G Model
PEC 5297 No. of Pages 6
ARTICLE IN PRESS
Patient Education and Counseling xxx (2015) xxx–xxx
Contents lists available at ScienceDirect
Patient Education and Counseling
journal homepage: www.elsevier.com/locate/pateducou

A case study of vocal features associated with galvanic skin response to stressors in a clinical interaction

Marie Postma-Nilsenová[a,*], Erik Holt[b], Lena Heyn[c], Kim Groeneveld[a], Arnstein Finset[b]

[a] Tilburg center for Cognition and Communication, Tilburg University, The Netherlands
[b] Department of Behavioural Science, Institute of Basic Medical Sciences, Faculty of Medicine, University of Oslo, Norway
[c] Lovisenberg Diaconal University College, Oslo,Norway

ARTICLE INFO

Article history:
Received 2 February 2015
Received in revised form 10 February 2016
Accepted 5 March 2016

Keywords:
Medical communication
Psychophysiology
Voice
Stress

ABSTRACT

Objective: We investigated vocal characteristics associated with physiologically determined stressful episodes by means of post-hoc acoustic analyses of speech recorded in a clinical setting. Our research addressed the understudied question of which vocal features may serve as cues naturally occurring stress and is the first to explore this issue in a pitch accent language.

Methods: The vocal profile of a single female patient interacting with a physician was analyzed with standard speech analysis software for acoustic indicators of stress-related arousal determined by galvanic skin response measurements.

Results: Vocal jitter, representing an aspect of voice quality perceived as hoarseness, appeared to increase during and immediately after skin conductance response intervals. Skin conductance levels during the response intervals were negatively correlated with acoustic features used to approximate the perception of voice unsteadiness (slope and standard deviation of fundamental frequency).

Conclusion: An acoustic analysis of vocal properties of speech uttered during independently detected skin conductance response intervals revealed individual patterns for some acoustic features linked to stress in earlier studies.

Practice implications: Non-invasive methods of arousal detection in physician-patient communication based on acoustic analyses of vocal profiles may, in combination with other analyses, help identify stressful events and thus improve the process of medical information gathering and decision-making.

1. Introduction

Stress occurs frequently in doctor-patient communication; it is not only experienced by patients, but also by medical professionals conveying bad news [1]. Next to its negative impact on the doctor-patient relationship, the presence of stress may lead to cognitive overload and thus negatively influence the process of medical information gathering and decision-making [2]. Arguably, stress in the course of a medical interaction cannot always be easily detected due to social display rules, i.e., the socio-cultural norms on expressions of emotions that can override natural behavioral reactions.

Current techniques of stress, arousal and affect detection in physician-patient interactions involve manual coding schemes for verbal and nonverbal cues [3,4] combined with evaluations based on self-reports and measurements of psychophysiological responses. Of these three types of measurements, manual coding is likely to involve the highest number of man hours involved in annotations of behavioral cues [5]. With respect to self-reports, past studies have indicated that subjective estimates of stress, collected before or after an exposure to a stressor, cannot always be associated with actual physiological responses measured during the event itself, at least with respect to stress induced in a laboratory, e.g., by means of the Trier Social Stress Test (see Ref. [6] for a review). For psychophysiology, past studies have shown that bodily reactions accompanying acute stress, such as increase in heart rate, rapid blood flow, activation of sweat glands, and increase in the respiration rate [7] can be measured using modern technology sensors. Among these, the most commonly used method in research on medical interactions involves galvanic skin response measurements, which assess sweating of the skin [8,9].

* Corresponding author at: Tilburg center for Cognition and Communication, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.
E-mail addresses: Marie.Postma@tilburguniversity.edu, m.nilsenova@uvt.nl, m.nilsenova@gmail.com (M. Postma-Nilsenová).

http://dx.doi.org/10.1016/j.pec.2016.03.006
0738-3991/

Contrary to self-reports, psychophysiological methods offer the advantage of continuous assessment of affective responses [2].

Like psychophysiological measurements, vocal characteristics, such as the intensity of the voice, its pitch and harmonic energy, are important sources of information about speakers' affective states in terms of process tracing over a time line [10,11]. This is due to the mechanism of speech production, in particular, *phonation*, that may be influenced by physiological processes related to emotional states. Phonation is created by modifications of the airstream in the larynx, including vocal fold vibration; in turn, these modifications give rise to a complex sound consisting of a fundamental frequency *F0* (the acoustic correlate of pitch[1]) and harmonic overtones (multiples of *F0*). Any change in the airstream pressure, due to respiratory rate, or the tension of the respiratory and the laryngeal muscles that help stretch or relax the vocal folds [12–14], will affect phonation and give rise to vocal characteristics that can be perceptible to the human ear and can be reliably measured with currently available acoustic software [15,16]. For example, increased muscle tension associated with increased levels of stress would result in higher vocal pitch [17] that might give rise to the perception of a 'strained' voice [18] and be experienced by the speakers as 'a lump in the throat' and an actual muscle tension in the larynx [18]. Acoustic measurements of vocal characteristics thus possibly offer an alternative to or an enrichment of existing methods of arousal detection and thus help identify potentially stressful events in medical interactions. Post-hoc or even real-time acoustic analyses of short samples of speech might provide means to collect indicators of stressors in the communication process in a non-invasive low-cost way. Next to that, speech analysis is likely to reveal perceived stress levels in a less costly manner than manual annotations of behavioral cues and in a more reliable way than pre- and post-stressor self-reports for fragments collected during the stressful event [19].

The link between a physiological reaction to a stimulus and its reflection in the human voice can be illustrated on the example of high arousal emotional states. Arousal—an increased activity of the sympathetic nervous system (SNS) leads to an increased muscle tension and respiratory rate [20–22], which will provoke a change both in voice quality and pitch range [10,23]. Since arousal is present in a variety of emotional states (both with a positive and negative valence), vocal features associated with high arousal, such as high F0, might occur for very different emotions (e.g., anger and elation); similarly, vocal features linked to low arousal can be indicative of both sadness and boredom [10]. To wit, research findings for vocal cues to emotions [24] typically offer descriptions of overlapping clusters of features rather than unique emotion detection criteria. In many contexts, though, arousal detection from vocal output might be sufficient as a cue to a possible affective event associated with increased arousal levels (as in the case of stress detection), especially if it is combined with information about the context of the interaction, the semantic content of utterances, as well as psychophysiological measurements.

Past research shows that stress can be reflected in vocal parameters such as fundamental frequency (*F0*), jitter (irregularity in *F0* due to irregularities between successive glottal pulses), shimmer (amplitude irregularities) or speaking rate, due to the effect of increased heart rate and bronchodilation [26]. Depending on the context and the speaker, vocal indicators of stress may be somewhat individual [27,28,7]. Interestingly, as reported by Kurniawan et al. [7], in certain settings, the vocal parameters might be more successful predictors of acute stress than direct physiological measures which differ from person to person, with variances in age, gender, ethnicity, and hormonal cycle.

Currently, the specific speech characteristics associated with stress responses detected by physiological measurements are relatively understudied [19], in particular with respect to naturally occurring stressors that have not been induced by laboratory procedures. The first contribution of the research reported here concerns the identification of possible stress markers in speech collected during a naturally occurring clinical interaction. Second, existing studies of vocal indicators of stress so far did not include languages in which local (i.e., restricted to one or two syllables) changes in fundamental frequency provide important linguistic information regarding the meaning of words. In these so-called pitch-accented languages, typically described as languages with a 'singing quality', changes in *F0* can be used to distinguish between part-of-speech categories (nouns *vs.* verbs), morphological categories (singular *vs.* plural) or simply between different morphological stems. In Norwegian, the language of the patient explored in this study, *F0* changes on a syllable- or word-level may be quite elaborate and dependent both on the dialect of the speaker and the word context.[2] Although little is known about the vocal characteristics related to affective expressions in Norwegian, the use of *F0* as an affective marker is limited in languages with a high degree of tonality compared to non-tonal languages such as English.

The aim of this study was to explore the link between available galvanic skin response measurements and vocal features. Given that vocal analyses and skin conductance measurements are typically conducted independently of each other, the objective was to find a link between the two types of cues. We made use of computational techniques for speech analysis [16] in order to obtain an accurate representation of the changes in the speaker's vocal profile. Compared to observational analyses, the computational approach has the advantage of overcoming various problems in the human perception of vocal profiles, such as individual differences in sensitivity to different components of complex sounds [27] and contextual effects [28].[3]

Since increased arousal is typically linked to an increased skin conductance and heart rate, which, in turn, contributes to vocal jitter and *F0* increase [25], we expected that the levels of vocal jitter and *F0* would increase during skin conductance response intervals. In view of the perceptual association between an unsteady, trembling voice and stress [18], we also assumed that the standard deviation of *F0* and its slope, both acoustic measures related to the degree of local changes in the F0, would be higher during the skin conductance response intervals.

2. Method

In the study reported here, we analyzed a subset of the data collected by Refs. [29,30] at the University of Oslo, containing audio taped clinical interactions between physicians and patients in follow-up consultations for leukemia, lymphoma, multiple myeloma, or testicular cancer in an outpatient clinic. The patients whose recordings were included in the data set were all 18 years or older and possessed the cognitive skills to comprehend the information provided about the purpose of the study. During the sessions, their skin conductance (SC) measurements were recorded and analyzed following the procedures reported in Ref.

[1] From a production point of view, *F0* stands for the rate at which the vocal folds open and close. Acoustically, it is defined as the lowest periodic cycle component of the acoustic waveform. The *F0* contour, sometimes referred to as the 'melody' of an utterance, is the sequence of *F0* values across the utterance.

[2] The classical example involves the distinction between 'beans' and 'farmers' (*bønner* vs. *bønder*), which in spoken Norwegian is expressed by means of tone.

[3] For example, our interpretation of the intensity and voice quality of speech is typically relative to the prosodic properties of the context in which an utterance occurs.

Please cite this article in press as: M. Postma-Nilsenová, et al., A case study of vocal features associated with galvanic skin response to stressors in a clinical interaction, Patient Educ Couns (2016), http://dx.doi.org/10.1016/j.pec.2016.03.006