Available online at www.sciencedirect.com
ScienceDirect
Computer Speech and Language xxx (2014) xxx–xxx
COMPUTER SPEECH AND LANGUAGE
www.elsevier.com/locate/csl

Analysis of production characteristics of laughter☆

Vinay Kumar Mittal *, Bayya Yegnanarayana

Speech and Vision Laboratory, International Institute of Information Technology, Hyderabad 500032, India

Received 16 November 2013; received in revised form 14 May 2014; accepted 28 August 2014

Abstract

In this paper, the production characteristics of laughter are analysed at call and bout levels. Data of natural laughter is examined using electroglottograph (EGG) and acoustic signals. Nonspeech-laugh and laughed-speech are analysed in comparison with normal speech using features derived from the EGG and acoustic signals. Analysis of EGG signal is used to derive the average closed phase quotient in glottal cycles and the average instantaneous fundamental frequency (F_0). Excitation source characteristics are analysed from the acoustic signal using a modified zero-frequency filtering (mZFF) method. Excitation impulse density and the strength of impulse-like excitation are extracted from the mZFF signal. Changes in the vocal tract system characteristics are examined in terms of the first two dominant frequencies derived using linear prediction (LP) analysis. Additional excitation information present in the acoustic signal is examined using a measure of sharpness of peaks in the Hilbert envelope of the LP residual at the glottal closure instants. Parameters representing degree of change and temporal changes in the production features are also derived to study the discriminating characteristics of laughter from normal speech. Changes are larger for nonspeech-laugh than laughed-speech, with reference to normal speech.

Keywords: Laughter analysis; Nonspeech-laugh; Laughed-speech; Modified zero-frequency filtering; Dominant frequency; Closed phase quotient; EGG; Differenced EGG

1. Introduction

Laughter is a vocal-expressive communicative signal produced by human speech production mechanism, that occurs as either nonlinguistic event or interspersed with normal speech. Laughter signals can have widely varying acoustic features. Laughter characteristics are analysed usually at episode, bout, call or segment *levels*. An *episode* consists of two or more laughter bouts, separated by inspirations. A laugh *bout* is an acoustic event, produced during one exhalation, or inhalation sometimes. The period of laughter vocalization contains one or more laugh-cycles or laugh-pulses, called *calls*, interspersed with pauses. Calls are also referred to as notes or laugh-syllables. *Segments* reflect changes in the production mode within a call, that can be seen better in the components of spectrogram (Moore and Von Leden, 1958; Bachorowski et al., 2001). A laughter bout consists of three parts: *onset*, that has short steep laughter, *apex*, the vocalization part, and *offset*, the post-vocalization part in which the smile fades out smoothly (Ruch and

☆ This paper has been recommended for acceptance by T. Kawahara.

* Corresponding author. Tel.: +91 40 66531422; fax: +91 40 66531413; mobile: +91 9866648770.
E-mail addresses: vinay.mittal@iiit.ac.in, vkmittal1967@outlook.in (V.K. Mittal).

http://dx.doi.org/10.1016/j.csl.2014.08.004

Please cite this article in press as: Mittal, V.K., Yegnanarayana, B., Analysis of production characteristics of laughter. Comput. Speech Lang. (2014), http://dx.doi.org/10.1016/j.csl.2014.08.004

Ekman, 2001). Number of calls in a laugh bout (3–8) is limited by the dynamic change (reduction) in the volume of lungs. Typically up to four calls occur in a laugh bout (Provine and Yong, 1991; Rothganger et al., 1998). In this study, laughter is analysed at bout and call levels.

Categorization of laughter sounds was carried out in several studies in different ways. Laughter was categorized into three classes: *spontaneous* laughter, that occurs without restrain on its expression, *voluntary* laughter, a kind of faked laughter, and *singing* laughter, that has breathiness, aspiration and phonation with lesser resonance in trachea (Ruch and Ekman, 2001). Three types of laugh bouts were discussed in (Bachorowski et al., 2001): *song-like* laugh involving pitch modulation of voiced sounds (like in giggle or chuckle), *snort-like unvoiced call* with salient turbulence in nasal-cavity, and *unvoiced grunt-like* laugh including breathy pants and harsher cackles. Three classes of vowel quality *ha*, *he* and *ho* of laugh sounds were studied in (Provine and Yong, 1991; Provine, 2000). Laughter was also categorized as *voiced laughter*, that involves regular vocal fold vibration like in melodic song-like bouts and giggles, and *unvoiced laughter*, that includes open-mouth breathy sounds, closed-mouth grunts and nasal-snorts (Owren and Bachorowski, 2003). The continuum from speech to laugh was divided into *speech*, *speech-laugh* and *laugh* (Nwokah et al., 1999; Menezes and Igarashi, 2006). Laughter in dialogic interaction was categorized as: *speech-smile*, *speech-laugh* and *laughter* (Kohler, 2008). Four phonetic types of laughter were studied in (Campbell et al., 2005; Tanaka and Campbell, 2011): *voiced*, *chuckle*, *breathy ingression* and *nasal-grunt*.

Since laughter signal is produced by human speech production mechanism, its production characteristics can be analysed in terms of the excitation source and vocal tract system characteristics, like for normal speech. Significant changes apparently take place in the characteristics of the excitation source, during production of laughter. But, acoustic analyses of laughter have been carried out mostly using spectral and perceptual features (Bickley and Hunnicutt, 1992; Bachorowski et al., 2001; Truong and Leeuwen, 2007; Makagon et al., 2008). Features such as fundamental frequency, root mean square amplitude, time duration and formant structure were examined in the acoustic analysis (Bickley and Hunnicutt, 1992), to discriminate laughter and speech. Acoustic features such as F_0, number of calls per bout, spectrograms and formant clusters ($F2$ *vs* $F1$) were used for analysing temporal and source-filter effects of laughter (Bachorowski et al., 2001). In a study, rhythm (duration) and changes in F_0 were used as features for evaluating laughter bouts (Kipper and Todt, 2003). Pairwise feature combinations such as pitch-energy, global pitch-voicing, perceptual linear prediction-modulation spectrum were used for modeling laughter and speech (Truong and Leeuwen, 2005). Another study used the degree of variation in F_0, intensity and durational patterning (onset, main part, pause and offset) features for assessing naturalness of synthesized laughter (Lasarcyk and Trouvain, 2007). Source feature like glottal open quotient were considered along with spectral tilt (Menezes and Igarashi, 2006), but these features were derived (approximately) from differences in amplitudes of harmonics in the spectrum. Features such as instantaneous pitch period, strength of excitation, and their slopes and ratio were proposed for laughter analysis (Sudheer Kumar et al., 2009).

In the current study, we examine changes in the glottal excitation source characteristics and associated changes in the vocal tract system characteristics, during production of laughter. Laughter at bout and call levels is analysed. Production characteristics of the speech-laugh continuum are analysed in three categories: *normal speech (NS)*, *laughed-speech (LS)* and *nonspeech-laugh (NSL)*. Laughed-speech consists of (linguistic) speech interspersed with (nonlinguistic) laugh content to some degree. Only voiced nonspeech-laugh, produced spontaneously, is considered. Data consists of natural laugh responses. In each case, both electroglottograph (EGG) (Fant et al., 1985) and acoustic signals are examined. Changes in the glottal vibration characteristics are examined using features such as closed phase quotient in each glottal cycle (Mittal and Yegnanarayana, 2013a) and F_0 (i.e., $F_{0_{\mathrm{EGG}}}$), both derived using the differenced EGG signal (Gobl, 1988). Excitation source features are also extracted from the acoustic signal using a modification of the zero-frequency filtering (ZFF) method (Murty and Yegnanarayana, 2008). Features such as excitation impulse density and strength of excitation are derived. Changes in the vocal tract system characteristics are examined using the first two *dominant frequencies* (F_{D_1} and F_{D_2}) (Mittal and Yegnanarayana, 2013b), derived from the acoustic signal using linear prediction (LP) analysis (Makhoul, 1975). Production features are also examined using a sharpness measure (Seshadri and Yegnanarayana, 2009) of peaks in the Hilbert envelope of LP residual (Markel and Gray, 1982) of the acoustic signal around glottal closure instants. Voiced/nonvoiced decision (Dhananjaya and Yegnanarayana, 2010) is based on the framewise energy of the modified ZFF output signal. Parameters derived to measure the degree of changes and temporal changes in the production features are also explored to discriminate NS, LS and NSL.

The paper is organized as follows. Section 2 discusses details of the data collected for this study. The signal processing methods used for deriving the source and system characteristics are discussed in Section 3. Changes in

Please cite this article in press as: Mittal, V.K., Yegnanarayana, B., Analysis of production characteristics of laughter. Comput. Speech Lang. (2014), http://dx.doi.org/10.1016/j.csl.2014.08.004