Psychiatry Research 239 (2016) 253–258

Contents lists available at ScienceDirect

Psychiatry Research

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

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

Francesco Bernardini [a,b,*], Anya Lunden [c], Michael Covington [d], Beth Broussard [b], Brooke Halpern [b], Yazeed Alolayan [e], Anthony Crisafio [e], Luca Pauselli [a,b], Pierfrancesco M. Balducci [a,b], Leslie Capulong [b], Luigi Attademo [a], Emanuela Lucarini [a], Gianfranco Salierno [f], Luca Natalicchi [g], Roberto Quartesan [a,h], Michael T. Compton [b,i]

[a] School of Psychiatry, University of Perugia, Perugia, Italy
[b] Lenox Hill Hospital, Department of Psychiatry, New York, NY, United States
[c] College of William and Mary, Williamsburg, VA, United States
[d] The University of Georgia, Athens, GA, United States
[e] The George Washington University, Washington, DC, United States
[f] Community Mental Health Service, Magione, Italy
[g] Community Mental Health Service, Ponte San Giovanni, Perugia, Italy
[h] Division of Psychiatry, Department of Medicine, University of Perugia, Perugia, Italy
[i] Hofstra Northwell School of Medicine, Department of Psychiatry, Hempstead, NY, United States

ARTICLE INFO

Article history:
Received 6 July 2015
Received in revised form
15 March 2016
Accepted 23 March 2016
Available online 25 March 2016

Keywords:
Phonetics
Psycholinguistics
Psychotic disorders
Speech acoustics

ABSTRACT

This is the first cross-language study of the effect of schizophrenia on speech as measured by analyzing phonetic parameters with sound spectrography. We hypothesized that reduced variability in pitch and formants would be correlated with negative symptom severity in two samples of patients with schizophrenia, one from Italy, and one from the United States. Audio recordings of spontaneous speech were available from 40 patients. From each speech sample, a file of F0 (pitch) and formant values (F1 and F2, resonance bands indicating the moment-by-moment shape of the oral cavity), and the portion of the recording in which there was speaking ("fraction voiced," FV), was created. Correlations between variability in the phonetic indices and negative symptom severity were tested and further examined using regression analyses. Meaningful negative correlations between Scale for the Assessment of Negative Symptoms (SANS) total score and standard deviation (SD) of F2, as well as variability in pitch (SD F0) were observed in the Italian sample. We also found meaningful associations of SANS affective flattening and SANS alogia with SD F0, and of SANS avolition/apathy and SD F2 in the Italian sample. In both samples, FV was meaningfully correlated with SANS total score, avolition/apathy, and anhedonia/asociality.

1. Introduction

The negative symptoms of schizophrenia are very impairing to patients' lives and functioning and are a focus of considerable empirical and clinical attention. Measures of negative symptom severity are rarely used in clinical settings, and research-based measures of negative symptom severity rely on ratings by trained raters after in-depth interviews. A crucial step towards the comprehension and treatment of these symptoms involves the development of valid and reliable tools to efficiently measure them. Computerized acoustic analysis of patients' speech is a sophisticated and promising approach for measuring two cardinal negative symptoms of schizophrenia: blunted affect (which involves a diminished expression of emotion characterized in part by less variation of pitch and amplitude when speaking), and alogia (which involves a reduction in the quantity of speech and, by extension, impairment in producing it) (Cohen et al., 2008).

Many language abnormalities have been observed in schizophrenia patients, in syntactic, semantic, pragmatic, and phonetic domains of linguistics (Covington et al., 2005). Concerning the latter, the most often reported are flattened intonation or aprosody

* Correspondence to: Scuola di Specializzazione in Psichiatria, Università degli Studi di Perugia, Piazzale Lucio Severi, Edificio Ellisse, 8 piano, Sant'Andrea delle Fratte, 06132 Perugia PG, Italy.
E-mail address: francescobernardini78@yahoo.fr (F. Bernardini).

http://dx.doi.org/10.1016/j.psychres.2016.03.037
0165-1781/

(a kind of speech that is soft, emotionless, and lacking normal variation in speed, tone, and emphasis) (Spoerri, 1966; Cutting, 1985; Alpert et al., 1989; Rieber and Vetter, 1994); voice quality abnormalities, such as strained or choking voice (Spoerri, 1961, 1964, 1966); and pitch abnormalities (Stein, 1993). Although automated acoustic analysis of speech features has existed for decades, its application to research on serious mental illnesses has been modest (Cohen and Elvevåg, 2014). A study by Kring et al. (1994), conducted on a sample of medication-free patients with schizophrenia, showed a correlation between affective flattening and the acoustic index of vocal expressiveness. A study by Cohen et al. (2012) employed computerized acoustic analysis of speech produced by 48 outpatients with schizophrenia and mood disorders, showing that speech characteristics (average pause length, number of utterances, inflection, and emphasis) were significantly associated with severity of psychosis and negative symptoms. Furthermore, speech characteristics were shown to be for the most part highly stable over time.

We conducted a cross-language study of the effect of schizophrenia on speech, as measured by acoustic resonance frequencies (formants) that indicate the position and movement of the tongue, jaw, and lips when speaking. Specifically, our analysis relied on phonetic parameters assessed with sound spectrography. Many of the structural movements involved in speech production can be measured phonetically through spectral decomposition of sound waves into formants; i.e., a sound's component frequencies. Formants are the acoustic realization of vowels (as well as of some consonants), and the relative position of the first and second formants (F1, F2) allow humans to distinguish between different vowels in Italian and in English. We extracted the linguistic acoustic phonetic parameters of pitch (F0) and vowel resonances (jaw lowering/tongue height, F1; and tongue front/back position and/or lip rounding, F2) from the audio recordings of schizophrenia patients' speech.

We hypothesized that variability in formants would be negatively correlated with negative symptom severity in two samples of patients with schizophrenia, one from Italy, and one from the United States. Although primarily an exploratory study given the relatively limited literature on correlations between computer-assessed phonetic measures and negative symptom severity (and virtually no literature using the particular methodology that we employed), we had three *a priori* hypotheses based on a previous, similar study in the U.S. (Covington et al., 2012). For each hypothesis, we deemed, according to Cohen (1988), a potentially clinically meaningful effect size to be $r \geq 0.30$ or $r \leq -0.30$ (i.e., greater than a small effect). The magnitude of the correlation is of more interest than the p-value because this is an exploratory study with a small sample. We hypothesized that: (1) variability of F1 would be negatively correlated with negative symptom severity, (2) variability of F2 would be negatively correlated with negative symptom severity, and (3) correlations between phonetic indices and negative symptom severity would be driven by two of the Scale for the Assessment of Negative Symptoms (SANS; Andreasen, 1983) subscales: affective flattening/blunting, and alogia. We also hypothesized that these correlations would be similar among two samples of schizophrenia patients: one Italian speaking and one English speaking. Understanding any differences in associations of pitch and formant variability and negative symptom severity between Italian-speaking patients and English-speaking patients would be informative as there are differences in the phonology of Italian and English (e.g., differences in the size of the vowel inventory, differences in intonation contours, much less use of central vowels in Italian compared to English). Although not a phonetic measure per se, we also examined the extent of pauses within each speech recording by computing a variable termed "fraction voiced" (FV), which indirectly approximates the proportion of the file comprised of speaking as opposed to pauses.

2. Methods

2.1. Setting and sample

Data from the 20 Italian patients ($n=13$ males; $n=12$ inpatients) were collected in the Inpatient Unit of the Santa Maria della Misericordia Hospital in Perugia (Italy), and in two Community Mental Health Services for outpatients, one in Perugia and one in Magione (Italy). Schizophrenia patients who were native Italian speakers, without known or suspected mental retardation, and aged 18–60 years of age, were eligible.

Data from the 20 U.S. patients (again, $n=13$ males; $n=12$ inpatients) were collected, as part of a larger study, in four inpatient units (one in New York City, one in Long Island, and two in Washington, DC) and in four outpatient services (one in New York City, one in Long Island, and two in Washington, DC). Schizophrenia patients who were native English speakers, again without known or suspected mental retardation, and in this sample, ages 18–50 years, were eligible. Italian and U.S. patients were matched for gender and inpatient/outpatient status, and were matched for age by decade (e.g., 20–30, 30–40).

2.2. Data acquisition

Data from the 20 Italian patients were collected between August 2014 and February 2015. All procedures were approved by the Ethical Committee of the Umbria Region. Data from the 20 U.S. patients were collected between February 2013 and February 2015. All procedures were approved by the Institutional Review Boards of the George Washington University (Washington, DC) and the North Shore–LIJ Health System (New York). The research assessment began only after the patient gave written informed consent to participate. The diagnosis of schizophrenia was confirmed using the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I; First et al., 1998).

Symptom severity was rated, blinded to phonetic analysis findings that were not obtained until later in the project period, with the SANS, and, secondarily, with the Positive and Negative Syndrome Scale (PANSS; Kay et al., 1987), which also provided a measure of positive and general psychopathology symptoms. Assessors received intensive training on the PANSS and SANS before the beginning of patient recruitment. To assess inter-rater reliability, intraclass correlation coefficients (ICCs) were calculated using a two-way mixed (judges fixed) effects analysis of variance model (Shrout and Fleiss, 1979); four assessors (three from the U.S. sites and one from the Italian sites) were the fixed effect while 12 target ratings (12 patients interviewed in the U.S.) were the random effect. The ICC for the PANSS negative symptom subscale was 0.92 (95% CI: 0.80–0.97), and the ICC for the SANS total score was 0.97 (95% CI: 0.92–0.99).

Basic cognitive functioning was assessed with the Mini-Mental State Examination (MMSE; Cockrell and Folstein, 1988), to ensure that all participants had a MMSE score of >23 (suggesting that cognition was intact enough to particpate in the study).

Audio recordings of the patients' voice were taken during three elicitation activities for spontaneous speech, as follows. First, subjects were shown a line drawing and asked to describe the picture with as much detail as possible. They were allowed two minutes to speak, with at most one prompt from the interviewer. Second, they were asked to speak for two minutes (after up to 30 s of thinking about their intended answer) in response to, "Please describe what a perfect, most ideal day would be like for you." Third, they were asked to speak for two minutes (again, after up to