Computers and Electrical Engineering 000 (2017) 1–13

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Amused speech components analysis and classification: Towards an amusement arousal level assessment system☆

Kevin El Haddad*, Hüseyin Çakmak, Stéphane Dupont, Thierry Dutoit

TCTS lab - University of Mons, Belgium

A R T I C L E I N F O

Article history:
Received 3 March 2016
Revised 12 June 2017
Accepted 12 June 2017
Available online xxx

Keywords:
Amusement intensity level
Laughter
Smile
Paralinguistic
Affective computing
Machine learning application

A B S T R A C T

In this paper, we present our work on analysis and classification of smiled vowels, chuckling (or shaking) vowels and laughter syllables. This work is part of a larger framework that aims at assessing the level of amusement in speech using the audio modality only. Indeed all of these three categories occur in amused speech and are considered to contribute in the expression of different levels of amusement. We first analyze these three amused speech components on the acoustic level. Then, we improve a classification system we previously developed. With a limited amount of data and features, we are able to obtain good classification results with different systems. Among the compared systems, the best one achieved 82.8% of accuracy, therefore outperforming chance.

1. Introduction

Affective computing and more specifically, emotion recognition is currently one of the hottest research topics due to its potential in many different application areas. Applications involve Human-Computer Interactions (HCI), medical and social areas. Emotions can trigger expressions in different modalities, one of the most important being speech. Previous work on emotion recognition from speech mostly focused on classifying several types of emotions [1–3]. The work we present here rather focuses on a single emotion of positive valence: amusement. It is part of a larger framework of assessing accurately the amusement arousal level in speech using information from the audio modality only. In fact most of the previous work related the emotion arousal or intensity level estimation tackles the problem by using multimodal data. For instance Patwardhan and Knapp present work in [4] on estimating the intensity level of anger expressions using speech and motion capture data. Also, Yoshiko et. al. estimate anger intensity level using speech and linguistic data [5]. Dhall and Goecke propose an estimation of different levels of smile and laugh using also multimodal data in [6] (smiling and laughter in that work aren't necessarily expressions for amusement).

In this paper, we propose a preliminary analysis and classification work. Indeed, we present analyses of amused speech components and improved results on classifying these components compared to a previous classification attempt [7]. The ultimate goal being, as mentioned earlier, to estimate accurately the intensity level of amusement in speech using a single modality only, the audio.

☆ Reviews processed and approved for publication by Editor-in-Chief.

* Corresponding author.
E-mail addresses: kevin.elhaddad@umons.ac.be, kevin.el.haddad@gmail.com (K. El Haddad).

http://dx.doi.org/10.1016/j.compeleceng.2017.06.012

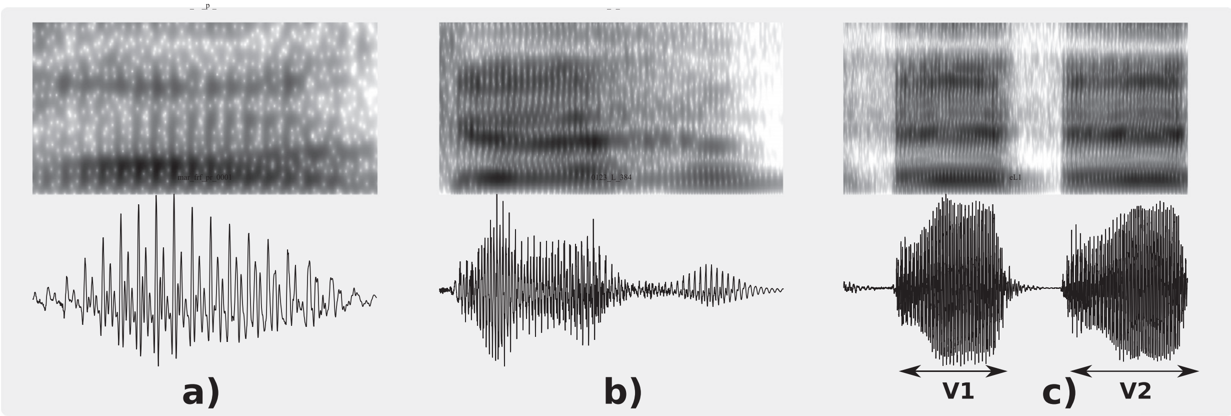

Fig. 1. Representation on the spectral (above) and time (below) domains of (a) a smiled vowel, (b) a chuckling vowel and (c) a laughter syllable.

Smiled vowels, chuckling (or shaking) vowels [8] and laughter syllables can all be found in amused speech and are therefore considered here as its components. We will refer to them as Amused Speech Components (ASC). We consider that two main dimensions constitute amusement in speech. The first one is the smile which is not only a visual expression, but also identifiable audibly [9,10]. The second one is laughter. Laughter interrupting and/or intermingling with speech causes what is called speech-laughs [11]. The estimation of the amusement arousal/intensity level in speech needs, by definition, the establishment of different levels. These levels should be based on the two main amused speech components previously mentioned, i.e. speech-smile and laughter (and/or speech-laugh). Indeed, our hypothesis is that amusement intensity level of an uttered sentence depends on the presence of these two components and is correlated with their intensities. The presence (or absence) of each ASC, and their combination in an uttered sentence, could be representative of an intensity level.

In our previous works regarding ASC [7,12], classification features were extracted based on observations made on data we collected. The efficiency of these features was tested in a classification task with different machine learning algorithms. Motivated by the good results obtained in these works, we present a more detailed analysis of the ASC, new discriminating feature sets, and improved classification results using these features.

The remaining of this paper is organized as follows. We first present the data collected for the purpose of this work in Section 2. In that section, we will start by giving a more detailed definition of the ASC. We will then present the data collection protocol, followed by the analysis of data. Section 3 will summarize our previous works and recall the initial ASC classification results obtained. Then, the new feature sets will be introduced in Section 4. The improved classification results obtained will be presented in Section 5. We will finally conclude and give our perspectives for future work in Section 6.

2. Amused speech components data

2.1. Description

Speech-smile is a term used to describe the alteration of speech due to smiling. As already mentioned, smiles can indeed be audibly identifiable in speech [9,10]. It is therefore possible to make use of this dimension in this work.

A prototypical laughter event is a sequence of fricatives and vowels. A laughter syllable is the succession of a fricative and a vowel (e.g. a “ha” sound). Please note that we use the term “fricative” here to refer only to the /*h*/ which is the most present fricative sound in laughter [13]. In this work, the pattern for laughter syllables can also be described as the succession of two fricatives because this pattern can be found in natural laughs and, in particular, in our data.

Chuckling or shaking vowels, as presented in [7,12], are vowels altered by some kind of tremolo in an amused sentence. This can be seen by comparing the smiled vowels, chuckling vowels and laughter syllable patterns. Fig. 1 shows the common temporal and spectral representations of a smiled vowel (a), a chuckling vowel (b) and (c) two consecutive laughter syllables found in our data. Some observations can be made out of these figures. First, a discontinuity in the spectral representation of the chuckling vowel can be noticed. A more obvious and accentuated discontinuity can be observed in the laughter syllables spectral representation separating the first vowel and the second one (V1 and V2 respectively, in Fig. 1(c)). Since, in the case of the laugh, this discontinuity is due to the fricative separating the two vowels and considering the fact that both phenomena occur in amused speech (and apparently at two different levels of amusement), the discontinuity in the case of the chuckling vowel must also be due to an air pulse.

After a comparison of the chuckling and smile vowels in the temporal domain, it seems like the chuckling vowel pattern is formed of a sequence of two vowels separated by a “breathy vowel” (which would be located where the discontinuity is). The conclusion drawn from these observations is that the tremolo-like sound perceived in chuckling vowels is produced by the alteration of a vowel (probably a speech-smile vowel) by a muscular mechanism similar to the one producing laughter (please refer to Ruch and Ekman [13] for more details on laughter production).

Please cite this article as: K. El Haddad et al., Amused speech components analysis and classification: Towards an amusement arousal level assessment system, Computers and Electrical Engineering (2017), http://dx.doi.org/10.1016/j.compeleceng.2017.06.012