Schizophrenia Research xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Schizophrenia Research

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

The aprosody of schizophrenia: Computationally derived acoustic phonetic underpinnings of monotone speech

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

[a] *Columbia University College of Physicians & Surgeons, Department of Psychiatry, New York, NY, USA*
[b] *College of William and Mary, Department of English, Linguistics Program, Williamsburg, VA, USA*
[c] *The George Washington University Milken Institute School of Public Health, Department of Epidemiology and Biostatistics, Washington, DC, USA*
[d] *Case Western Reserve University, Department of Neurology, Cleveland, OH, USA*
[e] *Lenox Hill Hospital, New York, NY, USA*
[f] *The George Washington University School of Medicine and Health Sciences, Washington, DC, USA*
[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, GA, USA*

ARTICLE INFO

Article history:
Received 21 April 2017
Received in revised form 11 January 2018
Accepted 14 January 2018
Available online xxxx

Keywords:
Acoustic resonance
Aprosody
Linguistics
Negative symptoms
Phonetics
Phonology
Psychosis
Schizophrenia

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.
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, patients with aprosody differed significantly from controls in Task 5 in both unadjusted tests and those adjusted for sociodemographics. For the standard deviation (SD) of F1, no significant differences were found in adjusted tests. Regarding SD of F2, patients with aprosody had lower values than controls in Task 3, 4, and 5. For variation in intensity/loudness, patients with aprosody had lower values than patients without aprosody and controls across the five 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.

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.

* Corresponding author at: New York State Psychiatric Institute, 1051 Riverside Drive, Unit 100, New York, NY 10032, USA.
E-mail address: mtc2176@cumc.columbia.edu (M.T. Compton).

https://doi.org/10.1016/j.schres.2018.01.007
0920-9964/

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 et al. (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 et al. (1986), using his Voxcom technology, a computer-driven program that separates the voice signal into two channels, amplitude/loudness and frequency/pitch (subjects 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., 2016), 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; $t = 0.26$, $df = 200$, $p = .79$), gender (72.4% male v. 63.5%; $\chi^2 = 1.86$, $df = 1$, $p = .17$), or race (81.3% African American v. 76.0%; $\chi^2 = 1.74$, $df = 2$, $p = .42$).

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 ($n = 2$).

Please cite this article as: Compton, M.T., et al., The aprosody of schizophrenia: Computationally derived acoustic phonetic underpinnings of monotone speech, Schizophr. Res. (2018), https://doi.org/10.1016/j.schres.2018.01.007