State-Space Approach to Structural Representation of Perturbed Pitch Period Sequences in Voice Signals

Gabriel A. Alzamendi, Gastón Schlotthauer, and María E. Torres, *Entre Ríos, Argentina*

Summary: Objectives. The aim of this study was to propose a state space-based approach to model perturbed pitch period sequences (PPSs), extracted from real sustained vowels, combining the principal features of disturbed real PPSs with structural analysis of stochastic time series and state space methods.

Methods. The PPSs were obtained from a database composed of 53 healthy subjects. State space models were developed taking into account different structures and complexity levels. PPS features such as trend, cycle, and irregular structures were considered. Model parameters were calculated using optimization procedures. For each PPS, state estimates were obtained combining the developed models and diffuse initialization with filtering and smoothing methods. Statistical tests were applied to objectively evaluate the performance of this method.

Results. Statistical tests demonstrated that the proposed approach correctly represented more than the 75% of the database with a significance value of 0.05. In the analysis, structural estimates suitably characterized the dynamics of the PPSs. Trend estimates proved to properly represent slow long-term dynamics, whereas cycle estimates captured short-term autoregressive dependencies.

Conclusions. The present study demonstrated that the proposed approach is suitable for representing and analyzing real perturbed PPSs, also allowing to extract further information related to the phonation process.

Key Words: Perturbed pitch periods–Stochastic pitch model–Jitter–Structural time-series analysis–State-space models.

INTRODUCTION

Precise period perturbation assessment is one of the most difficult tasks in speech pathology and voice therapy.[1] This is because the perturbations arise in an unpredictable fashion and are usually concealed in the speech records. Specialists have not yet reached a full agreement on the nature and origin of these irregularities.[2] Indeed, it can be shown that irregularities arise even in a nonpathologic stable voice.[3,4] In the last few decades, this situation has drawn great attention from researchers and clinicians in the speech community. They have found that perturbations arise as a result of the combination of neurologic, biomechanical, aerodynamic, and acoustic sources throughout the speech production system.[5] Additionally, it has been argued that perturbations behave noticeably different in pathologic and nonpathologic voices.[6,7] In this work, we propose a method for analyzing and modeling a pitch period sequence (PPS), consisting of successive pitch periods extracted from a voice signal, which explicitly considers fluctuations and instantaneous perturbations.

Real PPSs are generally composed by identifiable structures presenting short- and long-term behavior.[8] Short-term structures carry information related to period perturbations, consisting of random cycle-to-cycle variations denominated jitter.[3,9] In objective voice analysis, there are diverse acoustical parameters conceived to quantify jitter, which can be classified as absolute measures (eg, perturbation factor or directional perturbation factor) or fundamental frequency-related measures (eg, jitter factor, jitter ratio, or coefficient of variation).[6] However, these objective features are highly sensitive to slow long-term components, normally associated with prosody information or intonation. Consequently, relative average perturbation features have been defined (eg, average absolute perturbation or period perturbation quotient).[6] Nevertheless, these parameters are inadequate where long-term components are strong. Therefore alternative methods are required for robust perturbation assessment. To solve this problem, we propose a state-space approach to PPS structural analysis, and we show that it allows to optimally separate jitter and long-term components.

Recent advances in PPS characterization have found great applicability in modern technology. It has been demonstrated that PPSs carry information belonging to the speaker itself (eg, identity, gender, or mood)[10–12] and related to the physiological condition of the speech production system (eg, vocal folds dynamic).[13] Therefore, theoretical models, considering jitter and long-term fluctuations, were developed and successfully used in several applications, eg, to enhance the naturalness of artificial voices in synthesis methods,[14] to synthesize expressive voices in human-computer interfaces,[11] to verify speakers in security systems,[10] and to simulate pathologic voices under controlled conditions.[15] Moreover, these models provide a theoretical framework to understand period perturbations.[5]

Nowadays, there is a great number of PPS models available in speech literature. The simplest one consists of a sequence of constant fundamental periods, not allowing the representation of aperiodic signals. On the other hand, versatile perturbed PPS models have been developed using simple stochastic laws. A straightforward strategy for jitter generation involves

Accepted for publication November 20, 2014.

From the Laboratorio de Señales y Dinámicas no Lineales, Facultad de Ingeniería de la Universidad Nacional de Entre Ríos and Consejo Nacional de Investigaciones Científicas y Técnicas, Entre Ríos, Argentina.

Address correspondence and reprint requests to Gastón Schlotthauer, Laboratorio de Señales y Dinámicas no Lineales, Facultad de Ingeniería, Universidad Nacional de Entre Ríos, cc 47, Suc 3 (3100) Paraná, Entre Ríos, Argentina. E-mail: gschlotthauer@conicet.gov.ar

Journal of Voice, Vol. 29, No. 6, pp. 682-692
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2014.11.007

a constant fundamental period perturbed by random noise. This approach has been applied in expressive speech synthesis for neutral voice transformation, where fundamental period and random noise depend on different emotions.[16] Moreover, Gaussian mixture models were applied to emotional speech synthesis, where different emotions were characterized considering long-term components as multimodal processes.[17] Recently, we developed a strategy to synthesize both normal and pathologic perturbed sustained vowels, where PPSs were obtained from a stochastic model based on jitter factor. This method proved to be useful for testing algorithms for fundamental frequency estimation[15] and for voice synthesis with high perceptual quality.[18]

All the methods mentioned previously assume that PPSs are independent and identically distributed stationary stochastic processes. Nevertheless, examination of real sequences demonstrates that these hypotheses are unrealistic and, as a consequence, previous methods are not able to suitable represent a real PPS. Schoentgen[19] has summarized the principal features of PPSs extracted from real normal voices. For the present work, some of those were considered:

- PPS presents Gaussian probability distribution;
- adjacent periods in a PPS are correlated where correlation degree varies with voice signals;
- there are structures that reinforce period correlation (microtremors);
- jitter size is small (0.1–1% relative to fundamental period);
- jitter appears to be a genuine stochastic phenomenon;
- meaningful statistics of jitter can be obtained from sustained vowel waveforms.

Considering the previously listed features, it is clear that more versatile models, able to represent complex structures, are required. The first attempt to understand period correlation made use of time series analysis methods based on autoregressive (AR) or AR moving average models.[20] These methods allowed to represent the existing strong correlation in both normal and pathologic voices, where model order depended on the analyzed voices.[12,21] Later, Ruinskiy and Lavner[14] proposed a jitter bank-based approach to characterize the relative amplitude and correlation in the PPSs, suitable for naturally hoarse voice synthesis. Despite the ability to represent correlation information, the previously mentioned methods assume that PPSs are stationary signals. Although it has been shown that short-term jitter is indeed a stationary process, PPSs are not necessarily stationary signals.[5]

Stochastic difference equations have demonstrated to be useful for modeling complicated random dynamics. Therefore, these methods provide the required theoretical framework for perturbed PPSs representation. Using this method, a jitter model able to represent aperiodic vocal folds oscillations was proposed,[19] and it was used to analyze the influence of glottal and external factors in period perturbations. Moreover, this model was applied in hoarse voice synthesis, showing that hoarseness strongly depends on jitter dynamics.[22,23] Additionally, artificial voices synthesized by this method were used to evaluate the effects of experience and training of voice pathologists in the correct identification of periods in perturbed sustained vowels, under controlled conditions of jitter[24] and additive noise.[25]

Other strategies for PPS modeling have been published in speech literature, eg, strategies based on biomechanical models,[13,26] spectral information of perturbation signals,[27] and nonlinear or chaotic signal processing.[28] Although most of previously mentioned methods have been successfully applied in voice synthesis tasks and theoretical perturbation modeling, only few of them can be applied to real PPS analysis. As far as the authors know, none of these methods incorporates the principal PPS features pointed out by Schoentgen[19] into the objective analysis of real voice signals. Therefore, in this article, we propose a state space-based approach to analyze and model PPSs extracted from real sustained vowels. State-space methods (SSMs) allow combining the PPS features with model-based, stochastic, time-series analysis. Within this framework, real PPSs are analyzed, and stochastic trend and cycle structures possessing a straightforward relationship with PPS features are estimated. In addition, the performance of this method is evaluated through statistical tests.

MATERIALS AND METHODS

PPS samples

In this section, the required procedure to obtain the PPSs and the principal materials used throughout this article are presented.

Voice database

In this work, the database (DB) developed by the Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab[29] is used, which includes sustained vowels /a/ of healthy individuals and patients with a wide range of voice disorders—eg, organic, neurologic, traumatic, and psychogenic. Voice samples are accompanied by detailed medical information gathered from tests and professional opinions. Only voices from healthy participants were considered. The participants were 21 males and 32 females, 38.81 ± 8.49 years and 34.16 ± 7.87 years, respectively. All samples in the DB were collected in a controlled environment and the duration of each signal was 3 seconds. The sampling rate and quantization level were 50 kHz and 16 bit, respectively.

PPS calculation

Voice samples were processed to extract PPSs with *Praat* software, developed by Boersma and Weenink[30] of the Institute of Phonetic Sciences, University of Amsterdam. *Praat* is one of the most widely used software in objective voice perturbation analysis. It applies a short-term analysis procedure, where pitch periods are obtained by waveform-matching methods. The technique applies autocorrelation analysis to estimate the location of fixed points in the glottal cycle, called pitch marks, where two consecutive waveforms look maximally similar. Therefore, pitch periods are calculated as the difference between consecutive pitch marks and the PPS {P_$_1$,P_$_1$,…,P_$_N$} is obtained, with N the number of elements in the sequence. In Figure 1, 20 milliseconds of a typical sustained vowel /a/

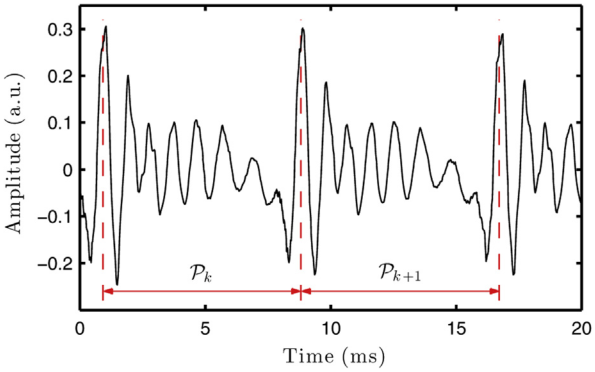

FIGURE 1. Pitch period sequences estimation. Solid line represents 20 milliseconds of a sustained vowel /a/, in arbitrary units (a.u.), and dashed lines indicate time pitch marks. Double arrows correspond to consecutive pitch periods.

are displayed in solid line and time pitch marks, calculated with *Praat*, are superimposed in vertical dashed line. Double-arrows represent consecutive pitch periods.

In Figure 2, circle marks represent a typical PPS obtained using the previously explained method, corresponding to a sustained vowel /a/ phonated by a healthy female. The random structure and the long-term fluctuations composing this PPS can be appreciated. This behavior was also observed in the remaining signals of the corpus. Although these signals correspond to stable sustained vowels, the assumption of a constant fundamental period—implying a constant fundamental frequency—is not realistic.

State space theory

SSMs, originally proposed by Kalman[31] from an engineering standpoint, have achieved great interest in diverse areas—ie, statistics, econometrics, medicine, and physics—and have been widely considered in applications like tracking, process control, forecasting, and failures detection, among others. In speech processing in particular, it has been used in fundamental frequency estimation,[32] formant and antiformant tracking,[33] and glottal source modeling,[34,35] among others. The main feature of SSMs is that it allows a stochastic model-based representation for nonstationary time series.[36] The SSMs considered in this article are presented in the following.

State-space model

A linear Gaussian state-space model (GSSM) is a useful and versatile instrument for representing stochastic time series. Mathematically, GSSMs are defined by a system of stochastic difference equations[31]:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{w}_k, \quad \mathbf{w}_k \sim \mathcal{N}(\mathbf{0},\mathbf{Q}), \\ \mathbf{z}_k &= \mathbf{H}\mathbf{x}_k + \mathbf{v}_k, \quad \mathbf{v}_k \sim \mathcal{N}(\mathbf{0},\mathbf{R}), \end{aligned} \tag{1}$$

where k is period index, $\mathbf{x}_k \in \mathbb{R}^p$ is the unobservable state vector and $\mathbf{z}_k \in \mathbb{R}^r$ is the measurable observations vector. Factors $\mathbf{A} \in \mathbb{R}^{p\times p}$, $\mathbf{B} \in \mathbb{R}^{p\times q}$, and $\mathbf{H} \in \mathbb{R}^{r\times p}$ are called state transition, error and measurements matrices, respectively. Finally, $\mathbf{w}_k \in \mathbb{R}^q$ and $\mathbf{v}_k \in \mathbb{R}^r$ correspond to state and measurement errors with covariance matrices $\mathbf{Q} \in \mathbb{R}^{q\times q}$ and $\mathbf{R} \in \mathbb{R}^{r\times r}$, respectively. Errors $\mathbf{w}_k$ and $\mathbf{v}_k$ are assumed to be mutually and serially independent and independent of the initial state $\mathbf{x}_0$.

Model (1) represents the common situation where actual state of a system, determined by $\mathbf{x}_k$, can not be directly observed or measured, but an ensemble of noisy measurements $\mathbf{z}_k$ is available. The first equation rules the state transitions into the system and is therefore called state transition equation, whereas the second one controls instantaneous measurement values and is called observation equation.[37] In general, time-dependent changes in model matrices are allowed.[31] Nevertheless, here we only consider model (1) ruled by a stochastic difference equation with constant coefficients.

Structural analysis

Structural time series analysis consists of decomposing a signal of interest into elements possessing a simple and straightforward interpretation. This framework provides a systematic strategy for model-based signal processing, helping with the understanding of underlying dynamics in complex processes.[36,38] For the present work, a structural analysis considering trend, cycle,

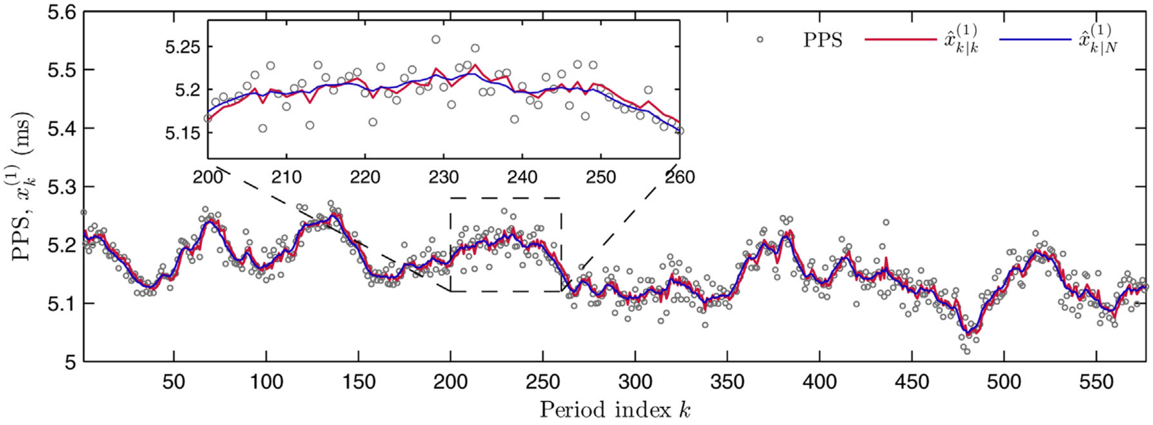

FIGURE 2. A typical PPS, in milliseconds, extracted from a sustained vowel /a/ phonated by a healthy female, along with the filtered $\widehat{x}_{k|k}^{(1)}$ and smoothed $\widehat{x}_{k|N}^{(1)}$ trend components. State estimates were calculated considering GSSM (I).

and irregular components is applied. According to this, a PPS is represented as follows:

$$P_\mathrm{k} = \mu_k + \psi_k + \varepsilon_k, \quad \varepsilon_k \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right), \tag{2}$$

where μ_k, ψ_k, and ε_k correspond to trend, cycle, and irregular components, respectively. Disturbance ε_k is assumed a Gaussian serially independent stochastic process. To ensure the required flexibility, each component is allowed to evolve stochastically over time.

Considering the features pointed out by Schoentgen, the previously defined structural components are associated with specific events in PPS. Trend μ_k represents the slow long-term fluctuations, resulting as a consequence of control mechanisms acting during phonation.[5] Cycle ψ_k models microtremors or any other phenomena reinforcing temporal correlation.[19] Finally, disturbance ε_k is associated with jitter phenomenon.

In this article, three different alternatives for modeling the trend component are considered. The simplest form is a local level $T(1)$ or random walk process, given by,

$$\mu_{k+1} = \mu_k + \eta_k, \quad \eta_k \sim \mathcal{N}\left(0, \sigma_\eta^2\right). \tag{3}$$

An alternative form corresponds to the so-called integrated random walk $T(2)$ process, defined as follows:

$$\begin{aligned} \mu_{k+1} &= \mu_k + \beta_k, \\ \beta_{k+1} &= \beta_k + \zeta_k, \quad \zeta_k \sim \mathcal{N}\left(0, \sigma_\zeta^2\right), \end{aligned} \tag{4}$$

where β_k represents the stochastic slope describing the trend change. Form $T(2)$ generates a smoother trend component than the one obtained with $T(1)$. Finally, a more general strategy is formulated combining the two trend models mentioned previously, a local linear trend $T(3)$ process, defined as follows:

$$\begin{aligned} \mu_{k+1} &= \mu_k + \beta_k + \eta_k, \quad \eta_k \sim \mathcal{N}\left(0, \sigma_\eta^2\right), \\ \beta_{k+1} &= \beta_k + \zeta_k, \quad \zeta_k \sim \mathcal{N}\left(0, \sigma_\zeta^2\right). \end{aligned} \tag{5}$$

It can be observed that, assuming $\sigma_\zeta^2 \to 0$ and $\beta_0 = 0$, this form reduces to $T(1)$. Otherwise, assuming $\sigma_\eta^2 \to 0$, Equation (5) reduces to $T(2)$. It is interesting to notice that, despite their simplicity, these forms successfully represent smooth, long-term, nonstationary, and temporal-dependent time series.[38]

The stochastic cycle component is represented considering an AR model $AR(\rho)$, given by the following equation:

$$\psi_{k+1} = -\mathrm{a}_1\psi_k - \mathrm{a}_2\psi_{k-1} - \ldots - \mathrm{a}_\rho\psi_{k-\rho+1} + \xi_k, \tag{6}$$

where ρ is model order, $\xi_k \sim \mathcal{N}(0, \sigma_\xi^2)$ and minus signs are used for convenience only. To represent a stochastic cycle component, the coefficients $\{\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_\rho\}$ need to be estimated to ensure that $AR(\rho)$ generates a wide-sense stationary process.[36]

Considering the structural representation (2), each PPS can be expressed in the form of a GSSM. In particular, observation $\mathbf{z}_k$ corresponds to element P_k in a PPS, for $k = 1,\ldots,N$. Given that each PPS is a univariate series; hereafter, the measurement dimension becomes $r = 1$. As an example, assume this sequence was generated by the $T(3)$ and the $AR(\rho)$ processes. Combining trend and cycle components in a vector form, the state vector in (1) can be defined as follows:

$$\begin{aligned} \mathbf{x}_k &= \begin{pmatrix} x_k^{(1)} & x_k^{(2)} & x_k^{(3)} & x_k^{(4)} & \ldots & x_k^{(p-1)} & x_k^{(p)} \end{pmatrix}^T \\ &= \begin{pmatrix} \mu_k & \beta_k & \psi_k & \psi_{k-1} & \ldots & \psi_{k-\rho+2} & \psi_{k-\rho+1} \end{pmatrix}^T. \end{aligned}$$

Then, considering (5) and (6), the state transition matrix can be formulated as:

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & -\mathrm{a}_1 & -\mathrm{a}_2 & \ldots & -\mathrm{a}_{\rho-1} & -\mathrm{a}_\rho \\ 0 & 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix},$$

and, similarly, error and observation matrices can respectively be defined through expressions:

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{pmatrix} \text{ and } \mathbf{H} = \begin{pmatrix} 1 & 0 & 1 & 0 & \ldots & 0 \end{pmatrix}.$$

Structural disturbances are gathered together, then state and measurement error vectors are defined by the following equation:

$$\mathbf{w}_k = \begin{pmatrix} \eta_k & \zeta_k & \xi_k \end{pmatrix}^T \text{ and } \mathbf{v}_k = \varepsilon_k,$$

with covariance matrices:

$$\mathbf{Q} = \begin{pmatrix} \sigma_\eta^2 & 0 & 0 \\ 0 & \sigma_\zeta^2 & 0 \\ 0 & 0 & \sigma_\xi^2 \end{pmatrix} \text{ and } \mathbf{R} = \sigma_\varepsilon^2,$$

respectively. These covariance matrices arise from the hypothesis that disturbance processes ε_k, η_k, ζ_k and ξ_k are serially and mutually independent.

It can be observed that state space and error dimensions, p and q respectively, are dependent on the components considered in the structural analysis. In Table 1, the compositions of the different structural models considered in this article are presented, sorted in order of increasing model complexity.

State filtering

State filtering, also called Kalman filtering, is an iterated forward method used for estimating the unobserved state vector $\mathbf{x}_k$, considering past and present system information.[31] Filtering concerns on the calculation of conditional expected values $\widehat{\mathbf{x}}_{k|k} = \mathrm{E}(\mathbf{x}_k|\mathbf{z}_1, \ldots, \mathbf{z}_k)$ for period index $k = 1,\ldots,N$. This is an optimal procedure, meaning that it computes the minimum mean square linear estimator conditioned on the given

TABLE 1.
Structural State Space Models

GSSM	Components	p	q
(I)	$T(1)$	1	1
(II)	$T(2)$	2	1
(III)	$T(3)$	2	2
(IV)	$T(3) + AR(2)$	4	3
(V)	$T(3) + AR(4)$	6	3
(VI)	$T(3) + AR(6)$	8	3
(VII)	$T(3) + AR(8)$	10	3

Notes: Composition of GSSMs considered in this work according to state space dimension p and error dimension q, where measurement dimension $r = 1$ owing to PPS is univariate.

observations $\{z_1,\ldots,z_k\}$. Because of its simplicity and robustness, it becomes a useful method in real-time applications.

In this article, we applied the so-called contemporaneous Kalman filter.[37,39] The algorithm consists of the following three steps:

1. Prediction:

$$\widehat{\mathbf{x}}_{k|k-1} = \mathbf{A}\widehat{\mathbf{x}}_{k-1|k-1},$$
$$\mathbf{P}_{k|k-1} = \mathbf{A}\mathbf{P}_{k-1|k-1}\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T,$$

2. Innovation:

$$\widehat{\mathbf{y}}_k = \mathbf{z}_k - \mathbf{H}\widehat{\mathbf{x}}_{k|k-1},$$
$$\mathbf{F}_k = \mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^T + \mathbf{R},$$

3. Correction:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}^T\mathbf{F}_k^{-1},$$
$$\widehat{\mathbf{x}}_{k|k} = \widehat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\widehat{\mathbf{y}}_k,$$
$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_{k|k-1},$$

where $\widehat{\mathbf{x}}_{k|k-1}$ and $\widehat{\mathbf{x}}_{k|k}$ are the *a priori* and *a posteriori* filtered state vectors, with covariance matrices $\mathbf{P}_{k|k-1}$ and $\mathbf{P}_{k|k}$, respectively. Factor $\widehat{\mathbf{y}}_k$ is the innovation or one-step ahead forecast error, with covariance matrix $\mathbf{F}_k$, and $\mathbf{K}_k$ is the so-called Kalman gain matrix. For initialization, the simplest strategy is to assume $\widehat{\mathbf{x}}_{0|0} \sim \mathcal{N}(\mathbf{x}_0, \mathbf{P}_0)$, with $\mathbf{x}_0 \in \mathbb{R}^p$ and $\mathbf{P}_0 \in \mathbb{R}^{p\times p}$ known. In this work, an alternative procedure called diffuse initialization is considered. Its theoretical background and main features are discussed further in the following.

Figure 2 shows, superimposed, the corresponding trend state estimate $\widehat{x}^{(1)}_{k|k}$, calculated through filtering procedure and considering GSSM (I). It can be seen that the estimates accordingly represent the slow long-term dynamics in the PPS. This is better represented in the enlarged portion displayed in Figure 2.

State smoothing

State smoothing is an alternative method for state vector estimation, which takes advantage of the entire observation sequence. In particular, in this article the so-called fixed interval smoothing method was considered.[37] This is an iterated backward procedure, which is applied after the forward filtering method was performed. Smoothing concerns the smoothed state vector calculation defined by the conditional expected values $\widehat{\mathbf{x}}_{k|N} = \mathrm{E}(\mathbf{x}_k|\mathbf{z}_1,\ldots,\mathbf{z}_N)$, for period index $k = 1,\ldots,N$. This is a noncausal method, which adds future information in the state estimation procedure and is only applicable in situations where stored signals are considered or in delay-tolerant real-time applications.[40]

Fixed interval smoothing algorithm consists of the following equations:

$$\mathbf{L}_k = \mathbf{A}(\mathbf{I} - \mathbf{K}_k\mathbf{H}),$$
$$\mathbf{r}_{k-1} = \mathbf{H}^T\mathbf{F}_k^{-1}\mathbf{y}_k + \mathbf{L}_k^T\mathbf{r}_k,$$
$$\mathbf{N}_{k-1} = \mathbf{H}^T\mathbf{F}_k^{-1}\mathbf{H} + \mathbf{L}_k^T\mathbf{N}_k\mathbf{L}_k,$$
$$\widehat{\mathbf{x}}_{k|N} = \widehat{\mathbf{x}}_{k|k-1} + \mathbf{P}_{k|k-1}\mathbf{r}_{k-1},$$
$$\mathbf{P}_{k|N} = \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{N}_{k-1}\mathbf{P}_{k|k-1}$$

where $\widehat{\mathbf{x}}_{k|N}$ corresponds to smoothed state vector, with covariance matrix $\mathbf{P}_{k|N}$. In this article, the factors $\mathbf{L}_k$, $\mathbf{r}_{k-1}$, and $\mathbf{N}_{k-1}$ are considered auxiliary parameters.[37]

In Figure 2, the trend state estimate $\widehat{x}^{(1)}_{k|N}$ calculated through smoothing method and considering GSSM (I) is shown. It can be seen that the smoothed trend presents a less fluctuating behavior than the filtered one. This phenomenon gave rise to the term state smoothing. Moreover, in the enlarged portion of Figure 2, similarities and differences between filtered and smoothed trends can be better appreciated. These features are analyzed in more detail in the following.

Diffuse state filtering and smoothing

As previously mentioned, filtering and smoothing methods depend on initial state vector $\mathbf{x}_0$. This information is totally or partially unknown; hence, these conditions need to be properly fixed or estimated. In other situations, an ensemble of similar signals is analyzed taking the same GSSM, making imperative to know or estimate appropriate initial state vector for each signal. Usually, the setting of $\mathbf{x}_0$ becomes a complicated and cumbersome task in practice. Fortunately, there is an analytical solution to the initialization problem that is both, easy to implement and computationally efficient.[39,40] It is based on a stochastic model for initial conditions given by the following equation:

$$\mathbf{x}_0 = \widehat{\mathbf{x}}_0 + \mathbf{T}\boldsymbol{\delta} + \mathbf{B}_0\mathbf{w}_0, \quad \mathbf{w}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_0), \tag{7}$$

where $\widehat{\mathbf{x}}_0 \in \mathbb{R}^p$ is known, $\boldsymbol{\delta} \in \mathbb{R}^s$ is unknown, and $\mathbf{w}_0 \in \mathbb{R}^{(p-s)}$. The selection matrices $\mathbf{T} \in \mathbb{R}^{p\times s}$ and $\mathbf{B}_0 \in \mathbb{R}^{p\times(p-s)}$ consist of columns of the identity matrix $\mathbf{I}_p$, under the restriction $\mathbf{T}^T\mathbf{B}_0 = 0$. Moreover, it is assumed that matrix $\mathbf{Q}_0 \in \mathbb{R}^{(p-s)\times(p-s)}$ is positive definite and known.

In model (7), unknown information is represented by the diffuse vector $\boldsymbol{\delta}$, assuming $\boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \kappa\mathbf{I}_s)$. Considering the diffuse model (7) and the condition $\kappa \to \infty$, an alternative filtering method independent of initial conditions is obtained. This method is called diffuse filtering. In practice, a few iterations d are required to lose the dependence of state estimates on κ, where generally $d \ll N$[40]. Estimates from $k = d + 1$ to N are obtained using the original Kalman filtering. Smoothed estimates are calculated from $k = N$ to $d + 1$ applying state smoothing procedure; whereas the remaining states are obtained through diffuse smoothing, a modified smoothing algorithm considering again the model presented in (7) and the condition $\kappa \to \infty$. Here, diffuse filtering and smoothing methods proposed by Koopman and Durbin[39] were implemented.

Parameter estimation

As can be observed in Equation (1), GSSMs depend on some unknown parameters $\mathbf{\Theta}$, which need to be estimated from observations. Here, the unknown parameters are the covariance matrices $\mathbf{Q}$ and $\mathbf{R}$, and the AR coefficients $\{a_1, a_2, \ldots, a_\rho\}$. Strategies for parameter estimation based on optimization procedures consisting on computationally maximize the associated log-likelihood function $\log \mathscr{L}(\mathbf{\Theta}|\mathbf{z}_1, \ldots, \mathbf{z}_N)$ have been proposed.[40,41] This approach requires solving the optimization problem:

$$\widehat{\mathbf{\Theta}} = \underset{\mathbf{\Theta} \in \mathscr{D}}{\arg\max} \log \mathscr{L}(\mathbf{\Theta}|\mathbf{z}_1, \ldots, \mathbf{z}_N), \tag{8}$$

where $\mathscr{D}$ is the domain of definition of $\mathbf{\Theta}$, and $\widehat{\mathbf{\Theta}}$ is the so-called maximum likelihood estimate. Generally, only relative values of $\log \mathscr{L}(\mathbf{\Theta}|\mathbf{z}_1, \ldots, \mathbf{z}_N)$ are important in this optimization procedure. Because of the diffuse methods, in this article, the diffuse version of log-likelihood function is applied.[37]

In the case of GSSMs (I), (II), and (III), the parameter estimation procedure is simple because the calculation of AR coefficients is not required. Therefore, covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ are estimated combining both expectation and maximization (E-M) algorithm and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. Otherwise, an iterated procedure was applied to solve problem (8), considering observations $\{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$ and a particular GSSM(j), where $j \in \{(\mathrm{IV}), \ldots, (\mathrm{VII})\}$. The flowchart of Figure 3 shows this procedure.

First considering the simplest GSSM (I), matrices $\mathbf{R}$ and $\mathbf{Q}$ are estimated applying the E-M algorithm, starting from random initial matrices, and state vector estimates are obtained through filtering and smoothing methods. Error $\widehat{e}_{k|N} = \mathbf{z}_k - \widehat{x}^{(1)}_{k|N}$ is computed, where $\widehat{x}^{(1)}_{k|N}$ is the trend component in the smoothed state vector. Given that GSSM (I) is not capable of representing the cycle component, this information is completely preserved in $\widehat{e}_{k|N}$. Therefore, parameters $\{a_1, a_2, \ldots, a_\rho\}$ are roughly estimated from $\widehat{e}_{k|N}$ through the linear prediction (LP) method, which ensures that these coefficients generate a wide-sense stationary AR model. Then matrix $\mathbf{A}$ in GSSM(j) is defined.

Then, matrices $\mathbf{R}$ and $\mathbf{Q}$ are adjusted again by E-M algorithm and the state vector estimates are obtained through filtering and smoothing methods, considering the GSSM (j). Error $\widehat{e}_{k|N} = \mathbf{z}_k - \widehat{x}^{(1)}_{k|N}$ is calculated, preserving cycle information. Next, $\{a_1, a_2, \ldots, a_\rho\}$ coefficients are estimated from $\widehat{e}_{k|N}$, through the LP method, and then matrix $\mathbf{A}$ in GSSM (j) is updated. The Euclidean distance between present and previously estimated AR coefficients is calculated and the convergence criterion is checked. If this distance is above a set threshold, additional parameter estimations are required. Otherwise, it is considered that the AR coefficient estimations are good enough and, finally, a precise covariance matrices calculation through BFGS method is performed.

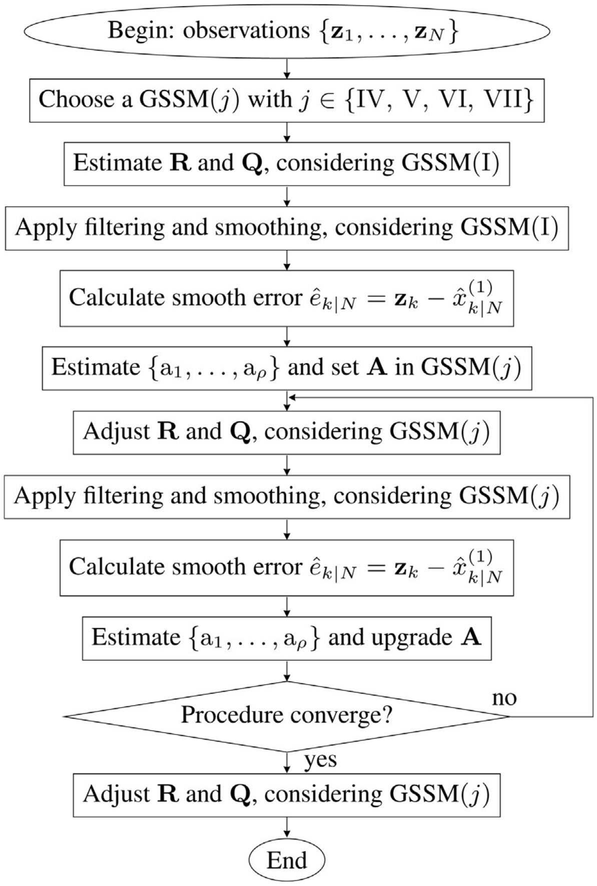
Begin: observations $\{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$
Choose a GSSM(j) with $j \in \{\mathrm{IV, V, VI, VII}\}$
Estimate $\mathbf{R}$ and $\mathbf{Q}$, considering GSSM(I)
Apply filtering and smoothing, considering GSSM(I)
Calculate smooth error $\hat{e}_{k|N} = \mathbf{z}_k - \hat{x}^{(1)}_{k|N}$
Estimate $\{a_1, \ldots, a_\rho\}$ and set $\mathbf{A}$ in GSSM(j)
Adjust $\mathbf{R}$ and $\mathbf{Q}$, considering GSSM(j)
Apply filtering and smoothing, considering GSSM(j)
Calculate smooth error $\hat{e}_{k|N} = \mathbf{z}_k - \hat{x}^{(1)}_{k|N}$
Estimate $\{a_1, \ldots, a_\rho\}$ and upgrade $\mathbf{A}$
Procedure converge?
no
yes
Adjust $\mathbf{R}$ and $\mathbf{Q}$, considering GSSM(j)
End

FIGURE 3. Flowchart describing the parameter estimation procedure for GSSMs (IV), (V), (VI), and (VII). The method is initialized with observations and the chosen GSSM, then covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ and AR coefficients $\{a_1, a_2, \ldots, a_\rho\}$ are optimally estimated.

The proposed optimization strategy has useful features. As it was previously stated, the covariance matrices are adjusted each iteration and, then, AR coefficients are estimated. This strategy allows separating the problem (8) into two simple and easy to solve subproblems and, as a consequence, the convergence of the optimization procedure is accelerated. Additionally, the LP method guarantees that the obtained AR coefficients generate a wide-sense stationary process, which is an essential requirement in time series analysis. Finally, this procedure has shown, in preliminary studies, to be robust with respect to random initialization of covariance matrices.

Diagnostic checking

Essential hypotheses concerning GSSMs in general, and their applicability to structural analysis in particular, were presented previously. It is assumed that state and measurement errors behave normally, with zero mean and constant covariance

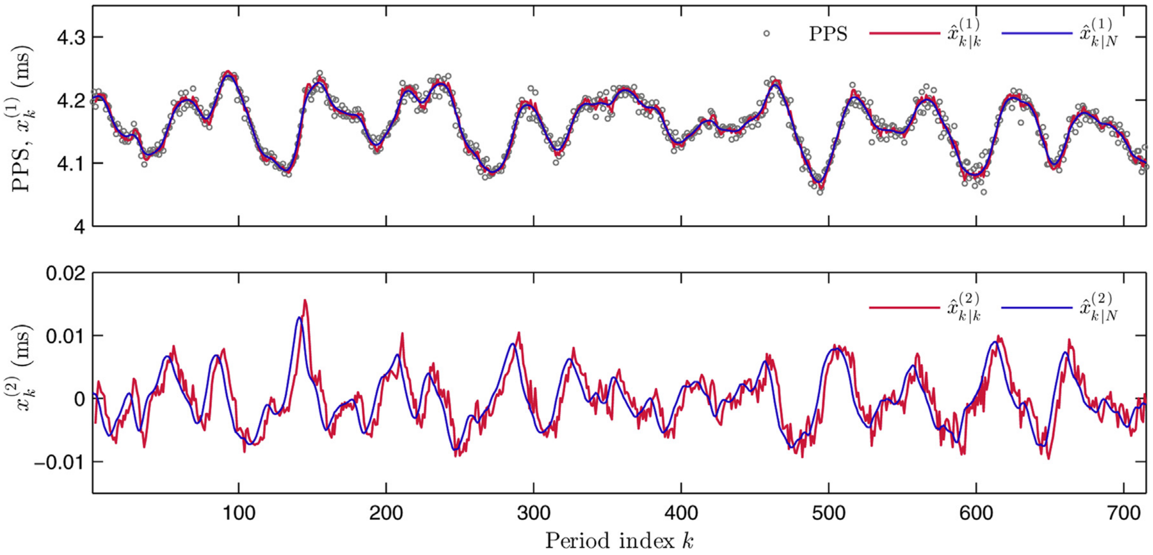

FIGURE 4. Structural analysis of a PPS. Top: PPS, in milliseconds, extracted from a sustained vowel /a/ phonated by a healthy female, along with the filtered $\hat{x}^{(1)}_{k|k}$ and smoothed $\hat{x}^{(1)}_{k|N}$ trend components. Bottom: filtered $\hat{x}^{(2)}_{k|k}$ and smoothed $\hat{x}^{(2)}_{k|N}$ estimates corresponding to stochastic slope in the trend. State estimates were calculated through SSMs and considering GSSM (II).

matrices and are serially and mutually independent. Under these assumptions and the fact that $r = 1$, the standard one-step ahead forecast error[38]:

$$\mathrm{e}_k = \frac{\hat{\mathbf{y}}_k}{\mathbf{F}_k^{1/2}}, \tag{9}$$

is a normally distributed and serially independent univariate sequence, with zero mean and unit variance. Therefore, the standard error (9) becomes a useful instrument to evaluate the capability of GSSM and structural analysis method for accurately representing PPSs. In this work, diagnostic tests are considered to statistically analyze the standard errors e_k obtained from the state space representation of the PPS corpus. Under the null-hypothesis, e_k is a normally distributed, homoscedastic, white random process with zero mean and unit variance.

RESULTS

In this section, SSMs for structural analysis are applied to PPS modeling, and the results are presented. Also, the proposed method is evaluated through statistical and graphical tools.

State space methods

In Figure 2, we show a PPS from a female participant and trend estimates calculated through filtering and smoothing methods, considering the simplest GSSM (I). Notice that this PPS corresponds to a sustained vowel /a/ presenting a fairly stable fundamental period. Furthermore, it can be observed that both filtered and smoothed states are suitable trend component representations, carrying information related to an individual itself and its abilities and skills to phonate.[6] Nevertheless, the flexibility and modeling capabilities of GSSM (I) are extremely limited, and therefore, this model is not suitable for representing the different PPS features presented in the Introduction appropriately.[19]

A different situation can be appreciated in Figure 4. At top, the PPS corresponding to another healthy female is plotted. Although it was also extracted from a sustained vowel, this sequence presents oscillations with abrupt transitions. This is a characteristic phenomenon of persons who are not able to phonate sustained vowels with a stable fundamental period[14] and occurs as a result of the action of control systems, both voluntary and involuntary, acting during phonation.[5] The proposed method was applied to represent this PPS, considering a GSSM (II). Trend components $\hat{x}^{(1)}_{k|k}$ and $\hat{x}^{(1)}_{k|N}$ calculated from filtering and smoothing methods, respectively, are shown superimposed at top of Figure 4. It can be concluded that both estimates represent fluctuations and transitions in the PPS accurately. At the bottom of Figure 4, components $\hat{x}^{(2)}_{k|k}$ and $\hat{x}^{(2)}_{k|N}$ are displayed corresponding to estimates of the stochastic slope in the trend. These estimates properly characterize the dynamics of this PPS and seem to anticipate abrupt transitions. Therefore, this result suggests that slope estimates are useful tools to analyze the stability of a sustained vowel. A time delay between filtered and smoothed states is appreciated in Figure 4. This phenomenon is analyzed in the following.

In Figure 5, structural analysis of a PPS corresponding to a healthy male and considering GSSM (V) is presented. At the top, this PPS is displayed, showing a stable dynamic with small oscillations. Superimposed trend component estimates $\hat{x}^{(1)}_{k|k}$ and $\hat{x}^{(1)}_{k|N}$ obtained through filtering and smoothing methods, respectively, are shown. At center, stochastic slope estimates $\hat{x}^{(2)}_{k|k}$ and $\hat{x}^{(2)}_{k|N}$ are displayed. It can be appreciated that the stability of the analyzed PPS is clearly represented by the combination of trend and slope estimates because trends properly represent the slow long-term dynamics in the PPS, although slope components present near-zero values corresponding to small transitions. At the

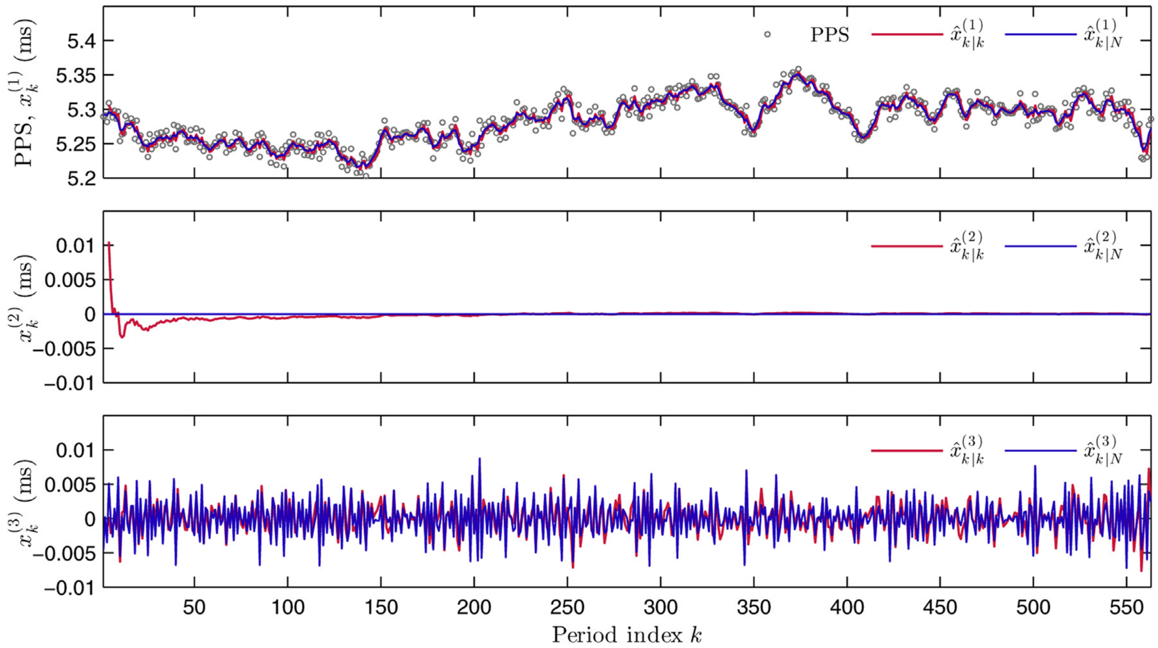

FIGURE 5. Structural analysis of a PPS. Top: PPS, in milliseconds, extracted from a sustained vowel /a/ phonated by a healthy male, along with the filtered $\hat{x}_{k|k}^{(1)}$ and smoothed $\hat{x}_{k|N}^{(1)}$ trend components. Center: filtered $\hat{x}_{k|k}^{(2)}$ and smoothed $\hat{x}_{k|N}^{(2)}$ estimates corresponding to stochastic slope in the trend. Bottom: filtered $\hat{x}_{k|k}^{(3)}$ and smoothed $\hat{x}_{k|N}^{(3)}$ estimates corresponding to the AR cycle component. State estimates were calculated through SSMs and considering GSSM (V).

bottom, filtered $\hat{x}_{k|k}^{(3)}$ and smoothed $\hat{x}_{k|N}^{(3)}$ cycle estimates, according to an $AR(4)$, are shown. By analyzing these sequences, it is noticeable that this PPS presents an AR component reinforcing the temporal correlation. This phenomenon corresponds to the features observed in real PPSs[19] and is properly captured by the proposed method. The remaining elements on filtered and smoothed state vectors do not add further information because they correspond to lagged version of cycle estimates.

As stated before, the PPS in Figure 4 shows a highly oscillatory dynamic, which is explained by trend and slope components. The PPS in Figure 5 instead exhibits a stable dynamic with small oscillations. In this case, the level is represented by the trend and a near-zero slope, whereas the oscillations are explained by the cycle component. Similar decompositions were obtained for the remaining PPSs in the corpus. Therefore, the proposed method seems to be effective and convenient for modeling complex PPS behaviors by means of simple stochastic components capturing the different features that can be involved.

The performance of filtering and smoothing methods can be compared by analyzing the estimates, obtained through these methods, and studying Figures 4 and 5. First, the noncausality in smoothed states can be appreciated in Figure 4. Because of the fact that only past and present information are considered in filtering, filtered states are lagging behind PPS. Smoothing method adds future information in state estimations and, therefore, no lag-time is observed. Second, the smoothing method generates more stable and less fluctuating state estimates than filtering at the expense of an increased computational cost. Third, by analyzing the filtered $\mathbf{P}_{k|k}$ and smoothed $\mathbf{P}_{k|N}$ state covariance matrices, we observed that smoothed states achieve the lowest variance, becoming more accurate estimates, than filtered ones. Finally, filtered states show a transient at the beginning, which is necessary to stabilize the estimation process. Particularly, this phenomenon can be easily appreciated at the center of Figure 5. By contrast, no transients are appreciated in smoothed states. Accordingly, the smoothed estimates are preferred for PPS modeling.

Statistical analysis

As stated previously, when a PPS is correctly modeled, given a GSSM, standard error e_k behaves as white Gaussian noise, with zero mean and constant variance. In this section, capabilities of the proposed method to model the PPS corpus are evaluated. For this purpose, every PPS was analyzed, taking different GSSM, and then the corresponding e_k was calculated. The central third of e_k, approximately 1 second, was considered for the

TABLE 2.
Statistical Analysis of Standard Errors e_k

GSSM	χ^2 (%)	*t* (%)	*H* (%)	*LB* (%)
(I)	86.79	100.00	86.79	41.51
(II)	90.57	100.00	81.13	41.17
(III)	88.68	100.00	83.02	50.94
(IV)	86.79	100.00	86.79	71.70
(V)	90.57	100.00	88.68	79.25
(VI)	86.79	100.00	84.91	84.91
(VII)	90.57	100.00	84.91	86.79

Notes: Percentage of the PPS corpus failing to reject the null hypotheses of normality (χ^2test), zero mean (*t* test), homoscedasticity (*H* test), and whiteness (*LB* test), according to different GSSMs. Significance level α = 0.05 for all tests

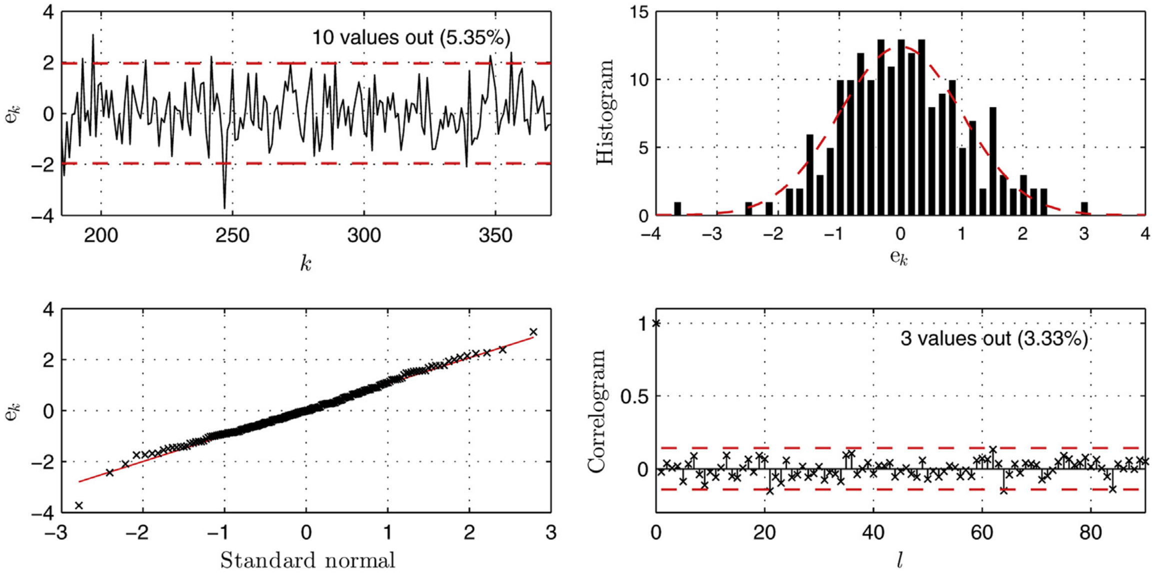

FIGURE 6. Subjective goodness of fit analysis. Top left: standard error e_k, central third, corresponding to PPS in Figure 5, along with a 95% standard Gaussian confidence interval. Top right: histogram of e_k and theoretical standard normal distribution. Bottom left: Normal plot comparing e_k versus the theoretical standard normal distribution. Bottom right: correlogram of e_k for lags 0,…,90, along with a 95% white process confidence interval.

evaluation because most voice signals showed the most stable behavior in this portion.

Different objective statistical tests were considered in the analysis of e_k. Table 2 summarizes the results achieved with these statistical tests according to the GSSMs presented in Table 1. First, nonparametric χ^2 test was applied taking as a null hypothesis that e_k is Gaussian, against the alternative that error is not Gaussian distributed.[42] In the second column of Table 2, the percentage of sequences in the corpus for which e_k fails to reject the null hypothesis is presented. This shows that assumption of Gaussian behavior is appropriate for most sequences, regardless the differences in GSSM structures. Next, zero mean assumption was examined, applying a traditional *t* test with zero mean as the null hypothesis. Third column of Table 2 presents the percentage of signals failing to reject this null hypothesis. It can be deduced that, for every Gaussian e_k, there is not enough evidence to reject the null hypothesis and, therefore, the zero mean assumption is considered appropriate.

Next, the standard error variances were evaluated considering the *H* test presented in the study by Harvey and Shephard,[36] taking homoscedasticity (constant variance) as a null hypothesis, against the alternative of heteroscedasticity. Fourth column of Table 2 presents the percentage of errors e_k which fails to reject the null hypothesis, considering the proposed GSSMs. As a result, it is concluded that homoscedasticity assumption is appropriate for most sequences. Finally, whiteness assumption (no temporal correlation) is analyzed according to the test proposed by Ljung and Box.[43] In this LB test, whiteness is assumed as a null hypothesis, against the alternative of e_k presenting some correlation. In the fifth column of Table 2, percentage of errors e_k of the PPS corpus failing to reject the null hypothesis is also reported. For GSSMs $\{(\mathrm{I}), \ldots, (\mathrm{III})\}$ more than half of error sequences rejected the null hypothesis because of some residual temporal correlation. In contrast, the AR structures incorporated in GSSMs $\{(\mathrm{IV}), \ldots, (\mathrm{VII})\}$ improve the representation capabilities of these models and, therefore, the portion of sequences failing to reject the null hypothesis increases considerably. In view of the previous analysis, it can be concluded that whiteness is the most restrictive assumption in PPS structural analysis.

Subjective methods for goodness of fit analysis were also considered. Graphical representations of standard error e_k allow evaluating the performance of the proposed method for PPS modeling. In Figure 6, different graphical representations are shown, corresponding to the PPS displayed in Figure 5. The top left subfigure shows error e_k (central third) obtained from the filtering method and considering GSSM (V), along with a 95% standard Gaussian confidence interval in dashed lines. Only 5.35% of the data (10 elements) are outside the confidence interval. The histogram of e_k is presented in the top right corner of the figure, along with the theoretical standard normal distribution function in dashed line. The bottom left subfigure shows a normal plot, where error e_k distribution is compared against the standard normal distribution. Here, the solid line represents the ideal situation where both distributions are equal. From these three subfigures, it can be appreciated that the probability distribution of error e_k is similar to the standard normal distribution, except for a few extreme values in both tails of the function, and that zero mean is a valid assumption.

Furthermore, the correlogram of e_k, for lags $l = 0, \ldots, 90$, is displayed in the bottom right subfigure, along with a 95% white process confidence interval in dashed line. Here, only 3.33% of the data (three elements) in the correlogram remains outside the confidence interval; therefore, there is not enough evidence to reject the assumption that the error e_k behaves like a white process. In summary, these graphical methods show that error e_k

TABLE 3.
Percentage of the PPS Corpus Correctly Modeled With the Proposed Methodology, According to Different GSSMs

GSSM	$\alpha = 0.05$ (%)	$\alpha = 0.01$ (%)
(I)	32.08	45.28
(II)	37.74	58.49
(III)	41.51	54.72
(IV)	54.72	73.58
(V)	66.04	81.13
(VI)	64.15	81.13
(VII)	69.81	81.13

Notes: Significance levels $\alpha = 0.05$ and $\alpha = 0.01$ were considered

TABLE 4.
Accumulated Percentage of the PPS Corpus Correctly Modeled With the Proposed Methodology, According to Progressive Increases in GSSMs Complexity ($p + q + r$)

Complexity	GSSM Set	$\alpha = 0.05$ (%)	$\alpha = 0.01$ (%)
3	{(I)}	32.08	45.28
4	{(I),(II)}	45.28	64.15
5	{(I),…,(III)}	47.17	64.15
8	{(I),…,(IV)}	66.04	77.36
10	{(I),…,(V)}	73.58	84.91
12	{(I),…,(VI)}	75.47	84.91
14	{(I),…,(VII)}	75.47	86.79

Notes: Significance levels $\alpha = 0.05$ and $\alpha = 0.01$ were considered.

can be considered a white normal process, white zero mean and unit variance, confirming that the PPS under analysis is correctly modeled by the proposed method, taking GSSM (V). These graphical methods allowed to further understand the results supplied by the statistical analysis presented in Table 2.

Table 3 presents the percentage of PPSs in the corpus failing to reject the null hypothesis for all statistical tests of Table 2 at the same time, depending on the GSSMs proposed. In this analysis, significance levels $\alpha = 0.05$ and $\alpha = 0.01$ were considered, where the first significance level is more restrictive than the second one. It can be appreciated that by increasing GSSM complexity, improving its modeling capabilities, more sequences could be correctly modeled but at the cost of considerable increasing the computational requirements for the SSMs. Moreover, when GSSM complexity increases, there are noticeable differences in the performances between GSSMs (III), (IV), and (V), but for the remaining models, these differences are not so evident. This phenomenon demonstrates the existence of a cycle component in most real PPSs, which are not correctly modeled by simple GSSMs. Nevertheless, this dynamic can be suitable represented by applying more versatile GSSMs.

In the previous analysis, the performance of each model was individually evaluated. Nevertheless, this experiment did not consider that some PPSs were correctly modeled only with one particular GSSM, whereas its representation became worse when the model changed (even with more complex GSSMs). Considering this situation, Table 4 introduces the accumulated percentage of correctly modeled PPSs, depending on increasing in GSSM complexity. Here, significance levels $\alpha = 0.05$ and $\alpha = 0.01$ were considered. In this context, complexity is given by the maximum total dimension $(p + q + r)$ in the embedded GSSM sets, presented in second column in Table 4. This parameter indicates the capability of each set to correctly represent PPSs. As expected, the percentage of correctly represented PPSs, for a given complexity level, is higher than the percentage provided by the previous analysis showed on Table 3. This occurs because each correctly modeled sequence is accumulated, regardless the GSSM. In addition, it can be appreciated, once again, the markedly difference in the performances between GSSMs (III), (IV), and (V) because of the incorporation of an AR component in the GSSM structure.

Finally, last row in Table 4 shows the total percentage of PPSs in the corpus correctly modeled in this work. It can be seen that more than 75% considering $\alpha = 0.05$, or more than 86% with $\alpha = 0.01$, of the corpus was correctly modeled. This result provide enough evidence to concluded that structural analysis, based on SSMs, is a suitable method for real PPS modeling.

CONCLUSIONS

In this work, we have proposed a state space-based method for the structural representation of perturbed PPSs. Stochastic Gaussian linear state space models were developed, considering the principal features observed on real period sequences and different complexity levels.

For each sequence, structural trend, slope, and cycle components were optimally estimated using the proposed approach. Although trend and slope estimates proved to properly represent slow long-term dynamics which are normally associated with prosody or intonation, cycle estimate preserved the AR component present in the sequences. The obtained results suggest that stochastic components could be associated with events occurring in phonation. In particular, the trend captures adaptation mechanisms in pitch periods, whereas cycle components effectively represent microtremors or any other phenomenon reinforcing temporal correlation. Objective statistical tests demonstrated that most signals in the considered corpus could be correctly represented. Therefore, the proposed method has shown to be a suitable strategy for perturbed PPSs modeling, allowing to generate statistical estimates carrying information related to the phonation process.

Specialists have extensively argued that short-term random period disturbances are important in clinical applications, especially in presence of pathologies. We have shown that the state space models become suitable theoretical frameworks to delve into this assertion because these models incorporate random perturbations into their structure. Future work will focus on optimal estimation and analysis of period perturbations, considering both healthy and pathologic signals. Also, the incorporation of the proposed method in natural voice synthesis systems to improve the perceptual quality of artificial voices will be considered.

Acknowledgment

This work was supported by the Argentinian institutions UNER under grant PID UNER 6136, CONICET, and ANPCyT under grant PICT-2012-2954. The authors would like to thank Dr. Paola Catalfamo Formento at the Universidad Nacional de Entre Ríos, Argentina, for her helpful advice in the manuscript preparation.

REFERENCES

1. Leong K, Hawkshaw MJ, Dentchev D, Gupta R, Lurie D, Sataloff RT. Reliability of objective voice measures of normal speaking voices. *J Voice*. 2013;27:170–176.
2. Kreiman J, Gerratt BR. Perception of aperiodicity in pathological voice. *J Acoust Soc Am*. 2005;117:2201–2211.
3. Titze IR. *Workshop on Acoustic Voice Analysis: Summary Statement*. Denver, USA: National Center for Voice and Speech; 1995.
4. Bonilha HS, Deliyski DD. Period and glottal width irregularities in vocally normal speakers. *J Voice*. 2008;22:699–708.
5. Titze IR. *Principles of Voice Production*. 2nd ed. Iowa, USA: National Center for Voice and Speech; 2000.
6. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech and Voice*. San Diego, USA: Singular Thomson Learning; 2000.
7. Velasco García MJ, Cobeta I, Martín G, Alonso-Navarro, Jimenez-Jimenez FJ. Acoustic analysis of voice in Huntington's disease patients. *J Voice*. 2011;25:208–217.
8. Silva DG, Oliveira LC, Andrea M. Jitter estimation algorithms for detection of pathological voices. *EURASIP J Adv Sig Pr*. 2009;2009:1–9. 9.
9. Titze IR, Liang H. Comparison of F0 extraction methods for high-precision voice perturbation measurements. *J Speech Hear Res*. 1993;36:1120–1133.
10. Farrus M, Hernando J. Using jitter and shimmer in speaker verification. *IET Signal Process*. 2009;3:247–257.
11. Govind D, Mahadeva Prasanna SR. Expressive speech synthesis: a review. *Int J Speech Technol*. 2013;16:237–260.
12. Schoentgen J, De Guchteneere R. Time series analysis of jitter. *J Phonetics*. 1995;23:189–201.
13. Fraile R, Kob M, Godino-Llorente JI, Sáenz-Lechón N, Osma-Ruiz VJ, Gutiérrez-Arriola JM. Physical simulation of laryngeal disorders using a multiple-mass vocal fold model. Biomed. *Biomed Signal Process Contr*. 2012;7:65–78.
14. Ruinskiy D, Lavner Y. Stochastic models of pitch jitter and amplitude shimmer for voice modification. In: *Proc IEEEI 2008*. IEEE; 2008:p. 489–493; Eliat, Israel.
15. Schlotthauer G, Torres ME, Rufiner HL. Pathological voice analysis and classification based on empirical mode decomposition. In: Esposito A, Campbell N, Vogel C, Hussain A, Nijholt A, eds. *Development of Multimodal Interfaces: Active Listening and Synchrony*. Heidelberg: Springer Berlin; 2010:364–381.
16. Cabral JP, Oliveira LC. Emovoice: a system to generate emotions in speech. In: *INTERSPEECH 2006*. Pittsburgh, PA, USA: International Speech Communication Association; 2006.
17. Wang L, Li A, Fang Q. A method for decomposing and modeling jitter in expressive speech in Chinese. In: *Proc. of Speech Prosody*. Dresden, Germany: International Speech Communication Association; 2006.
18. Alzamendi GA, Schlotthauer G, Rufiner HL, Torres ME. Evaluation of a new model for vowels synthesis with perturbations in acoustic parameters. *Lat Am Appl Res*. 2013;43:1–6.
19. Schoentgen J. Stochastic models of jitter. *J Acoust Soc Am*. 2001;109:1631–1650.
20. Endo Y, Kasuya H. A stochastic model of fundamental period perturbation and its application to perception of pathological voice quality. In: *Proc. of Fourth Int. Conference on Spoken Language ICSLP, 1996*. Philadelphia, PA, USA: International Speech Communication Association; 1996:p. 772–775.
21. Schoentgen J, De Guchteneere R. Predictable and random components of jitter. *Speech Comm*. 1997;21:255–272.
22. Fraj S, Grenez F, Schoentgen J. Synthetic hoarse voices: a perceptual evaluation. In: *Int. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications-MAVEBA*. Florence, Italy: Firenze University Press; 2009:p. 95–98.
23. Fraj S, Schoentgen J, Grenez F. Development and perceptual assessment of a synthesizer of disordered voices. *J Acoust Soc Am*. 2012;132:2603–2615.
24. DeJonckere P, Schoentgen J, Giordano A, Fraj S, Bocchi L, Manfredi C. Validity of jitter measures in non-quasi-periodic voices. Part I: perceptual and computer performances in cycle pattern recognition. *Logoped Phoniatr Vocol*. 2011;36:70–77.
25. Manfredi C, Giordano A, Schoentgen J, Fraj S, Bocchi L, Dejonckere P. Validity of jitter measures in non-quasi-periodic voices. Part II: the effect of noise. *Logoped Phoniatr Vocol*. 2011;36:78–89.
26. Titze IR, Story BH. Rules for controlling low-dimensional vocal fold models with muscle activation. *J Acoust Soc Am*. 2002;112:1064.
27. Aoki N, Ifukube T. Analysis and perception of spectral 1/*f* characteristics of amplitude and period fluctuations in normal sustained vowels. *J Acoust Soc Am*. 1999;106:423–433.
28. Zhang Y, Jiang JJ, Biazzo L, Jorgensen M. Perturbation and nonlinear dynamic analyses of voices from patients with unilateral laryngeal paralysis. *J Voice*. 2005;19:519–528.
29. Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab. Disordered Voice Database; 2009.
30. Boersma P, Weenink D. Praat: Doing Phonetics by Computer; 2013.
31. Kalman RE. A new approach to linear filtering and prediction problems. *Trans ASME J Basic Eng*. 1960;82:35–45.
32. Garner PN, Cernak M, Motlicek P. A simple continuous pitch estimation algorithm. *IEEE Signal Process Lett*. 2013;20:102–105.
33. Mehta DD, Rudoy D, Wolfe PJ. Kalman-based autoregressive moving average modeling and inference for formant and antiformant tracking. *J Acoust Soc Am*. 2012;132:1732–1746.
34. Jinachitra P, Smith III JO. Joint estimation of glottal source and vocal tract for vocal synthesis using Kalman smoothing and EM algorithm. In: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005*. New Palts, NY, USA: IEEE; 2005:p. 327–330.
35. Li H, Scaife R, O'Brien D. LF model based glottal source parameter estimation by extended Kalman filtering. In: *Proc. of the 22nd IET Irish Signals and Systems Conference*. Dublin, Ireland: IET; 2011.
36. Harvey AC, Shephard N. Structural time series models. In: Maddala GS, Rao CR, Vinod HD, eds. *Econometrics*. North Holland, Amsterdam: Elsevier Science Publishers; 1993:261–302.
37. Durbin J, Koopman SJ. *Time Series Analysis by State Space Methods*. 1st ed. New York, USA: Oxford Univ Pr (Sd); 2001.
38. Koopman SJ, Ooms M. Forecasting economic time series using unobserved components time series models. In: Clements MP, Hendry DF, eds. *The Oxford Handbook of Economic Forecasting*. Oxford: Oxford University Press; 2011:129–162.
39. Koopman SJ, Durbin J. Filtering and smoothing of state vector for diffuse state-space models. *J Time Anal*. 2003;24:85–98.
40. Koopman SJ. Exact initial Kalman filtering and smoothing for nonstationary time series models. *J Am Stat Assoc*. 1997;92:1630–1638.
41. Shumway RH, Stoffer DS. An approach to time series smoothing and forecasting using the EM algorithm. *J Time Anal*. 1982;3:253–264.
42. Kvam PH, Vidakovic B. *Nonparametric Statistics with Applications to Science and Engineering*. New Jersey, USA: John Wiley & Sons; 2007.
43. Ljung GM, Box GE. On a measure of lack of fit in time series models. *Biometrika*. 1978;65:297–303.