Non-linear dynamics of speech and voice in schizophrenia

Riccardo Fusaroli\textsuperscript{1,2,3}, Ethan Weed\textsuperscript{2,3,4}, Arndis Simonsen\textsuperscript{2,5}, Vibeke Bliksted\textsuperscript{2,5}

\textsuperscript{1}Center for Semiotics, Aarhus University, \textsuperscript{2}Interacting Minds Center, Aarhus University, \textsuperscript{3}Center of Functionally Integrative Neuroscience, Aarhus University, \textsuperscript{4}Linguistics, Aarhus University, \textsuperscript{5}Department of General Psychiatry, Aarhus University Hospital

Background

Patients with schizophrenia are often described as speaking monotonously and without emotion. Such anomalous speech and voice patterns are assessed in the diagnostic process and deeply impact the quality of everyday social interactions. It is therefore crucial to quantify these patterns and relate them to other symptoms.

Our aims were:

1) to achieve a more fine-grained understanding of the speech patterns in schizophrenia than has previously been achieved using traditional acoustic measures
2) to use supervised machine-learning to classify speech production as belonging to the control or the schizophrenia group
3) assess symptoms’ severity based solely on voice dynamics.

Methods (Cont’d)

We extracted measures of pause behavior and fundamental frequency (Fig 2), as well as measures of stability and regularity for both (RR, Det, L, LMax, Entr, Lam, Vmax, T1, T2, Trend) (Boersma, 2001; Marwan et al, 2007) (Fig 3). The most relevant features were selected via ElasticNet (Zhou & Hastie, 2005). Diagnosis was predicted using a 10-fold cross-validated discriminant function (Mahalanobis rule) (Rodriguez et al, 2012).

Accuracy was balanced using Variational Bayesian mixed-effects inference. SANS and SAPS scores were predicted using a 10-fold cross-validated multiple linear regression. Both analyses were iterated 1000 times to test for stability of results.

Results

Voice dynamics allowed the discrimination of patients with schizophrenia from healthy controls (Fig 4) with a balanced accuracy of 85.68% (p<0.00001, Confidence Intervals: 82.50% – 86.97%), a sensitivity of 81.27% and a specificity of 86.97%.

Results (Cont’d)

Voice dynamics explained 26.76% (measured as Adjusted R Square, p<0.000001) of the variance of SANS scores and 20.33% (p<0.00001) of SAPS scores (Fig 5). Between 20% and 44% of the variance of single factors was explained.

Discussion

When comparing them to healthy controls, schizophrenics’ voice was characterized as: (1) Slower and with longer pauses; (2) Less structured, that is, with fewer repetitions of f0 sequences; (3) More “stable”, that is, the same low level of regularity is kept constant over time, while the controls tend to increase and decrease the amount of regularities and repetitions over time.

The study points toward the usefulness of non-linear time series analyses techniques in picking out the subtle differences that characterize the unusual voice characteristics of people with schizophrenia and in relating them to the symptoms. Automated analysis of voice dynamics thus reveals potential for the assessment and monitoring of the disorder.

References


