18         |

3.2. Vocal Dynamic Visual Patterns

3.2.1. Qualitative measures
Generally female healthy voices showed various loops configuration (degrees 4 and 3), regular line trajectories with 4 and 3 degrees of regularity and a strong convergence (degree 3) while Reinke’s edema voices showed minor loops configuration (degrees 1 and 2), irregular trajectory lines (degrees 1 and 0) and weaker convergence (degrees 3 and 1).
Figures 5A and 5B show two examples of VDVP configuration. Figure 5A presents loops 3, regularity 4 and convergence 3 (L3R4C3) while Figure 5B shows loops 1, regularity 1 and convergence 1 (L1R1C1).

Results of qualitative analysis for loops (L), Regularity (R) and Convergence degree (C) for both groups are presented in table 2. The number in each parameter indicates the five different degrees presented in section 2.3.1.

Table 2. Results for the 23 healthy voices and 25 Reinke’s edema (RE) voices

| Loops | Healthy N=23 | RE N=25 | Regularity | Healthy N=23 | RE N=25 | Convergence | Healthy N=23 | RE N=25 |
|-------|-------------|---------|------------|-------------|---------|-------------|-------------|---------|
| L4    | 10          | 0       | R4         | 7           | -       | C4          | 2           | 1       |
| L3    | 9           | 1       | R3         | 4           | 2       | C3          | 14          | 9       |
| L2    | 3           | 9       | R2         | 6           | 4       | C2          | 5           | 6       |
| L1    | 1           | 15      | R1         | 6           | 14      | C1          | 2           | 9       |
| L0    | -           | -       | R0         | -           | 5       | C0          | -           | -       |

3.2.2. Quantitative measures
VDVP of healthy and Reinke’s edema were analyzed by means of ten Poincaré Section. Ten Poincaré sections were performed in each signal. Values of Axis Standard Deviation ASD for healthy voices were from 0,1859 to 0,3046 (mean 0,2435) and for Reinke’s edema voices were from 0,2383 to 0,4082 (mean 0,3109). Dispersion Standard Deviation DSD values for healthy voices were from 0,0127 to 0,0293 (mean 0,0208) and for Reinke’s edema were from 0,0199 to 0,0661 (mean 0,0365). For ASD and for SDS the T-Test showed statistical difference between the groups (P = <0,001). Power of performed test with alpha = 0,050: 1,000.

4. Discussion
The methods for assessing healthy and pathological voices are not new, but acoustic techniques used previously were based on linear models, and according Titze, this approach might not be indicate for aperiodic voices, as Reinke’s edema voices [2]. Other procedures, like video-laryngoscopy, analyzes the vocal folds movements, but are considered an invasive method. The authors used an acoustic nonlinear approach in order to assess the vocal dynamic in a non invasive way.
Results of perturbation measures showed statistical difference for jitter, this may be correlated with the characteristic functionality of chronic inflammatory disease, as Reinke’s edema. Shimmer showed no statistical significance, suggesting that traditional perturbation measures, may not adequately characterize these kind of pathology. Fundamental frequency was different for each groups, and according to Tsuji et al, females with Reinke’s edema usually present lower fundamental frequency due to the increasing volume of mass of the lesion [12]. Qualitative analysis of VDVP showed a differential dynamic for both groups. Due to fluid collection the vocal folds vibrate slower and in an irregular pattern producing a rough voice with perceptive noise. These characteristics appear in the VDVP as irregular trajectories and weak convergence. In healthy voices the vocal folds vibration is quasi periodic and produce regular trajectories with strong convergence.

We observed that voice signals with higher jitter and shimmer are correlated with VDVP configuration of irregular and disperse trajectories and with higher values of Standard Deviation ASD and Dispersion Standard Deviation in the Poincaré Section.

5. Conclusion

This paper shows that tools based on nonlinear dynamical methods, as phase space reconstruction and Poincaré section seem to be suitable techniques for voice signals characterization and analysis, because of the chaotic component of the human voice. The VDVP analysis was able to discriminate between different stages of Reinke’s edema evolution and healthy voices. Vocal Dynamic Visual Pattern qualitative analysis has shown a potential value to describe normal and Reinke’s edema voices. The standard healthy VDVP of vowel /a/ showed complex configuration and nearly periodic behavior, while Reinke’s edema voices presented simpler loops configuration, irregular trajectory lines and lower convergence degree. Values of Standard Deviation ASD and Dispersion Standard Deviation in the Poincaré Section were different for each group.

We consider that perturbation analysis associated with quantitative and qualitative analysis of the VDVP provide more information and represent non-invasive and helpful tools to describe objectively the vocal folds dynamic.

6. Acknowledgements

The authors acknowledge the Research Foundation of the State of São Paulo (FAPESP) for the financial support. We also acknowledge the School of Engineering of São Carlos (EESC/USP) for the research facilities and the use of its voice database.

7. References

[1] Marcotullio D, Magliulo G, Pezone T. (2002). Reinke’s and risk factors: clinical and histopathologic aspects. *Am J Otolaryngol*. 23:81-84.
[2] Titze IR (1995) Workshop on Acoustic Voice Analysis: Summary Statement. Denver, National Center for Voice and Speech.
[3] Pitsikalis B; Kokkinos I; Maragos P. (2003). Nonlinear Analysis of Speech Signals: Generalized Dimensions and Lyapunov Exponents. Proc. European Conf. on Speech Communication and Technology - Eurospeech Geneva, Switzerland.
[4] Jiang JJ, Zhang Y; MacCallum J; Sprecher A; Zhou L. (2009). Objective Acoustic Analysis of Pathological Voices from Patients with Vocal Nodules and Polyps. *Folia Phoniatri Logop*; 61:342–349.
[5] Scalassara PR, Dajer ME, Maciel CD, Guido CR, Pereira JC. (2009). Relative entropy measures applied to healthy and pathological voice characterization. Applied Mathematics and Computation 207:95–108 (2009).

[6] Montagnoli A.N., Pereira J.C., [Análise de Voz, Versão 4.10] Sistema de Auxílio à Análise Acústica da Voz. 2010.

[7] Fraser, AM; Swinney, HL. Independent coordinates for strange attractors from mutual information. Phys. Rev. Lett. 1986, vol. 33, pp. 1.

[8] Kantz H; Schreiber, T. Nonlinear Time Series Analysis. Cambridge University Press, Cambridge (1997).

[9] Packard NH, Crutchfield JP, Farmer JD, Shaw RS. Geometry from a time series. Phys Rev Lett. 1980;45:712.

[10] Takens F. Detecting strange attractors in turbulence. In: Rand DA, Young LS, editors. Lecture Notes in Mathematics. Berlin: Springer-Verlag; 1981:366–381.

[11] Titze IR, Baken R, Herzel H. Evidence of chaos in vocal fold vibration. In: Titze IR, editor. Vocal Fold Physiology: New Frontiers in Basic Science. San Diego, CA: Singular Publishing Group; 1993:143–188.

[12] Tsuji DH, Inamura R; Ubirajara LS. Fisiologia Da Laringe. In: Pinho, S; Tsuji, Dh; Bohadana, Sc. Fundamentos em laringologia e Voz. Ed. Revinter, Rio de Janeiro, 2006.