Behav Res
DOI 10.3758/s13428-015-0584-1

BRIEF COMMUNICATION

A psychometric investigation of "macroscopic" speech measures for clinical and psychological science

Alex S. Cohen[1] · **Tyler L. Renshaw**[1] · **Kyle R. Mitchell**[1] · **Yunjung Kim**[2]

Abstract The analysis of vocal expression is a critical endeavor for psychological and clinical sciences and is an increasingly popular application for computer–human interfaces. Despite this, and despite advances in the efficiency, affordability, and sophistication of vocal analytic technologies, there is considerable variability across studies regarding what aspects of vocal expression are studied. Vocal signals can be quantified in a myriad of ways, and their underlying structure, at least with respect to "macroscopic" measures from extended speech, is presently unclear. To address this issue, we evaluated the psychometric properties—notably, the structural and construct validity—of a systematically defined set of global vocal features. Our analytic strategy focused on (a)identifying redundant variables among this set, (b)employing principal components analysis (PCA) to identify nonoverlapping domains of vocal expression, (c)examining the degrees to which the vocal variables are modulated as a function of changes in speech task, and (d)evaluating the relationship between the vocal variables and cognitive (i.e., verbal fluency) and clinical (i.e., depression, anxiety, and hostility) variables. Spontaneous speech samples from 11 independent studies of young adults (>60 s in length), employing one of three different speaking tasks, were examined (N= 1,350). Confounding variables (i.e., sex, ethnicity) were statistically controlled for. The PCA identified six distinct domains of vocal expression. Collectively, vocal expression (defined in terms of these domains) was modulated as a function of speech task and was related to the cognitive and clinical variables. These findings provide empirically grounded implications for the study of vocal expression in psychological and clinical sciences.

Keywords Voice · Expression · Prosody · Acoustic · Frequency · Intensity

Vocal communication is a foundational tool for exchanging both explicit and implicit information between people and is an important indicator of trait-related individual differences (Roy, Bless, & Heisey, 2000) and emotional and cognitive states (Cohen, Dinzeo, Donovan, Brown, & Morrison, 2015; Giddens, Barron, Clark, & Warde, 2010; Sobin & Alpert, 1999). Moreover, vocal communication is affected in a broad array of mental illnesses, such as schizophrenia (Cohen, Alpert, Nienow, Dinzeo, & Docherty, 2008), depression (Cannizzaro, Harel, Reilly, Chappell, & Snyder, 2004), and anxiety (Cohen, Kim, & Najolia, 2013; Laukka et al., 2008). For these reasons, vocal communication is important for both clinical and psychological sciences. Regarding the assessment of vocal communication, objective technologies that employ computerized acoustic analysis of digitized speech samples have existed for decades. The use of these technologies is a boon for many reasons, not the least of which involves near perfect "interrater" reliability (assuming similar analytic procedures) and high test–retest reliability (assuming similar recording conditions and speaking tasks) in healthy (Shriberg et al. 2010) and communication-disordered (Kim, Kent, & Weismer, 2011) populations. Vocal analytic technologies are also important for computer–human interfaces—notably involving self-automation, emotion-assessment applications, and telemedicine software (Cohen & Elvevåg, 2014; Esposito & Esposito,

✉ Alex S. Cohen
acohen@lsu.edu

[1] Department of Psychology, Louisiana State University, 236 Audubon Hall, Baton Rouge, LA 70803, USA

[2] Department of Communication Sciences and Disorders, Louisiana State University, Baton Rouge, LA 70803, USA

Published online: 11 April 2015

2012; Krajewski, Batliner, & Golz, 2009). At present, there is a lack of consensus regarding the optimal measures of vocal expression for psychological and clinical sciences. We sought to redress this issue by evaluating psychometric properties of key measures extracted from a large database of spontaneous speech procured from young adults.

Although a large literature on vocal expression exists, establishing a consensus on which measures to employ is quite difficult. In part, this reflects complexity in how vocal expression is conceptualized more generally. Consider that aspects of vocal expression vary considerably as a function of a wide range of contextual and individual difference variables, including sex (Scherer, 2003), affective (Batliner, Steidl, Hacker, & Nöth, 2008; Sobin & Alpert, 1999; Tolkmitt & Scherer, 1986), arousal (Cohen, Hong, & Guevara, 2010; Johnstone et al., 2007), social (Nadig, Lee, Singh, Bosshart, & Ozonoff, 2010), speaking task (Huttunen, Keranen, Vayrynen, Paakkonen, & Leino, 2011; Scherer, 2003), and cognitive (Cohen, Dinzeo, et al., 2015) factors. Thus, the importance of an isolated speech measure may be relegated to specific contexts or groups of individuals. In effect, there is no isomorphic measure of vocal expressivity.

Even the structure of vocal measures is complicated and likely varies depending on whether "microscopic" or "macroscopic" aspects of speech are examined. Acoustic analysis is a technology that generally focuses on "micro" or brief vocal samples (e.g., utterances), in part, because it provides "high-resolution" information about the physical processes involved in communication (Kent & Kim, 2003; Shriberg et al., 2010). Theoretical-based structural models of vocal expression at this "micro" level of analysis exist and generally focus on the physical functions involved in speaking and on physical anomalies that occur as part of communication disorders (Kent & Kim, 2003). Conceptually driven taxonomies of emotional expression—for example, based on changes during relatively brief speaking epochs—have also been developed (Banse & Scherer, 1996; Scherer, 1986; Sobin & Alpert, 1999). A complementary, but different, approach to understanding speech focuses on "macroscopic" features of vocal communication involving extended speech samples (generally >30 s). This approach yields aggregate statistics characterizing how speech is produced and about absolute values and signal variability across entire speech samples. Although the aforementioned "micro" approaches can provide much more nuanced information about the signal, the latter approach is indispensable for capturing more stable phenomenon, particularly those requiring extended sampling. For example, some evidence suggests that pause behavior may become longer and more erratic as a function of "online" cognitive abilities being unavailable or restricted in some manner (Cohen, Dinzeo, et al., 2015). Hence, increased pause length may be a useful index of cognitive resource depletion due to fatigue (e.g., in airplane pilots; Huttunen et al., 2011) or of cognitive deficits more generally (e.g., in psychiatric or neurological disorders; Cohen & Elvevåg, 2014; Cohen, McGovern, Dinzeo, & Covington, 2014). Given the dynamic nature of pause behavior, analysis of brief vocal samplings is inadequate for approximating an individual's cognitive abilities. From a pragmatic perspective, "macro" analytic approaches are typically automated in a way that "microscopic" approaches aren't, since they focus on vocal features that are less influenced by individual outliers (and, hence, do not typically require manual data inspection). For these reasons, automated "macroscopic" methods are important for a broad range of psychological and clinical applications, and are the focus of this study.

It is worth providing a brief primer on acoustic analysis for readers with limited familiarity with this topic. "Macroscopic" speech indices typically focus on four different physical properties or signals (Alpert et al. 1986; Cohen et al. 2010; Cohen et al. 2009), (a)the fundamental frequency (i.e., F0)—the lowest frequency originating from the vocal folds that defines the subjectively defined vocal "pitch;" (b)the first formant frequency (F1), important for vowel expression that is shaped by vertical tongue articulation; (c)the second formant frequency (F2), also important for vowel expression that is shaped by horizontal and back-and-forth tongue articulation; and (d) intensity (i.e., volume). In terms of characterizing these variables, various measures of speech production—defined as the absence of signal (e.g., average pause length), the presence of signal (e.g., average utterance length), variability of the signal (e.g., standard deviation of pause or utterance length), and number of discrete signal events (e.g., number of pauses)—can be computed. Similarly, a seemingly infinite number of measures regarding the signals can be computed (e.g., mean, standard deviation, range) for the F0, F1, F2, and intensity values. Moreover, variability statistics can be computed across different epochs, such that variability can be examined on small time scales (e.g., signal perturbation; change on the order of assessment "frames"; 10–50 ms), within utterances (e.g., consecutive voiced frames; typically 250–1,500 ms), or across key sections of the speech sample, or in its entirety.

It should be clear that a massive number of acoustic variables can be computed from speech. This is an issue when evaluating findings across literatures in which inconsistency in variables is likely the rule rather than the exception. Consider a recent meta-analysis of published studies evaluating vocal deficits in patients with schizophrenia versus non-psychiatric controls (Cohen, Mitchell, & Elvevåg, 2014). Across 13 studies appropriate for review, a total of ten different variables were reported. These variables appeared an average of 2.5 times across the 13 studies, and each individual study reported an average of only 1.92 variables. Even more importantly, there were dramatic differences in the effect sizes reported across variables, even among those that were conceptually related, suggesting that some, but not all, of the variables are important for understanding this disorder.

Recent efforts employing sophisticated data reduction strategies have been conducted focusing on "macroscopic"-level speech data. Of particular note, Batliner et al. (2011) employed feature vector analysis of 4,000 speech (acoustic and linguistic) features from a large corpus of speech samples from child–robotic interactions to derive a more modest set of variables (e.g., duration of speech, intensity, pitch, formant spectrum, and voice quality). Other similar efforts, employing a range of statistical strategies, have been conducted (e.g., Schuller et al. 2007a, b; Vogt & Andre, 2005), and ongoing exchanges have been established to aid in organizing analytic approaches (e.g., Eyben, Weninger, Groß, & Schuller, 2013; Schuller et al. 2007a, b; Schuller et al. 2013). Although these efforts have provided critical insight for the application of acoustic technologies, information regarding the psychometrics of these variables, notably in terms of the incremental validity of individual variables, internal consistency, and factor structure, is not explicitly clear. Moreover, most prior studies have focused on the predictive power of acoustic variables in classifying emotional states—an important endeavor, in that emotional expression is closely tied to vocal expression. However, vocal expression is a function of a wide range of contextual, cognitive, and clinical variables, as well, and exploring how vocal expression is tied to these variables is a critical compliment to understanding their validity.

In the present study, we sought to evaluate the psychometric properties of global acoustic features of spontaneous speech as a function of speech task (i.e., tapping different cognitive and contextual functions) and clinical (i.e., hostility, depression, anxiety) and neuropsychological (i.e., verbal fluency) variables. We focused on global vocal signals, involving F0, F1, F2, and intensity, and vocal production variables, using a systematically defined and limited set of variables. Our selection process is elaborated on in the Method section. It is noteworthy that the number of variables examined is not necessarily important (Batliner et al., 2006); the performance of a classification system using 32 vocal features was similar to that of another using 1,000 features. Our focus on a limited set of features facilitated a more qualitative evaluation of the potential overlap, independence, and redundancy of variables than could be achieved in studies analyzing large variable sets (e.g., by reporting zero-order correlation matrices). We subjected the nonredundant vocal variables to principal components analysis (PCA) for data reduction purposes. PCA was used because it is a relatively straightforward and interpretable analytic strategy that has been employed in studies of brief vocal utterances (Slavin & Ferrand, 1995; Yamashita et al. 2013). The resulting factors were subjected to validity analysis—examining (a)the degree to which the factor scores changed as a function of different speaking tasks, (b)their relative associations with a clinical neuropsychological test of speech production (i.e., the semantic verbal fluency test), and (c)their associations with measures of clinical symptomatology (i.e., depression, anxiety, and hostility).

Method

Participants

Data were aggregated from 11 separate studies conducted at large public universities. As part of these studies, participants were asked to provide spontaneous speech samples. Descriptive statistics and study data are provided in Table 1. In total, data were available for 1,350 undergraduate students who reported English as their primary language. Approximately two thirds of this sample was female (i.e., 65.58%) and three quarters was Caucasian (i.e., 78.65%). Each of the studies was approved by the appropriate Institutional Review Boards, and all participants provided written informed consent prior to beginning the study.

Speaking tasks

Across studies, participants were asked to produce speech in one of three different tasks that varied in topical scope, involving (1)"superficial" speech (discussing daily routines, hobbies and/or living situations), (2)"restricted" speech (discussing experiences and reactions to neutral images from the International Affective Picture System [IAPS]; Lang, Bradley, & Cuthbert, 2005) (e.g., door, lamp), and (3)"introspective" speech (discussing autobiographical memories that were neutral in tone; e.g., life events or changes that were not inherently pleasant or unpleasant). For all tasks, instructions and stimuli presentation (e.g., IAPS slides) were automated on a computer and participants were encouraged to speak as much as possible. Research assistants read all instructions to the participants, but were not allowed to speak while the participant was being recorded. Additional information regarding the tasks is provided in Table 1. We expected to find systematic differences in vocal expression across the three tasks. On the basis of prior research from our lab using these tasks (Cohen et al., 2010), and from the extant literature more generally (Huttunen et al., 2011; Scherer, 2003), we reasoned that the restricted task was a particularly challenging task, at least cognitively, in that participants were asked to produce speech that was restricted in topic and in response to stimuli that was artificial in nature. Thus, we expected that vocal production and signal (e.g., F0) variability within speech would be shortest/lowest for this task. Conversely, the superficial and introspective tasks involved content that was general and more automatic (i.e., content that is relatively easy to retrieve from autobiographical stores) in nature, so we expected vocal production and signal for these tasks to be greater/higher than for the restricted task.

Table 1 Descriptive and study characteristics

Study	*N*	*N* Speech Samples	% Female	% Caucasian	Speech Length	Speech Task
1	79	1	49.9%	84.8%	200 s	Superficial
2	121	3	70.4%	86.4%	20 s	Restricted
3	227	1	68.1%	80.3%	100 s	Superficial
4	139	4	79.1%	79.9%	60 s	Introspective
5	154	1	68.0%	72.2%	90 s	Superficial
6	114	1	66.7%	68.3%	90 s	Superficial
7	122	1	65.3%	84.7%	200 s	Superficial
8	115	2	57.9%	82.1%	100 s	Superficial
9	96	4	69.0%	87.0%	60 s	Introspective
10	82	1	60.2%	79.5%	60 s	Superficial
11	101	4	66.8%	60.0%	60 s	Introspective

Acoustic analysis of speech

The Computerized Assessment of Natural Speech protocol (CANS), developed by our lab to assess vocal expression from spontaneous speech, was employed here. The speech was digitally recorded at 16 bits/s at a sampling frequency of 44100 Hz using headset microphones. The CANS protocol takes advantage of Praat software (Boersma & Weenink, 2013), a shareware program that has been used extensively in speech pathology and linguistic studies, as well as Macros, developed by our lab. Sound files are organized into "frames" for analysis, which for the present study was set at a rate of 100 per second. During each frame, F0, F1, F2, and intensity are quantified. On the basis of prior research examining optimization filters for measuring fundamental frequency in automated research (Vogel, Maruff, Snyder, & Mundt, 2009), we applied a low-pass (i.e., 75-Hz) and a high-pass (i.e., 300-Hz) filter. All of the frequency values were converted to semitones due to their nonlinear nature. As we noted in the introduction, a near limitless number of acoustic variables can be computed. We employed a systematic approach to defining our acoustic variables. This approach to characterizing vocal signals was informed by work from studies of clinical populations both from our lab (Cohen et al., 2008; Cohen et al., 2010; Cohen et al., 2009; Cohen, Morrison, Brown, & Minor, 2012) and others (e.g., Alpert et al., 1986; Cannizzaro et al., 2004; Laukka et al., 2008), and is utilized in communication sciences more generally (e.g., Johnstone et al., 2007; Tolkmitt & Scherer, 1986). Furthermore, this approach is consistent with the larger theoretical speech prosody literature (Huttunen et al., 2011; Scherer, 2003; Sobin & Alpert, 1999), and provides a straightforward conceptual framework for understanding signal properties. It is not, by any means, exhaustive (see the Discussion section for potential limitations).

For each of the F0, F1, F2, and intensity signals, we computed the mean and variability values. Variability was defined at three different temporal levels, in terms of (1)the "frame" (i.e., "Perturbation"—signal change in consecutive frames[1]), (2)"local" variability (within utterances), and (3)"global" variability across utterances (i.e., across the speech sample). Local and global variability was defined in terms of two commonly used computations, involving standard deviation and range scores (i.e., average difference between the highest and lowest values within utterances). Additionally, speech production was examined in terms of the presence (i.e., utterance) or absence (i.e., pause) of a signal—what we collectively refer to as *speech production*. "Utterances" were defined as an epoch of F0 signal greater than 150 ms in length with no contiguous pause greater than 50 ms, whereas "pauses" were defined in terms of at least 50 ms of signal absence. Both the mean and standard deviation values were computed for these variables. An additional summary variable of speech production (i.e., Silence Percent), was included in this study. These variables are listed in Table 2.

Verbal fluency

Measures of verbal fluency were included in three of the studies examined and were included here to evaluate the convergent validity of our speech production measures. For two of these studies, semantic fluency tests (i.e., fruits and vegetables) from the Repeatable Battery for the Assessment of Neuropsychological Status (Randolph, 1998) were administered. For the third study, a different semantic fluency (e.g., animal naming) test was used (Green et al., 2004). Data for

[1] Our "perturbation" measures are conceptually similar to the measures of jitter and shimmer that are reported in the extant communication sciences literature, in that they reflect variability on a temporally brief scale. For automation purposes, our measure was based on consecutive frames (10 ms), as opposed to a "cycle"-to-"cycle" basis. In this regard, our measure is meant to reflect subtle perturbation/variability in signal, as opposed to jitter and shimmer more generally.

Table 2 Vocal variables examined in this study

Variable	Description	Variable	Description
Pause Variables			
Silence Percent	Percentage of time without F0 signal	Pause Number	Total number of pauses
Pause *SD*	Standard deviation of pause length excluding the first and last pauses.	Pause Mean	Average pause length in milliseconds (ms), excluding the first and last pauses.
Utterance Variables			
Utterance Mean	Average utterance length in milliseconds (ms)	Utterance.*SD*	Standard Deviation of utterance length in milliseconds (ms)
Fundamental Frequency (F0) Variables			
F0 Mean	*M* computed within each utterance and averaged across all utterances	F0 *SD* Local	Average of SDs computed within each utterance.
F0 *SD* Global	*SD* of SDs computed within each utterance.	F0 Range Local	Average of range scores computed within each utterance.
F0 Range Global	*SD* of range scores computed within each utterance.	F0 Perturbation	Absolute value of average change in consecutively voiced frames within utterance
First Formant (F1) Variables			
F1 Mean	*M* computed within each utterance and averaged across all utterances	F1 *SD* Local	Average of SDs computed within each utterance.
F1 *SD* Global	*SD* of SDs computed within each utterance.	F1 Range Local	Average of range scores computed within each utterance.
F1 Range Global	*SD* of range scores computed within each utterance.		
Second Formant (F2) Variables			
F2 Mean	*M* computed within each utterance and averaged across all utterances	F2 *SD* Local	Average of SDs computed within each utterance.
F2 *SD* Global	*SD* of SDs computed within each utterance.	F2 Range Local	Average of range scores computed within each utterance.
F2 Range Global	*SD* of range scores computed within each utterance.		
Intensity Variables			
Intensity Mean	*M* computed within each utterance and averaged across all utterances	Intensity *SD* Local	Average of SDs computed within each utterance.
Intensity *SD* Global	*SD* of SDs computed within each utterance.	Intensity Range Local	Average of range scores computed within each utterance.
Intensity Range Global	*SD* of range scores computed within each utterance.	Intensity Perturbation	Absolute value of average change in consecutively voiced frames within utterance,

M mean, *SD* standard deviation, *Range* maximum– minimum

other types of verbal fluency (e.g., phonological fluency) and other speaking tasks more generally (e.g., repetition tests) were not collected as part of these studies, and hence were not available for analysis here. Scores were standardized by study to account for potential differences in raw scores across tests. On the basis of evidence that speech rate has been significantly associated with increased verbal fluency in children and adolescents (Martins, Vieira, Loureiro, & Santos, 2007), we predicted that greater verbal fluency ability would be associated with shorter pauses in speech. Note that empirical support for this notion is not overwhelming, particularly in that Martins et al. reported that verbal fluency was not associated with number of extended pauses (>4,000 ms).

Mental health symptoms

Clinical symptoms were measured using the Brief Symptom Inventory (BSI; Derogatis & Melisaratos, 1983), which measures a broad range of psychopathology during the past seven days. We were particularly interested in depression (i.e., "feeling no interest in things"), anxiety (i.e., "feeling tense or keyed up"), and hostility (i.e., "having urges to break or smash things") from this instrument, as these symptoms and emotional conditions have been related to vocal expression in other studies. Specifically, we predicted that both depression and anxiety would be associated with less speech production and speech variability (Cannizzaro et al., 2004; Cohen et al., 2013), that anxiety would also be associated with greater F0 perturbation variability (Laukka et al., 2008), and that hostility would be associated with greater signal variability (Sobin & Alpert, 1999). The BSI has well-documented psychometric properties and has been used in hundreds of published, peer-reviewed studies to date.

Analyses

Analyses were conducted in five steps. First, we evaluated demographic (i.e., sex, ethnicity) variables in their relationship to each of the 28 vocal variables in order to evaluate their potential impact on consequent analyses. Non-Caucasian groups were collapsed into a single group due to their relatively small sample sizes. Second, we computed a zero-order correlation matrix between the 28 vocal variables to identify redundancies (defined as r values >.85) that could be excluded from further analyses. Third, the remaining variables were subjected to PCA to further reduce the number of items. Because our data were potentially inter-correlated, oblique rotations (i.e., Promax) were used. Variables with notable cross-loadings (weights > .40 on multiple factors) or without loadings (weights < .40 on any factor) were excluded from the final PCA solution—though they were examined in the subsequent analyses. We reran the PCA in men and women and in Caucasians and non-Caucasians to ensure that the structure was not grossly invariant across sex and ethnicity groups. Fourth, we evaluated the degree to which vocal variables changed as a function of speaking task using both analyses of variance and logistic regression. For the latter analyses, the speaking task was entered as a dichotomous dependent variable (i.e., superficial vs. restricted, and introspective vs. restricted) in two separate regressions, and the PCA factors identified above were entered in a first (and single) step. In a second step, we entered the nonredundant vocal variables excluded from the final PCA solution. This second step allowed for the evaluation of whether the excluded vocal variables contributed meaningful variance to speaking tasks, above that made by the factor scores. Pseudo-R values (i.e., Cox & Snell, 1989) were used to evaluate the relative contributions of each step. Finally, we employed hierarchical regressions to determine the relative contributions that the vocal factors (Step 1) and the excluded vocal items (Step 2) made to the cognitive (i.e., verbal fluency) and clinical (i.e., depression, anxiety, hostility) variables. For normalization purposes, all "extreme scores" (i.e., <>3.5 standard deviations [SDs] from the mean) were converted to the closest value 3.5 SDs from the mean, and all variables were normalized (i.e., skew statistic< 2.0) prior to being included in the PCA. Because of the large sample size (and hence, high degree of statistical power), effect sizes are reported and evaluated when appropriate. For the fourth and fifth steps, the vocal variables were transformed to statistically control for sex and ethnicity (using standardized residuals computed from linear regressions). Unless otherwise noted, all variables are normally distributed.

Results

Demographic variables

Of the 28 variables examined, six were statistically different between men and women—though only four of these differences exceeded a small effect size (d > 0.20). Women had higher mean F0, F1, and F2 values (ds > 0.50) than men, less F0 perturbation (d = 0.42), and longer and more variable utterances (ds < 0.20) (all ps < .05).

Caucasians were statistically different from non-Caucasians on 17 of the 28 variables, although none of the consequent group differences exceeded a small effect size. Caucasians showed less variability in virtually all F0, F1, F2, and intensity variables (ps < .05). Sex and ethnicity were considered potential confounds for this study.

Redundancy analysis

Across the F0, F1, F2, and Intensity variables, the Range Local variables were redundant with the relevant SD Local scores, and thus were excluded from the PCA, since they were

judged to be less stable indicators of variability than the *SD* scores (e.g., computed on the basis of two vs. all data points in an utterance). The Pause and Utterance *SD* scores were also redundant with their relevant Mean scores, and the Intensity Range Global score was redundant with the Intensity *SD* Global. These scores were also excluded from the PCA. The nonredundant correlation values were included in a zero-order matrix (see Table 3).

Factor structure

An initial PCA was conducted on 21 variables, yielding six factors with eigenvalues greater than 1, explaining 71% of the variance. Inspection of the pattern matrices yielded six cross-loaded (i.e., multiple beta weights > .40) or nonfitting items (i.e., all beta weights < .40; i.e., Silence Percent, F0, F1, F2, Intensity Mean, and Utterance *M*), for which removal yielded a better fit. This yielded a six-factor solution explaining 78.55% of the variance. Separate PCAs computed for men and women (e.g., six factors explaining 77.78% and 76.26% of the variance, respectively) and Caucasians and non-Caucasians (e.g., six factors explaining 78.77% and 77.15% of the variance, respectively) revealed identical structures, suggesting that the factor structure is invariant across sex and ethnicity. These data are included in Table 4. The component correlation matrix suggested that the factors were relatively independent, with only two of the 21 possible correlations exceeding a value of .20 (r = .31, F1 and F2 variability; r = .29, F0 and F2 variability).

Vocal variables across speaking tasks Each of the PCA-based vocal domain scores was significantly different between the restricted and the superficial and introspective speaking tasks (see Fig. 1). The effect sizes for the Pause, F1, F2, and Intensity variables were in the large range (ds > 1.56), whereas those for F0 and Perturbation were more modest (ds > 0.39). Logistic regressions are presented in Table 5. The chi-square values for all steps were significant, though inspection of the pseudo-R-square values suggests that the six PCA domains explained the lion's share of the variance. As a function of restricted topical conditions, Pauses got shorter and F1 and Intensity Variability decreased. F2 and F0 Variability also decreased relative to the introspective and superficial speech, respectively. In the second step, F1 Mean decreased during the restricted condition for both regressions. As compared to the introspective speech conditions, the Utterance Means got longer, and Intensity increased.

Clinical and verbal fluency correlates of vocal variables Table 6 contains the results of the hierarchical regressions. The PCA-based domains made significant contributions to the variances in verbal fluency, depression, and

Table 3 Zero-order correlation matrix of nonredundant vocal variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. Silence	1.00																			
2. Pause *M*	.69	1.00																		
3. Pause *N*	–.24	–.38	1.00																	
4. Utterance *M*	–.63	–.12	–.23	1.00																
5. F0 *M*	–.17	.00	–.15	.26	1.00															
6. F0 *SD* L	–.20	–.02	–.03	.31	.41	1.00														
7. F0 *SD* G	–.09	–.02	.03	.12	.45	.84	1.00													
8. F0 Range G	–.26	–.09	–.13	.35	.45	.65	.78	1.00												
9. F0 Perturb	.31	.23	.00	–.26	–.35	.03	–.09	–.15	1.00											
10. F1 *M*	–.17	–.14	.15	.05	.30	.15	.16	.13	–.25	1.00										
11. F1 *SD* L	–.23	–.07	.10	.32	–.20	.13	.02	.12	.13	.03	1.00									
12. F1 *SD* G	.16	.00	.22	–.27	–.43	–.13	–.07	–.08	.20	–.01	.57	1.00								
13. F1 Range G	–.18	–.15	.21	.10	–.47	–.10	–.14	–.02	.14	–.13	.47	.60	1.00							
14. F2 *M*	–.12	–.05	–.02	.12	.52	.29	.30	.25	–.18	.51	.12	–.03	–.20	1.00						
15. F2 *SD* L	–.38	–.17	.08	.44	.07	.30	.18	.32	–.03	–.11	.48	.09	.28	–.16	1.00					
16. F2 *SD* G	–.15	–.23	.24	–.04	.12	.14	.21	.23	–.12	.05	.09	.18	.17	–.01	.51	1.00				
17. F2 Range G	–.37	–.28	.15	.30	.15	.24	.23	.35	–.11	–.06	.22	.10	.30	.00	.55	.73	1.00			
18. Intens *M*	–.06	.14	–.18	.26	–.05	.01	–.07	–.05	–.08	–.23	.10	–.14	.01	–.16	.14	–.07	.01	1.00		
19. Intens *SD* L	–.22	–.09	.11	.23	–.03	.10	.02	.11	.03	–.01	.26	.06	.22	.01	.24	.07	.16	.32	1.00	
20. Intens *SD* G	.00	–.02	.10	–.04	–.05	.00	.00	.01	.02	.00	.07	.11	.12	.01	.01	.04	.03	.04	.77	1.00
21. Intens Perturb	.23	.22	.05	–.18	.02	.09	.07	–.01	.44	.15	–.02	.07	.00	.10	–.13	–.16	–.18	–.41	.12	.02

M mean, *SD* standard deviation, *L* local, *G* global, *Perturb* perturbation, *Intens* intensity

Table 4 Pattern matrix from the principal components analysis

	F0 Variability	Intensity Variability	F2 Variability	F1 Variability	Signal Perturbation	Pauses
F0 *SD* Global	**.98**	–.06	–.04	–.05	.05	.11
F0 *SD* Local	**.92**	.01	–.01	.00	.11	.00
F0 Range Global	**.85**	.08	.03	.02	–.11	–.07
F2 *SD* Global	–.07	**.93**	–.14	–.06	.07	.14
F2 Range Global	.02	**.90**	–.03	.00	–.01	.05
F2 *SD* Local	.10	**.71**	.20	.05	–.04	–.18
F1 *SD* Local	.14	–.02	**.87**	.04	–.09	–.12
F1 *SD* Global	–.05	–.12	**.87**	–.09	.09	.13
F1 Range Global	–.13	.09	**.75**	.04	.01	.12
Intensity *SD* Global	–.05	–.07	–.05	**.95**	.00	.04
Intensity *SD* Local	.02	.06	.04	**.93**	.03	–.01
Intensity Perturbation	.14	–.06	–.09	.07	**.84**	.08
F0 Perturbation	–.08	.11	.12	–.04	**.82**	–.11
Pause number	.01	.02	.09	.01	.17	**.87**
Pause mean	–.06	–.06	–.02	–.02	.33	**–.69**

The variables loaded in each factor are in bold.

hostility, whereas the contributions of the additional vocal scores were only significant for the hostility measure. Inspection of the beta weights revealed that increased pauses were associated with better verbal fluency performance, that increased vocal perturbation was associated with more severe depressive symptoms, and that no specific vocal factor was associated with hostility—though perturbation increased at a trend level as a function of increased hostility. With respect to the additional variables, higher F1 Mean scores were associated with more hostility. In sum, the PCA-based factors explained a modest but meaningful amount of variance in the clinical and cognitive variables.

Discussion

Computerized measures of spontaneous speech are improving in sophistication, and their presence in psychological and clinical science is increasing. Despite this, studies of their psychometric properties, notably in terms of incremental and structural validity, have been lacking. In a large corpus of speech samples from young adults, we evaluated 28 systematically defined variables that have been used in the literature. Through redundancy analysis and PCA, we identified six broad domains tapping distinct aspects of vocal expression. These domains were robust across both sex and ethnicity. A validity analysis suggested that each of these variables was important in some fashion, either because the variable changed as a function of contextual factors (i.e., speech topic) or via its significant association with verbal fluency or clinical symptom measures. Going forward, our recommendation is that future studies of "macroscopic" levels of vocal expression include measures of each major vocal signal, notably F0, F1, F2, and intensity, as well as the lack of signal (i.e., pauses).

With one notable exception (i.e., Perturbation; see below), our specific measures of variability within a single signal provided little incremental validity over other signal-matched measures. Measures of signal variability based on range scores were generally redundant with those computed using standard deviation scores. Moreover, measures focused on local versus global features were highly correlated, as well, and in the case of the F0 and intensity signals, were almost redundant. Conceptually speaking, local and global measures tap different phenomenon—as, for example, an individual may show considerable variability within utterances with high levels of consistency across utterances. Nonetheless, the present findings suggest that, at least in studies of extended spontaneous speech, "macroscopic" measures of signal variability need not be separately reported or evaluated. Given the importance of F0 and intensity variability in clinical and psychological studies (e.g., Cannizzaro et al., 2004; Cohen et al., 2013; Laukka et al., 2008), further research confirming their potential redundancy is warranted.

The exception to the aforementioned redundancy between variability measures involved Signal Perturbation—a measure of signal variability occurring at the "frame" level. Generally, measures of F0 and Intensity perturbation were not highly correlated with their respective local and global variability measures, and perturbation measures emerged as a distinct factor in the PCA. Moreover, the Perturbation factor was associated with clinical measures in a relatively unique way. Of note, increasing signal perturbation was associated with both depression (significantly) and hostility (at a trend level).

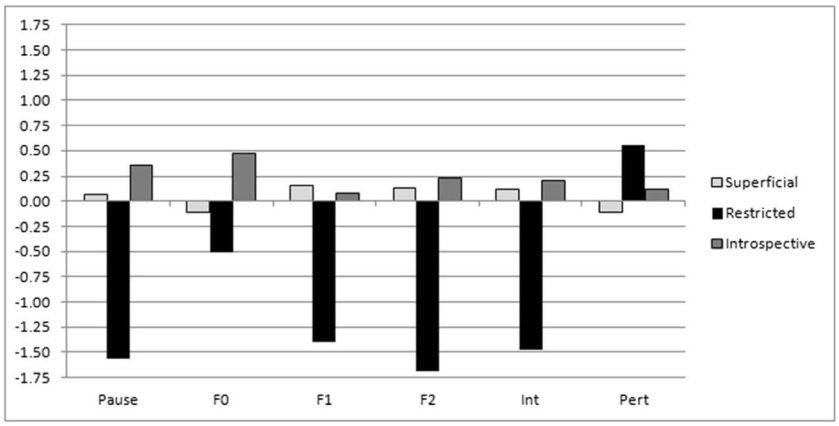

Fig. 1 Standardized differences in vocal domains as a function of speaking task

Interestingly, F0 perturbation has been associated with anxiety, or the so-called "jittery" voice (Fuller, Horii, & Conner, 1992). The reason for our null findings in this regard are not clear. Significant correlations notwithstanding, several factors likely attenuated the magnitude of these relationships. First, insofar as the samples examined here were not clinical in nature, the depression, anxiety, and hostility levels were likely restricted in range. Second, our measures of clinical symptoms were based on self-report, an important, but relatively circumscribed, domain of assessment. It is likely that depression, anxiety, and hostility ratings derived from clinical interviews would yield more comprehensive estimates of symptom severity. Finally, the clinical measure tapped symptoms from a one-week epoch, so clinical scores did not necessarily reflect symptomatology at the time of the vocal assessment.

Additional redundancies were observed between the mean and variability values for both Pause and utterance length. Similarly, Silence Percent was considered redundant (or at least, superfluous), in that it was correlated with a number of speech production and variability measures, and its inclusion in the PCA complicated the structure (i.e., contributing to dual loadings). Interestingly, Silence Percent was not related to speaking task or to cognitive or clinical correlates above and beyond the PCA factors, suggesting that this variable provides little incremental validity beyond other vocal measures.

The degree to which the mean values of F0, F1, F2, and intensity signals are informative above and beyond vocal variability and pause measures is a bit unclear. There were some isolated significant findings, such as that F1, Intensity, and Utterance Means changed as a function of speech topics. Collectively, however, the mean values explained very little

Table 5 Logistic regressions evaluating the relative contributions of vocal domains (identified from the PCA) and additional vocal variables (excluded from the PCA) to speaking task

	DV= Introspective v. Restricted Task			DV= Superficial v. Restricted Task		
	B (B_{SE})	Wald	Exp(B)	B (B_{SE})	Wald	Exp(B)
Step 1. Vocal Factors	χ^2= 387.83*, ΔR^2= .58			χ^2= 461.99*, ΔR^2= .37		
Pause	2.11 (0.32)	44.30*	8.24	1.27 (0.18)	49.75*	3.55
F0 variability	0.98 (0.32)	9.33*	2.65	–0.16 (0.20)	0.61	0.85
F1 variability	0.65 (0.34)	3.59	1.92	0.78 (0.21)	14.16*	2.18
F2 variability	0.63 (0.34)	3.38	1.88	1.06 (0.21)	25.60*	2.90
Intensity variability	0.87 (0.38)	5.13*	2.38	0.72 (0.22)	11.10*	2.06
Perturbation	0.02 (0.26)	0.00	1.02	–0.23 (0.18)	1.66	0.80
Step 2. Additional Vocal Variables	χ^2= 31.32*, ΔR^2= .03			χ^2= 86.25*, ΔR^2= .06		
Silence percent	1.31 (0.60)	4.80	3.71	–0.16 (0.45)	0.13	0.85
Utterance mean	0.71 (0.43)	2.77	2.04	–0.80 (0.36)	4.79*	0.45
F0 mean	–0.95 (0.59)	2.62	0.39	–0.05 (0.30)	0.03	0.95
F1 mean	1.59 (0.44)	13.28	4.92	1.51 (0.29)	27.82*	4.54
F2 mean	–0.23 (0.40)	0.34	0.79	–0.17 (0.28)	0.36	0.84
Intensity mean	0.07 (0.33)	0.04	1.07	–0.61 (0.25)	6.13*	0.54

ΔR^2, Cox and Snell pseudo *R*-squared. * $p < .05$

Table 6 Hierarchical linear regressions evaluating the relative contributions of vocal domains (identified from the PCA) and additional vocal variables (excluded from the PCA) to cognitive and clinical variables

	DV = Verbal Fluency			DV = Depression			DV = Anxiety			DV = Hostility		
	B (B_{SE})	β	*t*	B (B_{SE})	β	*t*	B (B_{SE})	β	*t*	B (B_{SE})	β	*t*
Step 1: Vocal Factors	$\Delta r^2 = .05$, $\Delta F = 2.49^*$			$\Delta r^2 = .05$, $\Delta F = 2.28^*$			$\Delta r^2 = .03$, $\Delta F = 1.14$			$\Delta r^2 = .05$, $\Delta F = 2.28^*$		
Pause	0.27 (0.07)	.23	3.73*	–0.35 (0.56)	–.04	–0.62	–0.14 (0.49)	–.02	–0.29	0.32 (0.36)	0.05	0.88
F0	0.08 (0.08)	.07	1.10	–0.02 (0.51)	.00	–0.04	0.09 (0.45)	.01	0.19	–0.51 (0.33)	–0.10	–1.55
F1	0.03 (0.07)	.03	0.49	0.82 (0.54)	.10	1.51	0.48 (0.48)	.07	1.00	0.14 (0.35)	0.03	0.40
F2	–0.01 (0.08)	–.01	–0.13	–0.65 (0.54)	–.07	–1.21	–0.75 (0.48)	–.10	–1.57	–0.57 (0.35)	–0.10	–1.63
Intensity	0.00 (0.07)	.00	0.00	–0.52 (0.54)	–.06	–0.95	0.49 (0.48)	.06	1.01	0.01 (0.35)	0.00	0.03
Perturbation	–0.01 (0.07)	–.01	–0.12	1.07 (0.45)	.15	2.40*	0.25 (0.40)	.04	0.62	0.48 (0.29)	0.11	1.65
Step 2: Additional Vocal Variables	$\Delta r^2 = .02$, $\Delta F = .82$			$\Delta r^2 = .01$, $\Delta F = .59$			$\Delta r^2 = .03$, $\Delta F = 1.38$			$\Delta r^2 = .05$, $\Delta F = 1.99^*$		
Silence percent	–0.10 (0.27)	–.09	–0.37	2.09 (1.92)	.27	1.09	1.88 (1.69)	.27	1.11	–1.81 (1.22)	–0.36	–1.49
Utterance mean	0.01 (0.24)	.01	0.05	1.47 (1.64)	.19	0.90	1.58 (1.44)	.23	1.09	–1.43 (1.04)	–0.29	–1.37
F0 mean	0.05 (0.10)	.05	0.45	–0.22 (0.65)	–.04	–0.34	1.32 (0.57)	.27	2.32	1.03 (0.41)	0.28	2.51*
F1 mean	0.10 (0.07)	.11	1.41	0.56 (0.55)	.08	1.01	–0.31 (0.49)	–.05	–0.64	–0.10 (0.35)	–0.02	–0.28
F2 mean	–0.05 (0.09)	–.05	–0.49	0.04 (0.51)	.01	0.08	–0.20 (0.45)	–.04	–0.44	–0.05 (0.32)	–0.01	–0.14
Intensity mean	–0.06 (0.08)	–.05	–0.71	0.68 (0.53)	.10	1.27	0.28 (0.47)	.05	0.59	0.62 (0.34)	0.14	1.84

* $p < .05$

variance above and beyond the vocal factors identified in the PCA. F0 values tended to be higher in people with greater self-reported hostility. Generally speaking, F0 and Intensity mean values are associated with emotional and clinical states in many studies of vocal expression (Batliner et al., 2008; Batliner et al., 2006; Cannizzaro et al., 2004; Cohen et al., 2010; Cohen et al., 2013; Johnstone et al., 2007; Laukka et al., 2008; Sobin & Alpert, 1999; Tolkmitt & Scherer, 1986), so it is a bit surprising that these measures were not more highly associated in this study. Many prior studies have not controlled for sex and ethnicity in the same manner that we did, and it is the case that Mean F0 values are generally much higher in women. It is also the case that many prior studies have employed a "microscopic" level of analysis or have not examined spontaneous speech. In this manner, concerns can be raised that the relationship between these variables is confounded by demographic variables or is attenuated in speech, particularly involving extended speech samples.

Several limitations warrant mention. First, the sample was relatively homogeneous with respect to age, ethnicity, and education. Given that culture, age, and education can affect speech, it would be important to replicate the present findings in a more diverse sample. Second, our measures of convergent validity were by no means comprehensive. That is, our use of three different speech tasks does not begin to approximate the variety of contextual and speech factors that potentially influence speech outside the laboratory setting. Third, all of the speech samples examined in this study were produced as part of laboratory studies with limited opportunities for interaction with the research assistant. It would be important to evaluate the psychometric properties of spontaneous speech under more ecologically valid conditions. Finally, the acoustic measures examined in this study were by no means exhaustive, and it is possible that the factor structure and clinical correlates would differ if other measures were used. The variables examined here covered the major conceptual components of vocal analysis discussed in the literature (i.e., five signals across varying temporal levels)—though it remains an empirical question whether variables computed using other means might show different results.

The human voice offers an important window into the state of many psychological operations of an individual. Insofar as vocal samples can be easily obtained and their analysis can be automated, the application of vocal analysis, particularly involving spontaneous speech, has a near unlimited potential. A major obstacle in implementing vocal technologies involves analyzing and interpreting vocal signal—that is, which of the many variables that can be extracted from vocal signal should be used? The present data suggest that a limited number of domains are important for vocal analysis of extended speech samples and, importantly, offer an empirically derived structure for future research and technological applications.

References

Alpert, M., Homel, P., Merewether, F., Marz, J., & Lomask, M. (1986). Voxcom: A system for analyzing natural speech in real time. *Behavior Research Methods, Instruments, & Computers, 18,* 267–272. doi:10.3758/BF03201035

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70,* 614–636. doi:10.1037/0022-3514.70.3.614

Batliner, A., Steidl, S., Hacker, C., & Nöth, E. (2008). Private emotions versus social interaction: A data-driven approach towards analysing emotion in speech. *User Modeling and User-Adapted Interaction, 18,* 175–206. doi:10.1007/s11257-007-9039-4

Batliner, A., Steidl, S., Schuller, B., Seppi, D., Laskowski, K., Vogt, T., & Aharonson, V. (2006). Combining efforts for improving automatic classification of emotional user states. In *Proceedings of IS-LTC* (pp. 240–245). Ljubljana, Slovenia: Slovenian Language Technologies Society.

Batliner, A., Steidl, S., Schuller, B., Seppi, D., Vogt, T., Wagner, J., & Amir, N. (2011). Whodunnit—searching for the most important feature types signalling emotion-related user states in speech. *Computer Speech & Language, 25,* 4–28. doi:10.1016/j.csl.2009.12.003

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.3.59). Retrieved from www.praat.org/

Cannizzaro, M., Harel, B., Reilly, N., Chappell, P., & Snyder, P. J. (2004). Voice acoustical measurement of the severity of major depression. *Brain and Cognition, 56,* 30–35. doi:10.1016/j.jneuroling.2006.04.001

Cohen, A. S., Alpert, M., Nienow, T. M., Dinzeo, T. J., & Docherty, N. M. (2008). Computerized measurement of negative symptoms in schizophrenia. *Journal of Psychiatric Research, 42,* 827–836. doi:10.1016/j.jpsychires.2007.08.008

Cohen, A. S., Dinzeo, T. J., Donovan, N. J., Brown, C. E., & Morrison, S. C. (2015). Vocal acoustic analysis as a biometric indicator of information processing: Implications for neurological and psychiatric disorders. *Psychiatry Research, 226,* 235–241. doi:10.1016/j.psychres.2014.12.054

Cohen, A. S., & Elvevåg, B. (2014). Automated computerized analysis of speech in psychiatric disorders. *Current Opinion in Psychiatry, 27,* 203–209. doi:10.1097/YCO.0000000000000056

Cohen, A. S., Hong, S. L., & Guevara, A. (2010). Understanding emotional expression using prosodic analysis of natural speech: Refining the methodology. *Journal of Behavioral Therapy and Experimental Psychiatry, 41,* 150–157. doi:10.1016/j.jbtep.2009.11.008

Cohen, A. S., Kim, Y., & Najolia, G. M. (2013). Psychiatric symptom versus neurocognitive correlates of diminished expressivity in schizophrenia and mood disorders. *Schizophrenia Research, 146,* 249–253. doi:10.1016/j.schres.2013.02.002

Cohen, A. S., McGovern, J. E., Dinzeo, T. J., & Covington, M. A. (2014). Speech deficits in serious mental illness: A cognitive resource issue? *Schizophrenia Research, 160,* 173–179. doi:10.1016/j.schres.2014.10.032

Cohen, A. S., Minor, K. S., Najolia, G. M., & Lee Hong, S. (2009). A laboratory-based procedure for measuring emotional expression from natural speech. *Behavior Research Methods, 41,* 204–212. doi:10.3758/BRM.41.1.204

Cohen, A. S., Mitchell, K. R., & Elvevåg, B. (2014). What do we really know about blunted affect and alogia?: A meta-analysis of objective assessments. *Schizophrenia Research, 159,* 533–538. doi:10.1016/j.schres.2014.09.013

Cohen, A. S., Morrison, S. C., Brown, L. A., & Minor, K. S. (2012). Towards a cognitive resource limitations model of diminished expression in schizotypy. *Journal of Abnormal Psychology, 121,* 109–118. doi:10.1037/a0023599

Cox, D. R., & Snell, E. J. (1989). *The analysis of binary data* (2nd ed.). London: Chapman and Hall.

Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine, 13,* 595–605. doi:10.1017/S0033291700048017

Esposito, A., & Esposito, A. M. (2012). On the recognition of emotional vocal expressions: Motivations for a holistic approach. *Cognitive Processing, 13*(Suppl. 2), S541–S550.

Eyben, F., Weninger, F., Groß, F., & Schuller, B. (2013). Recent developments in opensmile, the Munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM International Conference on Multimedia* (pp. 835–838). New York, NY: ACM. doi:10.1145/2502081.2502224

Fuller, B. F., Horii, Y., & Conner, D. A. (1992). Validity and reliability of nonverbal voice measures as indicators of stressor-provoked anxiety. *Research in Nursing and Health, 15,* 379–389. doi:10.1002/nur.4770150507

Giddens, C. L., Barron, K. W., Clark, K. F., & Warde, W. D. (2010). Beta-adrenergic blockade and voice: A double-blind, placebo-controlled trial. *Journal of Voice, 24,* 477–489. doi:10.1016/j.jvoice.2008.12.002

Green, M. F., Nuechterlein, K. H., Gold, J. M., Barch, D. M., Cohen, J., Essock, S., & Marder, S. R. (2004). Approaching a consensus cognitive battery for clinical trials in schizophrenia: The NIMH-MATRICS conference to select cognitive domains and test criteria. *Biological Psychiatry, 56,* 301–307. doi:10.1016/j.biopsych.2004.06.023

Huttunen, K., Keranen, H., Vayrynen, E., Paakkonen, R., & Leino, T. (2011). Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights. *Applied Ergonomics, 42,* 348–357. doi:10.1016/j.apergo.2010.08.005

Johnstone, T., van Reekum, C. M., Bänziger, T., Hird, K., Kirsner, K., & Scherer, K. R. (2007). The effects of difficulty and gain versus loss on vocal physiology and acoustics. *Psychophysiology, 44,* 827–837. doi:10.1111/j.1469-8986.2007.00552.x

Kent, R. D., & Kim, Y. J. (2003). Toward an acoustic typology of motor speech disorders. *Clinical Linguistics and Phonetics, 17,* 427–445. doi:10.1080/0269920031000086248

Kim, Y., Kent, R. D., & Weismer, G. (2011). An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria. *Journal of Speech, Language, and Hearing Research, 54,* 417–429. doi:10.1044/1092-4388(2010/10-0020)

Krajewski, J., Batliner, A., & Golz, M. (2009). Acoustic sleepiness detection: Framework and validation of a speech-adapted pattern recognition approach. *Behavior Research Methods, 41,* 795–804. doi:10.3758/BRM.41.3.795

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2005). *International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual* (Technical Report No. A-6). Gainesville, FL: University of Florida, Center for Research in Psychophysiology.

Laukka, P., Linnman, C., Åhs, F., Pissiota, A., Frans, Ö., Faria, V., & Furmark, T. (2008). In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech. *Journal of Nonverbal Behavior, 32,* 195–214. doi:10.1007/s10919-008-0055-9

Martins, I. P., Vieira, R., Loureiro, C., & Santos, M. E. (2007). Speech rate and fluency in children and adolescents. *Child Neuropsychology, 13,* 319–332. doi:10.1080/09297040600837370

Nadig, A., Lee, I., Singh, L., Bosshart, K., & Ozonoff, S. (2010). How does the topic of conversation affect verbal exchange and eye gaze? A comparison between typical development and high-functioning autism. *Neuropsychologia, 48,* 2730–2739. doi:10.1016/j.neuropsychologia.2010.05.020

Randolph, C. (1998). *RBANS Manual—Repeatable battary for the assessment of neuropsychological status*. San Antonio, TX: Psychological Corp.

Roy, N., Bless, D. M., & Heisey, D. (2000). Personality and voice disorders: A multitrait–multidisorder analysis. *Journal of Voice, 14,* 521–548. doi:10.1016/S0892-1997(00)80009-0

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99,* 143–165. doi:10.1037/0033-2909.99.2.143

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40,* 227–256. doi:10.1016/S0167-6393(02)00084-5

Schuller, B., Batliner, A., Seppi, D., Steidl, S., Vogt, T., Wagner, J., & Aharonson, V. (2007a). The relevance of feature type for the automatic classification of emotional user states: Low level descriptors and functionals. In *Proceedings of INTERSPEECH 2007* (pp. 2253–2256). Baixas, France: International Speech Communication Association.

Schuller, B., Seppi, D., Batliner, A., Maier, A., & Steidl, S. (2007b). *Towards more reality in the recognition of emotional speech. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2007 (ICASSP 2007) (Vol. 4, pp. IV-941–IV-944)*. Piscataway, NJ: IEEE Press.

Schuller, B., Steidl, S., Batliner, A., Vinciarelli, A., Scherer, K., Ringeval, F., . . . Kim, S. (2013). The INTERSPEECH 2013 Computational Paralinguistics Challenge: Social signals, conflict, emotion, autism. In F. Bimbot et al. (Eds.), *Proceedings of INTERSPEECH 2013* (pp. 148–152). Baixas, France: International Speech Communication Association.

Shriberg, L. D., Fourakis, M., Hall, S. D., Karlsson, H. B., Lohmeier, H. L., McSweeny, J. L., & Wilson, D. L. (2010). Perceptual and acoustic reliability estimates for the Speech Disorders Classification System (SDCS). *Clinical Linguistic and Phonetics, 24,* 825–846. doi:10.3109/02699206.2010.503007

Slavin, D. C., & Ferrand, C. T. (1995). Factor analysis of proficient esophageal speech: Toward a multidimensional model. *Journal of Speech and Hearing Research, 38,* 1224–1231. doi:10.1044/jshr.3806.1224

Sobin, C., & Alpert, M. (1999). Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research, 28,* 167–365. doi:10.1023/A:1023237014909

Tolkmitt, F. J., & Scherer, K. R. (1986). Effect of experimentally induced stress on vocal parameters. *Journal of Experimental Psychology: Humman Percepttion and Performance, 12,* 302–313. doi:10.1037/0096-1523.12.3.302

Vogel, A. P., Maruff, P., Snyder, P. J., & Mundt, J. C. (2009). Standardization of pitch-range settings in voice acoustic analysis. *Behavior Research Methods, 41,* 318–324. doi:10.3758/BRM.41.2.318

Vogt, T., & André, E. (2005). Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. In *Proceedings of the IEEE International Conference on Multimedia and Expo, 2005 (ICME 2005)* (pp. 474–477). Piscataway, NJ: IEEE Press.

Yamashita, Y., Nakajima, Y., Ueda, K., Shimada, Y., Hirsh, D., Seno, T., & Smith, B. A. (2013). Acoustic analyses of speech sounds and rhythms in Japanese- and English-learning infants. *Frontiers in Psychology, 4*(57), 1–10. doi:10.3389/fpsyg.2013.00057