HHS Public Access

Author manuscript

Psychiatry Res. Author manuscript; available in PMC 2016 March 30.

Published in final edited form as:

Psychiatry Res. 2015 March 30; 226(1): 235–241. doi:10.1016/j.psychres.2014.12.054.

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], and **Sean C. Morrison**[a]

[a]Louisiana State University, Department of Psychology

[b]Rowan University, Department of Psychology

[c]Louisiana State University, Department of Communication Sciences and Disorders

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, 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.

Keywords

Emotion; prosody; affect; expression; speech; load; cognition; working memory; mild cognitive impairment; schizophrenia; depression

Send correspondence to, Alex S. Cohen, Ph.D., Louisiana State University, Department of Psychology, 236 Audubon Hall, Baton Rouge, LA 70803, Phone: (225) 578-7017, Fax: (225) 578-4125, acohen@lsu.edu.

Conflicts of Interest: There are no conflicts of interest to report.

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 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; Cohen et al., 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 milliseconds), 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 emotionallyneutral 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.

2.0 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 [SD] = 3.73) with a range of 18 to 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-second 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-second 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 millisecond Inter-Stimulus Intervals (ISI). The target to distracter ratio was set at 50% (16 of 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 throughout the task. This task involved responding to stimuli when consecutively appearing visual symbols on a computer screen were identical. In addition to attentional vigilance, participants must update the target each trial "on-line" using working memory resources. A total of six different visual symbols (identical to those in the medium-load task) were used, approximately one-third of which reflected targets (10 of 32). Stimuli were presented at 2000 millisecond ISIs. Otherwise, the high-load task was identical to the medium-load task. Visual symbols, as opposed to verbal or alpha-numeric characters, were employed. Based on Baddeley's model of attention (Baddeley, 1992), visual and verbal short-term memory abilities have some separate resources but are commonly limited by a central "executor". For each task, participants underwent training practice blocks without the narrative component to become familiar with the task. Block order was randomized across participants, and blocks were separated by a 30-second interval during which participants were asked to "relax and breathe deeply". Because participants lacked feedback regarding their performance during the high and medium load tasks in terms of accuracy on the cognitive tasks or verbal output, it was conceivably possible for participants to focus their efforts on one, but not both aspects of the dual task. To guard against this possible confound, participants producing speech less than 10% for the duration of the task (n = 29; 10.32% of the sample) or poor performance (< 25% accuracy) on either the cognitive task (n = 27; 9.51% of the sample) were excluded from the analyses.

2.3 Computerized Assessment of Prosody

The Computerized assessment of Affect from Natural Speech protocol (CANS), developed by our lab to assess vocal expression from natural speech, was employed here. Speech was digitally recorded at 16 bits per second at a sampling frequency of 44,100 Hertz using headset microphones. The CANS procedure was performed in two distinct step; each of which were automated and "batched". The first step involved analyzing the recordings using Praat (Boersma and Weenink, 2006), a shareware program that has been used extensively in speech pathology and linguistic studies. The Praat system organizes sound files into "frames" for analysis, which for the present study was set at a rate of 100 per second. During each of these frames, frequency and volume was quantified. The second step of analysis involved Macro programs developed by our lab to extract variables of interest. For this study, we focused on eight variables, based on our prior work (Cohen et al., 2008; Cohen et al., 2010; Cohen et al., 2009; Cohen et al., 2012) and the larger extant speech prosody literature (Boersma and Weenink, 2006; Scherer, 2003). Four of these variables pertained to vocal expression and four to vocal variability (see Table 1 for information on the variables employed here, and how they were computed). Based on prior research examining optimization filters for measuring fundamental frequency in automated research (Vogel et al., 2009), we applied a low (i.e., 75 Hertz) and high (i.e., 300 Hertz)-pass filter. Because the use of fixed versus non-fixed microphones was not consistent across recruitment sites, there is potential variability in mean intensity within subjects across conditions. Thus, this measure was not examined in this study. There is no reason to think other measures were affected by this; as F0 signal quality and intensity variability are independent of mean intensity.

2.4 Syntactic and Semantic Complexity

Research assistants trained in syntactic and semantic coding conventions coded the speech samples from the 149 Louisiana participants. Because of the time-intensive nature of this analysis, we limited syntactic and semantic coding to the baseline and high-load experimental conditions; those theorized to show the most dramatic differences under conditions of cognitive load. Inter-rater reliability between coders based on analysis of a random selection of 10% of the transcripts was excellent (ICC = 0.94). Syntactic complexity was measured by coding instances of conjoined and embedded structures. Conjoined structures included verb phrases united by a coordinating conjunction. For example from this study's transcripts, "we had a random roommate but she has become good friends with all of us" is a single conjoined structure using the coordinating conjunction "but." Embedded structures included verb phrases joined by a subordinating conjunction, those that were a part of a relative clause, or an infinitival phrase. The instances were then divided by the number of complete and intelligible utterances to create a Complex Structure Ratio. Similar measures have been used previously to study deteriorating syntactic complexity in aging adults particularly in dual task activities (Kemper et al., 2011)

To measure semantic complexity, we calculated Type Token Ratio (TTR), the number of different words compared to total number of words in a sample (Kuder, 2008) analyzed using Systematic Analysis of Language Transcripts (Miller and Iglesias, 2010) computer software. Generally, the higher the TTR, the more complex the language sample (e.g. less repetition of words or more diverse language forms). Furthermore, TTR has been used to measure semantic complexity in children (Peets, 2009); normal adults (Silverman, 1977) and adults with mental disorder and neurological injury (Manschreck et al., 1985).

2.5 Analyses

We conducted the data analysis in four phases. First, we examined whether vocal expression was related to sex, age and ethnicity to identify variables that may influence our main analyses of interest. Additionally, we examined whether there were differences in vocal variables or participant characteristics across study site. Second, we examined cognitive test performance in the medium and high-load conditions. Third, we employed repeated measures ANOVAs and ANCOVAs to examine whether vocal expression changed across the baseline, medium and high load conditions. Finally, we employed repeated measures ANOVAs and ANCOVAs to determine whether syntactic and semantic complexity changed across the baseline and high load conditions. If we found that syntactic and semantic complexity reduced as a function of cognitive demands (as we expected), we planned to repeat the third set of analyses controlling for these variables. As noted below, these analyses were unnecessary. All significance tests reported here are two-tailed and all variables are normally distributed unless otherwise noted.

3.0 RESULTS

3.1 The relationship between prosodic and descriptive/procedural variables

There were no significant differences in any of the variables as a function of speech topic order. Age was significantly related to frequency perturbation for the baseline, medium and high load conditions (*r's*[223] > 0.14, *p*'s < 0.03), such that older participants showed greater perturbation. There were no significant differences of note between ethnic groups (i.e., Caucasian versus non-Caucasian) in any of the dependent variables examined here. With respect to sex, women tended to show briefer pauses, and showed greater F0 variability (p's < 0.05) than men. Women also showed less Intensity variability for the high and baseline conditions (p's < 0.05). Finally, there were a number of differences across data collection sites; such that participants from New Jersey tended to show less Silence, longer Pauses, and greater levels of Intensity Perturbation. The New Jersey sample also had greater male representation than the Louisiana sample (64% & 32% respectively; $x^2 = 23.04$, $p < 0.001$), which may account for these differences. To address the potential confounding effects of these variables, all relevant analyses were recomputed controlling for age**,** sex and study site. No notable changes were observed.

3.2 Zero-Order Correlations

Zero-order correlations are included in Table 2. None of the variables were considered redundant, and few were inter-correlated at even a medium effect size level (i.e., *r* value exceeding 0.30). All eight variables were retained for subsequent analysis.

3.3 Cognitive performance

Means and standard deviations for performance on the medium and high-load conditions were computed. Performance for the medium-load condition was largely at ceiling (M ± SD = 98% ± 3 % correct) with little variability in scores across participants. In contrast, performance for the high-load condition was significantly lower (77% ± 8% correct; $t[22] = 29.72$, $p < 0.001$).

3.4 Change across conditions

Data for these analyses are presented in Figures 1 and 2. Results suggest that, as cognitive demands increased, total silence time increased, mean pauses became longer, and the total number of utterances decreased. Interestingly, utterances did not change significantly in length of in variability of length. With respect to vocal variability measures, cognitive demands were associated with decreases in perturbation for both frequency and intensity, and were associated with decreased F0 variability.

3.4 Measures of Syntactic and Semantic Complexity

Data for these analyses are omitted from table or figure format for space concerns. For type-token ratio, the repeated measures ANOVA main effect was not significant (Wilk's Lambda $F[1, 147] = 2.04$, $p = 0.16$, $\eta^2 = 0.01$). For structural complexity, the repeated measures ANCOVA main effect was significant (Wilk's Lambda $F[1, 147] = 58.51$, $p < 0.001$, $\eta^2 = 0.30$). Evaluation of the descriptive statistics for the baseline (0.79 ± 0.50) and high (0.96 ±

0.54) conditions suggested that syntactic complexity increased as a function of increasing processing load. Thus, in contrast to our expectations, increased processing demands were associated with increased syntactic complexity.

4.0 DISCUSSION

When under conditions of relatively pronounced cognitive load, healthy adults in this study produced less speech overall that was characterized by increased silence, longer means, fewer utterances, less perturbation of F0 and Intensity signals and less F0 variability more generally. In total, there were significant reductions in six of eight variables examined here. These results did not appear to be due to differences in age, ethnicity or recruitment site. Moreover, these findings were not simply an artifact of decreased syntactic or semantic complexity, as these features either remained similar or increased as a function of increasing cognitive loads. To our knowledge, these data provide the most compelling evidence to date that vocal aspects of natural speech are related to information processing.

It may seem intuitive to the reader that measures of vocal productivity were reduced under states of relatively high cognitive load. After all, the cognitive substrata underlying speech production are many, including but not limited to verbal fluency, long-term autobiographical recall, attention and language abilities (Valle-Lisboa et al., 2014). It is less intuitive that signal variability declined as a function of restricted cognitive resources. It is well established that vocal expressivity, *vis a vis* prosody, is heavily dependent on right parietal cortical structures {Borod, 1993 #4901;Ross, 2008 #4902}, though its relationship to other structures, notably those involved in attention and working memory, is less known. The present findings suggest that prosodic expression is also dependent on working memory and attention systems, and that evaluation of signal variability can be offer important insight into information processing resources.

Moving forward, we believe there are two ways that automated computerized analysis of natural vocal expression can be applied to neurology and psychiatry settings. A first application is as a screening tool or complimentary diagnostic tool for assessing the presence of overt cognitive disorders. While it is the case that vocal expression within individuals is highly variable over time and across contexts, it does stand to reason that vocal analysis of natural speech could be used as a case identification or cross-sectional diagnostic tool to complement existing neuropsychological tests. Applying acoustic analysis for this purpose will require considerable work in developing norms of vocal characteristics, and defining “signature” abnormalities that characterize distinct disorders. A second, and perhaps more promising application for acoustic analysis involves longitudinal tracking of individuals that are either experiencing or at risk for experiencing cognitive difficulties, for example, in older adults experiencing mild cognitive impairments. The logic behind this endeavor is that information processing capacity is a potentially important and functionally relevant index of many illness states, and that vocal analysis, insofar as it can be developed into a reliable and valid proxy of information processing capacity, is easy to conduct and repeatable in a way other assessments aren’t. Assessment of natural speech offers many practical advantages over standard neuropsychological tests, for example, employing data capture over telephone, video and smartphone technologies. Moreover, given that vocal analysis is highly sensitive (with measurement potentially occurring at the millisecond level), it has the potential to detect subtle changes in information processing capacity in a way not possible with standard neuropsychological measures. For example, patients could be asked to provide speech samples to a phone call center that automatically analyzes speech, compares it to samples on file, and provides feedback regarding clinical relevance. Note that standard neuropsychological measures (e.g., verbal fluency) tapping language functions suffer from profound practice effects even after a few administrations (Lezak, 2012). While admittedly in its infancy in terms of psychometric evaluation, vocal analysis offers considerable promise over existing measures as a proxy for understanding information processing.

As yet, it remains to be seen whether vocal analysis can provide meaningful data regarding the processing load of an individual at a particular moment (i.e., that long pauses, fewer utterances and lower signal perturbation reflect state of high cognitive-load); an endeavor complicated by the fact that speech is affected by a range of contextual factors. In this regard, it seems critical to develop procedures for eliciting natural speech that are controlled in some manner, so as to minimize the influence of contextual variables and to help ensure that speech is occurring under a state of high processing load. It would seem that there are many potential ways of standardizing and controlling the conditions while maintaining the ecological validity and repeatability of the assessment (e.g., probes assessing “what did you do yesterday”). Similarly, induction of high processing load can be practically accomplished many different ways, such as through the use of dual task designs (e.g., finger-tapping while talking), asking challenging questions (e.g., “what did you have for lunch last Tuesday”), or employing standardized verbal-based clinical assessments (e.g., Serial sevens). Moreover, it is as yet unclear whether induction of high load states is even necessary for individuals that show high levels of cognitive impairment, as producing fluent speech alone may be sufficient to exhaust their processing resources. Finally, it remains to be seen which vocal variables are most sensitive to changes in information processing abilities; particularly as expressed in neurological and psychiatric disorders. The present study evaluated eight distinct variables, though it would be important to more meaningfully understand how these specific variables may be related to specific cognitive processes. Practically speaking, it would be helpful to reduce the number of variables, or to identify a single proxy variable. Exploring ways of procuring speech that can be efficiently applied to a wide array of settings, offer a flexible platform that can be applied to a wide array of populations with varying functional levels while maintaining ecological validity is a challenge for the next step in this line of research.

Some limitations warrant mention. First, the present sample was not particularly diverse in terms of age or ethnicity. Although there is no obvious reason for why the link between cognitive resources and vocal expression would differ across cultural groups, this remains to be empirically evaluated. Relatedly, while the present sample reflected participants from two geographically distinct regions, other potentially important geographical groups were not examined in this study. Second, there was limited information about individual difference variables, and it is unclear whether the relationships between vocal expression and cognitive resources systematically varies across individuals. One could imagine, for example, that certain key cognitive abilities (e.g., working memory, attention) may mitigate the effects of cognitive load. Executive functions, particularly involving the ability to strategize and update behavior, may also play a role. Third, we did not evaluate the effects of exogenous (e.g., caffeine, prescription or illicit medications) or endogenous (e.g., fatigue) confounding variables that may have attenuated the effects of this study. Finally, this study was an analog study, meant to demonstrate "proof of concept", but did not focus on patients with neurodegenerative or psychiatric disorders. This is a matter for future research.

The present study reflects a promising first step in the evaluation of acoustic analysis for understanding neurological and psychiatric disorders. The present data provides experimental evidence that vocal expression, in particular, pause behavior and subtle perturbations in frequency and intensity, is linked to information processing load in healthy adults. These findings dovetail those reported in studies linking information processing to a range of other behaviors in both clinical and nonclinical populations (Barch and Berenbaum, 1996; Calderon et al., 2001; Granholm et al., 2007; Huntley and Howard, 2010; James et al., 2013; Puh et al., 2007). It would be important to replicate the present findings in clinical populations and to determine the optimal methods for eliciting speech under conditions of high processing load. We believe the clinical and research potential of a biometric tool that is inexpensive, automated, uninvasive and ecologically valid is quite large.

Acknowledgments

The authors wish to acknowledge the participants for their participation, lab members for their help processing and collecting data.

Funding: Funding for this study was provided by a Louisiana Board of Regents and National Institute of Mental Health (R03 MH092622) grant to the primary author. The funding agencies had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

References

Alpert M, Merewether F, Homel P, Marz J. Voxcom: A system for analyzing natural speech in real time. Behavior Research Methods, Instruments & Computers. 1986; 18:267.

Baddeley A. Working memory. Science. 1992; 255:556–559. [PubMed: 1736359]

Baddeley, AD. Working Memory. Oxford University Press; New York: 1986.

Barch D, Berenbaum H. The relationship between information processing and language production. Journal of Abnormal Psychology. 1994; 103:241–251. [PubMed: 8040493]

Barch DM, Berenbaum H. Language production and thought disorder in schizophrenia. Journal of Abnormal Psychology. 1996; 105:81–88. [PubMed: 8666714]

Boersma, P.; Weenink, D. Praat: doing phonetics by computer (Version 4.4.05). 2006.

Calderon J, Perry RJ, Erzinclioglu SW, Berrios GE, Dening TR, Hodges JR. Perception, attention, and working memory are disproportionately impaired in dementia with Lewy bodies compared with Alzheimer's disease. Journal of Neurology, Neurosurgery and Psychiatry. 2001; 70:157–164.

Cohen AS, Alpert M, Nienow TM, Dinzeo TJ, Docherty NM. Computerized measurement of negative symptoms in schizophrenia. Journal of Psychiatric Research. 2008; 42:827–836. [PubMed: 17920078]

Cohen AS, Hong SL, Guevara A. Understanding emotional expression using prosodic analysis of natural speech: refining the methodology. Journal of Behavior Therapy and Experimental Psychiatry. 2010; 41:150–157. [PubMed: 20022000]

Cohen AS, Minor KS, Najolia GM, Lee Hong S. A laboratory-based procedure for measuring emotional expression from natural speech. Behavior Research Methods. 2009; 41:204–212. [PubMed: 19182141]

Cohen AS, Morrison SC, Brown LA, Minor KS. Towards a cognitive resource limitations model of diminished expression in schizotypy. Journal of Abnormal Psychology. 2012; 121:109–118. [PubMed: 21553936]

Decety J, Lamm C. Human empathy through the lens of social neuroscience. Scientific World Journal. 2006; 6:1146–1163. [PubMed: 16998603]

Granholm E, Verney SP, Perivoliotis D, Miura T. Effortful cognitive resource allocation and negative symptom severity in chronic schizophrenia. Schizophrenia Bulletin. 2007; 33:831–842. [PubMed: 16956985]

Green MF, Kern RS, Braff DL, Mintz J. Neurocognitive deficits and functional outcome in schizophrenia: Are we measuring the “right stuff”? Schizophrenia Bulletin. 2000; 26:119–136. [PubMed: 10755673]

Huntley JD, Howard RJ. Working memory in early Alzheimer’s disease: a neuropsychological review. Internation Journal of Geriatric Psychiatry. 2010; 25:121–132.

Huttunen K, Keranen H, Vayrynen E, Paakkonen R, Leino T. Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights. Applied Ergonomics. 2011; 42:348–357. [PubMed: 20832770]

James A, James C, Thwaites T. The brain effects of cannabis in healthy adolescents and in adolescents with schizophrenia: a systematic review. Psychiatry Research. 2013; 214:181–189. [PubMed: 24139960]

Jansma JM, Ramsey NF, de Zwart JA, van Gelderen P, Duyn JH. fMRI study of effort and information processing in a working memory task. Human Brain Mapping. 2007; 28:431–440. [PubMed: 17133397]

Johnstone T, van Reekum CM, BÃ¤nziger T, Hird K, Kirsner K, Scherer KR. The effects of difficulty and gain versus loss on vocal physiology and acoustics. Psychophysiology. 2007; 44:827–837. [PubMed: 17608801]

Kemper S, Herman RE, Nartowicz J. Different effects of dual task demands on the speech of young and older adults. Aging, Neuropsychology, and Cognition. 2005; 12:340–358.

Kemper S, Schmalzried R, Herman R, Mohankumar D. The effects of varying task priorities on language production by young and older adults. Experimental Aging Research. 2011; 37:198–219. [PubMed: 21424957]

Kuder, SJ. Teaching students with language and communication disabilities. 3. Pearson/A & B; Boston: 2008. Language and literacy in the school years.

Laeng B, Waterloo K, Johnsen SH, Bakke SJ, Lag T, Simonsen SS, Hogsaet J. The eyes remember it: oculography and pupillometry during recollection in three amnesic patients. Journal of Cognitive Neuroscience. 2007; 19:1888–1904. [PubMed: 17958490]

Lezak, MD. Neuropsychological assessment. 5. Oxford University Press; Oxford; New York: 2012.

Manschreck TC, Maher BA, Hoover TM, Ames D. Repetition in schizophrenic speech. Language and Speech. 1985; 28(Pt 3):255–268. [PubMed: 3836315]

Miller, J.; Iglesias, A. Systematic analysis of language transcripts (SALT) Instructional Version 2010. SALT Software, LLC; 2010.

Nadig A, Lee I, Singh L, Bosshart K, Ozonoff S. How does the topic of conversation affect verbal exchange and eye gaze? A comparison between typical development and high-functioning autism. Neuropsychologia. 2010; 48:2730–2739. [PubMed: 20493890]

Peets KF. Profiles of dysfluency and errors in classroom discourse among children with language impairment. Journal of Communication Disorders. 2009; 42:136–154. [PubMed: 19185881]

Plass, JL.; Moreno, R.; Branken, R. Cognitive load theory. Cambridge University Press; New York, NY US: 2010.

Puh U, Vovk A, Sevsek F, Suput D. Increased cognitive load during simple and complex motor tasks in acute stage after stroke. International Journal of Psychophysiology. 2007; 63:173–180. [PubMed: 16814889]

Scherer KR. Vocal communication of emotion: A review of research paradigms. Speech Communication. 2003; 40:227–256.

Silverman G. The effect of topic and repetition on type-token ratios of spoken monologues. Language and Speech. 1977; 20:232–239. [PubMed: 613184]

Smith C. Topic transitions and durational prosody in reading aloud: Production and modeling. Speech Communication. 2004; 42:247–270.

Sobin C, Alpert M. Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. Journal of Psycholinguistic Research. 1999; 28:167–365.

Tolkmitt FJ, Scherer KR. Effect of experimentally induced stress on vocal parameters. Journal of Experimental Psychology: Human Perception. 1986; 12:302–313.

Tombu MN, Asplund CL, Dux PE, Godwin D, Martin JW, Marois R. A Unified attentional bottleneck in the human brain. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:13426–13431. [PubMed: 21825137]

Tucek DC, Mount WM, Abbass HA. Neural and Speech Indicators of Cognitive Load for Sudoku Game Interfaces. Lecture Notes in Computer Science Volume. 2012; 7663:210–217.

Valle-Lisboa JC, Pomi A, Cabana A, Elvevag B, Mizraji E. A modular approach to language production: models and facts. Cortex. 2014; 55:61–76. [PubMed: 23517653]

Vogel AP, Maruff P, Snyder PJ, Mundt JC. Standardization of pitch-range settings in voice acoustic analysis. Behavior Research Methods. 2009; 41:318–324. [PubMed: 19363172]

Yin B, Ruiz N, Chen F, Kwawaja MA. Automatic cognitive load detection from speech features. Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces. 2007

Author Manuscript

Author Manuscript

Author Manuscript

- Our study employed an experimental design with a dual attention task and narrative speaking task in a large adult sample to evaluate whether acoustic analysis of natural speech can serve as a biometric indicator of cognitive functioning.
- Our results provide compelling evidence of a link between information processing resources and aspects of vocal expression, including vocal production and vocal expression.
- Computerized vocal assessment holds promise for providing important about information processing in individuals, and can reflect an automated, inexpensive and uninvasive biometric measure directly translatable to the clinic.

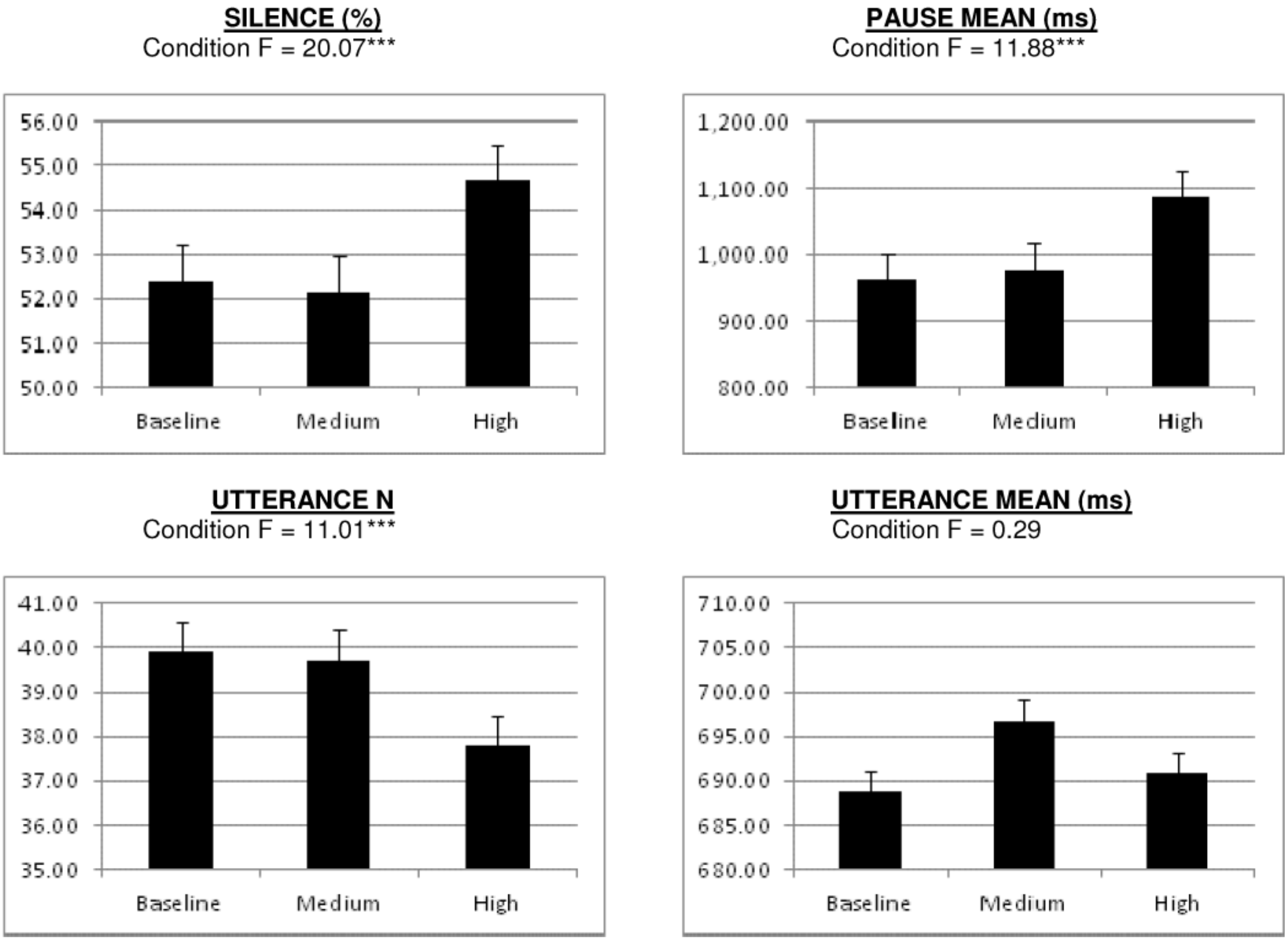

Figure 1.

Comparison of speech production measures for baseline-, medium- and high-load conditions with the results of each mixed-design analysis of variance. *** = $p < 0.001$.

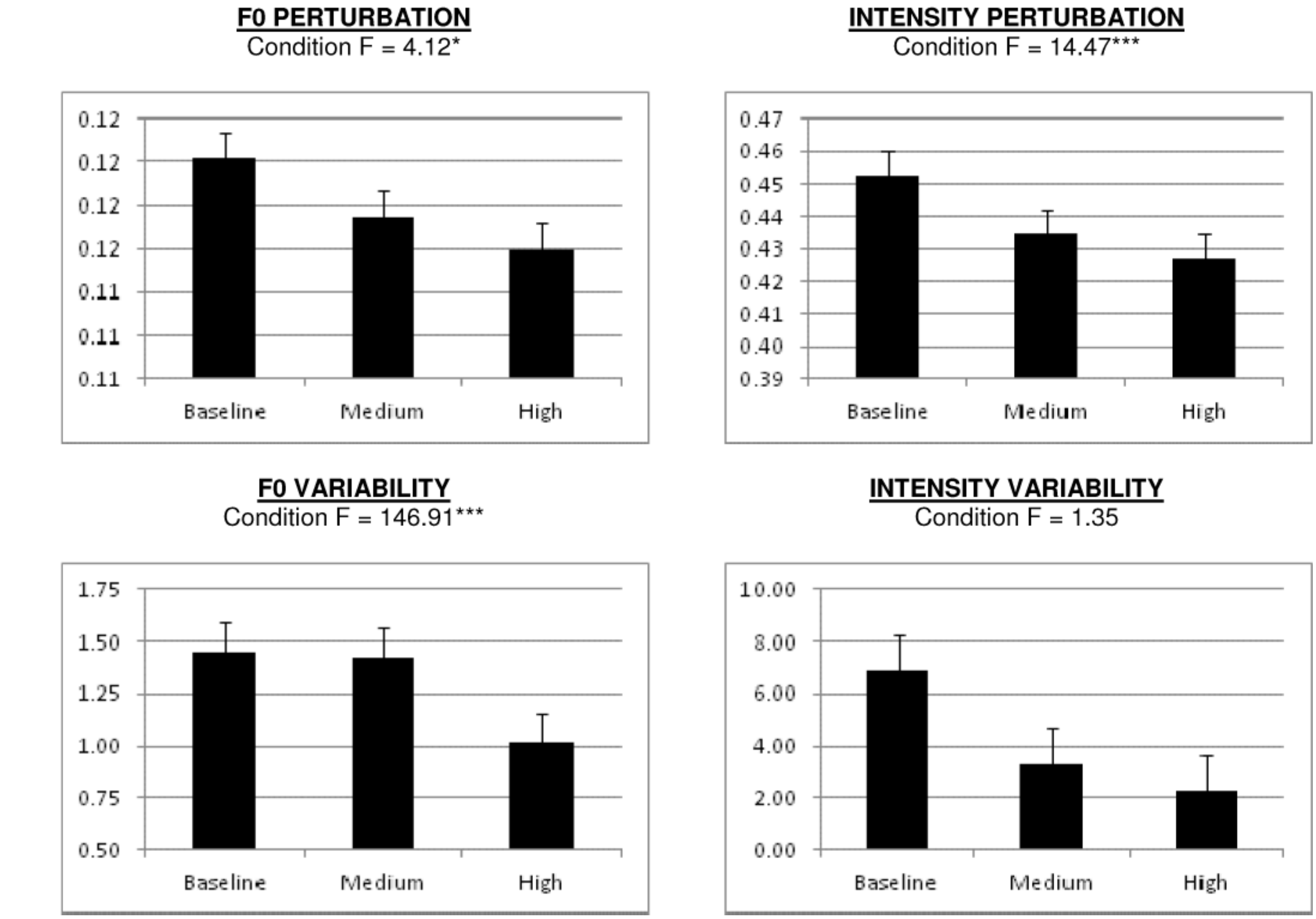

Figure 2.

Comparison of speech variability measures for baseline-, medium- and high-load conditions with the results of each mixed-design analysis of variance. * = $p < 0.01$,*** = $p < 0.001$.

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Computerized acoustic analysis variables examined in this study.

Variable	Description	Increasing scores reflect…	Changes to Raw data	Units of Measure
		Speech Production		
Silence Percent	Percentage of time not speaking	Less percentage of time speaking	None	Percentage
Pause Mean	Average pause length in milliseconds (ms), excluding the first and last pauses.	Longer average pauses	None	Milliseconds
Utterance Number	Total number of utterances (> 150 ms)	More utterances	Computed, per 60,000 milliseconds	Number of Utterances per minute of speech
Utterance Mean	Average utterance length in milliseconds (ms)	Longer average utterances	None	Milliseconds
		Speech Variability		
Frequency Perturbation	Absolute value of average change in consecutively voiced frames within utterance, averaged across utterances.	Increasing levels of perturbation in F0 signal	None	Semitones
Local Intonation	SD of F0 values computed within each utterance and averaged across all utterances	Higher F0 variability within utterances.	Converted to semitones,	Semitones, per second of average speech
Intensity Perturbation	Absolute value of average change in consecutively voiced frames within utterance, averaged across utterances.	Increasing levels of perturbation in intensity signal	None	Decibels
Local Emphasis	SD of Intensity values computed within each utterance and averaged across all utterances	Higher intensity variability within utterances.	Divided by average utterance length in seconds	Decibels, per second of average speech

Table 2

Zero-order correlation matrix between vocal variables from the baseline condition.

	1.	2.	3.	4.	5.	6.	7.
1. Silence	1.00	–	–	–	–	–	–
2. Pause M	0.03	1.00	–	–	–	–	–
3. Utterance N	−0.14*	−0.53*	1.00	–	–	–	–
4. Utterance M	−0.63*	−0.08	−0.49*	1.00	–	–	–
5. F0 SD	−0.16*	−0.02	−0.09	0.21*	1.00	–	–
6. F0 Perturbation	0.10	0.18*	0.02	−0.22*	−0.01	1.00	–
7. Intensity SD	−0.06	−0.05	−0.02	0.07	−0.02	−0.10	1.00
8. Intensity Perturbation	−0.10	0.09	0.21*	−0.15*	0.11	0.42*	−0.14*

* = $p < .05$