Available online at www.sciencedirect.com

ScienceDirect

Computer Speech and Language 28 (2014) 1233–1253

COMPUTER SPEECH AND LANGUAGE

www.elsevier.com/locate/csl

Data-driven detection and analysis of the patterns of creaky voice☆

Thomas Drugman [a,*], John Kane [b], Christer Gobl [b]

[a] *TCTS Lab, University of Mons 31, 7000 Mons, Belgium*

[b] *Phonetics and Speech Laboratory, School of Linguistic, Speech and Communication Sciences, Trinity College Dublin, Ireland*

Received 12 April 2013; received in revised form 3 December 2013; accepted 7 March 2014

Available online 15 March 2014

Abstract

This paper investigates the temporal excitation patterns of creaky voice. Creaky voice is a voice quality frequently used as a phrase-boundary marker, but also as a means of portraying attitude, affective states and even social status. Consequently, the automatic detection and modelling of creaky voice may have implications for speech technology applications. The acoustic characteristics of creaky voice are, however, rather distinct from modal phonation. Further, several acoustic patterns can bring about the perception of creaky voice, thereby complicating the strategies used for its automatic detection, analysis and modelling. The present study is carried out using a variety of languages, speakers, and on both read and conversational data and involves a mutual information-based assessment of the various acoustic features proposed in the literature for detecting creaky voice. These features are then exploited in classification experiments where we achieve an appreciable improvement in detection accuracy compared to the state of the art. Both experiments clearly highlight the presence of several creaky patterns. A subsequent qualitative and quantitative analysis of the identified patterns is provided, which reveals a considerable speaker-dependent variability in the usage of these creaky patterns. We also investigate how creaky voice detection systems perform across creaky patterns.

Keywords: Creaky voice; Vocal fry; Irregular phonation; Glottal source

1. Introduction

This paper presents an empirical investigation of the temporal excitation patterns related to the voice quality often referred to as creaky voice. Creaky voice is a raspy or croaking quality of the voice generally produced with a very low pitch and often with highly irregular periodicity (Laver, 1980). Creaky voice is used for a variety of functions in spoken communication and, hence, presents both an opportunity and a challenge (because of its distinctive acoustic characteristics) for speech technology.

1.1. Terminology

One major difficulty with studying creaky voice, and indeed voice quality in general, is the problem of reconciling the variation in the terminology used in the literature. Many studies use the terms *irregular phonation* (Slifka, 2006;

☆ This paper has been recommended for acceptance by R.K. Moore.

* Corresponding author. Tel.: +32 65374749.

E-mail addresses: thomas.drugman@umons.ac.be (T. Drugman), kanejo@tcd.ie (J. Kane), cegobl@tcd.ie (C. Gobl).

http://dx.doi.org/10.1016/j.csl.2014.03.002

0885-2308/

Surana and Slifka, 2006a; Vishnubhotla and Espy-Wilson, 2006; Böhm et al., 2010) or *glottalisation* (Dilley et al., 1996; Redi and Shattuck-Hufnagel, 2001). However, both these terms are rather broad and indeed cover other classes of phonation (or laryngeal activity) aside from what we consider as *creaky voice*. In this paper, we interpret creaky voice based solely on the auditory criterion "a rough quality with the additional sensation of repeating impulses" as is done in Ishi et al. (2008a) (which is close to that in Laver, 1980). Note that the ". . .sensation of repeating impulses" part clearly discriminates the voice quality from harsh voice. Any speech region displaying such a quality will be treated in this paper as creaky voice. Note that such a criterion will contain both creak and creaky voice, and, hence, will not apply the discrimination of the two used in Laver (1980). The term vocal fry (perhaps used more by American researchers; Laver, 1980) is often used in the literature (Hollien and Wendahl, 1968; Ishi et al., 2008a; Wolk and Abdelli-Beruh, 2012), and is likely to corresponding closely to our working definition of creaky voice.

1.2. Creaky voice in speech communication

Creaky voice has been studied in relation to various functions in speech communication, and most commonly with phrase or sentence boundary marking (Surana and Slifka, 2006; Drugman et al., 2013). Similarly creaky voice has been associated with turn-yielding in Finnish (Ogden, 2001). However, creaky voice is likely to be also implicated in a range of speech functions other than boundary marking. It has been studied in relation to hesitations in Swedish (Carlson et al., 2006), and also creaky voice-like properties have been observed as an allophonic variant of word medial oral stops (Zue and Laferriere, 1979; Crystal and House, 1988). Creaky voice has also been investigated in terms of emotion and affectively coloured speech (Yanushevskaya et al., 2005; Gobl and Ní Chasaide, 2003; Ishi et al., 2008b) and is likely to be a significant correlate of subtle variation in levels of activation and formality of the speaking setting (Kane et al., 2011). The use of creaky voice has also recently been shown to be increasingly common for young American females (Wolk and Abdelli-Beruh, 2012) and has also been linked to the portrayal of social status (Yuasa, 2010).

1.3. Creaky voice in speech technology

As a consequence of its implication in these various roles, the proper handling of creaky voice, and the distinctive acoustic characteristics associated with it, has a significant importance for speech technology. For speech synthesis (see e.g. Silen et al., 2009; Drugman et al., 2012), this could result in improved naturalness for speakers who use creaky voice and also for the development of expressive speech synthesis. As it has been shown that listeners are sensitive to creaky voice in terms of recognising a speaker's identity (Böhm and Shattuck-Hufnagel, 2007), it is likely beneficial for speaker recognition systems (Espy-Wilson et al., 2006; Elliot, 2002) to exploit information to do with creaky voice. Detection of creaky voice may also benefit developments in emotion recognition and conversational analysis.

1.4. Physiology

Although the focus of this paper is on the acoustic characteristics of creaky voice, we include here a brief outline of some of the physiological characteristics reported as being associated with creaky voice. Laver (1980) provides one of the more comprehensive descriptions of creaky voice. Here he describes creaky voice as involving low subglottal pressure, high levels of adductive laryngeal tension (i.e. the muscular tension involved in bringing the vocal folds together) and typically involves low levels of longitudinal vocal fold tension (probably the main physiological parameter utilised for pitch variation). Edmondson and Esling (2006) provide some additional physiological evidence, in particular details on the presence of *ventricular incursion*. This involves the ventricular folds pushing down and covering the *true* vocal folds, causing an increased mass. This has the consequence of lowering the frequency of vibration and often causing secondary vibrations above the glottis (Moisik and Esling, 2011).