Journal of Phonetics 54 (2016) 80–97

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Technical Article

Characterization of laryngealization as irregular vocal fold vibration and interaction with prosodic prominence

Leonardo Lancia [a,*], Daniel Voigt [b], Georgy Krasovitskiy [c]

[a] Aix-Marseille Université, CNRS, LPL-UMR7309, Aix-en-Provence, France
[b] Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[c] University of Oxford, Oxford, UK

ARTICLE INFO

Article history:
Received 2 September 2014
Received in revised form
29 July 2015
Accepted 24 August 2015

Keywords:
Vocal-fold vibration
Irregularity
Laryngealization
Recurrence analysis
Wavelet-based functional mixed models

ABSTRACT

This paper introduces an original variant of recurrence analysis to quantify the degree of regularity of vocal fold vibration as captured by electroglottography during phonation. The proposed technique is applied to the analysis of laryngealized phonation as this phonation type typically shows irregular vibration cycles. The reliability of this approach is validated with synthetic vocal fold vibration signals, demonstrating that it permits measuring the regularity of vocal fold vibration, unaffected by changes in fundamental frequency. The method is also applied to real electroglottographic signals recorded at the onset of vowel-initial nonsense words produced in a speeded repetition task by five female German speakers. Results show that the degree of laryngealization during the production of word-initial vowels is modulated by the presence of stress (with stressed vowels being less laryngealized). Due to its robustness to changes of *F*0, the proposed technique proves to be a suitable tool for studying vocal fold regularity in concatenated speech. Its applications are not limited to the study of glottalization, since the degree of regularity of vocal fold vibration has paralinguistic functions and is a clinically relevant measure of voice pathologies.

1. Introduction

Voicing is produced by a complex interplay of the articulatory organs, typically resulting in repeated opening and closure of the two vocal folds (Fant, 1971; Titze, 1994). These oscillation cycles are subject to random variations with regard to their shape, period length and amplitude. If these deviations are relatively small and the corresponding waveform displays quasi-periodic properties, the quality of the voice is commonly referred to as “modal” (Laver, 1980). Along with this phonation type, other categories exist which have been subsumed under the collective term “non-modal phonation” (Gordon & Ladefoged, 2001). These are characterized by certain acoustic, physiological and perceptual differences from the modal phonation state. Although potential articulatory and acoustic dimensions of variation linked to the perception of different voice qualities have been identified in the literature, conclusive quantitative characterizations of these diverse voice qualities are yet to be developed. Gordon and Ladefoged (2001) propose, for example, that voice qualities can be distinguished at the articulatory level on a continuum “defined in terms of the aperture between the arytenoid cartilages” which is maximal in the production of voiceless sounds, decreases in breathy voicing, decreases further in modal voicing and is even smaller in creaky voicing. Given that the vocal folds are expected to spend more time in contact when the arytenoid cartilages are adducted, under the hypothesis proposed by Gordon and Ladefoged it is reasonable to characterize quantitatively different voice qualities via articulatory or acoustic measures that depend on the open quotient of the voicing cycles (i.e. the duration of the time interval during which the vocal folds are open normalized with respect to the duration of the cycle). However, the results reported by Gerratt and Kreiman (2001) suggest that a multidimensional characterization based on various articulatory and/or acoustic features may better reflect listeners' perception of differences between voice qualities. Other authors propose that variation in the regularity of voicing can differentiate voice qualities both in linguistic and paralinguistic terms and can be a

* Corresponding author.
E-mail addresses: leonardo.lancia@blri.fr (L. Lancia), daniel_voigt@eva.mpg.de (D. Voigt), georgy.krasovitskiy@wolfson.ox.ac.uk (G. Krasovitskiy).

0095-4470/$ - see front matter © 2015 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2015.08.001

distinguishing feature between normal and pathological voice conditions, e.g., Parkinson's disease (Little, McSharry, Roberts, Costello, & Moroz, 2007) or Reinke's edema (Lim, Choi, Kim, & Choi, 2006; Matar, Portes, Lancia, Legou, & Baider, 2014).

A particular type of non-modal voice, usually characterized by irregular vocal fold vibration is laryngealized voicing, also known as laryngealization or glottalization (Blankenship, 2002). Glottalization is commonly associated with a constriction at the glottal level, occasionally accompanied by a general constriction of the larynx (often leading to adduction of the ventricular folds) and a narrowing of the pharyngeal stricture. This typically results in voicing cycles displaying longer closed phases and shorter open phases. However, glottal constriction can also lead to damping of the voice signal, cycle-to-cycle amplitude modulations and diplophonia (Edmondson & Esling, 2006). Extreme constriction causes a complete interruption of the voicing signals and, when released, produces a burst of acoustic energy characteristic of a plosive consonant that is called a glottal stop.

Laryngealization can express specific characteristics or emotional states of the speaker (Gobl & Ní Chasaide, 2003; Henton & Bladon, 1987). As a vocalic feature, it has a distinctive function in the phoneme inventories of various languages (Gordon, 2012; Ladefoged & Maddieson, 1996), while in others it can signal features of adjacent phonemes (e.g. Kohler, 1994 for German; Gobl & Ní Chasaide, 1999 for American English; Cho, Jun, & Ladefoged, 2002 for Korean). When oral stops occur between voiced sounds (e.g. vowels or nasal consonants) in colloquial speech, these can be replaced by a glottal stop (Kohler, 1994 for German, Docherty & Foulkes, 2005 for British English and Roberts, 2006 for American English varieties). Furthermore, laryngealization can characterize prosodic and morphemic boundaries (e.g. Kohler, 1994 and Pompino-Marschall, & Żygis, 2011 for German; Dilley, Shattuck-Hufnagel, & Ostendorf, 1996 for American English; Garellek, 2014, for American English and Spanish and Fougeron, 2001 for French). In these studies, vowel-initial words are reported to be often preceded by a glottal stop or produced with a degree of laryngealization. Moreover, it has been shown for American English that the likelihood of word-initial laryngealization increases in words produced at the left boundary of prosodic constituents, and that it increases with the position of the prosodic constituent in the prosodic hierarchy (cf. Dilley et al., 1996; Garellek, 2014). A common assumption in the literature is that prosodically conditioned laryngealization signals interruptions of the speech chain, thus aiding segmentation and lexical access (cf. Dilley et al., 1996; Slifka, 2006 and references therein). Moreover, laryngealization introduces limits on the control of *F*0 and consequently produces complex interactions with other phonological features such as lexical tones (e.g. Silverman, 1997 for Mixtec; Lancia, Avelino, & Voigt, 2013 for Yalálag Zapotec) or vowel quality (e.g. Pompino-Marschall, & Żygis, 2011 and Lancia & Grawunder, 2014 for German).

The principal aim of this paper is the illustration of a new method to quantify the regularity of voicing. Unlike the extant approaches, our method is insensitive to modulations of *F*0, robust to changes in the configuration of the oral articulators and does not rely on the detection of individual oscillation cycles. It is thus well suited to the analysis of phonation regularity in connected speech with continuously changing *F*0, and enables the analysis of voicing characterized by high degrees of irregularity, where the vocal fold vibration cycles are difficult to identify. The resulting measure is applied to characterize laryngealized phonation. Even though it is commonly acknowledged that this phonation type results in less regular vocal fold vibration than modal voicing, it has been proposed that the distinguishing feature of laryngealization is an increase in complexity (Berry, 2001) involving laryngeal and supralaryngeal structures (e.g., ventricular folds, tongue root, pharynx walls). We argue, however, that the complex yet still relatively regular vibratory patterns reported in many studies of laryngealized voice (e.g. Berry, 2001; Neubauer, Mergell, Eysholdt, & Herzel, 2001, Moisik & Esling, 2014), though found in sustained phonation, are less common in connected speech. While the former condition provides sufficient time for the articulators to be stably set up in the configuration required for laryngealization, in the latter, this process is naturally transient and the resulting phonation exhibits greater irregularity in the shape and duration of the oscillation cycles (cf. Slifka, 2006) compared to modal voicing.

In the next section we set out the theoretical motivation and describe the proposed recurrence-based method. The method is validated through the analysis of synthetic signals mimicking vocal fold vibration as recorded with electroglottography (EGG), where amplitude and temporal variability are systematically altered. In the following section (Section 3) the proposed measure is applied to quantify vowel laryngealization in real EGG signals recorded while German speakers repeated, at sustained speech rate, vowel-initial disyllabic nonsense words with stress on the initial or on the final syllable. We also investigate the interactions between laryngealization of the vowels surrounding the glottal constriction and the presence of prosodic prominence on those vowels.

2. Measuring regularity in vocal fold vibration

2.1. Background

Numerous approaches and techniques have been proposed for the study of non-modal voice qualities, each with its own objectives and assumptions. Several quantitative measures derived from acoustic speech recordings (e.g.: H1–H2, H1–A3, etc.) rest on the assumption that a shorter closed phase of the vocal fold oscillation cycle produces an increase in spectral tilt (Hanson, Stevens, Kuo, Chen, & Slifka, 2001). However, these measures need to be corrected according to formant values and bandwidths (Iseli & Alwan, 2004), and are therefore affected by artifacts due to formant extraction. Moreover, most acoustic methods focus on a single feature of laryngealization, namely the asymmetry between the open and the closed phases of the voicing cycle, disregarding additional articulatory characteristics which may potentially affect the signal. Other acoustic measures, such as jitter and shimmer, assume relatively long episodes of stationary voice signal (Karnell, 1991; Scherer, Vail, & Guo, 1995), and are therefore only of limited utility in the analysis of running speech. Another method of quantifying laryngealization is based on the analysis of the oral airflow as measured with a pneumotach mask (Slifka, 2006; Rothenberg, 1977). This technique is based on the idea that the glottal

airflow can be reconstructed by removing from the measured airflow signal the effects of the filter through inverse filtering (Rothenberg, 1973). In practice the reconstructed signal is expected to deviate from the actual glottal airflow because the signals analyzed are not stationary and because the effects of sub-glottal interactions are not completely removed from the signal (cf. Löfqvist, Koenig, & McGowan, 1995).

In addition, EGG signals can also be used to capture the relevant characteristics of vocal fold vibration (Fabre, 1958; Baken, 1992). This approach is noninvasive, independent of the configuration of the oral articulators, economical and simple to implement. However, most extant EGG-based methods rely on automatic detection of *F*0 and/or on the detection of precise articulatory events in the individual voicing cycles (Rothenberg & Mahshie, 1988; Lucero & Koenig, 2000; Mooshammer, 2010).

This may prove problematic when analyzing irregular voicing cycles, as the presence of irregularity and the potential increase in complexity (for instance due to the interaction between vocal and ventricular folds) hinders the detection of *F*0 and disrupts the usual relations between properties of the articulatory system and properties of the EGG signal (e.g. the correspondence between the point at which vocal fold opening occurs and the point at which the negative peak of the first derivative of the EGG signal is observed). A review of the methods traditionally employed to quantify regularity in vocal fold vibration is offered by Lucero and Koenig (2000).

In what follows, we present a novel quantitative method to reliably measure regularity of vocal fold vibration as captured by EGG. The proposed technique is based on a variant of Recurrence Quantification Analysis (RQA), originally developed to study features of multidimensional dynamics underlying the shape of quasi-periodic univariate trajectories and applied to various dynamical systems to examine the regularity of their behavior (Marwan, Romano, Thiel, & Kurths, 2007; Webber & Zbilut, 1994); a related method based on estimating the recurrence of the dynamics underlying acoustic voicing signals was proposed by Little, McSharry, Roberts, Costello, and Moroz (2007). The implementation proposed by these authors, however, rests on the assumption that the analyzed waveform is stationary, i.e., making it applicable only to sustained phonation. To overcome this limitation, an original version of Recurrence Analysis was recently proposed (Lancia, Fuchs, & Tiede, 2014; Lancia, Avelino, & Voigt, 2013) and is further improved and systematically tested in the present paper.

2.2. Recurrence Quantification Analysis

RQA (Marwan, Romano, Thiel, & Kurths, 2007) provides a specific set of tools to study the recurrent behavior of a nonlinear dynamical system. When applied to the analysis of voicing cycles, this method does not depend on the localization of the onsets and offsets of vocal fold vibration cycles and has thus been applied to acoustic voicing signals in which the detection of these time-points is often unreliable due to the complex subharmonic structure and significant irregularity (Little et al., 2007). With voicing signals, however, this approach is sensitive to changes of *F*0 and to changes in the position of the oral articulators. Consequently, it can only be reliably applied to sustained phonation (Little et al., 2007). Although changes in the articulator position do not constitute an issue when analyzing EGG signals, the issue of variable *F*0 remains. Building on Lancia et al. (2013, 2014), we modify RQA to make it robust to *F*0 modulation and thus enable voicing analysis in connected speech.

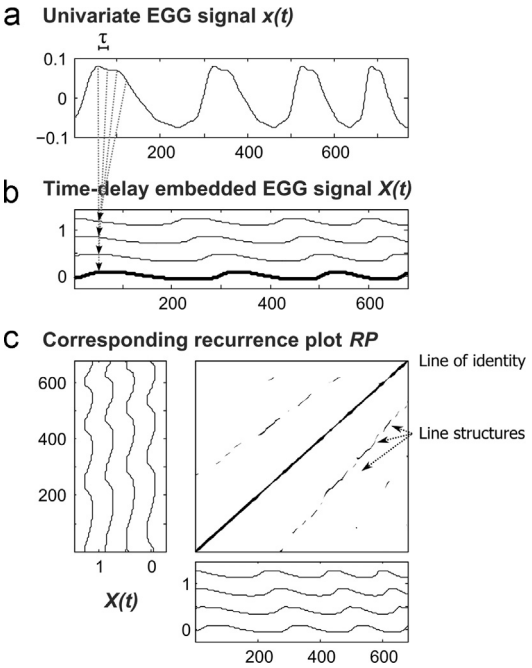

Fig. 1. Individual RQA steps of constructing an RP from an EGG signal. Panel (a): Original univariate EGG time series. Panel (b): Time-delay embedding of the signal scaled to standard deviation of 1 and centered around its mean value ($\tau = 30$, $M = 4$). Panel (c): Resulting RP obtained with $\varepsilon = 0.1$ with different line structures indicating recurrent portions of the trajectory. Isolated dots indicate random matches between values. The reconstructed time series is plotted along the sides of the RP.

2.2.1. Basic RQA algorithm applied to EGG signals

Dynamical systems are mathematical abstractions that permit modeling of processes which affect one or more variables changing over time. These variables constitute the dimensions of the system's state space. At each instant, the system is fully characterized by its position in this space and by a law of change by which the current state of the system depends on its previous states. Vocal fold vibration may be considered as the product of the interaction between variables which, in practice, cannot be directly measured during speech production. Strictly speaking, EGG signals represent electrical conductance values dependent in some linear or nonlinear way on the behavior of the multidimensional system driving the behavior of the vocal folds (Rothenberg & Mahshie, 1988). This is a common problem in the study of real dynamical systems; the classical solution is to reconstruct a multidimensional time series, assumed to be equivalent to that of the underlying system in its state space, on the basis of the observed univariate signal. Starting from the univariate EGG signal $x(t)$, we can reconstruct such a multidimensional time series $X(t)$ by stacking a given number of time-delayed copies of the original time series (Takens, 1981). On the first dimension, the value of the reconstructed time series X at time t corresponds to the value of the observed time series x at the same time, while the value on the ith dimension at time t corresponds to the value of the observed time series at time $t + (i - 1)\cdot\tau$, where τ is the delay parameter (see Fig. 1b). If the reconstructed trajectory has M dimensions, the vector formed by the M values observed at time t is called a *state vector* as it represents the system's state at that point in time. In order to perform time-delay embedding, we must specify (i) the number of dimensions M and (ii) the time-lag τ. The number of dimensions M should be equal to the double of the number of underlying dimensions plus one (for the noise component, cf. Kantz & Schreiber, 2004). This quantity is usually determined with a *false nearest neighbors* (FNN) algorithm (Kennel, Brown & Abarbanel, 1992), giving the maximum number of dimensions that result in a small number of false nearest neighbors states. False nearest neighbors are defined as state vectors that are nearest neighbors only when embedding the time series with a too small number of dimensions. The time lag τ is chosen in such a way as to minimize the dependence between the reconstructed dimensions. To this end, the mutual information function between the original signal and its time-delayed copies is computed. The time lag is set equal to the delay at which the first minimum of the mutual information function is observed. The application of these methods to the time series in Fig. 1a yields $M = 4$ and $\tau = 30$ resulting in the reconstructed time series in Fig. 1b.

Given a reconstructed multidimensional time series $X(t)$, the portions of the trajectory repeated over time are represented graphically in a recurrence plot (RP), a square graph with a side length equal to the signal length T (cf. Fig. 1c). A black dot at the RP coordinate (i, j) indicates that the state of the system at time $t = i$ is close to its state at time $t = j$. While an isolated dot can be considered a random match, a diagonal line (a continuous line with slope 1) indicates that a portion of the waveform is repeated. If

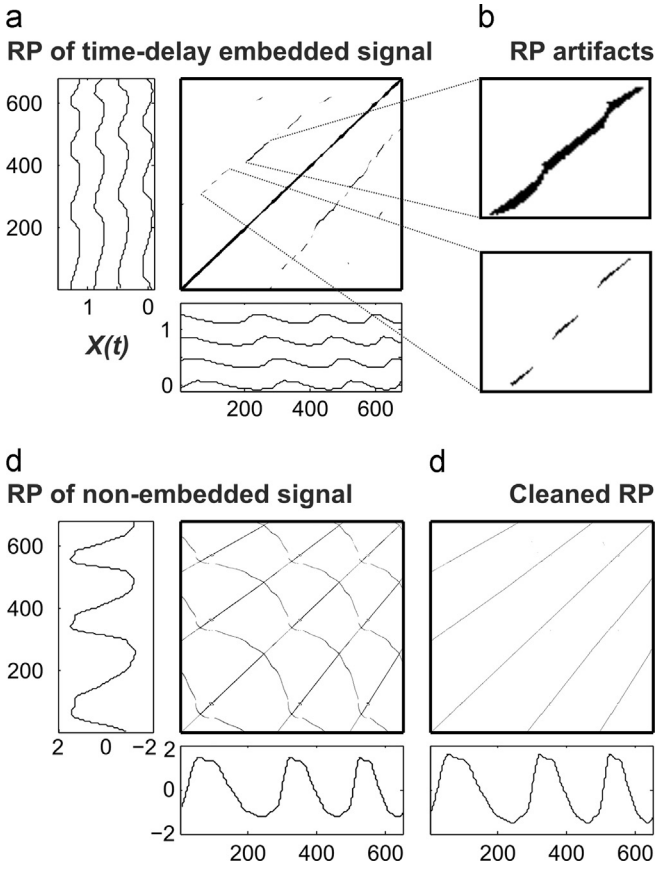

Fig. 2. RP artifacts caused by embedding and changes in *F*0. Panel (a): Same RP as in Fig. 1. Panel (b): Two RP regions magnified to show differences in the thickness of different line structures and deviations from a straight line with slope 1. Panel (c): RP of the same signal without embedding – when the number of embedding dimensions is too small (in this case it is equal to 1, line structures with negative slope appear). Panel (d): RP resulting from the cleaning algorithm presented in Lancia et al. (2013) applied to the RP from panel (c).

such a line is found between coordinates (i, j) and $(i + \delta, j + \delta)$, the section spanning from time $t = i$ to time $t = i + \delta$ is considered equal to that between time $t = j$ and $t = j + \delta$. The main diagonal of the RP contains the results of the comparison of each state vector with itself $(i = j)$, and the vertical distance between each dot and the main diagonal represents the time interval between occurrences of two similar state vectors and is referred to as *return time*.

The RP is derived from the reconstructed time series $X(t)$ by comparing each state vector to all the other state vectors of the trajectory, employing a certain distance measure (Euclidian or maximum norm are commonly used) and a comparison criterion. In the simplest comparison criterion, two state vectors are considered equal (yielding a black dot on the RP) if the distance obtained is lower than a fixed threshold ε. The comparison operation can be expressed as

$$RP_{i,j} = \theta(\varepsilon - \|X(i), X(j)\|)$$

where $RP_{i,j}$ is the value at RP coordinate (i, j), $\theta(\cdot)$ is a Heaviside step function – equal to one if its argument is positive and equal to zero in any other case – and $\|X(i), X(j)\|$ is a distance function (or norm) between the state vectors $X(i)$ and $X(j)$. If the *fixed amount of neighbors* (FAN) criterion is adopted, a different threshold ε is applied to each RP column so that the total number of dots per column is fixed (and corresponds to a given proportion of the length of the time series); if the *fixed recurrence rate* (FRR) criterion is adopted, the threshold remains the same for all columns and is determined in such a way that the total number of RP dots matches a predefined value (corresponding to a given proportion of the squared length of the time series).

The main quantitative indices are derived from counts of dots located on diagonal lines (Marwan et al., 2007). The measure of repeatability of the waveform, known as *percentage of determinism* (%DET) is given by the number of such dots expressed as a percentage of the total number of dots. Besides %DET, the mean and maximum lengths of the diagonal lines are often considered too (cf. Marwan et al., 2007).

2.2.2. Problems with voicing signals

Applying the basic RQA algorithm described above to EGG signals from connected speech raises three problems related to changes in *F*0 and in the dimensionality of the underlying signals, which directly affect the reliability of the derived measures. The first issue concerns the slopes of the resulting lines in the RP. The slope depends on the ratio between the underlying signal's rate of change during different repetitions of the same basic shape, and curved lines are produced when the acceleration of one of the two repetitions changes over time (Facchini, Kantz, & Tiezzi, 2005; Marwan & Kurths, 2005; Marwan, 2011). However, since classic RQA indices are derived from the lengths of straight lines with slope 1, these measures fail to adequately capture potential recurring patterns if time-varying *F*0 is introduced. In practical terms, this means that the resulting RQA indices underestimate the actual amount of recurrence when *F*0 varies. The second issue is related to the thickness of the RP line structures (see Fig. 2b). Assuming that the state vector $X(t)$ is equal to $X(t + dt)$ and the reconstructed time series X changes slowly, it is likely that $X(t)$ is also equal to $X(t + dt + 1)$, as $X(t + dt) \approx X(t + dt + 1)$. This implies the presence of a dot not only at coordinate $(X(t), X(t + dt))$, but also at $(X(t), X(t + dt + 1))$. With decreasing *F*0, the lines in the RP therefore tend to become thicker and the derived RQA measures depending on the number of dots in the RP tend to overestimate recurrence. The third issue arising from the application of RQA to EGG signals of connected speech concerns the parameterization of the time-delay embedding procedure. The delay parameter τ depends on the signal's frequency of oscillation. Consequently, in the presence of *F*0 modulations, parameter values which may be appropriate for one portion of the signal may prove inappropriate for another portion of the same signal. The effect of *F*0 modulations on time delay-embedding introduces gaps in the line structures present in the RP (such as those visible in the bottom panel of Fig. 2b). Due to time-delay embedding, information observed at different time points in the original univariate time series is grouped in the same state vector of the reconstructed time series. In the original time series, the time points grouped in the same state vector are spaced by a constant interval (τ). If the period of the repeated cycles is constant, the observed values will be grouped consistently through the repetitions of the same basic cycle. However, if the period changes from one cycle to another, values which are grouped in the same state vector when observed in one period will be split in different state vectors when observed in the following periods. As a result, with changing *F*0 the obtained state vectors will not match across periods, even if the cycles are repeated perfectly.

Finally, the number of dimensions M to be reconstructed through time-delay embedding may also change over time. As mentioned above, non-modal vocal fold vibration modes are characterized by increased complexity compared to modal voicing (Berry, 2001). The impressionistic observations by Edmonson and Esling (2006) show that different portions of the larynx may interact with the vocal folds to produce different voice qualities and the configuration of the larynx may also vary within the same voice quality, meaning that the dimensionality of the underlying dynamics may be subject to change over time. Consequently, an appropriate number of dimensions to reconstruct the underlying trajectory cannot be determined as that number would vary along the signal.

2.2.3. Modified RQA algorithm and new regularity measure

One potential solution to the problems set out above is based on the observation that an RP obtained from a univariate signal without performing time-delay embedding contains all the information present in the corresponding RP of the embedded multidimensional time series reconstructed from the same signal (Iwanski & Bradley, 1998). However, a side effect of omitting the time-delay embedding is an increased variability in the thickness of the line structures, meaning that new spurious dots which do not represent true repetitions of the signal may appear in the RP. Additionally, downward-pointing line structures (moving left to right) are expected when the RP represents a cyclic unidimensional trajectory (see Fig. 2c). The presence of these line structures indicates that a portion of the signal is similar to another portion of the same signal time-reversed. Since a unidimensional cyclic trajectory can only

move up and down, the first half of a cycle is similar not only to the first half of the following cycle, but also to the time-reversed second half of the same cycle.

Lancia, Fuchs, and Tiede (2014) have shown that such artifacts can be reliably identified and removed from the RP through an automatic procedure, leaving only the dots which represent true repetitions. Their method employs a modified version of the tracking algorithm proposed by Marwan, Thiel, and Nowaczyk (2002) with a combination of monotonic constraints to eliminate spurious dots (see Fig. 2c and d for an illustration). The tracking algorithm is applied to each group of connected dots and returns a line of constant thickness, equal to one point, which maintains the shape and orientation of the group of dots from which it is computed. Once continuous downward-pointing lines are removed, a variant of the DET measure, called *elastic determinism*[1] (EDET), can be computed as the ratio between the number of dots belonging to continuous lines – regardless of their slope and curvature – and the number of locations in the RP. An in-depth description of the cleaning algorithm can be found in Lancia et al. (2014). EDET has been shown to be robust to nonstationarity and capable of efficiently separating amplitude variability from temporal variability.

Although the algorithm proposed by Lancia et al. (2014) remedies the problems caused by skipping time-delay embedding, it was primarily designed as a measure of similarity between two repetitions of the same multidimensional time series. Hence, two potential weaknesses arise when measuring the regularity of a unidimensional signal containing multiple repetitions of a basic cycle. First, the approach is sensitive to changes in average *F*0, since the number of lines parallel to the main diagonal depends on the number of repetitions of the same pattern in the signal. When more cycles are included in the same RP (due to higher *F*0), EDET is expected to increase accordingly – even if the underlying signal does not become more regular. Second, if branching line structures are present in the RP obtained from several cycles of a unidimensional signal, and if a long downward-pointing branch is connected to a short upward-pointing branch, the proposed cleaning algorithm will produce an error as neither of the two branches will be retained.

We propose a number of modifications to the RQA algorithm which reduce its dependency on *F*0 (see Fig. 3 for an illustration). The basic idea is as follows: first, an RP with a very lax similarity criterion is built from the observed signal without time-delay embedding, denoted as RP^{lax}, and cleaned of dots belonging to downward lines and of dots inflating the thickness of upward lines. We assume that RP^{lax} contains the set of possible recurrences given the number of cycles of the observed signal and the shape of the basic cycle. Next, a second RP is built for the same signal with a strict similarity criterion, denoted as RP^{strict}. Based on these, we calculate the ratio between the number of dots in RP^{strict} and the number of dots in RP^{lax}. When computing the RPs, however, the number of branching points in the RP increases, as does the likelihood of tracking errors during cleaning. This can be avoided by limiting the regions of the RPs where potential recurrences might occur with a spatial RP mask.

The steps to implementing this approach are as follows:

Step 1. Defining spatial restrictions. To identify the regions of the RP where recurrences can potentially occur, a number of substeps are performed:

Step 1.a. Finding potential recurrences: First, an RP is constructed by comparing each value of the observed signal $x(t)$ with all its other values. It is denoted as RP^{all} and obtained using an FRR criterion and a threshold such that 50% of its space contains black dots. Although the FAN criterion (cf. Section 2.2.1) is usually considered more appropriate than FRR (cf. Marwan et al., 2007), as the former assumes that each point of the time series has potentially the same number of replications, it tends to introduce more horizontal structures in the RP. Since the RP obtained in step 1a) is used to define the regions in which recurrences are expected, the presence of horizontal structures may lead to inappropriate location of these regions. The resulting diagonal bands shown in Fig. 3a represent the regions where potential repetitions of the signal should be located. However, using such a high recurrence rate, yields a criterion which is too lax to reliably identify these recurrences, producing many redundant dots in the RP.

Step 1.b. Removing spurious structures: To remove the superfluous dots, groups of connected dots are identified through a connected component algorithm as implemented in the MATLAB software (cf. Sedgewick, 1998, for a description). A skeletonization algorithm (Kong & Rosenfeld, 1996) is then applied separately to each group. This produces a 'skeleton' of the given structure – a continuous line with constant thickness of one dot, which is equidistant from the boundaries of the shape. Also for this algorithm we used the implementation available in the MATLAB software. All redundant RP dots resulting from the previous step are thus removed, while still preserving the information on the repeated patterns (see Fig. 3b). However, the RP still contains undesirable branches pointing in the bottom-right direction. These are filtered by identifying the junction points of the skeleton branches and the turning points (points at which the branches change direction). Locations corresponding to junctions and turning points in the RP (see circles in Fig. 3b) are set to 0. The various branches are now separated by white dots and can be detected by applying the connected components labeling algorithm once again. Lines pointing downward are identified by comparing the abscissa of their leftmost and rightmost extremes and subsequently eliminated. The resulting cleaned RP^{all} gives a first approximation of the positions in the plot where recurrences can be expected.

Step 1.c. Expanding the structure: To allow for more reliable recurrence detection, all the lines composing the skeleton in the cleaned RP^{all} are submitted to morphological expansion, producing bands of quasi-constant thickness. Applied to a binary shape, morphological expansion (or dilation) produces another shape which contains more dots but preserves the skeleton of the original shape (Kong & Rosenfeld, 1996). The transformation is usually applied iteratively and the boundary of the shape is displaced away from the skeleton by one dot per iteration. We utilized the implementation available in the MATLAB software, setting the number of

[1] "Elastic" refers to the independence of the measure from variations of the signal's rate of change.

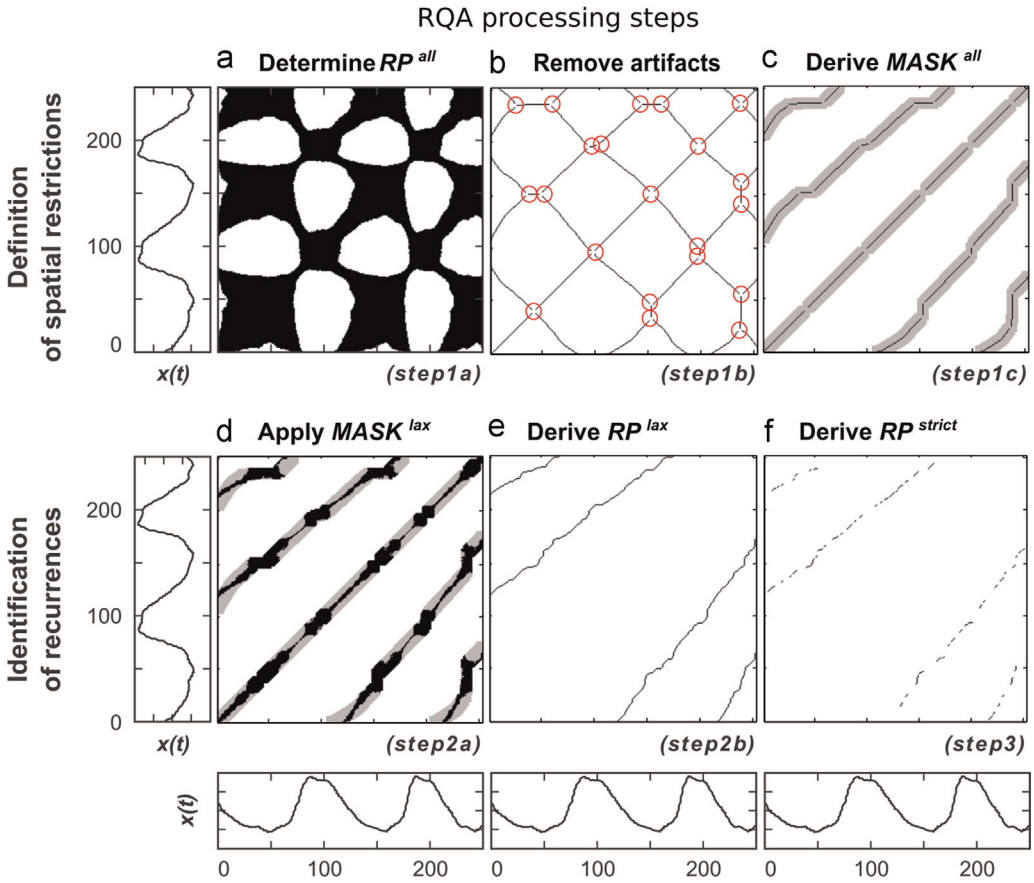

Fig. 3. Individual steps of the proposed RQA algorithm. Panel (a): RP with a high recurrence rate parameter obtained from time series $x(t)$ without embedding. Panel (b): Skeleton obtained from RP in panel (a) and the corresponding junction points (small gray circles) which are used to remove the branches pointing in the bottom right direction. Panel (c): The resulting masking bands (gray) which are obtained from expanding the remaining skeleton branches. Panel (d): RP with lax similarity criterion restricted to the bands obtained in panel (c). Panel (e): RP with lax similarity criterion after applying the DTW algorithm separately to each group of connected dots in panel (d). Panel (f): RP with strict similarity criterion restricted to the bands obtained in panel (e). The respective algorithmic steps performed in order to obtain the results displayed are indicated below the individual plots.

iterations to five. The bands in the resulting $MASK^{lax}$ (see gray bands in Fig. 3c) constitute the spatial restrictions imposed on the RPs before actual recurrences are identified in the subsequent steps.

Step 2.a. Determining baseline recurrence: A new RP is derived from the observed signal $x(t)$, again using an FRR criterion. In order to obtain a more precise location of the recurrence points in the RP, a smaller value needs to be used for the recurrence rate parameter than the value adopted in building RP^{all}. This parameter should, however, be large enough to guarantee the adoption of a lax similarity criterion. In the current implementation we set the recurrence rate so that the allowed number of recurrence is equal to exactly half of the dots contained in $MASK^{lax}$. To avoid the inclusion of dots located outside the bands of $MASK^{lax}$, corresponding values in the distance matrix used to generate the RP are set Infinity (see Fig. 3d).

Step 2.b. Skeletonization via dynamic time warping: Groups of connected dots in the obtained RP are isolated through the application of a connected component algorithm. Each group of connected dots individuates two portions of the time series and the corresponding rectangular region in the RP and the matrix of distance values. The rectangular region of distance matrix corresponding to each group of connected dots is modified so that the locations corresponding to white regions in the RP portion are set to infinity. The dynamic time warping algorithm (DWT, Sakoe & Chiba, 1978) is then applied to separately to each portion of the distance matrix modified in this way. DTW returns the coordinate of a line of thickness equal to one dot with positive slope connecting the lower left extreme with the top right extreme of the group of connected dots. Importantly, this algorithm maximizes the probability that the obtained line passes over locations corresponding to small distance values. The resulting RP is populated by lines of constant thickness (equal to one dot) with positive slope. The number of dots in the cleaned RP, denoted as RP^{lax} (see Fig. 3e), provides the first part of the regularity measure (see Step 4), representing the baseline rate of recurrence to which a more strictly defined set of dots is related.

Step 3. Determining strict recurrence: The final RP, denoted as RP^{strict} is constructed by adopting a fixed distance threshold ε^{strict} which is left as a free parameter to be regulated by the user (for the analyses discussed in the present paper, it is set to *10%* of the standard deviation of the analyzed signal portion). RP^{strict} is built by copying RP^{lax} and removing the dots corresponding to distance values higher than ε^{strict}. Finally, the isolated dots are individuated and discarded from RP^{strict}. The resulting cleaned RP^{strict} (Fig. 3f) contains only the set of dots considered to be true recurrences. Their number constitutes the second part of our regularity measure.

Step 4. Calculating the regularity measure: The ratio between the number of dots in the cleaned RP^{strict} and the number of dots in RP^{lax} is proposed as an index of regularity of the observed signal. It is denoted as *EDET ratio* and defined as follows:

$$EDET\ \ ratio = \frac{\sum_{i,j=1}^{T} RP_{i,j}^{\text{strict}}}{\sum_{i,j=1}^{T} RP_{i,j}^{\text{lax}}},$$

where T denotes the number of measured points (RP side length).

2.2.4. Implementation issues and parametrization

If the edges of the analyzed signal correspond to two peaks or two valleys, and if the rate of change of the signal is sufficiently slow at these points, boundary effects may be observed resulting in erroneous location of dots close to the lower left and top right corners of the RPs. This is easily resolved by moving the boundaries of the considered signal to the closest zero crossing (a point in time where the observed signal crosses the value zero). This choice implies a detrending of the signal (followed by a subtraction of the mean value), but this step is commonly included in the analysis of EGG signals to remove the low frequency components of the signal (i.e. those components whose frequency is lower than the fundamental frequency) as these are mainly determined by the vertical displacement of the larynx. In cases where the effects of the slow components are of interest, identification of the zero crossing points may be performed on the detrended signal while the recurrence analysis can be performed on the observed.

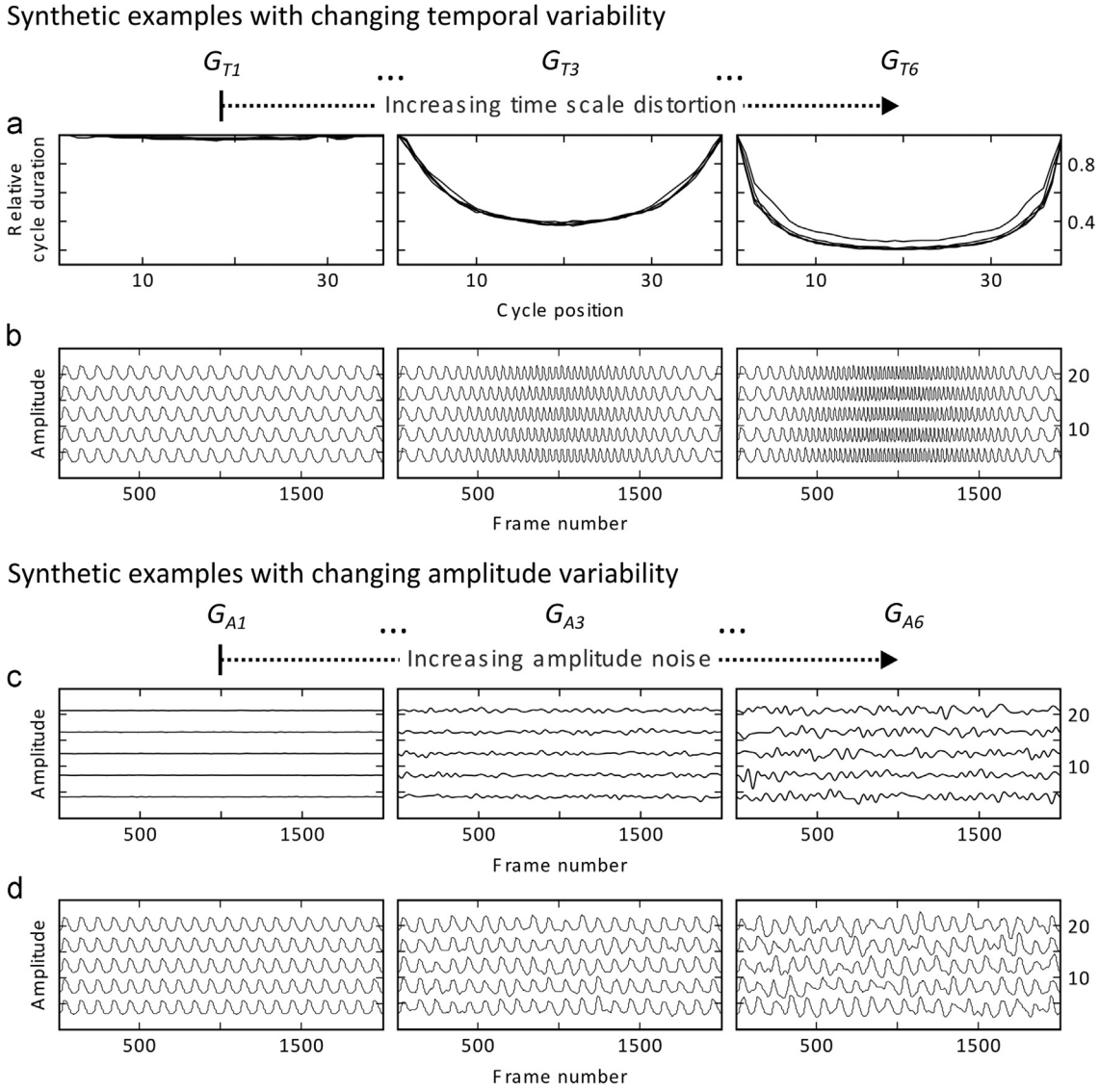

Fig. 4. Generation of synthetic EGG signals. Panel (a): Ratio between the duration of the first cycle and the duration of the following cycles over the position of the cycles in the synthetic signals; each panel shows five examples corresponding to five different synthetic signals from one particular group. Panel (b): Corresponding examples of time series with increasing time-scale deformation. Panel (c): Noise added to the synthetic signals over time (frame number). Panel (d): Examples of time series with increasing amplitude deformation. In each panel signals are shifted on the vertical axis for the purpose of visualization.

Since the algorithm proposed in Section 2.2.3 is applied to univariate trajectories, it is not necessary to define the time lag τ and the number of dimensions M here. The parameters which govern the functioning of the cleaning algorithm on the other hand are easily determined by general properties of the analyzed signals. It is worth noting that the choice of inappropriate values leads to undetected recurrences in both RP^{strict} and RP^{lax} which is likely to affect the sensitivity of the regularity measure but not to introduce biases. The threshold used to compute RP^{all} might be optimized by maximizing the number of isolated white regions in the RPs. The thickness of the bands in $MASK^{lax}$ is determined by the number of iterations in the application of the morphological expansion to the cleaned RP^{all} (cf. Fig. 3c). To preserve all potential line structures, this cannot exceed *50*% of the minimum expected period length. The parameter with the strongest impact on the final result is the threshold ε^{strict} used to compute the final recurrence plot. Smaller values imply greater sensitivity to irregularity. Schinkel, Dimingen, and Marwan (2008) propose employing a ε value that better separates the classes of time series under investigation. In comparing the regularity of modal and laryngealized voicing signals the optimal value of ε would therefore be the one that best separates laryngealized voicing from modal voicing.

2.3. Simulations with synthetic data

To demonstrate the reliability of the proposed RQA method in the study of EGG signals, simulations were run using artificially generated data exhibiting controlled variability in the temporal and spatial domains. Our aim is to demonstrate the capability of the RP measures to reliably quantify changes in the amplitude of the analyzed signal, unaffected by temporal variability potentially related to changes of *F*0.

2.3.1. Data generation and test conditions

Each time series used in the simulations was composed of 20 repetitions of the same EGG cycle, extracted from a recording of a German female speaker (see next section for details). This basic oscillation was then used to build tokens with different degrees of amplitude and temporal variability. In total, 12 groups of oscillations were generated: groups $G_{T1\ldots T6}$ comprise the examples with varying temporal properties and constant low amount of amplitude variability (see Fig. 4a and b, where, for the sake of clarity, only groups G_{T1}, G_{T3} and G_{T6} are shown), while groups $G_{A1\ldots A6}$ comprise those with varying amplitude and constant low amount of temporal variability (see Fig. 4c and d, where only groups G_{A1}, G_{A3} and G_{A6} are shown). Each group contains 30 trains of oscillations (i.e. 30 uninterrupted sequences of voicing cycles).

The variability in the data was progressively increased by a constant amount from group G_{T1} to G_{T6} in the temporal domain, and from G_{A1} to G_{A6} in the amplitude domain. To introduce variability, we followed the method proposed by Lucero (2005) and modified by Lancia et al. (2014). Temporal nonstationarity was introduced by changing the frequency of oscillation over the duration of each signal. In each time series, the period durations follow a parabolic shape with a minimum in the middle of the waveform, so that cycle durations gradually decrease toward the middle of the waveform, returning to the initial value toward the end (Fig. 4a). In the signals with the greatest temporal nonstationarity, the fastest cycle is approximately five times shorter than the slowest. In the time series with the smallest amount of temporal nonstationarity, the duration of the fastest cycle is about 0.98 times the duration of the slowest. To obtain a more realistic *F*0 variation across tokens of the same group, random oscillations were added after the deformation function had been applied to each token. Amplitude variability was introduced by adding randomly generated values to each signal, smoothly changing over time (Fig. 4c). The amount of amplitude variability was quantified by the root mean square error obtained by comparing the modified and the original waveforms. The amount used to generate the synthetic signals ranges from 0.1 to 0.6 times the standard deviation of the original train of EGG cycles. Fig. 4b and d show some examples of the resulting synthetic signals from groups G_{T1}, G_{T3}, G_{T6} and from groups G_{A1}, G_{A3}, G_{A6} to illustrate the range of introduced variabilities. Note that in the stimuli with the greatest temporal variability the rate of change of *F*0 is such that during the time-interval required to produce 10 voicing cycles at the slowest frequency (the half of a train of oscillations) *F*0 changes by 27.9 semitones. If we assume an initial frequency of 100 Hz, the rate of change of *F*0 amounts to 279 semitones per second, greatly exceeding the maximum rates observable in real speakers (cf. Xu & Sun, 2002 and references therein).

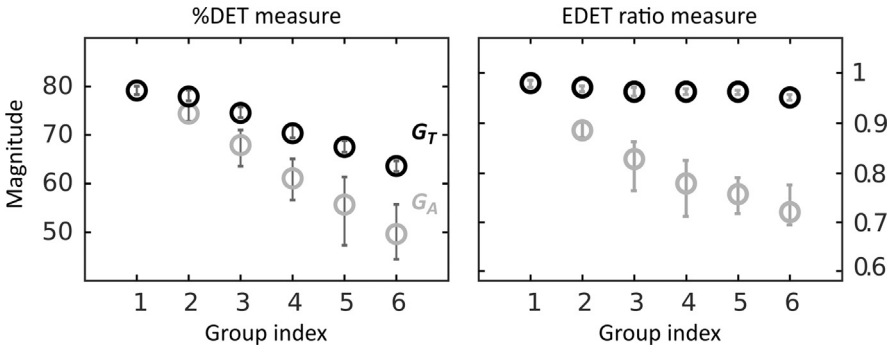

Fig. 5. Simulation results. The black circles represent mean values obtained from synthetic signals with increasing temporal variability from left to right (i.e., corresponding to groups $G_{T1\ldots T6}$). The gray circles represent mean values obtained from signals with increasing amplitude variability (i.e., groups $G_{A1\ldots A6}$). The bars indicate the standard deviations associated with each group.

The synthetic EGG signals were subsequently analyzed using the proposed novel EDET ratio and the classic %DET index based on the standard RQA approach with time-delay embedding. To calculate the %DET measure, we embedded the signals with a time delay τ, equal to 1/4 of the fastest cycle observed across the signals and $M = 3$ dimensions. All time series were partitioned into sliding windows with a window length such as to include at least two of the slowest cycles observed across the signals and a step size equal to 1/10 of the window length. Each window was then submitted to each RQA variant, yielding the corresponding %DET and EDET ratio measures. For each signal, a mean value was calculated from the observed regularity measures over all resulting RP windows. As shown in Fig. 5, the classic measure %DET decreases as amplitude variability increases; however, it is also sensitive to changes in *F*0. The EDET ratio decreases when the amplitude variability of the synthetic EGG signals increases but remains nearly constant over the whole range of values tested for temporal variability.

3. Real speech data: laryngealization and lexical stress at the onset of vowel-initial non-sense words

Having demonstrated the sensitivity of the EDET ratio measure through simulations, we proceed to apply the proposed RQA method to real EGG recordings collected during repeated production, at a sustained speech rate, of nonsense words starting and ending with a vowel. In this context, German speakers often produce a glottal constriction resulting in the laryngealization of the word-initial vowel or in an audible glottal stop at the juncture of two words (cf. Pompino-Marschall & Żygis, 2011). Furthermore, modulations of regularity are expected to depend on whether the stress is on the initial or on the final syllable. In previous studies on laryngealization before vowel-initial words, evidence for this maneuver is more commonly found before vowel-initial words when these are stressed, located at the left edge of an intonational phrase or when they carry a pitch accent. This is in line with Slifka's (2006) hypothesis that the presence of a glottal adduction gesture is an efficient way to rapidly create a discontinuity in the speech stream. Indeed, in the above cases, the discontinuity would signal the beginning of a new prosodic constituent, helping listeners focus their attention on the following acoustic material. On the other hand, the presence of a glottal constriction in these contexts is expected to modify the production of initial vowels in such a way as to contrast the effects of at least stress and pitch accents. Stress is expected to enhance the amplitude of vocal fold vibration, which is damped close to glottalization, and, in the presence of a pitch accent, to increase *F*0 excursion (cf. Mooshammer, 2010 for a recent review), which is reduced by glottalization; pitch accents require some control of *F*0 which is in part hindered by glottalization (Blankenship, 2002; Lancia et al., 2013; Silverman, 1997). Investigating the interaction between glottalization and prosodic factors may therefore potentially yield new insights on how speakers resolve the conflicting demands of different phonological factors on the physical systems involved in speech production. The following sections focus on the combined effects of stress and accent on the vowel preceding or following glottalization.

3.1. Speakers and procedure

We asked five female German speakers to repeatedly utter, at a sustained speech rate, vowel-initial and consonant-initial disyllabic nonsense words with stress on the first or on the second syllable (/ˈaska/, /as'ka/, /'kaska/, /kas'ka/). We selected the low back vowel and the velar voiceless plosive preceding the vowels because a retracted tongue position should favor the production of glottalization (Moisik & Esling, 2011), while the additional fricative consonant should prevent verbal transformation effects (e.g. from

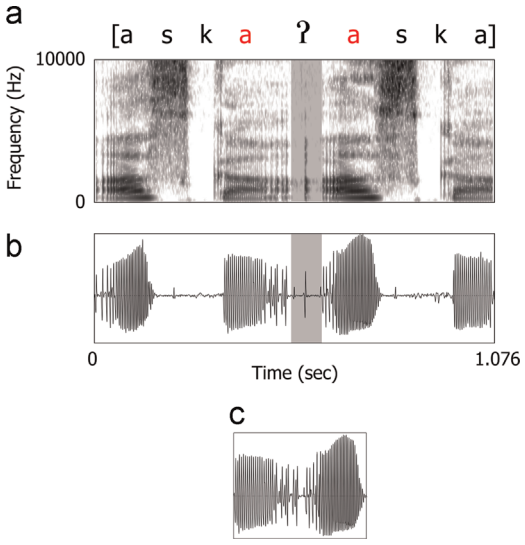

Fig. 6. Extraction and preprocessing of the portion of EGG signal containing the final vowel of a nonsense word and the initial vowel of the following nonsense word. Panel a: Acoustic spectrogram of two repetitions of the nonsense word /'aska/. Panel b: Corresponding filtered EGG signals. The gray bands indicate the interval of time corresponding to the intervocalic glottal stop. This interval of time is excluded from the analyzed portion of EGG signal (represented in panel c).

/aka/ to /ka:/) that may occur at sustained speech rate (cf. De Jong, Lim, & Nagao, 2002; Tuller & Kelso, 1991). Vocal fold vibrations were recorded through an EG-2 electroglottograph from Glottal Enterprises at 48,000 Hz. To control for speech rate, we asked the speakers to utter each nonsense word without interruption for a period of 10 s, in pace with a visual metronome blinking at 4 Hz.

3.2. Data processing and analyses

The recorded EGG signal was first band pass filtered (cut off freqs.: 50 and 1000 Hz) in order to remove the slow changes induced by vertical displacement of the larynx and rapid fluctuations mainly due to measurement noise. Vocalic portions of the EGG were automatically detected by identifying the peaks of its time-varying RMS energy (computed on hamming windows, 25 ms long, sliding through the EGG signal with a step of 3 ms) and subsequently corrected by hand. Each sequence of two consecutive RMS energy peaks identified two consecutive vowels. Given the locations of these two peaks (tP1 and tP2) and the location of the minimum of the RMS energy (tM) in the time-interval between the two peaks, the offset of the first vowel was located at the first instant included in the time-interval between tP1 and tM where the RMS energy was less than 20% of RMS energy observed at tP1. Similarly, the onset of the following vowel was located at the last instant in the time-interval between tM1 and tP1 where the RMS energy was less than 20% of the value observed at tP2. We then extracted several portions from each sequence of repetitions, beginning with the final vowel of a nonsense word and ending with the initial vowel of the following nonsense word. Where the two vowels were separated by a silent gap or by a silent gap followed by the burst of a consonant release (either oral or glottal), this section was removed from the extracted interval (cf. Fig. 6).

As can be observed in Fig. 6, the amplitude of the filtered EGG waveform slowly changes at the vowel boundaries, This occurs regardless of whether an oral consonant separates the vowels in the recorded speech stream or not. Slow amplitude modulations affect our measure of regularity, with the consequent reduction of potential differences between the regularity observed in vowels extracted from consonant-initial nonsense words and those extracted from vowel-initial nonsense words. To reduce the amplitude drop at the juncture between consecutive vowels, we followed a simple demodulation procedure. First, a new signal was built with length equal to that of the original waveform and with amplitude constantly equal to 0. The time-varying RMS energy of the filtered EGG signal was computed again using the shortest possible time-window which contained always two cycles of vocal fold vibration (length 20 ms), sliding trough the signal with a step of 1 ms. This curve was then scaled to vary between 0 and 1. In each time window, the amplitude values of the original waveform were divided by the RMS value corresponding to that window and added at the corresponding time-steps to the signal originally equal to 0.

Next, the resulting signal was windowed (window length = 15 ms, step = 3 ms) and recurrence analysis performed on each window. The $\varepsilon^{\text{strict}}$ parameter was set to 10% of the standard deviation of the portion of signal considered to build each RP and was thus varied between time windows. This yields a time varying measure of regularity for each pair of consecutive vowels. Before submitting the EDET ratio time series to statistical analyses, each time series was linearly interpolated to reach a constant length of 100 frames. However, the interpolation was conducted separately on portions of time series extracted from the same syllable (first or second) and produced with the same stress condition (+/−). By normalizing the time scales of vowels from same position and stress condition we aimed at aligning the location of the intervocalic juncture between EDET ratio curves from consonant-initial and vowel-initial nonsense words, while retaining the differences in the location of the intervocalic juncture between stress-initial and stress-final nonsense words.

The EDET ratio was complemented by a time-varying measure of the open quotient (henceforth OQ) from each portion of EGG signal. Additionally, the time-varying *F*0 and the time-varying RMS energy (window length: 25 ms, time-step: 3 ms) were computed from the acoustic signal corresponding to each extracted portion of EGG signal. These measures were intended to provide additional evidence for laryngealization and to verify that the speakers produced prominence at the expected locations (under the assumption that loudness and *F*0 are correlates of stress and of pitch accents in German, cf. Mooshammer, 2010).

The computation of the OQ was based on the detection of three distinct time points in each voicing cycle. To locate the peaks of the voicing cycles, we analyzed the band-pass filtered EGG signal. To detect the opening instants, we analyzed the original EGG signal low-pass filtered with a cut off frequency of 1000 Hz. Finally, to detect the closing instants we analyzed the first derivative of the low-pass filtered EGG signal. EGG peaks were located with the peak-finding algorithm provided within the MATLAB software. In each vowel, peaks whose amplitudes were smaller than the mean vowel-specific peak amplitude minus twice its standard deviation were removed. Next, the average and the standard deviation of the inter-peak distance (i.e. the length of the time-interval between consecutive peaks) were calculated. If the distance between each pair of consecutive peaks was smaller than the mean inter-peak distance minus twice the corresponding standard deviation, the peak corresponding to the smallest amplitude value was discarded. In each portion of band-pass filtered EGG signal delimited by two consecutive peaks we also located the minimum value.

In each time interval delimited by a minimum and by the following peak value, we located a closing instant by identifying the maximum value of the EGG derivative. We did not attempt to locate the opening instant because of the weak relation between this articulatory event and characteristics of the EGG signals or of its derivative (cf. Baken, 1992; Colton & Conture, 1990). Instead, following Rothenberg and Mahshie (1988), in each time-interval delimited by a peak value and by the following minimum value, we detected the last point in time when the amplitude of the low-pass filtered EGG signal exceeded 35% of the range of values observed in that time interval. This strategy does not aim to locate the true instant in time at which the vocal folds reach their maximal aperture but rather a point in time which depends on the location of the true opening instant. However, consistently with previous works, we denote this time point as the opening instant. The cycle duration was computed as the distance in time between a closing instant and the closing instant of the following cycle. For each opening instant (Ot_i) an *OQ* value was computed (OQ_i) as the duration of the time

interval delimited by the opening instant Ot_i and the following closing instant (Ct_{i+1}) expressed as a proportion of the duration of the time-interval delimited by the closing instants preceding and following the opening instant considered (Ct_i and Ct_{i+1} respectively):

$$OQ_i = \frac{Ct_{i+1} - Ot_i}{Ct_{i+1} - Ct_i},$$

Due to the strategy adopted to locate the opening instant, this measure should not be considered as a measure of the open quotient *strictu sensu*, but rather as a measure depending on this quantity.

The RMS signals, *F*0 signals and OQ signals were linearly interpolated following the same procedure as adopted for the EDET ratio signals.

3.2.1. Statistical analyses

We employed wavelet-based functional mixed models (Morris & Carroll, 2006) to estimate the time course of the effects of laryngealization on the EDET ratio measure during vowel production and to assess how these are modulated by stress. To this end, each time series corresponding to the temporal evolution of the EDET ratio measure during the production of two consecutive vowels was submitted to discrete wavelet transform (Mallat, 1989). The resulting coefficients, representing the curve in the wavelet space, were taken as dependent variables in a mixed linear model (Pinheiro & Bates, 2000). The fixed factors of the model were: (i) presence vs. absence of the initial oral consonant (reference level: no initial oral consonant); (ii) stress position (on the vowel following the boundary between consecutive nonsense words or on the vowel preceding that boundary; reference level: post-boundary position): (iii) the position of the nonsense word in the chain of utterances speakers produced without interruptions (henceforth referred to as the position in the experimental trial). The last factor was centered around its mean and scaled to standard deviation of one. Due to centering, the reference level of this factor corresponds to the middle of the experimental trial; therefore, the effects of the other factors and of their interactions refer to the nonsense words uttered in the middle of the experimental trial. The three-way interactions of the model predictors were also considered. The following random terms were used: a random intercept for each speaker and a speaker-specific random slope for each of the simple fixed effects in the model.

Under a Bayesian approach, the model parameters (quantifying the effects of both fixed and random factors) are treated as random variables. The aim of the analysis is to derive a probability distribution describing how likely the possible combinations of values of the model parameters are given the observed data. Bayesian analysis is informed by the knowledge we have of the behavior of the parameters under investigation before we see the data. This knowledge is summarized in the prior distribution, describing how probability (plausibility) is distributed among the possible choices for the values of the model parameters before taking into account the data at hand. This information is then combined with the information contained in the data to produce a

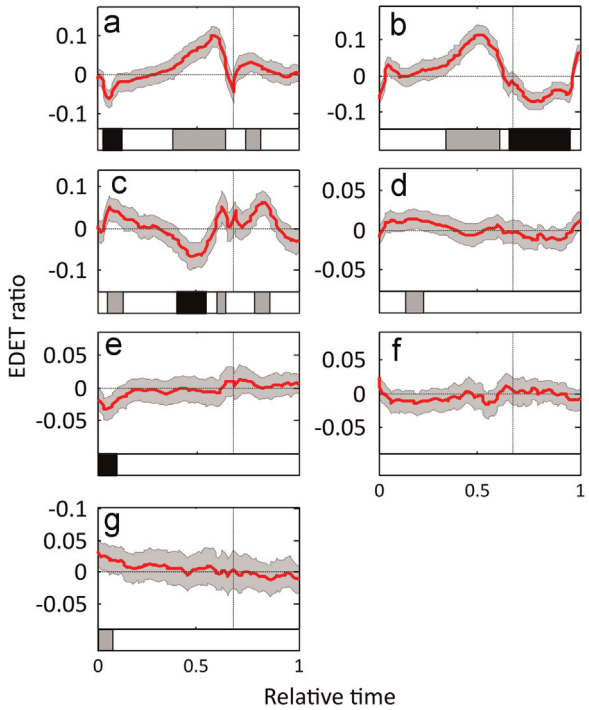

Fig. 7. Evolution over time of the effects tested by the statistical model (continuous red lines, color online) and their credible intervals (gray bands). The boxes below each panel indicate time-intervals where the corresponding effect is significantly positive (light-gray boxes) or significantly negative (black boxes). Panel (a): Effect of initial oral consonant. Panel (b): Effect of stress position. Panel (c): Effect of the interaction between stress position and presence of an initial oral consonant. Panel (d): Effect of the position of the nonsense words in the experimental trial. Panel (e): Effect of the interaction between position of the nonsense words in the experimental trial and presence of an initial oral consonant. Panel (f): Effect of the interaction between position of the nonsense words in the experimental trial and stress position. Panel (g): Effect of the interaction between position of the nonsense words in the experimental trial, presence of an initial oral consonant and stress position. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

posterior distribution, which expresses how probable the possible combinations of values of the model parameters are given both the prior distribution and the observed data. In practical applications, large samples from the posterior distribution are obtained via Markov chain Monte Carlo simulations and are used to derive the inferences of interest. When principled *a priori* assumptions on the values of the model parameters are not available, appropriate vague (i.e. non-informative) prior distributions are adopted. A spike and slab distribution with a peak around zero and medium to heavy tails is an appropriate vague prior distribution for time series data changing over time on a time-scale much slower than the data sampling rate (cf. Morris & Carroll, 2006).

Given that, as a result of the wavelet transform, each observed time series is represented by a vector of wavelet coefficient values, the obtained posterior distribution corresponds to the distribution of the effects of the model predictors on the values of the wavelet coefficients describing the observed time series. This distribution is transformed via the inverse wavelet transform to obtain the posterior distribution of the fixed factor effects on the observed time series. From the obtained distribution, we extracted time-varying average values of the fixed factor effects and their credible intervals indicating the extension of 95% of the posterior distribution at each point in the normalized time, where 0 corresponds to the beginning and 1 to the end of each sequence of consecutive vowels (cf. Fig. 7). In the following, an effect is considered significant at a given time-point if the time-varying credible interval of its posterior distribution does not include 0 at that time point. Note that if we asses statistical significance from credible intervals, stating that an effect is significant means that it is considered plausibly different from 0 given the data at hand. This interpretation of statistical significance is different from that adopted by frequentist approaches where a significant effect is one that, if the experiment is repeated many times, is expected to be observed with a probability which is equal to or higher than one minus the significance threshold adopted.

3.2.2. *EDET ratio results*

In Fig. 7 we display the time-varying mean effects of the tested predictors on the observed time series as well as their 95% credible intervals.

The model estimates of the mean curves for sequences of vowels extracted from consonant-initial and vowel-initial nonsense words, with stress on the first or on the second syllable are presented in Fig. 8. These are obtained by combining the evolution over time of the effects tested, shown in Fig. 7.

The continuous line in Fig. 8a represents the evolution over time of the EDET ratio during the production of vowels extracted from vowel-initial nonsense words with stress on the post-boundary vowel. This line corresponds to the model intercept. The dashed line in the same panel represents the evolution over time of the EDET ratio when nonsense words start with an oral consonant. This is obtained by adding to the model intercept the evolution over time of the effect of the presence of the initial oral consonant (displayed in Fig. 7a). By denoting each curve with the label of the corresponding panel displayed in italic typeface and by adopting the

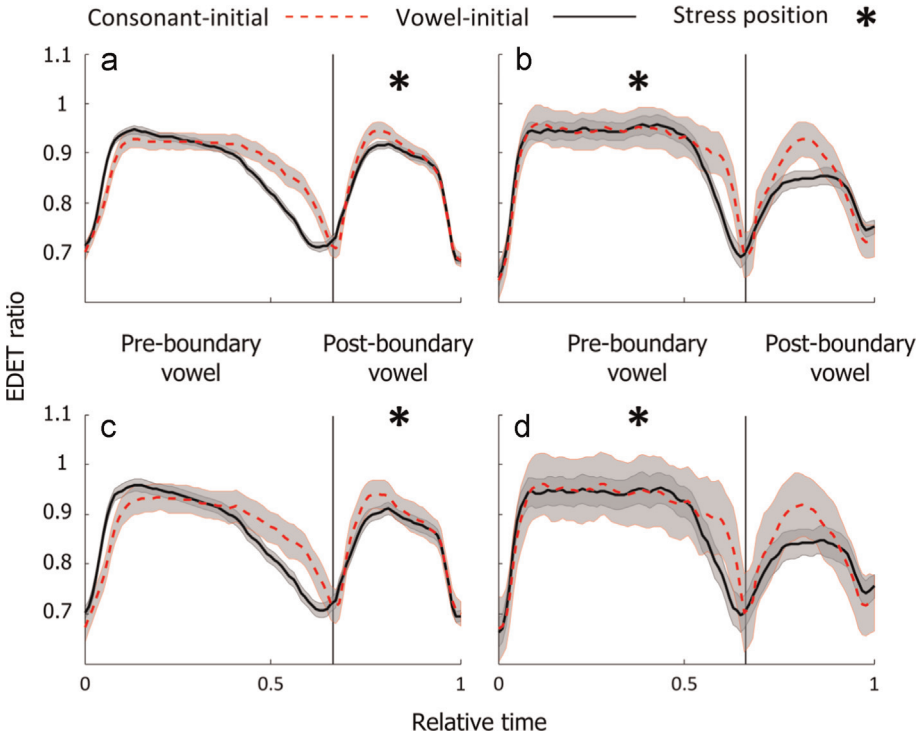

Fig. 8. Time course of vocal fold vibration regularity in the repeated speech task (continuous black lines and dashed red lines, color online) and their credible intervals (gray bands). The vertical continuous lines indicate the position of the boundary separating consecutive nonsense-words. Dashed red lines (color online): vowels extracted from consonant-initial nonsense words. Continuous black lines: vowels extracted from vowel-initial nonsense words. Left column: estimates from nonsense words with stress on the post-boundary vowel. Right column: estimates from nonsense words with stress on the pre-boundary vowel. Stress position is indicated by an asterisk. Top row: estimates from nonsense words in the middle of the experimental trial. Bottom row: estimates from nonsense words produced after ca 6 repetitions from the middle of the experimental trial. Stress position is indicated by an asterisk. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

convention that bold type face indicates continuous (black) lines from Fig. 8 while normal typeface indicates dashed (red) lines from the same figure, we obtain: *Fig. 8a*=***Fig. 8a***+*Fig. 7a*. This effect is significantly positive before and after the vowel juncture. Voicing regularity decreases at the offset of the pre-boundary vowel in both kinds of nonsense words, but the decrease starts sooner in vowel-initial nonsense words than in consonant-initial nonsense words. Moreover the peak regularity observed during the production of the post-boundary vowel is higher in consonant-initial nonsense words than in vowel-initial nonsense words.

The continuous line in Fig. 8b represents values of EDET ratio from vowel-initial nonsense words with stress on the pre-boundary vowel (***Fig. 8b***=***Fig. 8a***+*Fig. 7b*). Fig. 7b shows that when stress moves onto the pre-boundary vowel the regularity of voicing increases significantly during the pre-boundary vowel and drops significantly during the post-boundary vowel. In other words, when a vowel is stressed it is produced with more regular voicing.

The dashed line in Fig. 8b represents EDET ratio values from consonant-initial nonsense words with stress on the pre-boundary vowel (***Fig. 8b***=***Fig. 8b***+*Fig. 7a*+*Fig. 7c*). Note that during a large portion of the pre-boundary vowel the effect of the interaction between the presence of the initial consonant and the presence of stress on the pre-boundary vowel (Fig. 7c) is the reversed version of the effect of the consonant alone (Fig. 7a). This means that the difference in regularity between pre-boundary vowels followed by oral consonants and those followed by a vowel is significantly reduced (or canceled) when stress is located on the preboundary vowel. However this difference is still observed (and it is actually strengthened) at the end of the pre-boundary vowel where the curve in Fig. 7c is significantly positive. In the time interval corresponding to the production of the post-boundary vowel, the effect of the interaction is significantly positive, meaning that when the stress is on the pre-boundary vowel, the post-boundary vowel in nonsense words beginning with an oral consonant is produced with more regular voicing than in those beginning with a vowel.

Since the values of the factor representing the position of the nonsense words in the experimental trials were normalized, the effect in Fig. 7d represents the time-varying difference in regularity between nonsense words produced in the middle of the experimental trial (its reference level) and those produced after 6.3 nonsense words (the standard deviation of the distribution of the positions of the nonsense words in the experimental trials). The continuous line in Fig. 8c corresponds to the estimated EDET ratio trajectories for nonsense words produced late in the experimental trials (***Fig. 8c***=***Fig. 8a***+*Fig. 7d*). No significant effect of nonsense-

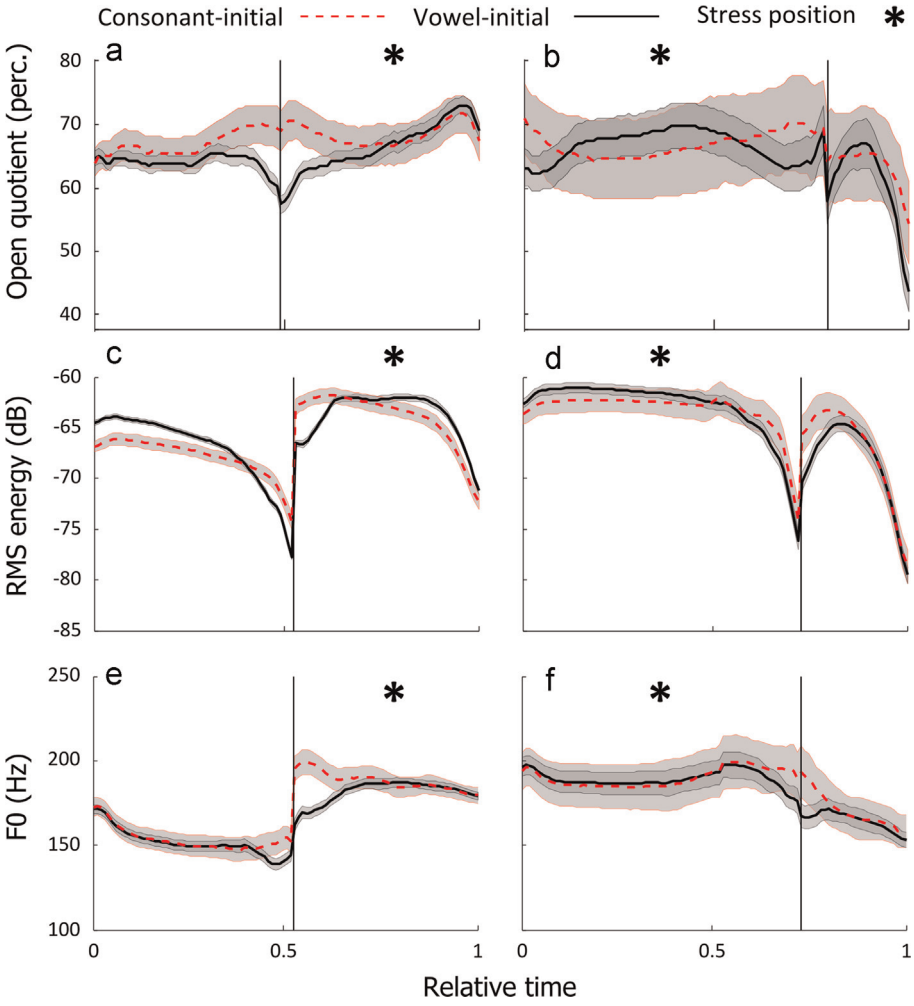

Fig. 9. Time course of estimated OQ, RMS energy and *F*0 values (continuous black lines and dashed red lines, color online) and their credible intervals (gray bands). The vertical continuous lines indicate the position of the boundary separating consecutive nonsense words. Dashed red lines (color online): vowels extracted from consonant-initial nonsense words. Continuous black lines: vowels extracted from vowel-initial nonsense words. Top-row panels: OQ estimates. Middle-row panels: RMS energy estimates. Bottom-row panels: *F*0 estimates. Stress position is indicated by an asterisk. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

word position is observed around the juncture of the two vowels.

The dashed line in Fig. 8c represents EDET ratio values from consonant-initial nonsense words produced relatively late in the experimental trial (***Fig. 8c*** = ***Fig. 8c*** + *Fig. 7a* + *Fig. 7e*). The interaction between the presence of the initial consonant and the effect of the position of the nonsense word in the experimental trial (Fig. 7e) is not significant in the time-interval surrounding the juncture between the vowels, suggesting that differences between EDET ratio trajectories from vowel-initial and consonant-initial nonsense words are not affected by the position of the nonsense words in the experimental trial.

The continuous line in Fig. 8d corresponds to the estimated EDET ratio trajectory for vowel-initial nonsense words with stress on the pre-boundary vowel produced relatively late in the experimental trial (***Fig. 8d*** = ***Fig. 8c*** + *Fig. 7b* + *Fig. 7f*). The interaction between the position of the nonsense word in the experimental trial and the presence of stress on the pre-boundary vowel (Fig. 7f) is not significant in the region around the vowel boundary. Therefore the differences in regularity due to the location of stress on the post-boundary vowel do not change significantly across the experimental trial.

The dashed line in Fig. 8d reflects the evolution over time of the EDET ratio for consonant-initial nonsense words with stress on the pre-boundary vowel produced relatively late in the experimental trial (*Fig. 8d* = ***Fig. 8d*** + *Fig. 7a* + *Fig. 7c* + *Fig. 7e* + *Fig. 7g*). The triple interaction between position of the nonsense word in the experimental trial, presence of the initial consonant and presence of stress on the pre-boundary vowel (Fig. 7g) does not prove significant in the region surrounding the juncture between the vowels; therefore, we cannot conclude that the observed difference between consonant-initial and vowel-initial nonsense words with stress on the pre-boundary vowel changes significantly along the experimental trials.

In summary, in both vowel-initial and consonant-initial nonsense words, we observe a lowering of regularity at the juncture between the vowels. This indicates that the amplitude of the waveform continues to decrease at the vowel boundaries despite the application of the preprocessing steps aimed at flattening the time-varying RMS energy of the EGG signal or that the voicing cycles at the right edge of the pre-boundary vowel display different amplitude values to the voicing cycles at the left edge of the post-boundary vowel. When stress is on the post-boundary vowel, in sequences of vowel-initial nonsense words the regularity of voicing decreases faster at the offset of the pre-boundary vowel and increases more slowly during the onset of the post-boundary vowel compared to consonant-initial nonsense words. When the stress is on the pre-boundary vowel, during the production of the post-boundary vowel the regularity of voicing is significantly lower in vowel-initial nonsense words than in consonant-initial ones.

The longer drop in regularity of vocal fold vibration at the juncture of vowels extracted from vowel-initial nonsense words indicates a glottal stop or a glottal constriction. The interaction with stress suggests that, when stress falls on the initial syllable of a nonsense word (i.e. on the post-boundary vowel in Figs. 7–9), speakers release the glottal constriction to levels appropriate for modal voicing earlier than when the stress falls on the final syllable of the nonsense words (i.e. on the pre-boundary vowel in Figs. 7–9). In the latter case the consequences of laryngealization extend over the entire duration of the initial vowels of the nonsense words.

3.2.3. OQ, RMS energy and F0

The model estimates for the evolution over time of OQ, *F*0 and RMS energy curves extracted from nonsense words produced in the middle of the experimental trials are shown in Fig. 9 together with the 95% credible intervals of the relative posterior distributions (obtained by combining the posterior distributions of fixed factors as illustrated in the modeling of the EDET ratio trajectories). The estimated OQ, RMS energy and *F*0 curves are arranged in the top-row, the middle-row and in the bottom-row panels respectively. The estimates for nonsense words with stress on the pre-boundary vowel are displayed in the panels on the right, while the estimates for nonsense words with stress on the post-boundary vowel are presented in the panels on the left.

The curves in Fig. 9a reveal that when stress occurs on the post-boundary vowels, the open quotient is lower, in the time interval surrounding the intervocalic boundary, if no oral consonant separates the vowels. This is consistent with the presence of laryngealization at the intervocalic boundary. This difference is not visible in Fig. 7b, displaying the estimates for vowels extracted from nonsense words with stress on the pre-boundary vowel.

Comparing the behavior of stressed and unstressed vowels on the same side of the intervocalic boundary from Fig. 9c and d, peak RMS energy appears to be higher for stressed vowels. Similar results are obtained for *F*0 (cf. Fig. 9e and f), showing higher *F*0 values for stressed than for unstressed vowels. It is therefore reasonable to assume that the speakers were able to produce stress and accent at the expected locations. Around the intervocalic juncture, both RMS energy and *F*0 are lower in vowel-initial nonsense words (although the effect is stronger for nonsense words with stress on the post-boundary vowel, it is significant in both cases). Both lower *F*0 and lower intensity can be interpreted as resulting from the production of laryngeal constriction and are therefore compatible with the presence of laryngealization.

4. Discussion

RQA was applied in this paper to investigate laryngealization, thus providing quantitative means of characterizing a non-modal voice quality for which adequate measurement tools have hardly been available up to now. This dearth of methods can be attributed to the heterogeneous nature of the articulatory maneuvers underlying the voicing activity commonly labeled as laryngealized, creaky or glottalized. The influential view proposed by Ladefoged (1973) and further developed by Gordon and Ladefoged (2001) maps modal and laryngealized voicing onto one phonetic dimension characterized by variable degrees of constriction at the glottal level. Many, however, make qualitative distinctions among various types of glottalization (e.g., Dilley et al., 1996; Edmondson & Esling, 2006, Moisik & Gick, 2013). Evidence collected in these studies suggests that higher degrees of constriction can be attained by

recruiting an increasing number of laryngeal and pharyngeal structures (including the tongue root). Different articulatory configurations can also be responsible for qualitative differences between the vibration patterns observed in laryngealization: firstly, changes in glottal constriction cause alterations of the mass and density of the folds; secondly, the upward movement of the larynx favors contact between the vocal and ventricular folds, so that these structures interact during the production of voicing (Edmondson & Esling 2006; Esling & Moisik, 2012; Moisik & Gick, 2013). Further degrees of constriction can result from interaction with the aryepiglottic folds, and/or narrowing of the epiglottis and of the pharyngeal passage.

Capturing the effects of these complex interactions is an extremely difficult task. Firstly, laryngealization can manifest itself in a wide range of patterns (e.g., decreased open quotient, diplophonia, amplitude modulations, etc.); therefore, we do not know a priori which feature to measure. Secondly, recurring patterns produced in laryngealized voicing in running speech are particularly unstable. Thirdly, we cannot exclude that different patterns of laryngealization follow each other or alternate in the same vowel. This may be due to the transient nature of laryngealized voicing or to the complexity of the underlying biomechanics (cf. Moisik & Esling, 2014). The many interactions involved in the production of laryngealization, however, introduce variability in vocal fold dynamics which can be used to characterize different degrees of glottal constriction. This variability is reliably captured by the EDET ratio as shown by our analyses of the interaction between laryngealization and prosodic prominence before vowel-initial nonsense words. These analyses reveal a systematic presence of laryngealization at the juncture between the final vowel of a nonsense word and the following vowel-initial nonsense word and a systematic effect of prominence, which seems to act as a blocking factor for the spreading of irregularity of vocal fold vibration over the duration of a vowel. To produce laryngealization and prosodic prominence, speakers have to fulfill conflicting requirements, since the articulatory maneuvers associated with laryngealization hinder the production of the stronger vocal fold vibration and higher *F*0 characteristic of accented vowels. Considering that stressed vowels are more resistant to coarticulation than unstressed vowels (De Jong, Beckman, & Edwards, 1993), their resistance to the spreading of glottal constriction is to be expected.

On the other hand, several authors report more evidence of laryngealization before or during the production of stressed word-initial vowels (cf. Garellek, 2014; Redi & Shattuck-Hufnagel, 2001), whereas in our data irregularity of voicing extends over the entire duration of the post-boundary vowels only if these are unstressed. Our results are not necessarily inconsistent with the observations of Rodgers (1999) whose results suggest that while glottal stops occur more frequently before phrase-initial stressed vowels, glottalization is observed more often in the production of phrase-initial unstressed vowels. Although we did not analyze the occurrence of full glottal stops in the present paper, we observed that, in the production of nonsense words with the stress on the initial syllable (the post-boundary syllable in our segmentation), the glottal constriction gesture is anticipated and overlaps more with the final vowel of the preceding nonsense word (the post-boundary vowel). It is likely that the anticipation of the glottal constriction gesture before word-initial stressed vowels underlies also the production of Roger's data. Indeed, the glottal constriction gesture is expected to occur earlier in the production of a vowel-initial word preceded by a full glottal stop and not presenting audible signs of glottalization during the initial vowel than in the production of a vowel-initial word not preceded by a glottal stop but presenting audible signs of glottalization during the initial vowel.

Another feature of our data which seems to diverge from previous findings is the cross-speaker stability of the presence of laryngealization and of the effects of the manipulated factors. Indeed, the presence of word-initial laryngealization and its dependence on stress and on the position of the word in the prosodic hierarchy are reported to be strongly speaker-dependent (Pierrehumbert & Talkin, 1992; Redi & Shattuck-Hufnagel, 2001). These differences may be due to the particular experimental paradigm adopted (whose choice was motivated by the reported stability of the behaviors elicited with this task, cf. De Jong et al., 2002). Alternatively, it may be argued that our measure captures features of laryngealization not captured with other approaches. The data analyzed in this paper do not allow us to test these two interpretations because of the potential presence of factors specific to the repeated speech task pushing speakers to reliably produce laryngealization regardless of the presence of stress and accent on the following vowel. Independently from the nature of the factors causing the systematic presence of laryngealization before vowel-initial nonsense words in the discussed data, our results indicate that when forced to produce both word-initial laryngealization and accent on the word-initial vowel, speakers can mediate between the conflicting requirements of these two maneuvers by anticipating the glottal constriction gesture so as to overlap more with the final vowel of the preceding word.

5. Conclusions

We presented a novel RQA method that permits a reliable quantification of phonation regularity in concatenated speech. By analyzing synthetic trains of oscillations, we demonstrated the reliability of our method even in the presence of strong *F*0 and amplitude variability, a condition where conventional analytical approaches based on cycle detection are unreliable (Baken, 1992; Colton & Conture, 1990). Analyses of EGG signals from laryngealized stressed and unstressed vowels suggest that speakers can modulate the degree of laryngealization when required by the production of prosodic prominence. Since the amount of variability in vocal fold vibration is affected by numerous other factors, the application of the proposed RQA method can be extended to other areas of research. Little et al. (2007) have already demonstrated the utility of recurrence analysis in the diagnosis of Parkinson's disease. Using our novel technique, the RQA approach can be further applied to the analysis of running speech in various voice disorders. Motor patterns in running speech are generally more complex than those underlying sustained phonation. Since the effects of Parkinson disease increase with the complexity of motor patterns, it is reasonable to expect earlier diagnoses if running speech is analyzed. More widely, by showing that it is possible to obtain interpretable patterns from recurrence plots representing the unfolding

over time of complex non stationary voicing signals, our contribution paves the way to a strongly inductive (i.e. bottom-up) approach to quantitative analysis of the dynamics underlying time series from speech production measurements.

Acknowledgments

This work, carried out within the Labex BLRI (ANR-11-LABX-0036) and ASLAN (ANR-10-LABX-0081), has benefited from support from the French government, managed by the French National Agency for Research (ANR), under the program title “Investissements d'Avenir” (ANR-11-IDEX-0001-02 and ANR-11-IDEX-0007).

References

Baken, R. J. (1992). Electroglottography. *Journal of Voice*, *6*(2), 98–110.
Berry, D. A. (2001). Mechanisms of modal and non-modal phonation. *Journal of Phonetics*, *29*(4), 431–450.
Blankenship, B. (2002). The timing of non-modal phonation in vowels. *Journal of Phonetics*, *30*(2), 163–191.
Cho, T., Jun, S. A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of phonetics*, *30*(2), 193–228.
Colton, R. H., & Conture, E. G. (1990). Problems and pitfalls of electroglottography. *Journal of Voice*, *4*(1), 10–24.
Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, *24*(4), 423–444.
De Jong, K., Beckman, M. E., & Edwards, J. (1993). The interplay between prosodic structure and coarticulation. *Language and Speech*, *36*(2–3), 197–212.
De Jong, K. J., Lim, B. J., & Nagao, K. (2002). Phase transitions in a repetitive speech task as gestural recomposition. *IULC Working Papers*, 2.
Docherty, G. J., & Foulkes, P. (2005). Glottal variants of /t/ in the Tyneside variety of English. In W. J. Hardcastle, & J. M. Beck (Eds.), *A figure of speech. A festschrift for John Laver* (pp. 173–197). New York: Routledge.
Edmondson, J. A., & Esling, J. H. (2006). The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopic case studies. *Phonology*, *23*(2), 157.
Esling, J. H., & Moisik, S. R. (2012). Laryngeal aperture in relation to larynx height change: An analysis using simultaneous laryngoscopy and laryngeal ultrasound. In D. Gibbon, D. Hirst, & N. Campbell (Eds.), *Rhythm, melody and harmony in speech: Studies in honour of Wiktor Jassem* (pp. 117–128). Poznan: Polskie Towarzystwo Fonetykzne.
Fabre, P. (1958). Etude comparée des glottogrammes et des phonogrammes de la voix humaine. *Annuaire Oto-rhino Laryngologie*, *75*, 767–775.
Facchini, A., Kantz, H., & Tiezzi, E. (2005). Recurrence plot analysis of nonstationary data: The understanding of curved patterns. *Physical Review E*, *72*(2), 021915.
Fant, G. (1971). *Acoustic theory of speech production*. Berlin: Walter de Gruyter.
Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, *29*(2), 109–135.
Garellek, M. (2014). Voice quality strengthening and glottalization. *Journal of Phonetics*, *45*, 106–113.
Gerratt, B. R., & Kreiman, J. (2001). Toward a taxonomy of non-modal phonation. *Journal of Phonetics*, *29*(4), 365–381.
Gobl, C., & Ní Chasaide, A. (1999). Voice source variation in the vowel as a function of consonantal context. In W. J. Hardcastle, & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (pp. 122–143). Cambridge: Cambridge University Press.
Gobl, C., & Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, *40*(1), 189–212.
Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics*, *29*(4), 383–406.
Gordon, M. (2012). The phonetics and phonology of non-modal vowels: A cross-linguistic perspective. In *Annual meeting of the Berkeley Linguistics Society* (Vol. 24(1), pp. 93–105).
Hanson, H. M., Stevens, K. N., Kuo, H.-K. J., Chen, M. Y., & Slifka, J. (2001). Towards models of phonation. *Journal of Phonetics*, *29*(4), 451–480.
Henton, C. G., & Bladon, A. (1987). Creak as a sociophonetic marker. In L. Hyman, & C. Li (Eds.), *Language, speech and mind: Studies in honour of Victoria A. Fromkin* (pp. 3–29). London: Routledge.
Iseli, M., & Alwan, A. (2004). An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation. In *Proceedings of IEEE international conference on acoustics, speech, and signal processing* (*ICASSP'04*) (Vol. 1, p. I-669).
Iwanski, J. S., & Bradley, E. (1998). Recurrence plots of experimental data: To embed or not to embed?. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *8*(4), 861–871.
Kantz, H., & Schreiber, T. (2004). *Nonlinear time series analysis*, Vol. 7. Cambridge: Cambridge University Press.
Karnell, M. P. (1991). Laryngeal perturbation analysis: Minimum length of analysis window. *Journal of Speech, Language and Hearing Research*, *34*(3), 544.
Kennel, M. B., Brown, R., & Abarbanel, H. D. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, *45*(6), 3403–3411.
Kohler, K. J. (1994). Glottal stops and glottalization in German. *Phonetica*, *51*(1-3), 38–51.
Kong, T., Y., & Rosenfeld, A. (1996). *Topological algorithms for digital image processing*. Elsevier Science, Inc..
Ladefoged, P. (1973). The features of the larynx. *Journal of Phonetics*, *1*(1), 73–83.
Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.
Lancia, L., Avelino, H., & Voigt, D. (2013). Measuring laryngealization in running speech: Interaction with contrastive tones in Yalálag Zapotec. In *Proceedings of interspeech 2013* (pp. 602–606). Lyon, France.
Lancia, L., Fuchs, S., & Tiede, M. (2014). Application of concepts from Cross-Recurrence Analysis in speech production: An overview and a comparison to other nonlinear methods. *Journal of Speech, Language and Hearing Research*, *26*, 1–16.
Lancia, L., & Grawunder, S. (2014). Tongue–larynx interactions in the production of word initial laryngealization over different prosodic contexts: A repeated speech experiment. In *Proceedings of 10th international seminar on speech production* (pp. 241–244). Cologne, Germany.
Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.
Lim, J.-Y., Choi, J.-N., Kim, K.-M., & Choi, H.-S. (2006). Voice analysis of patients with diverse types of Reinke's edema and clinical use of electroglottographic measurements. *Acta oto-laryngologica*, *126*(1), 62–69.
Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A., & Moroz, I. M. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *BioMedical Engineering OnLine*, *6*(1), 23.
Löfqvist, A., Koenig, L. L., & McGowan, R. S. (1995). Vocal tract aerodynamics in /aCa/ utterances: Measurements. *Speech Communication*, *16*(1), 49–66.
Lucero, J. C., & Koenig, L. L. (2000). Time normalization of voice signals using functional data analysis. *The Journal of the Acoustical Society of America*, *108*, 1408–1420.
Lucero, J. C. (2005). Comparison of measures of variability of speech movement trajectories using synthetic records. *Journal of Speech, Language, and Hearing Research*, *48*(2), 336–344.
Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(7), 674–693.
Marwan, N., Thiel, M., & Nowaczyk, N. R. (2002). Cross recurrence plot based synchronization of time series. *Nonlinear Processes in Geophysics*, *9*(3/4), 325–331.
Marwan, N. (2011). How to avoid potential pitfalls in recurrence plot based data analysis. *International Journal of Bifurcation and Chaos*, *21*(04), 1003–1017.
Marwan, N., Romano, M., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, *438*(5), 237–329.
Marwan, N., & Kurths, J. (2005). Line structures in recurrence plots. *Physics Letters A*, *336*(4), 349–357.
Matar, N., Portes, C., Lancia, L., Legou, T., & Baider, F. (2014). Acoustic correlates of masculinity in voices of Lebanese women with Reinke's edema. In *Proceedings of 10th international seminar on speech production*. Cologne, Germany (pp. 273–276).
Morris, J. S., & Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(2), 179–199.
Moisik, S. & Esling, J. H. (2011). The 'whole larynx' approach to laryngeal features. In *Proceedings of the 17th international congress of phonetic sciences* (*ICPhS 17*). Hong Kong (pp. 1406–1409).
Moisik, S. & Gick, B. (2013). The quantal larynx revisited. In *Proceedings of meetings on acoustics* (Vol. 19(1), pp. 060163).
Moisik, S. R., & Esling, J. H. (2014). Modeling the Biomechanical Influence of Epilaryngeal Stricture on the Vocal Folds: A Low-Dimensional Model of Vocal–Ventricular Fold Coupling. *Journal of Speech, Language, and Hearing Research*, *57*(2), S687–S704.
Mooshammer, C. (2010). Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German. *The Journal of the Acoustical Society of America*, *127*(2), 1047–1058.

Neubauer, J., Mergell, P., Eysholdt, U., & Herzel, H. (2001). Spatio-temporal analysis of irregular vocal-fold oscillations: Biphonation due to desynchronization of spatial modes. *The Journal of the Acoustical Society of America*, *110*, 3179–3192.
Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. J. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 90–117). Cambridge: Cambridge University Press.
Pinheiro, J. C., & Bates, D. M. (2000). *Linear mixed-effects models: Basic concepts and examples*. Springer.
Pompino-Marschall, B., & Żygis, M. (2011). Glottal marking of vowel-initial words in German. In *Proceedings of the 17th international congress of phonetic sciences (ICPhS 17)*. (pp. 1626–1629).
Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, *29*(4), 407–429.
Roberts, J. (2006). As old becomes new: Glottalization in Vermont. *American Speech*, *81*(3), 227–249.
Rodgers, Jonathan. 1999. Three influences on glottalization in read and spontaneous German speech. *Arbeitsberichte des Instituts fur Phonetik und digitale Sprachverarbeitung der Universitat Kiel (AIPUK)* (Vol. 34, pp. 177 – 284).
Rothenberg, M. (1973). A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *The Journal of the Acoustical Society of America*, *53*, 1632–1645.
Rothenberg, M. (1977). Measurement of airflow in speech. *Journal of Speech, Language and Hearing Research*, *20*(1), 155–176.
Rothenberg, M., & Mahshie, J. J. (1988). Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech, Language, and Hearing Research*, *31*(3), 338–351.
Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *26*(1), 43–49.
Scherer, R. C., Vail, V. J., & Guo, C. G. (1995). Required number of tokens to determine representative voice perturbation values. *Journal of Speech, Language and Hearing Research*, *38* (6), 1260–1269.
Schinkel, S., Dimingen, O., & Marwan, N. (2008). Selection of recurrence threshold for signal detection. *The European Physical Journal Special Topics*, *164*(1), 45–53.
Sedgewick, R. (1998). *Algorithms in C*. Addison-Wesley.
Silverman, D. (1997). Laryngeal complexity in Otomanguean vowels. *Phonology*, *14*(2), 235–261.
Slifka, J. (2006). Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice*, *20*(2), 171–186.
Takens, F. (1981). Detecting strange attractors in turbulence. In D. A. Rand, & L.-S. Young (Eds.), *Dynamical systems and turbulence, lecture notes in mathematics*, *Vol. 898* (pp. 366–381). Berlin: Springer.
Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs: Prentice Hall.
Tuller, B., & Kelso, J. S. (1991). The production and perception of syllable structure. *Journal of Speech, Language, and Hearing Research*, *34*(3), 501–508.
Webber, C. L., & Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology*, *76*(2), 965–973.
Xu, Y., & Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *The Journal of the Acoustical Society of America*, *111*(3), 1399–1413.