Psychiatry Research 226 (2015) 235–241

Contents lists available at ScienceDirect

Psychiatry Research

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

Vocal acoustic analysis as a biometric indicator of information processing: Implications for neurological and psychiatric disorders

Alex S. Cohen[a,*], Thomas J. Dinzeo[b], Neila J. Donovan[c], Caitlin E. Brown[c], Sean C. Morrison[a]

[a] *Louisiana State University, Department of Psychology, United States*
[b] *Rowan University, Department of Psychology, United States*
[c] *Louisiana State University, Department of Communication Sciences and Disorders, United States*

ARTICLE INFO

Article history:
Received 4 March 2014
Received in revised form
11 August 2014
Accepted 30 December 2014
Available online 14 January 2015

Keywords:
Emotion
Prosody
Affect
Expression
Speech
Load
Cognition

ABSTRACT

Vocal expression reflects an integral component of communication that varies considerably within individuals across contexts and is disrupted in a range of neurological and psychiatric disorders. There is reason to suspect that variability in vocal expression reflects, in part, the availability of “on-line” resources (e.g., working memory, attention). Thus, understanding vocal expression is a potentially important biometric index of information processing, not only across but within individuals over time. A first step in this line of research involves establishing a link between vocal expression and information processing systems in healthy adults. The present study employed a dual attention experimental task where participants provided natural speech while simultaneously engaged in a baseline, medium or high nonverbal processing-load task. Objective, automated, and computerized analysis was employed to measure vocal expression in 226 adults. Increased processing load resulted in longer pauses, fewer utterances, greater silence overall and less variability in frequency and intensity levels. These results provide compelling evidence of a link between information processing resources and vocal expression, and provide important information for the development of an automated, inexpensive and uninvasive biometric measure of information processing.

1. Introduction

Deficits in cognition reflect a critical facet of a wide range of central nervous system (CNS) disorders, such as strokes, neurodegenerative disorders and developmental disorders and are one of the most critical features for understanding functioning in severe mental illness (Green et al., 2000). While most clinical studies of basic cognition focus on discrete abilities such as language, memory and psychomotor speed (Lezak, 2012), there is growing interest in processing capacity related to the conscious control of attention, concentration, working memory and other “on-line” resources. According to cognitive load (Plass et al., 2010) and information processing theories (Baddeley, 1986; Tombu et al., 2011), resources within the CNS available for engaging in motivated activities/behaviors are finite and are of fixed capacity. Thus, they reflect a “bottleneck” for CNS operations more generally (Tombu et al., 2011). When this capacity is exceeded, either because of task complexity or demands from competing tasks, performance for these operations is impaired. There is substantial support for this general notion, as increased processing load is associated with reduced performance on a range of learning, motor, and other activities (e.g., Kemper et al., 2005; Plass et al., 2010) within healthy adults. Emerging data suggests that abnormal processing capacity is also important for understanding neurological and psychiatric conditions, such as Alzheimer's disease (Huntley and Howard, 2010), various dementias (Calderon et al., 2001), stroke (Puh et al., 2007), substance use (James et al., 2013) and schizophrenia (Granholm et al., 2007). Traditionally, processing capacity is measured using dual task methodologies that impose a substantial processing burden by requiring individuals to perform two effortful tasks simultaneously. Recent technological advancements in biometric analysis have complimented this effort and provided highly sensitive measures of information processing, for example, through the use of pupillometry (Granholm et al., 2007; Laeng et al., 2007) and functional Magnetic Resonance Imaging (Jansma et al., 2007). At issue is the translation of these methods and measures to patient care and assessment, as they tend to be time-consuming and complicated to administer, as well as expensive. This article evaluates the potential use of a different biometric measure of information processing, involving the use of automated

---

[*] Correspondence to: Louisiana State University Department of Psychology 236 Audubon Hall Baton Rouge, LA 70803.
Tel.: +1 225 578 7017; fax: +1 225 578 4125.
E-mail address: acohen@lsu.edu (A.S. Cohen).

http://dx.doi.org/10.1016/j.psychres.2014.12.054
0165-1781/

computerized acoustic analysis of natural vocal expression. This reflects a technology that is objective, inexpensive, automated, unobtrusive to procure and analyze, and reliant on data that is almost ubiquitously available (i.e., human speech).

By way of introduction for readers with limited experience in acoustic analysis, the most commonly analyzed vocal indices involve two distinct signals (Alpert et al., 1986; Cohen et al., 2010, 2009): the fundamental frequency (i.e., F0) – the lowest frequency originating from the vocal folds that defines the subjectively-defined vocal "pitch", and intensity (i.e., volume). Acoustic analysis often involves "speech production" – typically defined as the presence or absence of F0, and can be quantified in a number of ways – such as average pause length, total percentage of time in silence, number of utterances and average utterance length. Beyond vocal production, looking at variability of F0 and Intensity signals can be important. Variability is often quantified in terms of variability within vocal utterances (defined as blocks of speech with F0 signal), but can also be examined on very small time scales (e.g., change on the order of assessment "frames"; 10–50 ms), the latter of which is often referred to as signal "perturbation" or jitter/shimmer.

Acoustic properties of speech reflect key variables for understanding human behavior (Decety and Lamm, 2006). Vocal expression is highly variable across individuals and across contexts, and is influenced by a number of individual difference, for example, sex (Scherer, 2003), affective (Sobin and Alpert, 1999; Tolkmitt and Scherer, 1986), arousal (Cohen et al., 2010; Johnstone et al., 2007), social (Nadig et al., 2010) and speaking task (Scherer, 2003) factors, to name a few. Emerging evidence suggests that vocal expression is also linked to cognitive variables as well. Of note, a number of correlational studies have documented links between acoustic properties of natural speech and state measures of cognitive stress, for example, in how vocal expression in air pilots changes as a function of demanding flight conditions. The findings of this literature are not entirely consistent, though many studies report that vocal characteristics *increase* as a function of increased task demands (e.g., Huttunen et al., 2011). Interpretation of these studies is confounded in that cognitive and emotional/arousal demands are conflated, thus making it is difficult to determine specific factors that may be modulating vocal expression. Experimental studies, which provide the ability to control for extraneous factors such as arousal, of processing load and vocal expression (i.e., reading text) have been conducted, though the vast majority of these employ "reading" as opposed to natural speaking tasks (e.g., Tuček et al., 2012; Yin et al., 2007). Not surprisingly, the results of these studies are also variable, though the relevance of these studies to the question at hand is unclear given that the functional and neurobiological processes involved in verbalizing text is very different than those involved in freely generated natural speech (Smith, 2004).

To our knowledge, only a few published studies have examined how acoustic vocal features change as a function of experimentally manipulated processing load in healthy adults. Barch and Berenbaum (1994) analyzed the natural speech of 50 young adults engaged in two counterbalanced standardized interviews, one with a simultaneous cognitive task and one without, and found that word counts decreased as a function of processing demands. Note that these results were replicated in a psychiatric patient sample (Barch and Berenbaum, 1996). Cohen et al. (2012) conducted computerized acoustic analysis of natural speech in healthy individuals with psychometric schizotypy (e.g., a personality organization putatively underlying schizophrenia) and controls while engaged in various dual tasks, and found that broad indices of vocal production and variability decreased while participants were under heavy cognitive load. At issue with these studies is the use of word count and global measures of speech; measures that lack sophistication and thus, yield a limited understanding of, for example, in how words are produced (e.g., longer pauses, longer utterances, fewer utterances) or conveyed (e.g., how F0 or intensity changes at the utterance or perturbation levels). This is a critical limitation in that development of indices of impaired information processing based on vocal analysis (particularly, those that can be extracted through automated analysis) hinges on identifying facets of vocal expression most affected by information processing demands.

The purpose of this project was to evaluate whether aspects of natural vocal production and variability modulate as a function of experimentally-induced cognitive load in a large sample of healthy adults using much more sophisticated measures of voice than prior studies. For the present study, healthy adults were asked to provide speech on emotionally-neutral topics while engaging in separate baseline, medium and high load cognitive tasks. We hypothesized that increased processing load would be associated with *decreased* vocal expression – defined in terms of reduced speech production (i.e., more silence, longer pauses, fewer utterances and shorter utterances) and reduced speech variability (i.e., less F0 variability and perturbation and less Intensity variability and perturbation). In order to evaluate whether the effects of processing load were specific to vocal expression, we also measured syntactic and semantic complexity – the level of sophistication regarding the sentence structure. After all, changes in vocal production and variability could simply reflect participants producing less complicated speech under conditions of cognitive load. The resulting vocal expression was compared across conditions and analyzed using a variety of automated computerized programs assessing acoustic, syntactic and semantic-related variables.

2. Methods

2.1. Participants

Participants were 134 males and 147 females recruited from one of two large public universities. We selected two different regions of the United States for data collection in order to improve the generalizability of our results, as speaking characteristics can vary as a function of cultural and geographical location. Data collection sites were based in Louisiana in the southeastern United States ($n=149$) and in New Jersey ($n=132$) in the northeastern United States. Participants' average age was 19.92 years (standard deviation [S.D.]=3.73) with a range of 18–64. The sample was predominantly Caucasian (72.5%), with some African-American (13.5%), Hispanic (5.2%) and Asian-American (4.8%) representation. This study was approved by the appropriate Institutional Review Boards and all participants provided written informed consent prior to beginning the study.

2.2. Cognitive-load narrative task

Participants were seated in front of a computer monitor and performed three separate 60-s narrative tasks – a "baseline", medium and high cognitive-load condition. Speech samples involved "free" speech on one of three topics: hobbies, goals and living conditions. All instructions were printed on the computer screen and read to participants, for example "What kinds of hobbies do you have? You can discuss any hobby that you can think of, such as sports, walking, watching TV or anything else you can think of". These topics were selected because they were valence-neutral and open-ended. Participants were encouraged to speak for the duration of the 90-s task. The order of the topics was randomized across participants. Participants were encouraged to speak for the duration of the task and to provide as much speech as possible per both written and oral instructions. The medium and high load conditions employed a "dual attention" design where participants provided free-speech while simultaneously engaging in a cognitive task. During the baseline condition, participants provided their narratives without any competing task. The medium and high-load tasks were modeled after zero and one-back tasks commonly used in studies of cognition; with the one-back requiring increased attentional and working memory abilities compared to the zero-back task. During the medium-load task, participants were asked to provide their narratives while engaging in a continuous performance test. This task involved providing responses to target visual stimuli (i.e., pressing the space bar) appearing on the computer screen while inhibiting responses to distracter stimuli. Six different stimuli (e.g., @, #, $, %, & and *) were presented randomly at 500, 1000, 1500 and 2000 ms Inter-Stimulus Intervals (ISI). The target to distracter ratio was set at 50% (16–32). During the high-load task, participants provided their narratives while performing a very similar continuous performance task, with the modification that the target would change