Ambulatory Vocal Acoustics, Temporal Dynamics, and Serious Mental Illness

Alex S. Cohen,[a] Taylor L. Fedechko,[a] Elana K. Schwartz,[a] Thanh P. Le[a]
[a] Department of Psychology, Louisiana State University

Peter W. Foltz[b]
[b] Institute of Cognitive Science, University of Colorado

Jared Bernstein,[c] Jian Cheng[c]
[c] Analytic Measures Inc, Palo Alto, CA

Terje B. Holmlund,[d] Brita Elvevåg[d,e]
[d] Department of Clinical Medicine, University of Tromsø – the Arctic University of Norway
&
[e] The Norwegian Centre for eHealth Research,
University Hospital of North Norway, Tromsø, Norway.

Correspondence: Alex S. Cohen, Louisiana State University, Department of Psychology, 236 Audubon Hall, Baton Rouge, LA 70803, United States of America. Phone: (225) 578-7017, Fax: (225) 578-4125, Email: acohen@lsu.edu.

Submitted as a brief article to Journal of Abnormal Psychology

Disclosures and Acknowledgements: This project was funded by grant #231395 from the Research Council of Norway awarded to Brita Elvevåg.
Word Count of Abstract: 138

Word count of main text, references, abstract and tables: 4,781
Number of figures: 1
Number of tables: 3

Key words: ambulatory assessment, vocal expression, serious mental illness, psychiatric symptoms, ecological momentary assessment,

Abstract

Acoustic analysis of vocal expression offers a potentially inexpensive, unobtrusive, and highly sensitive biobehavioral measure of serious mental illness (SMI)-related issues. Despite literature documenting its use for understanding SMI, prior studies have largely ignored that vocal expression is highly dynamic within individuals over time. We employed ambulatory vocal assessment from SMI outpatients to understand links between vocal expression, SMI symptoms, and affective states. Vocal samples were analyzed using a validated acoustic analysis protocol. Overall, vocal expression was not directly related to SMI symptoms but changed as a function of state and state by symptom interactions. The results suggest that a) vocal expression fails to modulate across changing affective states in individuals with active SMI symptoms, b) this lack of modulation may be commonly associated with many SMI symptoms, and c) vocal analysis can accommodate temporal dynamics.

General Scientific Summary: Acoustic analysis of vocal expression offers a potentially inexpensive, unobtrusive, and highly sensitive biobehavioral measure of serious mental illness (SMI)-related issues. Despite literature documenting its use for understanding SMI, prior studies have largely ignored that vocal expression is highly dynamic within individuals over time. This manuscript attempts to close this gap by employing ambulatory vocal assessment from SMI outpatients to understand links between vocal expression, SMI symptoms, and affective states.

Ambulatory Vocal Acoustics, Temporal Dynamics, and Serious Mental Illness

Serious mental illness (SMI), defined as functional debilitation due to psychosis and mood disorders (per the US Alcohol, Drug Abuse and Mental Health Administration Reorganization Act [ADAMHA] of 1992), is a public health crisis. The last decade has seen improved technologies for remote tracking of patients' symptoms including technologies that can potentially improve the accuracy and ecological validity of diagnosis and symptom assessment as well as help reduce potentially catastrophic and expensive events (Insel & Cuthbert, 2015). Most ambulatory technologies gather patient self-reported data, but collecting objective "bio-behavioral" data circumvents concerns about the accuracy/integrity of patient self-report and provides more sophisticated and higher "resolution" data streams (De Vos & Debener, 2014; Tahmasian, Khazaie, Golshani, & Avis, 2013, Holmlund et al., 2018). The present study examined the utility of ambulatory-based vocal acoustic analysis for understanding psychiatric symptoms.

Computerized vocal analysis (i.e., acoustic analysis) has existed for nearly a century. It is potentially inexpensive to conduct, can be automated and conducted in "real-time," and involves data that can be unobtrusively and passively collected (Ben-Zeev et al., 2017). Acoustic analysis focuses on the frequency, quality, and intensity of sounds produced by air pumped from the lungs, through the vocal folds and larynx, and articulated through the tongue, palate, cheek, and other structures. Vocal signals are multi-determined, reflecting involvement from a broad range of systems (e.g., cortical, limbic/striatal, psychomotor) and a myriad of functions (e.g., cognitive, social, physiological, arousal, affective, linguistic) (Kemmerer, 2015; Scherer, 1989). Hence, vocal features can be important for understanding a broad range of issues, for example, socio-expressive deficits in autism, suicidality, negative symptoms in schizophrenia, psychomotor retardation and anhedonia in depression, pressured speech in mania, psychomotor agitation and

emotional dysregulation in personality disorders, cognitive dysfunctions in major neurocognitive disorders, and abnormal social expression in psychopathy (Cohen & Elvevåg, 2014; Cummins et al., 2015). To date, vocal acoustics have primarily been evaluated as a measure of SMI diagnosis or clinical episode, though its empirical support as a "precision medicine" tool has generally been underwhelming. This is because findings often do not replicate across individual studies (e.g., specific vocal features often show highly variable clinical correlates across studies), and because the magnitude of effects are often underwhelming. Consider recent meta-analyses of vocal acoustics associated with depression, suicide, and psychosis (Cohen, Mitchell, & Elvevåg, 2014; Cummins et al., 2015; Cohen, Mitchell, et al., 2016) where aggregate effect sizes were surprisingly modest in magnitude (e.g., acoustic measures of blunted affect (d = -.36 in Cohen, Mitchell, & Elvevåg, 2014; range of d values = .33 to -.92)).

The present study evaluated the degree to which vocal acoustics were related to a broad range of psychiatric symptoms (i.e., depression, affect, mania, and positive and negative symptoms of schizophrenia; measured using validated clinical rating scales), clinical state (measured using self-report scales completed by the patient at the time of vocal data acquisition), and their interaction. Guided by the extant literature (e.g., Cummins et al., 2015), we hypothesized that depressive and negative symptoms would be related to vocal acoustics (i.e., "flatter" and more sparse acoustics), but only as a function of state-related affect variables (i.e., lower positive affect, higher negative affect). We also hypothesized that manic and agitation symptoms would be related to vocal acoustics (i.e., more variable and greater production of acoustics), but only as a function of higher positive and negative affect. We examined vocal acoustics and clinical state over a five-day period from data collected using a smart device application (Holmlund et al., 2018) developed by our research team to monitor clinical state using a broad range of behavioral inputs. This study

focused on acoustic analysis and self-reported clinical state in our sample comprised stable outpatients with a broad range of SMI diagnoses.

Methods

Participants (Table 1). Participants (N = 25) were stable outpatients meeting US federal definitions of SMI per the ADAMHA Reorganization Act. All were receiving treatment for an SMI from a multi-disciplinary team and were living in a group home facility. Approximately two-thirds of the sample met criteria for schizophrenia (n = 16), one-third met criteria for major depressive disorder (n = 8), and one individual met for bipolar disorder (n= 1). Two-thirds of the sample had a history of psychosis (n = 17). Participants were free from major medical or other neurological disorders that would be expected to impair compliance with the research protocol. Though substance use was endorsed by the participants, only one individual reported substance use concerns within the last year, as indicated by a clinically-relevant AUDIT/DUDIT score (Berman, Bergman, Palmstierna, & Schlyter, 2005; Bush, Kivlahan, McDonell, Fihn, & Bradley, 1998). Exclusion of this individual from the main analyses did not meaningfully change the results.

[Insert Table 1 about here]

***Clinical Measures*.** Structured clinical interviews (SCID; First, Spitzer, Gibbon & Williams, 2002) were conducted by doctoral students under supervision of a licensed psychologist (AS Cohen). Psychiatric symptoms were measured using the Expanded Brief Psychiatric Rating Scale (BPRS; Lukoff, Nuechterlein, & Ventura, 1986). We used a factor solution (Kopelowicz, Ventura, Liberman, & Mintz, 2007) with some minor modifications to attain acceptable internal consistency (> .70). Diagnoses and symptom ratings reflected consensus from the research team.

Ambulatory Assessment. Participants were asked to complete a series of tasks that were presented via a smart device application we developed over five consecutive daily testing sessions (Holmlund et al., 2018). These tasks required listening, watching, speaking, and touching to interact with the smart device for approximately 20 minutes overall. For each testing session, participants completed self-report state assessments using a digital slider and a digitally recorded speaking task via the smart device. They were paid $5 per session. Generally, the testing sessions were self-directed. Participants were asked to find a quiet place to complete testing at a time and place of their choosing. Study staff provided daily instructional and technical support as needed.

Vocal data was collected as part of an active interaction with the smart device application using a standardized task. This approach to measure acoustics contrasts with passive recording approaches (e.g., sampling short epochs throughout the day), which carry privacy and other legal obstacles to clinical implementation (Mehl, Pennebaker, Crow, Dabbs, & Price, 2001). For the speaking portion of each testing session, participants responded to a fairly structured, but open-ended probe (e.g., "give a step-by-step explanation of how you boil an egg" and "exactly how would you get from where you are right now to the grocery store?") requiring a moderate cognitive load and broad cognitive abilities (k = 1 per session). A different probe was provided each session. Responses were not coded for accuracy in this study. This task was selected because it required a modest amount of speech production – in contrast to more open-ended tasks (e.g., "how are you doing?") that potentially encouraged brief, single word, and superficial responses (e.g., "fine"). Approximately 10% of the speech samples were excluded because insufficient speech was recorded (i.e., less than two utterances recorded).

"State" affect was self-reported using a digital slider coded on a scale from 0 to 100, with increasing scores reflecting increasing intensity/frequency of the state. During each testing session, participants provided responses to approximately five positive (PA) and five negative (NA) affect-related sliders drawn from a larger list of PA (i.e., hopeful, calm, appreciated, strong, concentration, happy, energetic) and NA (i.e., anxious, frustrated, afraid, sad, stressed, angry, in pain, helpless) states derived from a commonly used self-report scale (Watson & Clark, 1999). The individual sliders showed excellent internal consistency with the summary PA (ICC = .91) and NA (ICC = .91) scores.

Acoustic Analysis. Acoustic analysis was conducted using the Computerized assessment of Affect from Natural Speech (Cohen, Lee Hong, & Guevara, 2010; Cohen, Renshaw, Mitchell, & Kim, 2016; CANS). Digital audio files were organized into "frames" for analysis (i.e., 100 per second). During each frame, basic speech properties were quantified, including fundamental frequency (i.e., frequency or "pitch") and intensity (i.e., volume). We present data for five commonly used measures of acoustics derived from our prior Principal Component Analysis of 1350 nonpsychiatric adults (Cohen et al., 2015) and 309 patients with SMI (Cohen, Mitchell, et al., 2016). These variables are presented in Table 2. Optimization filters for measuring F0 were used (i.e., low = 75 Hz and high = 300 Hz) (Vogel, Maruff, Snyder, & Mundt, 2009). Because of the nonlinear nature of the hertz frequency scale, F0 values were converted to semitones.

[Insert Table 2 about here]

Analyses. We conducted preliminary analyses to understand our data, including: a) zero-order correlations of acoustic, symptom, and state variables, and b) intra-class correlation coefficients and correlations. Next, we conducted multi-level modelling to evaluate the degree to which demographics, psychiatric symptoms, ambulatory self-report state, and symptom by self-report state interactions were related to ambulatory-based acoustic variables (dependent variables). Participant and session were included as random effects in the model. Model fit was evaluated by comparing the full model to that of random intercepts using chi-square statistics. Symptoms were grand mean centered, and state and acoustic variables were group mean centered (by testing session). Values exceeding 3.5 SDs from the grand mean for all variables were Winsorized with values of 3.5 SD from the grand mean. The state PA and NA, and their interactions with symptoms, showed acceptable multicollinearity (i.e., VIF < 10). Coefficient significance was evaluated based on p-values from the Wald-statistic, from a likelihood ratio test and the 95% confidence intervals not overlapping zero. Significant interactions were probed using simple slope and intercept analysis of coefficient values computed at two levels (-1 SD, +1 SD) using t-tests. Unless otherwise noted, all variables were normally distributed (i.e., skew values < 2.0). The analyses and plots used the R "lme" and "sjPlot" packages.

Results

Data considerations. On average, participants completed 4.5 of five sessions, which reflects 90% of all sessions completed (data for 112 of 125 possible sessions were examined). In total, 17 of 25 participants completed all testing sessions. Of the eight who did not complete all 5 sessions, half missed just one session, 3 missed two sessions, and one missed three sessions. Time of day of session completion was generally not significantly related to any of the study dependent

or independent variables, though increasing time of day was associated with less state PA (r [114] = .21, p = .02) and greater number of utterances (r[114] = .25, p = .007). Including time of day did not significantly change interpretation of any of the models computed in this study. The acoustic measures showed fair to good stability, and the state PA and NA measures showed relatively higher stability (Table 1).

Correlations. The vocal production measures were significantly correlated, but not redundant, with each other (*r* = -0.73). The vocal variability measures showed modest inter-correlation (range of *r*'s = 0.01 to 0.43). Vocal production and vocal variability measures showed varying levels of correlation with each other (range of *r*'s =0.03 to 0.69). The state PA and NA measures were inversely correlated with each other (*r* = -0.57).

Symptom and State PA/NA Markers of Ambulatory Acoustics (Table 3). In virtually none of the models (i.e., two of 20) were symptoms independently related to ambulatory acoustic measures. In contrast, both state and symptom by state interactions were significantly associated with a wide variety of acoustic measures. This primarily involved affect and mania/agitation, but not positive/negative, symptoms. Somewhat unexpectedly, state NA and PA were associated with similar changes in vocal acoustics. Inspection of the coefficient valences (i.e., negative or positive) revealed that increased state NA and PA were associated (though not necessarily significantly) with longer pause times, fewer utterances, higher pitch, and greater intonation and emphasis.

[Insert Table 3 about here]

With two exceptions (see below), simple slope analysis (Figure 1) suggested that state PA and NA were associated with less quantity and more variable speech, but only in patients with less severe affective or manic symptoms. In contrast, the vocal expression of patients with high levels of affective and mania/agitation symptoms was relatively independent of state affect. For example, in patients with low levels of affective symptoms (-1 SD), increasing state PA was associated, at a trend level or better, with longer pauses (b(SD) = 0.38 (0.17), t = 2.29), fewer utterances (b(SD) = -0.45 (0.17), t = 2.66), and more intonation (b(SD) = 0.56 (0.18), t = 3.10) (see Figure 1). For patients with high levels of affective symptoms (+1 SD), increasing state PA was not associated with changes in acoustic variables (*t*'s < 1.22). Similar patterns were observed for state NA and for mania/agitation symptoms. Pitch was the exception, where significant slopes were observed for patients with high levels of affective symptoms, though this was statistically significant for state NA (b(SD) = -0.38 (0.12), t = 3.10) and not PA (b(SD) = -0.12 (0.10), t = 1.14) interactions. Negative symptoms were the other exception. For patients with more severe negative symptoms, state NA was associated with decreased utterances (b(SD) = -0.39 (0.18), t = 2.13) and greater emphasis (b(SD) = 0.34 (0.19), t = 1.84).

[Insert Figure 1]

To evaluate a more refined set of negative symptoms, we re-ran the multilevel models with the BPRS single item "blunted affect" entered as the symptom term. The results were different than for either affect or mania/agitation symptoms (see Table 3 and Figure 1). First, most of the coefficients were in the opposite direction as seen for affect or mania/agitation, with state affect being associated with decreased pause times, more utterances, and less emphasis. The interaction

terms were also different, with vocal changes reflecting more severe blunted affect interacting with state affect. Of note, increasing blunted affect (i.e., + 1 SD) and state NA were associated with fewer utterances (b(SD) = -0.45 (0.18), t = 2.44) and greater emphasis (b(SD) = 0.41 (0.19), t = 2.16), but not greater pause times (b(SD) = 0.30 (0.18), t = 1.70). Increasing blunted affect and state PA were significantly associated with greater intonation (b(SD) = 0.28 (0.21), t = 2.77), but not pause length (b(SD) = 0.31 (0.20), t = 1.57). Low levels of blunted affect (i.e., -1 SD) and state PA/NA were generally associated (at a trend or better) with acoustic variables, with state PA/intonation, state NA/utterance number.

Discussion

It has been proposed by us and others that acoustic analysis of ambulatory vocalizations can be understood in relation to SMI symptoms. Despite decades of research, this endeavor has lagged in many respects. This lack of progress reflects, in part, a lack of appreciation for the temporal dynamics of vocal acoustics within individuals. Using ambulatory-based acoustic analysis of voice recorded from a relatively structured speaking task in participants' home environment, we were able to evaluate the consistency of acoustic signal over time and its relationship to clinically-rated symptoms and state affect. Our results suggest that acoustic signals were somewhat stable over time within patients. Consistent with much prior research (e.g., Cohen et al., 2014; Cummins et al., 2015), acoustic variables were not, in and of themselves, highly related to psychiatric symptoms. However, abnormalities in acoustic variables were associated with state by symptom interactions. Hence, the claim that vocal recordings are useful for understanding mental illness was supported, but only so far as vocal signal is tracked over time.

What can ambulatory vocal expression meaningfully tell us about SMI-related symptoms and states, and how does this inform potential assessment? In answering this question, there are two important but unexpected findings worth noting. First, affective and mania/agitation symptoms showed similar (albeit indirect) relationships to vocal expression variables, which is surprising because these symptoms (e.g., depression/mania) are generally considered orthogonal, if not diametrically opposed with each other. Second, state PA and NA both showed similar moderating effects on vocal expression in patients despite, again, potentially reflecting polar-opposite ends of an affective valence spectrum (Russell, 1980). Hyperarousal of, for example, physiological, cognitive, and affective systems, is potentially common to each of these seemingly opposing symptoms/states. Symptoms captured by the BPRS affect (e.g., anxiety, hostility) and mania/agitation symptom clusters potentially reflect hyperarousal, as do both state PA and NA (Cacioppo, Gardner, & Berntson, 1997). It is well known that vocal characteristics modulate as a function of psychological and physiological arousal (Scherer, 1989) and rely on neural regions important to arousal (Harel, Cannizzaro, Cohen, Reilly, & Snyder, 2004). The pattern of interactions suggests that this vocal expression-arousal link was disrupted in some manner such that symptomatic patients were unusually "nonreactive." Reduced reactivity in arousal/activation has been noted across a range of systems as characteristic of psychiatric disorders, for example, in reduced striatal/amygdala reactivity in schizophrenia (Taylor, Phan, Britton, & Liberzon, 2005), physiological/striatal responsivity in depression (Bylsma, Morris, & Rottenberg, 2008), and striatal response/PFC connectivity in mania (Schreiter et al., 2016). Hence, a lack of vocal modulation across changing states may be a potential indicator of worsening state and a potential focus for future ambulatory-based acoustic analysis methods. Negative symptoms reflect a potential exception, as they were abnormally "reactive" as a function of state - though these results

should be considered preliminary since the BPRS is considered a sub-optimal measure of clinically-rated negative symptoms.

Several limitations warrant mention. First, the present acoustic analyses involved a relatively structured speaking task procured once per day at a time convenient to the participant. It is unclear whether the changes in acoustic signal seen in our participants reflect a "signature" observed for most speaking tasks and how this might change as a function of daily routine (e.g., sleep). Second, the present sample included relatively stable psychiatric patients over a relatively brief temporal epoch. Different samples, and different measures of symptoms, may have yielded different results. Of particular note, the sample was limited with respect to mania symptoms. Finally, the present study focused on an extreme aspect of the human population, namely a transdiagnostic SMI sample. Given that the most important findings from this sample involved within-individual variability, the present findings are still important. Note that the use of a traditional nonpsychiatric "control" group would not be particularly informative for understanding vocal expression in psychiatric populations as they differ in many essential, yet tangential to psychiatric illness, respects (e.g., life experience) from our psychiatric group (Miller & Chapman, 2001).

Based on the present study, the process by which vocal characteristics change as a function of affective state in patients with active SMI symptoms is disrupted in some manner. In terms of future research, the nature of this state by symptom interaction on vocal expression needs to be replicated and understood better. Practically speaking, this would involve a more thorough understanding of the reliability and contextual variability of vocal measures for measuring SMI-related processes. Increasing the number of assessments, and the contextual variability of the assessments across, for example, time of day and positive and negative affective states, will be

important for this endeavor. Given that acoustic signals will likely vary as a function of context, time, and other factors, “nontraditional” reliability metrics may be needed (e.g., Generalizability Theory or MLM; Calamia, 2018). The nature of the speaking task can potentially affect acoustic expression, presumably through cognitive, affective/valence arousal, and motivational aspects of the task. Hence, it will be important to explore different speaking tasks, as they likely have different informational value regarding individual’s symptom state. Considering the speaking task is also important for pragmatics of data collection, as speaking tasks that are more naturalistic, for example, using passive recording of individuals, versus those that are more structured will yield different amounts of speech and potentially different test-retest and tolerability.

References

Ben-Zeev, D., Brian, R., Wang, R., Wang, W., Campbell, A. T., Aung, M. S. H., … Scherer, E. A. (2017). CrossCheck: Integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse. *Psychiatric Rehabilitation Journal*, *40*(3), 266–275. https://doi.org/10.1037/prj0000243

Berman, A. H., Bergman, H., Palmstierna, T., & Schlyter, F. (2005). Evaluation of the Drug Use Disorders Identification Test (DUDIT) in criminal justice and detoxification settings and in a Swedish population sample. *European Addiction Research*, *11*(1), 22–31. https://doi.org/10.1159/000081413

Bush, K., Kivlahan, D. R., McDonell, M. B., Fihn, S. D., & Bradley, K. A. (1998). The AUDIT alcohol consumption questions (AUDIT-C): An effective brief screening test for problem drinking. *Archives of Internal Medicine*, *158*(16), 1789–1795. https://doi.org/10.1001/archinte.158.16.1789

Bylsma, L. M., Morris, B. H., & Rottenberg, J. (2008). A meta-analysis of emotional reactivity in major depressive disorder. *Clinical Psychology Review*. https://doi.org/10.1016/j.cpr.2007.10.001

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review*, *1*(1), 3–25. https://doi.org/10.1207/s15327957pspr0101_2

Calamia, M. R. (2018). Practical Considerations for Evaluating Reliability in Ambulatory Assessment Studies. *Psychological Assessment, In Press*.

Cannizzaro, M., Harel, B., Reilly, N., Chappell, P., & Snyder, P. J. (2004). Voice acoustical

measurement of the severity of major depression. *Brain and Cognition*, *56*(1), 30–35. https://doi.org/10.1016/j.bandc.2004.05.003

Cohen, A. S., Dinzeo, T. J., Donovan, N. J., Brown, C. E., & Morrison, S. C. (2015). Vocal acoustic analysis as a biometric indicator of information processing: Implications for neurological and psychiatric disorders. *Psychiatry Research*, *226*(1), 235–241. https://doi.org/10.1016/j.psychres.2014.12.054

Cohen, A. S., & Elvevåg, B. (2014). Automated computerized analysis of speech in psychiatric disorders. *Current Opinion in Psychiatry*, *27*(3), 203–209. https://doi.org/10.1097/YCO.0000000000000056

Cohen, A. S., Lee Hong, S., & Guevara, A. (2010). Understanding emotional expression using prosodic analysis of natural speech: Refining the methodology. *Journal of Behavior Therapy and Experimental Psychiatry*, *41*(2), 150–157. https://doi.org/10.1016/j.jbtep.2009.11.008

Cohen, A. S., Mitchell, K. R., Docherty, N. M., & Horan, W. P. (2016). Vocal expression in schizophrenia: Less than meets the ear. *Journal of Abnormal Psychology*, *125*(2), 299–309. https://doi.org/10.1037/abn0000136

Cohen, A. S., Mitchell, K. R., & Elvevåg, B. (2014). What do we really know about blunted vocal affect and alogia? A meta-analysis of objective assessments. *Schizophrenia Research*, *159*(2–3), 533–538. https://doi.org/10.1016/j.schres.2014.09.013

Cohen, A. S., Renshaw, T. L., Mitchell, K. R., & Kim, Y. (2016). A psychometric investigation of “macroscopic” speech measures for clinical and psychological science. *Behavior Research Methods*, *48*(2), 475–486. https://doi.org/10.3758/s13428-015-0584-1

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A

review of depression and suicide risk assessment using speech analysis. *Speech Communication*, *71*, 10–49. https://doi.org/10.1016/j.specom.2015.03.004

De Vos, M., & Debener, S. (2014). Mobile eeg: Towards brain activity monitoring during natural action and cognition. *International Journal of Psychophysiology*, *91*(1), 1–2. https://doi.org/10.1016/j.ijpsycho.2013.10.008

Depp, C. A., Mausbach, B., Granholm, E., Cardenas, V., Ben-Zeev, D., Patterson, T. L., … Jeste, D. V. (2010). Mobile interventions for severe mental illness: Design and preliminary data from three approaches. *Journal of Nervous and Mental Disease*, *198*(10), 715–721. https://doi.org/10.1097/NMD.0b013e3181f49ea3

First, M. B., Spitzer, R.L, Gibbon M., and Williams, J. B. . (2002). Structured Clinical Interview for DSM-IV-TR Axis I Disorders-Patient Edition (SCID-I/P, 1/2007 revision). In *Biometrics Research*. Retrieved from http://www.scid4.org/revisions/november_2001_02.htm

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). "Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis." *Autism Research*. https://doi.org/10.1002/aur.1678

Gillespie, S., Moore, E., Laures-Gore, J., Farina, M., Russell, S., & Logan, Y. Y. (2017). Detecting stress and depression in adults with aphasia through speech analysis. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* (pp. 5140–5144). https://doi.org/10.1109/ICASSP.2017.7953136

Greenberg, T., Chase, H. W., Almeida, J. R., Stiffler, R., Zevallos, C. R., Aslam, H. A., … Phillips, M. L. (2015). Moderation of the relationship between reward expectancy and prediction error-related ventral striatal reactivity by anhedonia in unmedicated major

depressive disorder: Findings from the EMBARC study. *American Journal of Psychiatry*, *172*(9), 881–891. https://doi.org/10.1176/appi.ajp.2015.14050594

Harel, B. T., Cannizzaro, M. S., Cohen, H., Reilly, N., & Snyder, P. J. (2004). Acoustic characteristics of Parkinsonian speech: A potential biomarker of early disease progression and treatment. *Journal of Neurolinguistics*, *17*(6), 439–453. https://doi.org/10.1016/j.jneuroling.2004.06.001

Hashim, N. W., Wilkes, M., Salomon, R., Meggs, J., & France, D. J. (2017). Evaluation of Voice Acoustics as Predictors of Clinical Depression Scores. *Journal of Voice*, *31*(2), 256.e1-256.e6. https://doi.org/10.1016/j.jvoice.2016.06.006

Holmlund, T. B., Foltz, P. W., Cheng, J., Bernstein, J., Cohen, A. S., & Elvevåg, B. (in press). Moving psychological assessment out of the controlled laboratory setting : Practical challenges. *Psychological Assessment.*

Insel, T., & Cuthbert, B. (2015). Brain disorders? Precisely. Precision medicine come to psychiatry. *Science*, *348*(6234), 499–500. https://doi.org/http://dx.doi.org/10.1126/science.aab2358

Insel, T. R. (2009). Translating scientific opportunity into public health impact: A strategic plan for research on mental illness. *Archives of General Psychiatry*, *66*(2), 128–133. https://doi.org/10.1001/archgenpsychiatry.2008.540

Insel, T. R. (2017). Digital phenotyping: Technology for a new science of behavior. *JAMA - Journal of the American Medical Association*, *318*(13), 1215–1216. https://doi.org/10.1001/jama.2017.11295

Kemmerer, D. (2015). *Cognitive neuroscience of language*. *Cognitive neuroscience of language.* Retrieved from

http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc11&NEWS=N&AN=2015-00295-000

Kopelowicz, A., Ventura, J., Liberman, R. P., & Mintz, J. (2007). Consistency of brief psychiatric rating scale factor structure across a broad spectrum of schizophrenia patients. *Psychopathology*, *41*(2), 77–84. https://doi.org/10.1159/000111551

Laukka, P., Linnman, C., Åhs, F., Pissiota, A., Frans, Ö., Faria, V., … Furmark, T. (2008). In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech. *Journal of Nonverbal Behavior*, *32*(4), 195–214. https://doi.org/10.1007/s10919-008-0055-9

Lukoff, D., Nuechterlein, H., & Ventura, J. (1986). Manual for the expanded brief psychiatric rating scale. *Schizophrenia Bulletin*, *12*, 594.

Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, and Computers*, *33*(4), 517–523. https://doi.org/10.3758/BF03195410

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, *110*(1), 40–48. https://doi.org/10.1037//0021-843X.110.1.40

Mundt, J. C., Vogel, A. P., Feltner, D. E., & Lenderking, W. R. (2012). Vocal acoustic biomarkers of depression severity and treatment response. *Biological Psychiatry*, *72*(7), 580–587. https://doi.org/10.1016/j.biopsych.2012.03.015

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Scherer, K. R. (1989). Vocal correlates of emotional arousal and affective disturbance. In

Handbook of social psychophysiology. Wiley handbooks of psychophysiology (pp. 165–197).

Schreiter, S., Spengler, S., Willert, A., Mohnke, S., Herold, D., Erk, S., … Bermpohl, F. (2016). Neural alterations of fronto-striatal circuitry during reward anticipation in euthymic bipolar disorder. *Psychological Medicine*, *46*(15), 3187–3198. https://doi.org/10.1017/S0033291716001963

Skodda, S., Grönheit, W., & Schlegel, U. (2011). Intonation and speech rate in parkinson's disease: General and dynamic aspects and responsiveness to levodopa admission. *Journal of Voice*, *25*(4). https://doi.org/10.1016/j.jvoice.2010.04.007

Tahmasian, M., Khazaie, H., Golshani, S., & Avis, K. T. (2013). Clinical application of actigraphy in psychotic disorders: A systematic review. *Current Psychiatry Reports*, *15*(6). https://doi.org/10.1007/s11920-013-0359-2

Taylor, S. F., Phan, K. L., Britton, J. C., & Liberzon, I. (2005). Neural response to emotional salience in schizophrenia. *Neuropsychopharmacology*, *30*(5), 984–995. https://doi.org/10.1038/sj.npp.1300679

Vogel, A. P., Maruff, P., Snyder, P. J., & Mundt, J. C. (2009). Standardization of pitch-range settings in voice acoustic analysis. *Behavior Research Methods*, *41*(2), 318–324. https://doi.org/10.3758/BRM.41.2.318

W. Adams, Z., McClure, E. A., Gray, K. M., Danielson, C. K., Treiber, F. A., & Ruggiero, K. J. (2017). Mobile devices for the remote acquisition of physiological and behavioral biomarkers in psychiatric clinical research. *Journal of Psychiatric Research*, *85*, 1–14. https://doi.org/10.1016/j.jpsychires.2016.10.019

Watson, D., & Clark, L. (1999). The PANAS-X Manual for the Positive and Negative Affect

Schedule-Expanded Form. *Iowa Research Online*, *277*(6), 1–27.

https://doi.org/10.1111/j.1742-4658.2010.07754.x

Table 1. Descriptive and demographic data for participants (n = 25)

	M (SD)	Range	ICC	Average r
Demographic Variables				
Age	49.68 (10.38)	30.00 – 67.00		
% Female	52%			
% Caucasian	36%			
% African-Am.	52%			
Education (years)	12.28 (1.43)	11.00 – 16.00		
Brief Psychiatric Rating Scale [a, b]				
Affective	2.07 (0.96)	1.00 – 5.25	0.74	0.45
Agitation	1.56 (0.63)	1.00 – 3.75	0.76	0.43
Positive	2.21 (1.18)	1.00 – 5.50	0.77	0.25
Negative	2.07 (0.96)	1.00 – 5.50	0.79	0.55
Acoustic Variables [c, d]				
Recording Length	20.65 (13.67) sec	1.35 – 60.00 sec	0.75	0.39
Pause Mean	10.77 (9.66) sec	1.71 – 38.11 sec	0.67	0.32
Utterances (N)	10.22 (7.97)	1.00 – 30.70	0.87	0.55
Pitch	86.26 (3.28)	80.33 – 92.17	0.89	0.63
Intonation	2.30 (0.87)	1.05 – 4.23	0.79	0.45
Emphasis	4.76 (0.97)	2.76 – 6.80	0.75	0.37
State Affect Variables [c, d, e]				
Negative Affect	35.48 (28.33)	0.05 – 96.53	0.92	0.70
Positive Affect	72.08 (24.06)	10.29 – 100.00	0.91	0.67

[a.] Data reflect average of items in factor; Possible range = 1 to 7. [b.] ICC and average r measure internal consistency of items, [c.] ICC and average r measure temporal stability (i.e., values over 5 administration), [d.] Data averaged over time within participants, [e.] Possible range: 0.00 – 100.00.

Table 2. Vocal properties and variables examined in this study

Name	Definition	Function
	"SPEECH PRODUCTION"	
Pause Mean	Average silence between voiced utterance (in seconds)	Pauses between vocal units
Number of Utterances	Number of voicings bounded by silence	Speech quantity
	"SPEECH VARIABILITY"	
"Pitch"	Average fundamental frequency (F0; in semitones)	Frequency of vocal fold vibrations
Intonation	SD of F0 within each utterance, averaged across utterances	Variability in F0
Emphasis	SD of intensity within each utterance, averaged across utterances (in decibels)	Variability in intensity/volume

Table 3. Multi-level modelling of ambulatory acoustic measures (dependent measures) as a function of psychiatric symptoms, state affect, and their interactions.

	Pause Length	Number of Utterances	Pitch	Intonation	Emphasis
			Affective Symptoms		
Model Fit	X^2 = 16.10*	X^2 = 13.50*	X^2 = 27.00*	X^2 = 11.10+	X^2 = 2.31
Affective Symptoms	0.13 (0.13)	-0.21 (0.17)	0.06 (0.15)	-0.08 (0.14)	0.07 (0.17)
State NA	0.50 (0.24)*	-0.46 (0.23)	0.76 (0.21)*	0.30 (0.26)	0.01 (0.27)
State PA	0.64 (0.24)*	-0.67 (0.24)*	0.52 (0.22)*	0.91 (0.27)*	0.21 (0.28)
Affective Sx x State NA	-0.24 (0.10)*	0.15 (0.09)+	-0.37 (0.08)*	-0.10 (0.11)	0.01 (0.11)
Affective Sx x State PA	-0.25 (0.08)*	0.22 (0.08)*	-0.21 (0.07)*	-0.29 (0.09)*	-0.09 (0.09)
			Mania/Agitation Symptoms		
Model Fit	X^2 = 15.90*	X^2 = 8.52	X^2 = 31.20*	X^2 = 11.10+	X^2 = 7.47
Manic/Agitation Symptoms	0.29 (0.22)	-0.22 (0.30)	0.33 (0.24)	-0.09 (0.25)	0.63 (0.28)*
State NA	0.71 (0.28)*	-0.57 (0.29)+	1.10 (0.25)*	0.25 (0.31)	0.54 (0.32)+
State PA	0.64 (0.26)*	-0.54 (0.27)*	0.68 (0.23)*	0.96 (0.29)*	0.07 (0.30)

Manic/Agitation Sx x State NA	-0.43 (0.16)*	0.25 (0.15)+	-0.63 (0.13)*	-0.09 (0.17)	-0.30 (0.17)+
Manic/Agitation Sx x State PA	-0.33 (0.12)*	0.25 (0.12)+	-0.33 (0.11)*	-0.41 (0.13)*	-0.02 (0.14)

Positive Symptoms

Model Fit	$X^2 = 1.80$	$X^2 = 4.22$	$X^2 = 10.20$	$X^2 = 3.01$	$X^2 = 9.81$
Positive Symptoms	-0.08 (0.11)	0.13 (0.13)	0.25 (0.12)+	-0.03 (0.12)	0.13 (0.13)
State NA	0.03 (0.31)	-0.36 (0.29)	0.01 (0.27)	0.27 (0.33)	0.75 (0.31)*
State PA	-0.09 (0.26)	-0.17 (0.25)	-0.01 (0.23)	0.43 (0.28)	0.02 (0.27)
Positive Sx x State NA	-0.02 (0.10)	0.09 (0.09)	-0.06 (0.09)	-0.08 (0.11)	-0.27 (0.10)*
Positive Sx x State PA	-0.01 (0.08)	0.06 (0.07)	-0.01 (0.07)	-0.11 (0.08)	-0.02 (0.08)

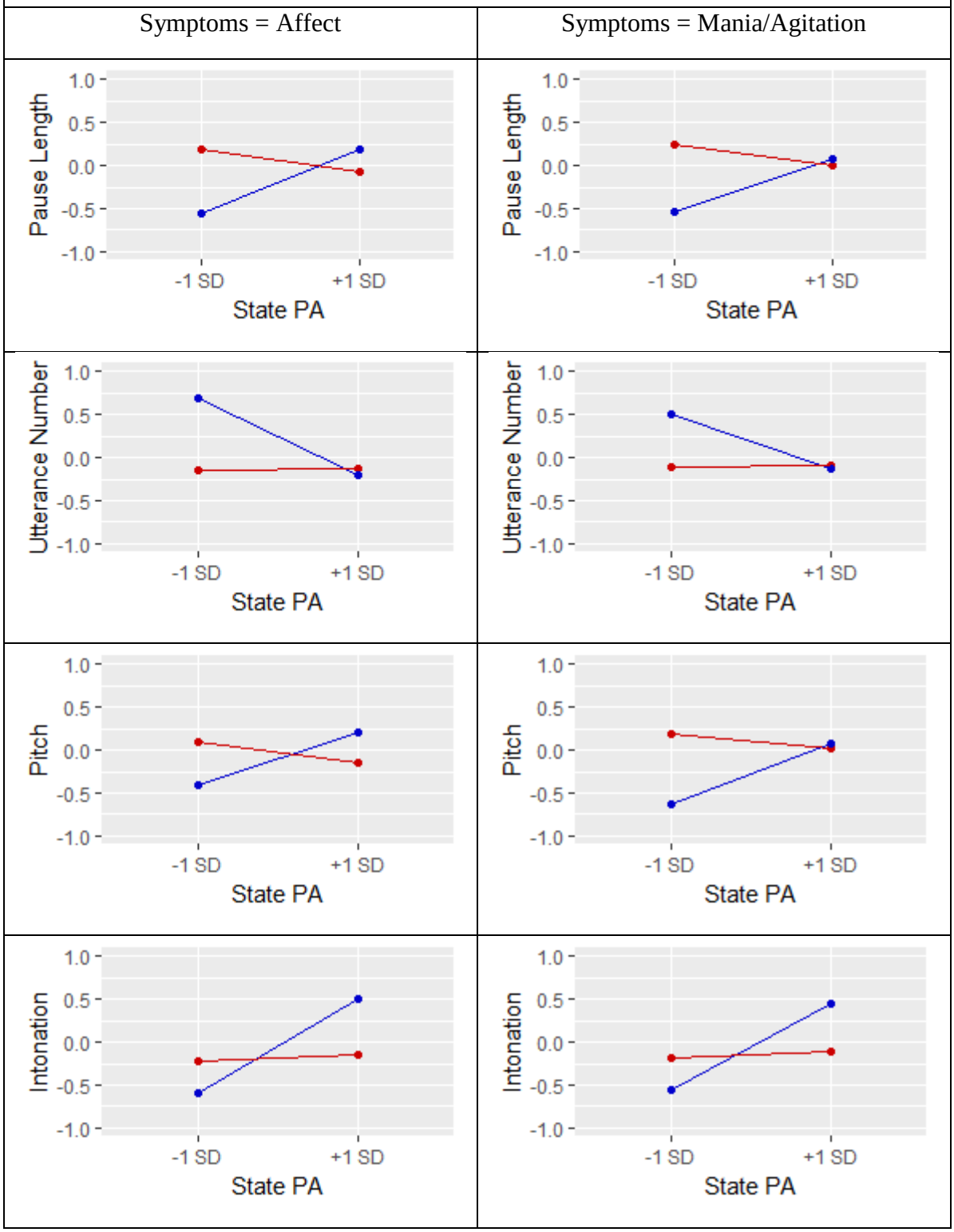
Negative Symptoms

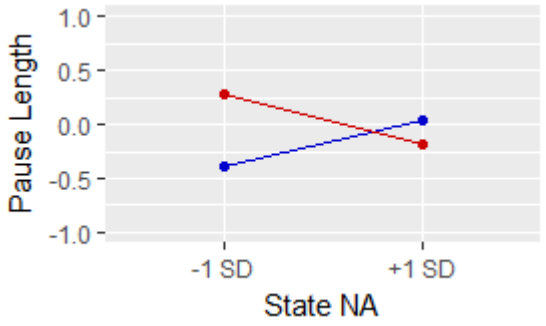
Model Fit	$X^2 = 3.29$	$X^2 = 13.70$*	$X^2 = 11.90$+	$X^2 = 3.59$	$X^2 = 10.20$
Negative Symptoms	0.05 (0.13)	-0.13 (0.16)	-0.38 (0.15)*	-0.15 (0.15)	-0.09 (0.16)
State NA	-0.45 (0.29)	0.78 (0.28)*	-0.05 (0.27)	-0.10 (0.31)	-0.75 (0.31)*
State PA	-0.34 (0.32)	0.26 (0.29)	-0.03 (0.28)	-0.33 (0.34)	0.07 (0.32)

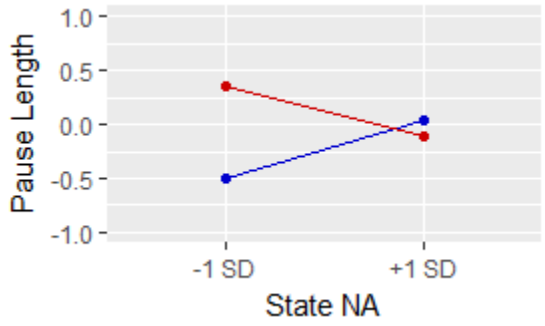
Negative Sx x State NA	0.20 (0.14)	-0.45 (0.14)*	-0.03 (0.13)	0.06 (0.15)	0.39 (0.15)*
Negative Sx x State PA	0.13 (0.16)	-0.15 (0.14)	-0.03 (0.14)	0.22 (0.17)	-0.06 (0.16)
Blunted Affect Symptoms					
Model Fit	$X^2 = 8.83$	$X^2 = 14.40$*	$X^2 = 14.60$*	$X^2 = 9.86$	$X^2 = 9.04$
Blunted Affect Symptoms	-0.02 (0.09)	-0.04 (0.12)	-0.32 (0.11)*	-0.08 (0.11)	-0.09 (0.12)
State NA	-0.55 (0.22)*	0.56 (0.21)*	-0.29 (0.20)	-0.16 (0.24)	-0.60 (0.24)*
State PA	-0.54 (0.21)*	0.29 (0.21)	-0.25 (0.20)	-0.47 (0.22)*	-0.07 (0.23)
Blunted Affect Sx x State NA	0.26 (0.10)*	-0.35 (0.10)*	0.11 (0.09)	0.11 (0.11)	0.31 (0.11)*
Blunted Affect Sx x State PA	0.25 (0.10)*	-0.18 (0.10)+	0.10 (0.10)	0.33 (0.11)*	0.02 (0.11)

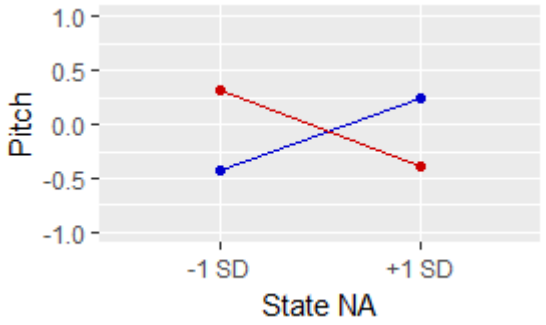
Note. Coefficient values with asterisks are statistically significant, on convergence between wald-statistic, likelihood ratio test and confidence intervals not including 0, * = $p < 0.05$, + = $p < .10$, X^2 = chi-square change statistic comparing the full model to the random-intercepts only model, NA = Negative Affect, PA = Positive Affect, Sx = Symptoms.

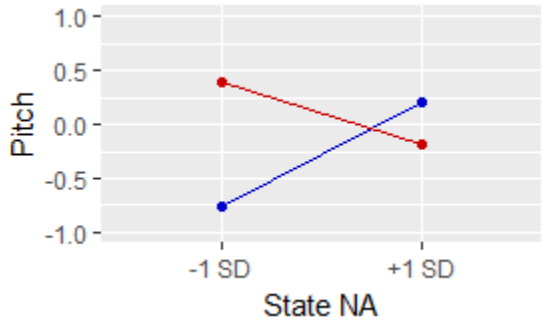
Figure 1. Simple slope plotted for acoustic variables (y-axis) for patients with high (+1 SD; Red line) versus low (-1 SD; Blue line) levels of symptoms as a function of state positive or negative affect (x-axis).





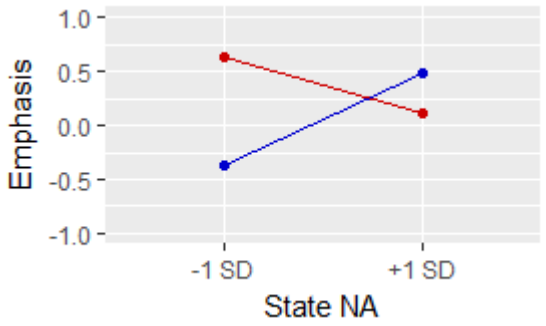


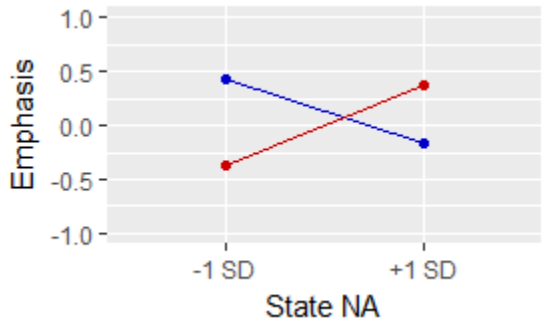


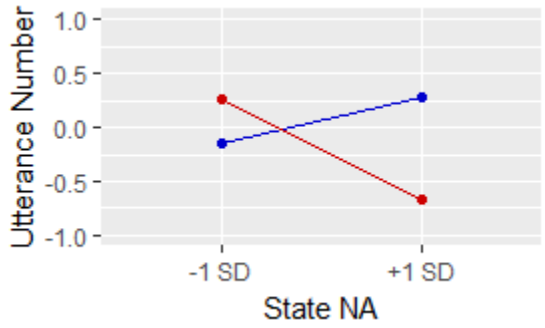


Symptoms = Positive Symptoms

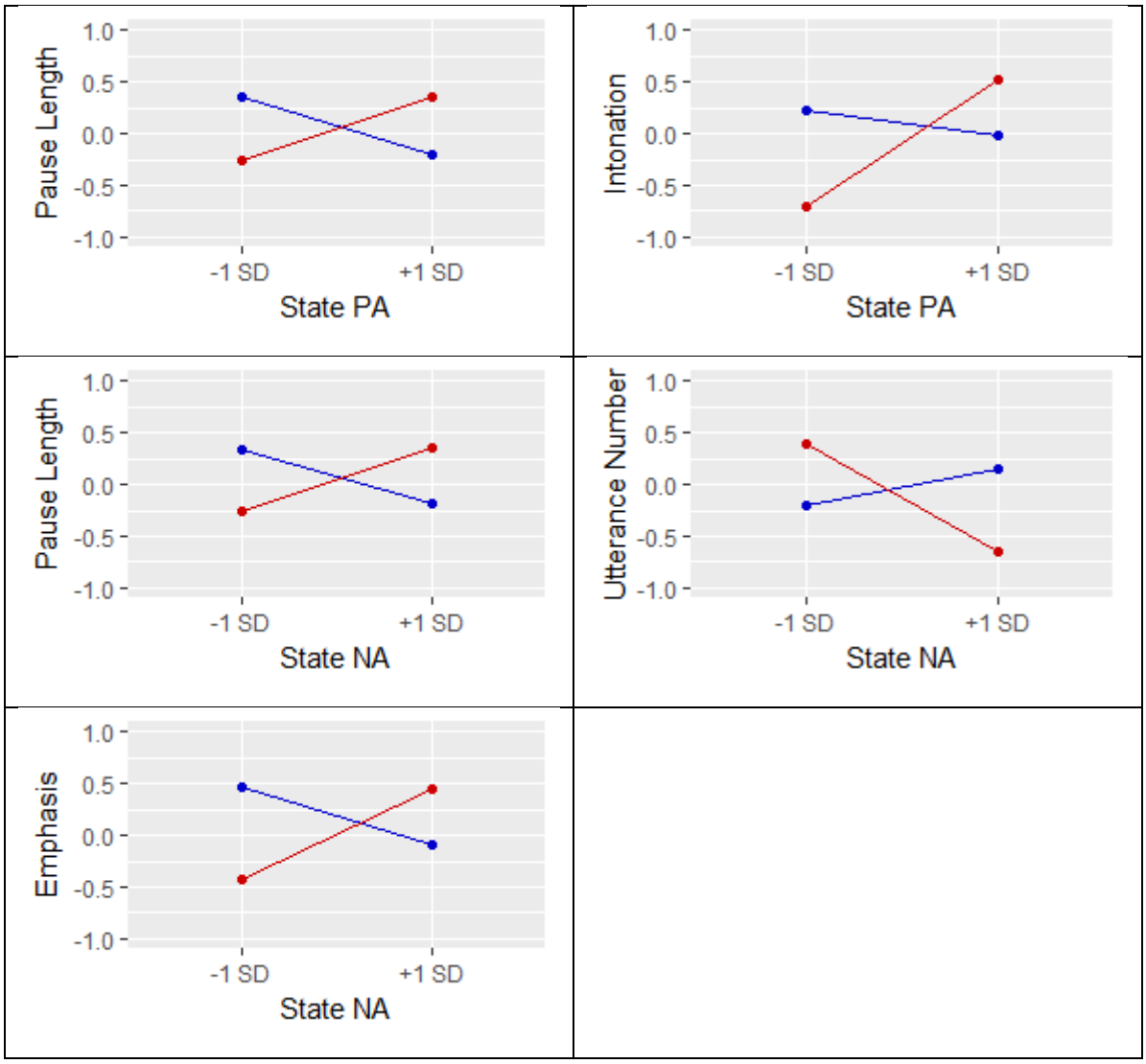
Symptoms = Negative Symptoms







Symptoms = Blunted Affect


Pause Length
1.0
0.5
0.0
-0.5
-1.0
-1 SD
+1 SD
State PA
Intonation
1.0
0.5
0.0
-0.5
-1.0
-1 SD
+1 SD
State PA
Pause Length
1.0
0.5
0.0
-0.5
-1.0
-1 SD
+1 SD
State NA
Utterance Number
1.0
0.5
0.0
-0.5
-1.0
-1 SD
+1 SD
State NA
Emphasis
1.0
0.5
0.0
-0.5
-1.0
-1 SD
+1 SD
State NA