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