



Published in final edited form as:

Schizophr Res. 2012 December ; 142(1-3): 93–95. doi:10.1016/j.schres.2012.10.005.

Phonetic Measures of Reduced Tongue Movement Correlate with Negative Symptom Severity in Hospitalized Patients with First-Episode Schizophrenia-Spectrum Disorders

Michael A. Covington^a, S.L. Anya Lunden^b, Sarah L. Cristofaro^c, Claire Ramsay Wan^c, C. Thomas Bailey^a, Beth Broussard^d, Robert Fogarty^a, Stephanie Johnson^c, Shayi Zhang^a, and Michael T. Compton^{d,*}

^aInstitute for Artificial Intelligence, The University of Georgia, Athens, Georgia, United States

^bProgram in Linguistics, The University of Georgia, Athens, Georgia, United States

^cDepartment of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia, United States

^dDepartment of Psychiatry and Behavioral Sciences, The George Washington University School of Medicine and Health Sciences, Washington, District of Columbia, United States

Abstract

Background—Aprosody, or flattened speech intonation, is a recognized negative symptom of schizophrenia, though it has rarely been studied from a linguistic/phonological perspective. To bring the latest advances in computational linguistics to the phenomenology of schizophrenia and related psychotic disorders, a clinical first-episode psychosis research team joined with a phonetics/computational linguistics team to conduct a preliminary, proof-of-concept study.

Methods—Video recordings from a semi-structured clinical research interview were available from 47 first-episode psychosis patients. Audio tracks of the video recordings were extracted, and after review of quality, 25 recordings were available for phonetic analysis. These files were de-noised and a trained phonologist extracted a 1-minute sample of each patient's speech. WaveSurfer 1.8.5 was used to create, from each speech sample, a file of formant values (F0, F1, F2, where F0 is the fundamental frequency and F1 and F2 are resonance bands indicating the moment-by-moment shape of the oral cavity). Variability in these phonetic indices was correlated with severity of Positive and Negative Syndrome Scale negative symptom scores using Pearson correlations.

© 2012 Elsevier B.V. All rights reserved.

*Corresponding author. Department of Psychiatry and Behavioral Sciences, The George Washington University School of Medicine and Health Sciences, 2120 L Street, N.W., Suite 600, Washington, D.C. 20037. Tel: 202-741-3554. Fax: 202-741-2891 / mcompton@mfa.gwu.edu (MT Compton).

Conflicts of Interest

The authors know of no conflicts of interest pertaining to this research or this article.

Contributors

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

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Results—A measure of variability of tongue front-to-back position—the standard deviation of F2—was statistically significantly correlated with the severity of negative symptoms ($r=-0.446$, $p=0.03$).

Conclusion—This study demonstrates a statistically significant and meaningful correlation between negative symptom severity and phonetically measured reductions in tongue movements during speech in a sample of first-episode patients just initiating treatment. Further studies of negative symptoms, applying computational linguistics methods, are warranted.

Keywords

Aprosody; First-episode psychosis; Linguistics; Phonetics; Schizophrenia

1. Introduction

Flattened intonation or aprosody—speech that is soft, emotionless, and lacking normal variations in speed, tone, and emphasis—is a negative symptom of schizophrenia (Spoerri, 1966; Cutting, 1985; Alpert et al., 1989; Covington et al., 2005). Reduced facial muscle activity is also associated with schizophrenia (Schneider et al., 1990; Trémeau et al., 2005), and such facial bradykinesia likely contributes to altered speech quality (Cannizarro et al., 2005). Using his VOXCOM technology from the mid-1980s, Alpert and colleagues (1997, 2000, 2002) correlated negative symptoms with vocal/acoustic parameters such as pauses, slowed rate, and aprosody. We conducted an analysis of deeper phonetic parameters (beyond rate and productivity), assessed with sound spectrography. Many of the movements involved in producing speech can be measured phonetically through spectral decomposition of sound waves into *formants* (resonance bands produced by shape of the oral cavity). Formants allow humans to recognize many speech sounds, especially vowels. The first two formants, F1 and F2, roughly indicate tongue height/mouth opening and tongue front-to-back position. Their absolute values depend on the sound being made and the individual's physical characteristics, but what interested us was their variability. We hypothesized that variability of F1 and F2 would be lower in patients with greater negative symptom severity. Because chronic antipsychotic exposure can induce “secondary” negative symptoms, a first-episode psychosis sample is particularly enlightening.

2. Method

2.1. Setting and sample

Data were collected as part of a larger study based at inpatient psychiatric units at a large, public-sector, urban hospital and a crisis stabilization unit in a neighboring suburban county. Both sites serve a predominately African American, low-income, socially disadvantaged population. English-speaking first-episode patients (without prior outpatient treatment for psychosis lasting >3 months, or prior hospitalization for psychosis >3 months before index hospitalization) aged 18–40 years were eligible.

2.2. Data acquisition

Data were collected between August 2008 and July 2010. All procedures were approved by the university's Institutional Review Board. The research assessment began after the patient was clinically stabilized, typically around hospital day 4. Virtually all were assessed within a week of initiating antipsychotic treatment. Diagnoses of psychotic disorders and substance use disorders were made using the *Structured Clinical Interview for DSM-IV Axis I Disorders* (SCID; First et al., 1998).

Symptom severity was rated, blinded to phonetic analysis findings, with the *Positive and Negative Syndrome Scale* (PANSS; Kay et al., 1987). Assessors received intensive training on the PANSS before and during the initial months of the study; thereafter, as a team they conducted roughly biweekly in-person patient interviews to allow for ongoing training, supervision, and monitoring of quality/consistency of PANSS interviews and rating. To assess inter-rater reliability, intraclass correlation coefficients (ICCs) were calculated using a two-way mixed (judges fixed) effects analysis of variance model; three assessors were the fixed effect while 12 target ratings were the random effect (Shrout and Fleiss, 1979). The ICC for the negative symptom subscale was 0.92 (95% CI: 0.86–0.96).

Video/audio recordings with a mean length of 31.5 ± 1.55 minutes—obtained for the purposes of rating movement abnormalities such as dyskinesias and stereotypies (Fantès et al., 2012)—were taken during the PANSS interview. Audio quality of many recordings was insufficient for this secondary phonetic analysis. Two linguists independently rated intelligibility and audio quality of 47 available audio files and selected the 25 highest quality files. Constant background noise was removed using *GoldWave 5.58*. A phonologist extracted a 1-minute sample of continuous speech by starting five minutes into the recording, extracting a segment while removing passages where the interviewer spoke, other noises were in the background, or the patient was merely saying “um-hm” or similar very short responses. Soft or unintelligible utterances were retained and silences were kept if they joined two utterances; unusually long silences (>7 seconds) were removed.

2.3. Data analysis

A computational linguist used *WaveSurfer 1.8.5* to create from each speech sample a file of formant values (F0, F1, F2, where F0 is the fundamental frequency or the pitch at which the vocal folds are vibrating, and F1 and F2 are resonance bands indicating moment-by-moment shape of the oral cavity). These values were sampled every 10 milliseconds, coding F0 as 0 if vocal folds were not vibrating based on spectral analysis. Using a computer program written locally in Microsoft C#, formant data were processed to compute summary measures for all 10-millisecond samples with nonzero F0. $Mean \log_{10}(F0)$, indicating average vocal pitch on a logarithmic scale (which is how humans perceive pitch in speech intonation), was not expected to correlate with negative symptoms. SD_N of $\log_{10}(F0)$ indicated pitch variability, $SD_N(F1)$ variability of tongue height or mouth opening, and $SD_N(F2)$ variability of tongue front-to-back position. These latter two measures were expected to inversely correlate with negative symptom severity. SD_N means “standard deviation for N rather than N–1” because the purpose was to summarize rather than estimate. Descriptive statistics, kurtosis/skewness, histograms, and Kolmogorov-Smirnov tests revealed that the four phonetic measures and PANSS negative symptom subscale score were approximately normally distributed. Correlations between phonetic measurements and PANSS negative symptoms were assessed with Pearson correlation coefficients using *IBM SPSS Statistics 19*.

3. Results

The 25 patients were 23.8 ± 4.4 years of age and had completed 11.8 ± 1.9 years of education. Nineteen (76%) were male, 22 (88%) were African American, 23 (92%) were single/never married, and 17 (68%) lived with family before hospitalization. Based on the SCID, 12 (48%) were diagnosed with schizophrenia, two with schizophreniform disorder, two with delusional disorder, two with schizoaffective disorder (depressive type), one with brief psychotic disorder, and six with psychotic disorder not otherwise specified. Among 23 completing the SCID substance use disorder module, seven (30%) had current alcohol abuse or dependence and eight (35%) had cannabis abuse or dependence. Eighteen (72%) were prescribed risperidone (mean dose, 4.1 ± 2.3 mg), with four also taking diphenhydramine

50mg and one benztropine 1mg. Among the remaining seven, two were prescribed ziprasidone 80mg, two paliperidone (3mg and 6mg), and one haloperidol 5mg; two refused medications.

Mean log₁₀(F0) was not significantly correlated with negative symptom severity ($r=-0.071$, $p=0.73$), as predicted. *SD_N of log₁₀(F0)* also was not significantly correlated with negative symptomatology ($r=-0.107$, $p=0.61$). There was a suggestion of a negative correlation between *SD_N(F1)* and negative symptoms ($r=-0.339$, $p=0.098$), which would be considered a medium effect size, though it did not reach statistical significance. *SD_N(F2)* was significantly and meaningfully—at nearly a “large” effect size of $r=0.50$, according to Cohen (1988)—associated with negative symptom severity ($r=-0.446$, $p=0.03$). Scatterplots for *SD_N(F1)* and *SD_N(F2)* are shown in Figure 1. As a check, *SD_N(F2-F1)*, sometimes used as an alternative measure of tongue front-back position, gave virtually identical results ($r=-0.437$, $p=0.03$).

Correlations between *SD_N(F2)* and the seven individual negative symptom items revealed four significant associations: blunted affect ($r=-.398$, $p=0.05$), emotional withdrawal ($r=-.423$, $p=0.04$), poor rapport ($r=-.404$, $p=0.05$), and lack of spontaneity and flow of conversation ($r=-.523$, $p=0.007$.) *SD_N(F2)* was also associated with the relevant general psychopathology item of motor retardation ($r=-0.488$, $p=0.01$), though it was not correlated with items pertaining to depression or the PANSS positive symptom subscale score. Among 18 patients taking risperidone, *SD_N(F2)* was not correlated with risperidone dosage ($r=-0.018$, $p=0.94$).

4. Discussion

Two aims of clinical speech signal processing are to use speech to differentiate cohorts (e.g., healthy controls versus those with Parkinson’s disease; Tsanas et al., 2012) and to monitor illness severity (e.g., Parkinson’s symptom severity; Tsanas et al., 2010). Might it be possible to use speech signal processing to monitor severity of negative symptoms in schizophrenia? We found a correlation between WaveSurfer-derived phonetic measures of reduced front-back tongue movement (and possibly tongue height movement) and interview-based negative symptom severity and motor retardation. Furthermore, front-back tongue movement was not associated with depressive or positive symptom severity, or antipsychotic dosage.

Alpert and colleagues (1997) noted that acoustic analyses reveal processes not apparent to the clinician that might prove useful for clinical assessment and research. While our small correlational study is a modest step, future assessment tools based on objective computational linguistics technology could augment clinical ratings. Successful phonetic screening for Parkinson’s disease and depression is being demonstrated (Mundt et al., 2007; Tsanas et al., 2010). For example, Tsanas and associates (2010) replicated Parkinson’s disease severity scores with clinically useful accuracy via straight-forward speech sampling collected with an at-home testing device for Internet-enabled measurement. Such an approach could someday be feasible for psychotic disorders.

Several methodological limitations are noteworthy. This was a small study using recordings of less-than-ideal quality based on speech that was not entirely spontaneous, but rather elicited in a semi-structured interview. Because we used only 25 files, with only a 1-minute sample of continuous speech being processed, larger studies should collect specialized continuous speech samples. Analyses may be easier if a standard text is read such that particular sounds are investigated. Another approach would be to focus on vowels (long used in speech signal analysis) from continuous speech (Sapir et al., 2010) or sustained vowels (Tsanas et al., 2010). Despite limitations, results suggest that a computational

linguistics approach to phonetic abnormalities in psychotic disorders could advance understandings of negative symptoms.

Acknowledgments

Role of the Funding Source

This work was supported by grant R01 MH081011 from the National Institute of Mental Health to the last author. The funding source had no role in data analyses, the writing of the manuscript, or the decision to submit it for publication.

The authors express gratitude to the participants involved in this study and the clinicians at the referral sites who identified eligible and interested patients. They also thank William A. Kretschmar, Jr. and Erin Ament for valuable suggestions.

References

- Alpert M, Kotsaftis A, Pouget ER. At issue: Speech fluency and schizophrenic negative signs. *Schizophr Bull.* 1997; 23:171–177. [PubMed: 9165627]
- Alpert M, Rosen A, Welkowitz J, Sobin C, Borod J. Vocal acoustic correlates of flat affect in schizophrenia: Similarity to Parkinson's disease and right hemisphere disease and contrast to depression. *Br J Psychiatry.* 1989; 154:51–56.
- Alpert M, Rosenberg SD, Pouget ER, Shaw RJ. Prosody and lexical accuracy in flat affect schizophrenia. *Psychiatry Res.* 2000; 97:107–118. [PubMed: 11166083]
- Alpert M, Shaw RJ, Pouget ER, Lim KO. A comparison of clinical ratings with vocal acoustic measures of flat affect and alogia. *J Psychiatr Res.* 2002; 36:347–353. [PubMed: 12127603]
- Cannizzarro MS, Cohen H, Rappard F, Snyder PJ. Bradyphrenia and bradykinesia both contribute to altered speech in schizophrenia: A quantitative acoustic study. *Cogn Behav Neurol.* 2005; 18:206–210. [PubMed: 16340393]
- Cohen, J. *Statistical Power Analysis for the Behavioral Sciences.* 2. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
- Covington MA, He C, Brown C, Naçi L, McClain JT, Fjordbak BS, Semple J, Brown J. Schizophrenia and the structure of language: The linguist's view. *Schizophr Res.* 2005; 77:85–98. [PubMed: 16005388]
- Cutting, J. *The psychology of schizophrenia.* Churchill Livingstone; Edinburgh: 1985.
- Fantes F, Ramsay Wan CE, Johnson S, Walker EF, Compton MT. Abnormal movements in first-episode nonaffective psychosis: Dyskinesias, stereotypies, and catatonic signs as distinct aspects of the clinical phenotype. 2012 (submitted).
- 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: 1998.
- Kay S, Fiszbein A, Opler L. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophr Bull.* 1987; 13:261–276. [PubMed: 3616518]
- Mundt JC, Snyder PJ, Cannizzarro MS, Chappie K, Geralt DS. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *J Neurolinguistics.* 2007; 20:50–64. [PubMed: 21253440]
- Sapir S, Ramig LO, Spielman JL, Fox C. Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech. *J Speech Lang Hearing Res.* 2010; 53:114–125.
- Schneider F, Heimann H, Himer W, Huss D, Mattes R, Adam B. Computer-based analysis of facial action in schizophrenic and depressed patients. *Eur Arch Psychiatry Clin Neurosci.* 1990; 240:67–76. [PubMed: 2149651]
- Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psycholog Bull.* 1979; 86:420–428.
- Spoerri T. Speaking voice of the schizophrenic patient. *Arch Gen Psychiatry.* 1966; 14:581–585. [PubMed: 5934868]

- Trémeau F, Malaspina D, Duval F, Corrêa H, Hager-Budny M, Coin-Bariou L, Macher JP, Gorman JM. Facial expressiveness in patients with schizophrenia compared to depressed patients and nonpatient comparison subjects. *Am J Psychiatry*. 2005; 162:92–101. [PubMed: 15625206]
- Tsanas A, Little MA, McSharry PE, Ramig LO. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Transactions on Biomed Engineering*. 2010; 57:884–893.
- Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Transactions on Biomed Engineering*. 2012; 59:1264–1271.

\$watermark-text

\$watermark-text

\$watermark-text

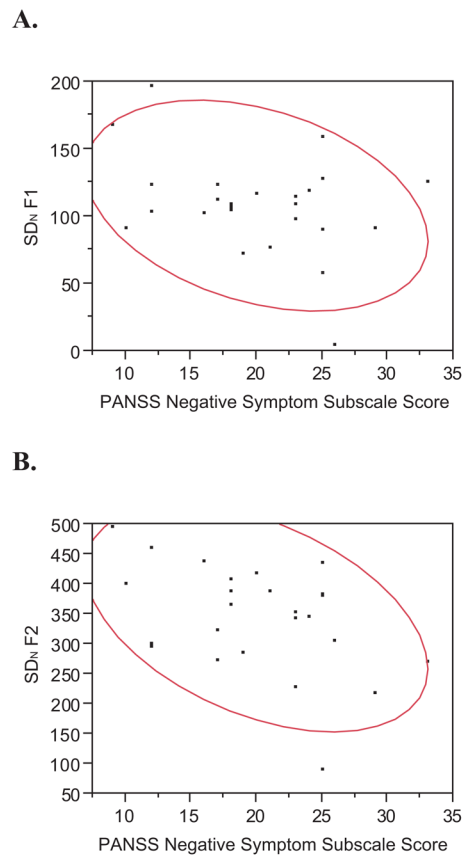


Figure 1. Scatterplots and 0.90 Density Ellipses for SD_N of F1 (A) and SD_N of F2 (B).