HHS Public Access

Author manuscript

Schizophr Res. Author manuscript; available in PMC 2019 July 30.

Published in final edited form as:

Schizophr Res. 2018 July ; 197: 392–399. doi:10.1016/j.schres.2018.01.007.

The Aprosody of Schizophrenia: Computationally Derived Acoustic Phonetic Underpinnings of Monotone Speech

Michael T. Compton, M.D., M.P.H.[a,*], **Anya Lunden, Ph.D.**[b], **Sean D. Cleary, Ph.D.**[c], **Luca Pauselli, M.D.**[a], **Yazeed Alolayan, M.B.B.S.**[d], **Brooke Halpern, L.M.H.C.**[e], **Beth Broussard, M.P.H., C.H.E.S.**[e], **Anthony Crisafio**[f], **Leslie Capulong**[e], **Pierfrancesco Maria Balducci, M.D.**[g], **Francesco Bernardini, M.D.**[h], **Michael A. Covington, Ph.D.**[i]

[a]Columbia University College of Physicians & Surgeons, Department of Psychiatry, New York, New York, U.S.

[b]College of William and Mary, Department of English, Linguistics Program, Williamsburg, Virginia, U.S.

[c]The George Washington University Milken Institute School of Public Health, Department of Epidemiology and Biostatistics, Washington, D.C., U.S.

[d]Case Western Reserve University, Department of Neurology, Cleveland, OH, U.S.

[e]Lenox Hill Hospital, New York, New York, U.S.

[f]The George Washington University School of Medicine and Health Sciences, Washington, DC, U.S.

[g]University of Perugia, Department of Medicine, Section of Psychiatry, Perugia, Italy

[h]Université Libre de Bruxelles, Erasme Hospital, Department of Psychiatry, Anderlecht, Belgium

[i]The University of Georgia, Institute for Artificial Intelligence, Athens, Georgia, U.S.

Abstract

Objective—Acoustic phonetic methods are useful in examining some symptoms of schizophrenia; we used such methods to understand the underpinnings of aprosody. We hypothesized that, compared to controls and patients without clinically rated aprosody, patients with aprosody would exhibit reduced variability in: pitch (F0), jaw/mouth opening and tongue height (formant F1), tongue front/back position and/or lip rounding (formant F2), and intensity/loudness.

[*]Corresponding Author: Michael T. Compton, M.D., M.P.H., New York State Psychiatric Institute, 1051 Riverside Drive, Unit 100, New York, NY 10032. Tel: 1 (646) 774-8762. mtc2176@cumc.columbia.edu.

Contributors

All authors contributed to the conceptualization and writing of this article, and all approved the final version for publication.

Conflicts of Interest

The authors know of no conflicts of interest pertaining to this research.
The authors report no financial relationships with commercial interests.

Methods—Audiorecorded speech was obtained from 98 patients (including 25 with clinically rated aprosody and 29 without) and 102 unaffected controls using five tasks: one describing a drawing, two based on spontaneous speech elicited through a question (Tasks 2 and 3), and two based on reading prose excerpts (Tasks 4 and 5). We compared groups on variation in pitch (F0), formant F1 and F2, and intensity/loudness.

Results—Regarding pitch variation, in unadjusted tests, patients with aprosody differed significantly from controls in Tasks 3 and 4; for Task 5, the difference was statistically significant in both unadjusted tests and those adjusted for sociodemographics. For the standard deviation (SD) of F1, the expected pattern was observed in the two reading tasks in adjusted tests. Regarding SD of F2, patients with aprosody had lower values than controls in unadjusted tests across all tasks; in adjusted tests the expected pattern was observed in the two spontaneous speech tasks.

Conclusions—Findings could represent a step toward developing new methods for measuring and tracking the severity of this specific negative symptom using acoustic phonetic parameters; such work is relevant to other psychiatric and neurological disorders.

Keywords

Acoustic resonance; Aprosody; Linguistics; Negative symptoms; Phonetics; Phonology; Psychosis; Schizophrenia

1. Introduction

Among individuals with schizophrenia, a diminution of normal behaviors and functions has been described since Bleuler (1911) and Kraepelin (1919); these anomalies are currently classified as “negative symptoms,” and include manifestations such as blunted affect, alogia, emotional withdrawal, poverty of speech and content of speech, aprosody, asociality or social withdrawal, anhedonia, amotivation, and avolition. Studies suggest a lifetime prevalence of prominent primary negative symptoms of 15–20%, increasing with age (Buchanan, 2007); however, many if not most persons with schizophrenia have some level of negative symptomatology. The negative symptoms are associated with even poorer functional outcomes and more reduced quality of life than the “positive symptoms” of hallucinations and delusions. Negative symptoms are cross-sectionally associated with poor social functioning and these symptoms also longitudinally predict social and occupational impairment; yet, despite their disabling nature and high prevalence, treatments are very limited (Murphy et al., 2006). Furthermore, the heterogeneity within the concept of “negative symptoms” complicates treatment planning and research.

Given the diversity of negative symptoms, lumping them into a single category impedes the ability of clinicians to track fluctuations in severity over time, and obscures research that could prove informative in understanding the fundamental components of psychotic disorders. As such, our field needs highly reliable, efficient, automated, and finely detailed measures of specific negative symptoms, unaffected by the major limitations of qualitative clinical ratings (e.g., mild, moderate, severe) and even research ratings based on in-depth clinical interviews (e.g., symptom severity scores of 1–5). Being able to objectively quantify the phenotypic characterization of a behavior will allow us to describe it with a dimensional

approach and possibly use it as a biomarker in an attempt to go beyond categorical diagnoses as given in DSM-5 (APA, 2013), and toward a new approach suggested by the National Institute of Mental Health through the Research Domain Criteria project (RDoC: https://www.nimh.nih.gov/research-priorities/rdoc/index.shtml). Along these lines, Deserno and colleagues (2017) proposed a computational approach to dissect mechanisms underlying different facets of negative symptoms using computational models based on behavioral and neuroimaging data. Several scientists have used automated facial or body analysis in an attempt to objectively identify and describe negative symptoms (Alvino et al., 2007; Hamm et al., 2011; Kupper et al., 2010; Wang et al., 2008).

Alpert and colleagues (1986), using his Voxcom technology, a computer-driven program that separates the voice signal into two channels, amplitude/loudness and frequency/pitch (objects also of our investigation), correlated negative symptoms with vocal/acoustic parameters. Patients with clinically rated “flat affect” were shown to have less vocal inflection and less overall speech production than patients with schizophrenia without flat affect (Andreasen et al., 1981). Acoustic analysis of patients’ spontaneous speech during interviews revealed that the duration of pauses was strongly correlated with the clinician’s impressions of flat affect and alogia (Alpert et al., 1997). In 2000, a study acoustically analyzing audiorecordings for fluency and two types of prosody (inflection and emphasis), showed that patients with flat affect spoke with less inflection and were less fluent compared to non-flat patients and a control group (Alpert et al., 2000). The same group (Alpert et al., 2002) also reported a weak correlation between the acoustic measures of vocal inflection and fundamental frequency (pitch) variance and negative symptoms as measured with the *Scale for the Assessment of Negative Symptoms* (SANS; Andreasen, 1983). Using the same technology, a study of medication-free patients with schizophrenia showed a correlation between affective flattening and the acoustic index of vocal expressiveness (Kring et al., 1994). Another study, involving 42 patients with chronic schizophrenia and 42 matched controls, revealed a close relationship between acoustic variables (Voxcom-derived) of patients’ speech and negative symptom severity (Stassen et al., 1995); such variables were also shown to predict at high reliability the severity of the negative syndrome at hospital release (Püschel et al., 1998).

In a prior study (Covington et al., 2012), we found, in 25 first-episode psychosis patients, a significant and meaningful correlation between negative symptom severity and variability of tongue front-to-back position (measured as formant F2, described in detail below in section 2.5). The same methodology was employed in a recent study (Bernardini et al., 2015), which compared the variability of specific natural speech phonetic parameters in two samples of 20 patients in Italy and 20 in the U.S. (the latter drawn from the current study), and analyzed their association with negative symptom severity. Meaningful correlations between negative symptom severity and variability in F2 and pitch were observed in the Italian sample. In the present study, we focus on *aprosody* specifically (rather than negative symptoms in general).

Prosody is the melodic line of speech produced by variations in pitch, rhythm, and stress of pronunciation (Wymer et al., 2002); thus, aprosody is an inability (or reduced ability) to produce such tone in speech. We collected diverse audiorecordings of spoken language in both patients and controls, to test four hypotheses: that, compared to controls, and compared

to patients with schizophrenia without clinically rated aprosody, patients with clinically rated aprosody would exhibit reduced variability in pitch (standard deviation (SD) of F0), mouth-opening (SD of F1), tongue-movement (SD of F2), and intensity/loudness. Such research could represent a step toward developing new methods for measuring and tracking the severity of this specific negative symptom using computationally derived acoustic phonetic parameters.

2. Methods

2.1. Setting and Sample

Ninety-eight patients with schizophrenia (or schizophreniform disorder or psychotic disorder not otherwise specified if they were first-episode or early-course patients) were recruited from sites in both Washington, D.C. (n=61, 62.2%), and New York City (n=37, 37.8%). In Washington, D.C., patients were enrolled from a Core Service Agency (CSA) that provides outpatient community mental health services in the Georgia-Petworth neighborhood (n=20, 20.4%); another CSA in northwestern D.C. (n=12, 12.2%); the inpatient psychiatric unit of a private, downtown, university-affiliated teaching hospital (n=15, 15.3%); and the inpatient psychiatric unit of a large community hospital in northwestern D.C. (n=14, 14.3%). In New York, patients were recruited from the inpatient psychiatric unit of a large community hospital in the Upper East Side of Manhattan (n=15, 15.3%), the outpatient mental health clinic of that hospital (n=3, 3.1%), an early intervention for psychosis service also affiliated with that hospital (n=2, 2.0%), an adult inpatient unit of a large psychiatric hospital in Queens (n=5, 5.1%), the outpatient mental health clinic affiliated with that hospital (n=11, 11.2%), and by referral from a social worker at a college who heard about the study (n=1, 1.0%).

Clinicians at the aforementioned sites referred potentially eligible patients. Native English-speaking patients, aged 18–50 years, with a Mini-Mental State Examination (MMSE; Folstein et al., 1975; Cockrell and Folstein, 1987) score of ≥24, were eligible for the study. Those with known or suspected mental retardation or dementia, a medical condition compromising ability to participate, a comorbid *Structured Clinical Interview for DSM-IV Axis I Disorders* (SCID; First et al., 1994) diagnosis of a mood disorder (or the presence of schizoaffective disorder), or inability to give informed consent were excluded. In the process of recruiting and enrolling the 98 patients, 105 patients were referred but deemed ineligible or refused to participate: English was not their first language (n=10), being older than 50 years (n=1), having a MMSE score <24 (n=8), known or suspected mental retardation (n=13), a medical condition compromising ability to participate (n=5), having an exclusionary diagnosis (n=20), lacking capacity to give informed consent (n=6), and being referred and eligible but refusing to participate (n=42). There were no significant differences between those who were eligible/enrolled and those who were ineligible or refused in terms of age (32.9±9.7 years v. 33.3±10.4; t=0.26, df=200, p=0.79), gender (72.4% male v. 63.5%; χ^2=1.86, df=1, p=0.17), or race (81.3% African American v. 76.0%; χ^2=1.74, df=2, p=0.42).

A total of 102 controls were recruited through advertisements placed in the *Washington Post* (n=4), *AM New York* (n=29), Craigslist (n=20), and *The Southwester* (n=1); by word-of-

mouth (n=14); and through flyers posted or handed out in public areas such as houses of worship, grocery stores, the YMCA, and various community centers (n=34). Eligible controls were native English-speaking and aged 18–50. Those with known or suspected mental retardation or dementia, a medical condition compromising ability to participate, or a SCID-based diagnosis of a psychotic or mood disorder were excluded. In the process of recruiting/enrolling controls, only three were excluded for being outside the target age-range (n=1), or having a SCID-based diagnosis of a depressive disorder (n=2).

2.2. General Procedures

Research assessments with patients were conducted in a quiet, comfortable office in the clinical settings mentioned above, and assessments with controls were conducted in research offices. Written informed consent, using Institutional Review Board-approved protocols, was obtained from all participants, who were reimbursed $80 upon completion of the 2–4-hour assessment.

In addition to the measures described in more detail below, basic sociodemographic information was obtained; the word-reading subtest of the *Wide-Range Achievement Test* (WRAT; Jastak and Wilkinson, 1984) was used to determine word recognition/literacy; and the psychotic and mood disorders modules of the SCID (First et al., 1994) were used to confirm and specify diagnoses among patients and to exclude such diagnoses among controls. Among patients, 80 had schizophrenia, and the SCID-based DSM-IV subtypes were as follows: paranoid (62, 63.3%), undifferentiated (12, 12.2%), and disorganized (5, 5.1%). Twenty-four patients (24.5%) were considered first-episode or early-course patients, meaning that they were assessed at the time of their first hospitalization or had been receiving services for up to three months in a specialty early-psychosis program. For those patients, some of whom might not yet have met criteria for schizophrenia, 15 (15.3%) had a diagnosis of psychotic disorder not otherwise specified, and three (3.1%) had schizophreniform disorder. Information about medications prescribed to patients at the time of the assessment was collected only for a subgroup of the sample (51, 52.0%) due to difficulties in accessing medical record data among outpatients at some sites. We calculated the chlorpromazine-equivalent dosage as a standardized measure (Andreasen et al., 2010; Leucht et al., 2015)

2.3. Measures of Aprosody among Patients

After an in-depth, semi-structured clinical interview, a trained research assessor scored the SANS (Andreasen, 1983), which is comprised of 25 items scored on a scale from 0=none to 5=severe. Although its items are commonly grouped into five subscales (affective flattening or blunting, alogia, avolition-apathy, anhedonia-asociality, and attention), we used only the “Lack of Vocal Inflections” item; its definition and rating scale are given in Table 1. We also administered the *Clinical Assessment Interview for Negative Symptoms* (CAINS; Forbes et al., 2010; Blanchard et al., 2010), a newer measure of negative symptoms that assesses both experiential and expressive deficits. We relied on the “Vocal Expression” item; its definition and rating scale are also shown in Table 1.

Using both SANS “Lack of Vocal Inflections” and CAINS “Vocal Expression” items (which were correlated at ρ=.71, $p<.001$), we created subsamples of patients with and without clinically rated aprosody. We first excluded patients with SANS “Lack of Vocal Inflections” scores of 1 (questionable) or 2 (mild), as well as patients with a CAINS “Vocal Expression” score of 1 (mild). *Patients with aprosody* ($n=25$) were those scoring 3–5 on the SANS item and 2–4 on the CAINS item. *Patients without aprosody* ($n=29$) included those scoring 0 on both items.

2.4. Acquisition of Speech Samples

A series of audiorecordings of participants’ speech was obtained using a Tascam DR-08 recorder set to the following specifications: (1) ENCODING: PCM 16-bit 44.1 kHz monaural, (2) LOW CUT: Low40 Hz, (3) REC EQ: Off, (4) microphone folded to a closed position, (5) built-in stand open, and (6) device placed on a table in front of the participant with the microphone about 12 inches from him or her. We used three elicitation tasks for spontaneous speech and two for reading aloud, as shown in Table 2.

2.5. Measures Derived from Acoustic Phonetic Analysis

From these five types of audiorecorded speech samples, we obtained a number of phonetic parameters. The sound files for each of the five tasks for each participant started and ended with the participant’s voice, excluding the assessor’s instructions. The assessor would sometimes provide a prompt (e.g., “Can you say anything more about that?”) up to two times if a participant stopped short of the two-minute mark for the first three speech elicitation activities. Despite the prompts, some participants’ speaking fell short of two minutes. Recordings longer than two minutes were not reduced. The recordings of twelve tasks (from ten subjects) were excluded because they were acoustically compromised and it was not possible to extract reliable information from them.

A linguist used the computer program VoiceSauce (Shue, 2010) to extract the phonetic linguistic parameter of pitch (F0). Pitch is only present when the vocal folds are vibrating, as they are for the articulation of vowels (and some consonants). The computer program WaveSurfer 1.8.8 (Medina and Solorio, 2006) was used to extract intensity readings. The pitch (F0) and intensity readings were taken every 10 milliseconds, and using those instances in which voicing (vocal fold vibration) was present, the following were calculated for each speaker, for each task. First, standard deviation of F0 (SD of F0) is variability in pitch (a larger number meaning a greater range of pitch during voicing). Second, variability in intensity/loudness was computed as an average of the average SDs of intensity changes over a 20-second window. For the third and fourth measures, the computer program Prosogram (Mertens, 2004) via Praat (Boersma and Weenink, 2015) was used to automatically delineate the vowels in the audiorecordings. The midpoint of the vowel resonances (F1 and F2) were extracted for every delineated vowel using a Praat script (Lennes, 2003), run on audiorecordings of female speakers with the setting of five formants expected in the first 5500 Hz, and on the audiorecordings of male speakers with the setting of five formants expected in the first 5000 Hz. The important resonances for vowels (F1 and F2) correspond to the shape of various parts of the vocal track during their articulation. Specifically, F1 indicates jaw/mouth opening and thus tongue height (tongue height goes

along with jaw opening; as the jaw drops, the tongue lowers), and F2 corresponds to tongue front/back position and/or lip rounding. For each speaker, for each task, we calculated SD of F1 and SD of F2. In order to identify datapoints that were outliers, all raw measurements were standardized for each speaker across all tasks, and datapoints that had a z-score >3.29 or <−3.29 were discarded. Outliers were generally of two kinds: spurious values from the automated measurements, and data points from the short interjections from the experimenter that were present in some tasks.

For F1 and F2, measurements were converted to Bark, a perceptual scale that is essentially linear at lower levels and logarithmic at higher levels, reflecting how human perception of Hz works (i.e., a greater increase in Hz is needed at higher levels to get the same perceptual increase). For intensity/loudness, we wrote a script that only used datapoints that were from definitely voiced segments ("definitely" found by not only an F0 (voicing) measurement for that datapoint but also for the ones before and after it). Specifically, we computed a moving average over a particular length of time, since neither very short nor very long variations are of interest (i.e., we were not interested in the difference between syllables in the same word, nor loudness changes over a period of minutes, which are probably due to the speaker getting farther away from the microphone or getting tired). Because this was not an *a priori* variable and was computed only after the conclusion of data collection, we could not be sure that the microphone/recorder was placed in the exact same position for participants recruited in Washington, D.C. (where several offices with different seating arrangements were used for assessments); variation in recorder placement could affect intensity/loudness. For this reason, only participants recruited from New York were included in the computation of this variable, as we were confident that the recorder was consistently placed in the same position at those sites.

2.6. Data Analyses

After examining distributional properties and descriptive statistics for all variables, as well as correlations among phonetic parameters, we compared patients with aprosody, patients without aprosody, and controls along a number of basic demographic characteristics. The three groups were then compared in terms of all the pitch, formant, and intensity/loudness variables across the five recorded tasks using the method of least squares to fit general linear models. Initial, unadjusted comparisons were followed by adjusted tests accounting for gender, race, ethnicity, marital status, living arrangement, employment status, and educational level. Post-hoc tests, with Bonferroni correction for multiple comparisons, were used to evaluate differences in the adjusted means across groups.

3. Results

Basic demographic characteristics of patients with aprosody (n=25), patients without aprosody (n=29), and controls (n=101) are shown in Table 3. Patients were more likely to be male, had a lower educational attainment, more commonly lived in a structured living arrangement or were homeless, and were less likely to be employed. Data on prescribed medications were available only for a subset of patients: 9 out of 25 (36.0%) with aprosody and 16 out of 29 (55.2%) without aprosody. The most prescribed antipsychotic in the first

group was risperidone, followed by olanzapine, clozapine, and quetiapine; in the group without aprosody, the three most prescribed antipsychotics were aripiprazole, olanzapine, and risperidone. The mean chlorpromazine-equivalent dosages in the two groups (321.7±164.2 mg and 374.9±284.3 mg) were not statistically different.

In advance of interpreting comparisons across the three groups, we examined correlations. As shown in Table 4, within each phonetic parameter, values were generally highly correlated across the five speech tasks. However, within each speech task, values across the four phonetic parameters were not highly correlated. This indicated that in subsequent analyses, although each phonetic parameter may quantify a different aspect of aprosody, values of each individual parameter across the five tasks tend to be highly correlated (i.e., within each phonetic parameter, like SD F0, there was little difference across the five tasks). We also explored correlations between chlorpromazine-equivalent dosage and the acoustic variables: (a) SD of F0 (range: −0.26 to −0.34, mean: −0.27); (b) SD of F1 (range: −0.08 to −0.27, mean: −0.20); (c) SD of F2 (range: −0.05 to −0.29, mean: −0.18); and (d) average SDs of intensity changes (range: −0.07 to 0.10, mean: 0.04).

Results from comparisons of pitch and formant variables across three groups (patients with aprosody, patients without aprosody, and controls) are shown in Table 5. Regarding pitch (F0) variation, in unadjusted tests, patients with aprosody differed from controls in Tasks 3 and 4, and for Task 5, the difference was statistically significant in both unadjusted (6.93 versus 8.28 in patients with aprosody and controls, respectively), and adjusted tests (6.84 versus 8.12, respectively, with patients without aprosody having a mean value of 7.59). In terms of SD of F1, patients *without* aprosody had lower values than controls in unadjusted tests across all tasks, though in adjusted tests, the expected pattern was observed in two of the tasks (e.g., Task 5: 0.93 versus 1.04, in patients with aprosody and controls, respectively, with patients without aprosody having a mean value of 0.96). Regarding SD of F2, patients with aprosody had lower values than controls in unadjusted tests across all tasks, though in adjusted tests the expected pattern was observed in two tasks (e.g., Task 3: 1.30 versus 1.44, in patients with aprosody and controls, respectively, with patients without aprosody having a mean value the same as controls, 1.44).

Results from comparisons of variation in intensity/loudness across the three groups are shown in Table 6, though the sample sizes are much smaller: eight patients with aprosody, seven patients without aprosody, and 40 controls. In unadjusted tests, patients with aprosody had the lowest values, and patients without aprosody had the highest values (with controls having intermediate values) across all five tasks. However, in adjusted tests, the expected pattern was more commonly observed (e.g., Task 5: scores across the three groups were: 4.10, 6.10, and 6.34, respectively). Figure 1 summarizes the statistically significant findings and respective η^2 effect sizes across the three groups, showing the largest effects for intensity/loudness.

4. Discussion

A priori, we would expect the assessment of “aprosody” in a patient to correlate specifically with speaker’s range of F0 (pitch), and perhaps also with intensity (loudness). We did indeed

find that patients with aprosody had a smaller F0 range than both patients without aprosody and healthy controls. However, there is another type of acoustic value that is impacted in patients: F2. We speculate that what has traditionally been termed “aprosody” may in fact be a collection of markedly different acoustic components of speech. The formant values F1 and F2 relate not to the melody of speech, but to the perceptual (and articulatory) differences between different vowel sounds. Further studies are needed to determine whether the lack of a difference in F1 between controls and patients with schizophrenia consistently fails to be present. F1 has a smaller range (than F2) to begin with, so it may be the case that compression along this dimension is harder to detect. The finding that patients with schizophrenia have a reduced range of F2 (a vowel differentiation measurement) as well of F0 and intensity strongly suggests that what is typically thought of as the lack of inflection has multiple distinct components. Further research is needed to determine whether some of these acoustic range compressions occurs more broadly in patients generally (as our study suggests intensity does) while other acoustic features are impacted only in patients with aprosody (as we found for F0 and F2).

Our approach is aligned with recent computerized analyses of patients’ speech, demonstrating that these new methods are promising as objective ways to evaluate negative symptoms. Cohen et al. (2008) examined the feasibility and validity of measuring flat affect, alogia, and anhedonia using widely available acoustic and lexical-analytic software. The computer-based inflection and speech-rate measures of natural speech samples—from 14 patients with clinically rated flat affect, 46 patients without flat affect, and 19 healthy controls—significantly discriminated flat-affect patients from controls, and the computer-based measure of alogia and negative emotion significantly discriminated the flat and non-flat patients in the same sample. Cohen and colleagues (2012) also employed computerized acoustic analysis of speech produced by 48 outpatients with schizophrenia and mood disorders, showing that speech characteristics were significantly associated with severity of psychosis and negative symptoms. Our results add to this emerging field of research—computationally derived acoustic differences between the groups suggest that quantifying aprosody through audiorecorded speech may be useful in developing tools for measuring and monitoring this specific “negative symptom.”

Although values of each individual parameter across the five tasks tend to be highly correlated, it appears that different types of prompts for obtaining audiorecorded speech may in fact produce some differences across phonetic parameters. For example, whereas loudness appeared to be blunted equally across all of our tasks, variation in both pitch (F0) and F1 were blunted most obviously in the reading tasks, and reduced variation in F2 was most apparent in the two spontaneous speech tasks. Future development of computational linguistic measures may need to rely on several means of eliciting speech to fully capture the various underpinnings of aprosody.

Despite the modest correlations that we observed between medication dosage and phonetic parameters—suggesting that a higher dosage of antipsychotic is associated with a reduced variability in the pitch, F1, and F2, but not in intensity/loudness—we were not able to adjust for medication dosage as doing so would have reduced the sample size in half. This

obviously represents a limitation when interpreting our results and future studies should fully consider the potential effects of medication dosage.

Our findings must be interpreted in light of several other methodological limitations. First, because we wanted to clearly define two subsamples of patients, those with questionable or mild aprosody were excluded from the analysis, resulting in a reduced sample size. Second, our computation of variation in intensity/loudness was possible only for a subset of the overall sample. Despite this, however, we found clear differences across groups, suggesting that variation in loudness (as opposed to pitch or vowel resonances) may be the phonetic parameter with the largest effect in detecting aprosody. Third, we did not measure "negative symptoms" (using the SANS and CAINS) among our controls, though aprosody is likely distributed along a continuum even in the general population. Fourth, our cross-sectional study could not investigate the longitudinal stability of findings; future studies should do so.

Findings suggest that computational methods can be employed to quantify very specific negative symptoms of schizophrenia. Similar work has shown that computational linguistic methods can differentiate patients with schizophrenia from unaffected controls (Martínez-Sánchez, 2015); our approach is to also dissect the linguistic underpinnings of specific symptoms; a specific negative symptom (aprosody) in this case. Such methods are likely applicable to diverse other psychiatric and neurological conditions. For example, quantitative features of vocal prosody in depressed participants have been shown to reveal changes in symptom severity over the course of depression (Yang et al., 2013), and automated detection of Parkinson's disease based on articulation, voice, and prosodic evaluations has also been demonstrated (Bocklet et al., 2011).

Acknowledgments

Role of the Funding Source

This work was supported by grant R21 MH097999-02 from the National Institute of Mental Health to the first author. The funding source had no role in data analyses, the writing of the manuscript, or the decision to submit it for publication.

Research reported in this publication was supported by National Institute of Mental Health grant R21 MH097999-02 ("Applying Computational Linguistics to Fundamental Components of Schizophrenia") to the first author. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or National Institute of Mental Health.

References

Alpert M, Kotsaftis A, Pouget ER. 1997; At issue: Speech fluency and schizophrenic negative signs. Schizophr. Bull. 23(2):171–177. [PubMed: 9165627]

Alpert M, Merewether F, Homel P, Marz J. 1986; Voxcom: A system for analyzing natural speech in real time. Behav. Res. Meth. Instruments Computers. 18(2):267–272.

Alpert M, Rosenberg SD, Pouget ER, Shaw RJ. 2000; Prosody and lexical accuracy in flat affect schizophrenia. Psychiatry Res. 97(2):107–118. [PubMed: 11166083]

Alpert M, Shaw RJ, Pouget ER, Lim KO. 2002; A comparison of clinical ratings with vocal acoustic measures of flat affect and alogia. J. Psychiatr. Res. 36(5):347–353. [PubMed: 12127603]

Alvino C, Kohler C, Barrett F, Gur RE, Gur RC, Verma R. 2007; Computerized measurement of facial expression of emotions in schizophrenia. J. Neurosci. Methods. 163(2):350–361. [PubMed: 17442398]

Andreasen, NC. The Scale for the Assessment of Negative Symptoms (SANS). The University of Iowa; Iowa City: 1983.

Andreasen NC, Alpert M, Martz MJ. 1981; Acoustic analysis: An objective measure of affective flattening. Arch. Gen. Psychiatry. 38(3):281–285. [PubMed: 7212958]

Andreasen NC, Pressler M, Nopoulos P, Miller D, Ho BC. 2010; Antipsychotic dose equivalents and dose-years: a standardized method for comparing exposure to different drugs. Biol. Psychiatry. 67(3):255–262. [PubMed: 19897178]

American Psychiatric Association (APA). Diagnostic and statistical manual of mental disorders. 5. American Psychiatric Publishing; Arlington, VA: 2013.

Bernardini F, Lunden A, Covington M, Broussard B, Halpern B, Alolayan Y, Crisafio A, Pauselli L, Balducci PM, Capulong L, Attademo L, Lucarini E, Salierno G, Natalicchi L, Quartesan R, Compton MT. 2016; Associations of acoustically measured tongue/jaw movements and portion of time speaking with negative symptom severity in patients with schizophrenia in Italy and the United States. Psychiatry Res. 239:253–258. [PubMed: 27039009]

Blanchard JJ, Kring AM, Horan WP, Gur R. 2010; Toward the next generation of negative symptom assessments: The collaboration to advance negative symptom assessment in schizophrenia. Schizophr. Bull. 7:291–299.

Bleuler, E. Dementia Praecox or the Group of Schizophrenias. Zinkin, J, translatorInternational Universities Press; New York: 1911.

Bocklet T, Nöth E, Stemmer G, Ruzickova H, Rusz J. 2011Detection of persons with Parkinson's disease by acoustic, vocal, and prosodic analysis. 2011 IEEE (Institute of Electrical and Electronics Engineers) Workshop on Automatic Speech Recognition and Understanding. :478–483.

Boersma, P; Weenink, D. [retrieved 4 April 2017] Praat: doing phonetics by computer [Computer program]. Version 5.4.22. 2015. from http://www.praat.org

Buchanan RW. 2007; Persistent negative symptoms in schizophrenia: An overview. Schizophr. Bull. 33(4):1013–1022. [PubMed: 17099070]

Cockrell JR, Folstein MF. 1987; Mini-Mental Status Examination (MMSE). Psychopharmacol. Bull. 24(4):689–692.

Cohen AS, Alpert M, Nienow TM, Dinzeo TJ, Docherty NM. 2008; Computerized measurement of negative symptoms in schizophrenia. J. Psychiatr. Res. 42(10):827–836. [PubMed: 17920078]

Cohen AS, Najolia GM, Kim Y, Dinzeo TJ. 2012; On the boundaries of blunt affect/alogia across severe mental illness: Implications for Research Domain Criteria. Schizophr. Res. 140(1):41–45. [PubMed: 22831770]

Covington MA, Lunden SL, Cristofaro SL, Ramsay Wan C, Bailey CT, Broussard B, Fogarty R, Johnson S, Zhang S, Compton MT. 2012; Phonetic measures of reduced tongue movement correlate with negative symptom severity in hospitalized patients with first-episode schizophrenia-spectrum disorders. Schizophr. Res. 142(1):93–95. [PubMed: 23102940]

Deserno L, Heinz A, Schlagenhauf F. 2017; Computational approaches to schizophrenia: A perspective on negative symptoms. Schizophr. Res. 186:46–54. [PubMed: 27986430]

First, MB, Spitzer, RL, Gibbon, M, Williams, JBW. Structured Clinical Interview for DSM-IV Axis I Disorders. Biometrics Research Department, New York State Psychiatric Institute; New York, NY: 1994.

Folstein MF, Folstein SE, McHugh PR. 1975; "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12(3):189–198. [PubMed: 1202204]

Forbes C, Blanchard JJ, Bennett M, Horan WP, Kring A, Gur R. 2010; Initial development and preliminary validation of a new negative symptom measure: The Clinical Assessment Interview for Negative Symptoms. Schizophr. Res. 124(1):36–42. [PubMed: 20869848]

Hamm J, Kohler CG, Gur RC, Verma R. 2011; Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders. J. Neurosci. Methods. 200(2):237–256. [PubMed: 21741407]

Jastak, S, Wilkinson, G. The Wide Range Achievement Test: Manual of Instructions. Jastak Associates; Wilmington, DE: 1984.

Kraepelin, E. Dementia Praecox and Paraphrenia. Huntington; New York: 1919.

Kring AM, Alpert M, Neale JM, Harvey PD. 1994; A multimethod, multichannel assessment of affective flattening in schizophrenia. Psychiatry Res. 54(2):211–222. [PubMed: 7761554]

Kupper Z, Ramseyer F, Hoffmann H, Kalbermatten S, Tschacher W. 2010; Video-based quantification of body movement during social interaction indicates the severity of negative symptoms in patients with schizophrenia. Schizophr. Res. 121:90–100. [PubMed: 20434313]

Lennes, M. Praat script. Modified by Dan McCloy, December 2011. 2003. Downloaded 14 January 2016. https://depts.washington.edu/phonlab/resources/getDurationPitchFormants.praat

Leucht S, Samara M, Heres S, Patel MX, Woods SW, Davis JM. 2014; Dose equivalents for second-generation antipsychotics: the minimum effective dose method. Schizophr. Bull. 40(2):314–326. [PubMed: 24493852]

Martínez-Sánchez F, Muela-Martínez JA, Cortés-Soto P, García Meilán JJ, Vera Ferrándiz JA, Egea Caparrós A, Pujante Valverde IM. 2015; Can the acoustic analysis of expressive prosody discriminate schizophrenia? Spanish J. Psychol. 18:e86.

Medina, E; Solorio, T. Wavesurfer: A tool for sound analysis. Departmental Technical Reports (CS). 2006. Paper 210. http://digitalcommons.utep.edu/cgi/viewcontent.cgi?article=1209&context=cs_techrep

Mertens, P. The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model; Speech Prosody 2004, International Conference; 2004.

Murphy BP, Chung YC, Park TW, McGorry PD. 2006; Pharmacological treatment of primary negative symptoms in schizophrenia: A systematic review. Schizophr. Res. 88(1):5–25. [PubMed: 16930948]

Püschel J, Stassen HH, Bomben G, Scharfetter C, Hell D. 1998; Speaking behavior and speech sound characteristics in acute schizophrenia. J. Psychiatr. Res. 32(2):89–97. [PubMed: 9694004]

Rapcan V, D'Arcy S, Yeap S, Afzal N, Thakore J, Reilly RB. 2010; Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia. Med. Engineering Physics. 32(9):1074–1079.

Shue, Y-L. Doctoral dissertation. University of California Los Angeles; 2010. The voice source in speech production: Data, analysis and models. http://www.seas.ucla.edu/spapl/voicesauce

Stassen HH, Albers M, Püschel J, Scharfetter CH, Tewesmeier M, Woggon B. 1995; Speaking behavior and voice sound characteristics associated with negative schizophrenia. J. Psychiatr. Res. 29(4):277–296. [PubMed: 8847655]

Wang P, Barrett F, Martin E, Milonova M, Gur RE, Gur RC, Kohler C, Verma R. 2008; Automated video-based facial expression analysis of neuropsychiatric disorders. J. Neurosci. Methods. 168(1):224–238. [PubMed: 18045693]

Yang Y, Fairbairn C, Cohn JF. 2013; Detecting depression severity from vocal prosody. IEEE Transactions on Affective Computing. 4(2):142–150. [PubMed: 26985326]

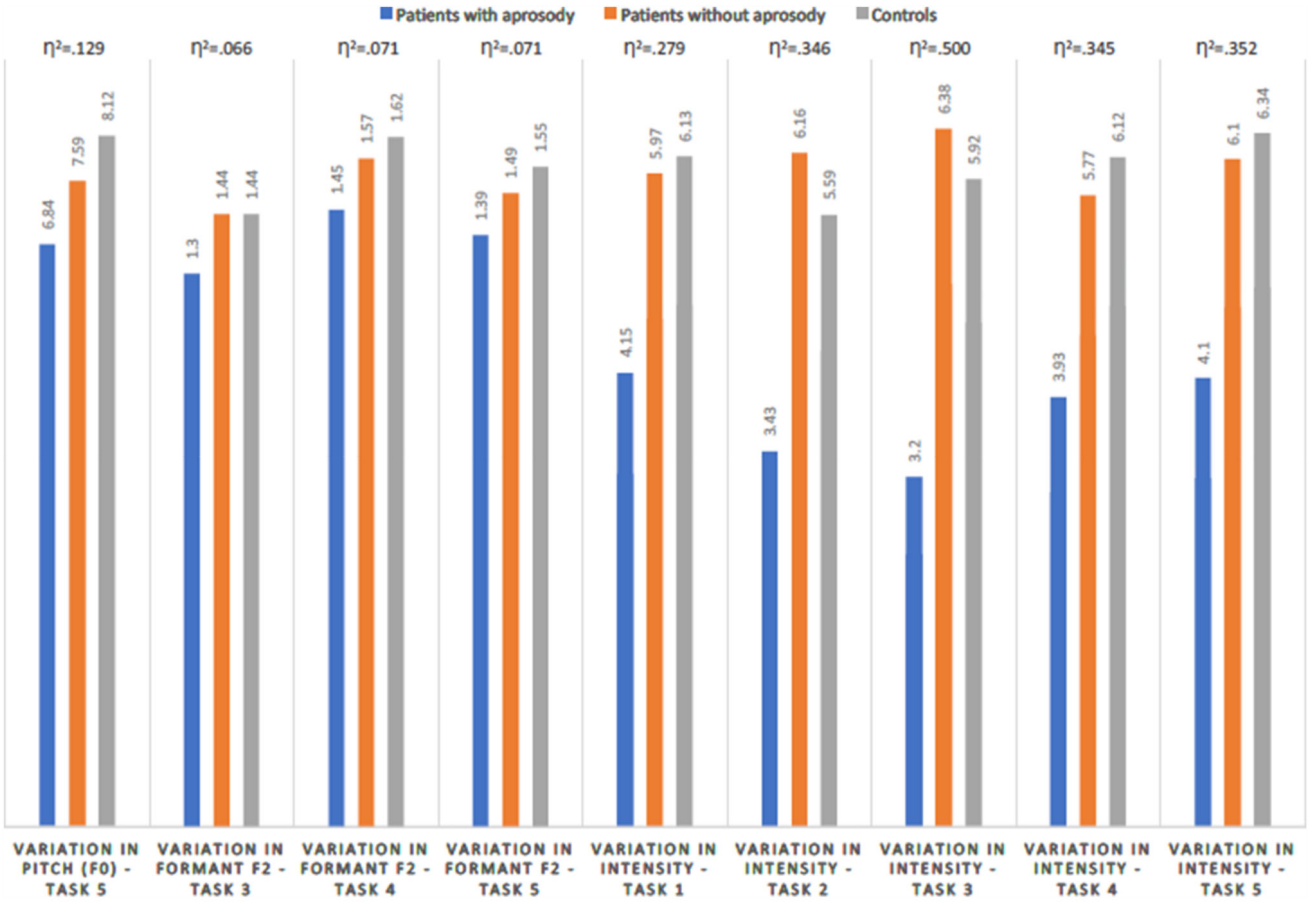

Figure 1.

Table 1

The Two Researcher-Rated Items Used to Derive Subsamples of Patients with and without Aprosody

SANS "Lack of Vocal Inflections" Item. *While speaking the subject fails to show normal vocal emphasis patterns. Speech has a monotonic quality, and important words are not emphasized through changes in pitch or volume. Subject also may fail to change volume with changes of subject so that he does not drop his voice when discussing private topics nor raise it as he discusses things which are exciting or for which louder speech might be appropriate*. Ratings: Not all all: Normal vocal inflections = 0 Questionable decrease = 1 Mild: Slight decrease in vocal inflections = 2 Moderate: Interviewer notices several instances of flattened vocal inflections = 3 Marked: Obvious decrease in vocal inflections = 4 Severe: Subject's speech is a continuous monotone = 5
CAINS "Vocal Expression" Item. *This item refers to prosodic features of the voice. This item reflects changes in tone during the course of speech. Speech rate, amount, or content of speech is not assessed*. Ratings: No impairment: *Within normal limits*. Normal variation in vocal intonation across interview. Speech is expressive and animated = 0 Mild deficit: *Mild decrease* in vocal intonation. Variation in intonation occurs with a limited intonation during a few parts of the interview = 1 Moderate deficit: *Notable decrease* in vocal intonation. Diminished intonation during several parts of the interview. Much of speech is lacking variability in intonation but prosodic changes occur in several parts of the interview = 2 Moderately severe deficit: *Significant lack* of vocal intonation with only a few changes in intonation throughout most of the interview. Most of speech is flat and lacking variability, only isolated instance of prosodic change = 3 Severe deficit: *Nearly total lack of* change in vocal intonation with characteristic flat or monotone speech throughout the interview = 4

Table 2

The Five Audiorecordings of Speech: Three Elicitation Tasks for Spontaneous Speech and Two for Reading Aloud

Task 1	Participants were shown a line drawing and asked to describe the picture with as much detail as possible. They were allowed two minutes to speak.
Task 2	Participants were asked to speak for two minutes (after up to 30 seconds of thinking and planning a response) in relation to, “Please describe what a perfect, most ideal day would be like for you.”
Task 3	Participants were asked to speak for two minutes (again, after up to 30 seconds of thinking/planning) in response to, “Please tell me about the scariest, most frightening experience you’ve ever had.”
Task 4	We obtained a recording of reading aloud—among those with a WRAT reading grade equivalent of ≥8 (n=141 (70.9%); specifically, 65 (66.3%) patients and 76 (75.2%) controls)—using the same neutral text passage as Stassen et al. (1995) and Rapcan et al. (2010), who provided us with their exact excerpt. This required approximately two minutes to read.
Task 5	Again, in those with a WRAT reading grade equivalent of ≥8, we obtained another recording of reading aloud, using the same emotionally stimulating text passage as Stassen et al. (1995) and Rapcan et al. (2010), who, again, provided their excerpt (Task 5). This also required about two minutes to read.

Table 3

Basic Demographic Characteristics of Patients with Aprosody (n=25), without Aprosody (n=29) and Controls (n=101)

	Patients with Aprosody	Patients without Aprosody	Controls	Test statistic, df, p
Age, in years	29.9±9.6	34.3±9.2	33.7±9.3	F=1.88; df=2; p=0.16
Gender, male	17 (68.0%)	23 (79.3%)	56 (55.4%)	χ^2=5.91; df=2; p=0.05
Ethnicity, non-Hispanic (n=154)	25 (100%)	27 (93.1%)	89 (88.1%)	χ^2=3.64; df=2; p=0.16
Race				χ^2=4.18; df=4; p=0.39
Black or African American	20 (80.0%)	24 (82.8%)	72 (71.3%)	
White or Caucasian	1(4.0%)	3 (10.3%)	17 (16.8%)	
Other	4 (16.0%)	2 (6.9%)	12 (11.9%)	
Marital Status (n=154)				χ^2=1.13; df=4; p=0.89
Single and never married	21 (87.5%)	26 (89.7%)	87 (86.1%)	
Married or living with a partner	2 (8.3%)	1 (3.4%)	9 (8.9%)	
Separated, divorced, or widowed	1 (4.2%)	2 (6.9%)	5 (5.0%)	
Years of education completed	12.4±2.5	12.5±2.7	14.0±2.5	F=6.75; df=2; p<0.01
Who the participant lived with, past month				χ^2=19.22; df=8; p=0.01
Alone	2 (8.0%)	12 (41.4%)	26 (25.7%)	
With parent, sibling or other family	13 (52.0%)	8 (27.6%)	44 (43.6%)	
With boyfriend/girlfriend or spouse/partner	2 (8.0%)	2 (8.0%)	13 (12.9%)	
With friends or roommates	1 (4.0%)	3 (10.3%)	12 (11.9%)	
Structured living arrangement or homeless	7 (28.0%)	4 (13.8%)	6 (5.9%)	
Has children (n=154)	9 (37.5%)	8 (27.6%)	48 (47.5%)	χ^2=3.93; df=2; p=0.14
Had a job during the past month	6 (24.0%)	12 (41.4%)	60 (59.4%)	χ^2=11.19; df=2; p<0.01
Chlorpromazine-equivalent dose, mg	321.7±164.2	374.9±284.3		t=0.51; df=23; p=.613

Table 4

Correlations among Variables

	Patients with aprosody	Patients without aprosody	Controls
Correlations within Individual Phonetic Parameter Values across the Five Speech Tasks, Each Cell Shows the Range (Mean) of 10 Correlations			
Variation in Pitch (F0)	.64–.86. (.74)	.75–.96 (.86)	.77–.94 (.82)
Variation in Formant F1	.77–.92 (.86)	.80–.94 (.87)	.83–.96 (.88)
Variation in Formant F2	.45–.86 (.63)	.70–.93 (.79)	.68–.92 (.75)
Variation in Intensity/Loudness	.36–.94 (.61)	.66–.99 (.83)	.53–.95 (.73)
Correlations among the Four Phonetic Parameters within the Five Speech Tasks, Each Cell Shows the Range (Mean) of 6 Correlations			
Task 1	−.08–.45 (.14)	.07–.80 (.32)	−.06–.33 (.14)
Task 2	−.51–.38 (−.09)	.18–.43 (.31)	.05–.38 (.28)
Task 3	−.57–.31 (−.02)	−.14–.27 (.11)	.15–.50 (.37)
Task 4	−.67–.13 (−.09)	.03–.75 (.27)	.08–.35 (.24)
Task 5	−.53–.19 (−.01)	.11–.73 (.31)	.07–.45 (.25)

Table 5

Aprosody-Related Acoustic Phonetic (Pitch and Formant) Measures in Patients with Aprosody (n=25), Patients without Aprosody (n=29), and Controls (n=101)

	Unadjusted/Crude Mean-by-Group Tests					*Adjusted Mean-by-Group Tests*[*]				
	A. Patients with aprosody	**B. Patients without aprosody**	**C. Controls**	**Sig.**	**Post-Hoc Test**	**A. Patients with aprosody**	**B. Patients without aprosody**	**C. Controls**	**Sig.**	**Post-Hoc Test**
Variation in Pitch (F0) by Task										
Task 1	6.77	6.73	7.24	n.s.		6.95	7.33	7.00	<.001	
Task 2	6.72	6.08	6.53	n.s.		6.92	6.73	6.29	<.001	
Task 3	5.90	6.06	6.84	n.s.		6.19	6.67	6.58	<.001	
Task 4	6.65	6.63	7.67	n.s.		6.63	7.30	7.50	<.001	
Task 5	6.93	6.91	8.28	<.05	a<c	6.84	7.59	8.12	<.001	A<C
Variation in Formant F1 by Task										
Task 1	0.91	0.84	1.02	<.05	B<C	0.95	0.88	1.01	<.01	
Task 2	0.99	0.89	1.03	n.s.		1.02	0.93	1.01	<.001	
Task 3	0.97	0.89	1.00	n.s.		1.03	0.93	0.98	<.001	
Task 4	0.85	0.84	0.97	<.05	a<c, b<c	0.86	0.89	0.96	<.001	
Task 5	0.92	0.90	1.05	<.05	b<c	0.93	0.96	1.04	<.001	
Variation in Formant F2 by Task										
Task 1	1.42	1.43	1.52	n.s.		1.44	1.45	1.51	n.s.	
Task 2	1.37	1.48	1.48	n.s.		1.38	1.50	1.47	<.05	
Task 3	1.28	1.43	1.45	<.01	A<B, A<C	1.30’	1.44	1.44	n.s.	a<b, A<C
Task 4	1.46	1.57	1.62	<.05	A<C	1.45	1.57	1.62	n.s.	A<C
Task 5	1.40	1.49	1.54	<.05	A<C	1.39	1.49	1.55	n.s.	A<C

[*]Adjusted for gender, race, ethnicity, marital status, living with, employment status, and education. To control for multiple comparisons Bonferroni adjustment was used for post hoc tests.

Sig. refers to the global F for the model. In adjusted analyses, demographic differences may drive overall model significance rather than linguistic variables.
Post-Hoc test:
Uppercase (A, B, C) indicates significant differences between means at p < .05.
Lowercase (a, b, c) indicates significant differences between means at p < .10 (but >.05).

Table 6

Variation in Intensity/Loudness across the Three Groups (n=55)

	Unadjusted/Crude Mean-by-Group Tests					*Adjusted Mean-by-Group Tests*[*]				
	A. Patients with aprosody (n=8)	**B. Patients without aprosody (n=7)**	**C. Controls (n=40)**	**Sig.**	**Post-Hoc Test**	**A. Patients with aprosody**	**B. Patients without aprosody**	**C. Controls**	**Sig.**	**Post-Hoc Test**
Task 1	4.15	6.39	6.06	<.001	A<B, A<C	4.15	5.97	6.13	<.01	A<B, A<C
Task 2	3.36	6.25	5.59	<.001	A<B, A<C	3.43	6.16	5.59	<.01	A<B, A<C
Task 3	3.19	6.47	5.91	<.001	A<B, A<C	3.20	6.38	5.92	<.001	A<B, A<C
Task 4	4.08	6.17	6.00	<.001	A<B, A<C	3.93	5.77	6.12	<.05	A<B, A<C
Task 5	4.23	6.53	6.22	<.001	A<B, A<C	4.10	6.10	6.34	<.05	A<C, A<B

[*]Adjusted for gender, race, ethnicity, marital status, living with, employment status, and education.

Post-Hoc test:
Uppercase (A, B, C) indicates significant differences between means at $p < .05$