medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Voice patterns as markers of schizophrenia: building a cumulative generalizable approach via a cross-linguistic and meta-analysis based investigation

Alberto Parola [a, b, c], Arndis Simonsen [b, d], Jessica Mary Lin [a, b], Yuan Zhou [e], Huiling Wang [f], Shiho Ubukata [g], Katja Koelkebeck [h,i], Vibeke Bliksted [b, d], Riccardo Fusaroli [a, b, l]

[a] Department of Linguistics, Cognitive Science and Semiotics, Aarhus University, Aarhus, Denmark

[b] The Interacting Minds Center - Institute of Culture and Society, Aarhus University, Aarhus, Denmark

[c] Department of Psychology, University of Turin, Turin, Italy

[d] Psychosis Research Unit - Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

[e] Institute of Psychology, Chinese Academy of Sciences, Beijing, China

[f] Department of Psychiatry, Renmin Hospital of Wuhan University, Wuhan, China

[g] Department of Psychiatry, Kyoto University, Kyoto, Japan

[h] LVR-Hospital Essen, Department of Psychiatry and Psychotherapy, Hospital and Institute of the University of Duisburg-Essen, Essen, Germany

[i] Center for Translational Neuro- & Behavioral Sciences (C-TNBS), University Duisburg-Essen, Germany

[l] Linguistic Data Consortium, University of Pennsylvania, USA

Correspondence concerning this article should be addressed to:

Alberto Parola, Jens Chr. Skous Vej 2, 8000 Aarhus C, Denmark

Phone number: +390116703065

E-mail: alberto.parola@cc.au.dk

Word counts abstract (max 250): 244

Word counts text body (max 4000): 3999

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Background and Hypothesis: Voice atypicalities are potential markers of clinical features of schizophrenia (e.g., negative symptoms). A recent meta-analysis identified an acoustic profile associated with schizophrenia (reduced pitch variability and increased pauses), but also highlighted shortcomings in the field: small sample sizes, little attention to the heterogeneity of the disorder, and to generalizing findings to diverse samples and languages.

Study Design: We provide a critical cumulative approach to vocal atypicalities in schizophrenia, where we conceptually and statistically build on previous studies**.** We aim at identifying a cross-linguistically reliable acoustic profile of schizophrenia and assessing sources of heterogeneity (symptomatology, pharmacotherapy, clinical and social characteristics). We relied on previous meta-analysis to build and analyze a large cross-linguistic dataset of audio recordings of 231 patients with schizophrenia and 238 matched controls (>4.000 recordings in Danish, German, Mandarin and Japanese). We used multilevel Bayesian modeling, contrasting meta-analytically informed and skeptical inferences.

Study Results: We found only a minimal generalizable acoustic profile of schizophrenia (reduced pitch variability), while duration atypicalities replicated only in some languages. We identified reliable associations between acoustic profile and individual differences in clinical ratings of negative symptoms, medication, age and gender. However, these associations vary across languages.

Conclusions: The findings indicate that a strong cross-linguistically reliable acoustic profile of schizophrenia is unlikely. Rather, if we are to devise effective clinical applications able to target different ranges of patients, we need first to establish larger and more diverse cross-linguistic datasets, focus on individual differences, and build self-critical cumulative approaches.

Keywords:

vocal analysis, psychosis, speech signal, digital phenotyping, prosody, negative symptoms

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

From its very first definitions, schizophrenia has been associated with voice atypicalities[1,2], qualitatively described in terms of e.g., poverty of speech, increased pauses, distinctive tone and intensity of voice. A recent systematic meta-analysis[3] indicates a plausible acoustic profile associated with schizophrenia. In this study, we assess how generalizable that profile is within a large cross-linguistic dataset, as well as sources of heterogeneity in vocal patterns of patients with schizophrenia.

Atypical voice patterns are included amongst negative symptoms of schizophrenia, such as alogia and blunted affect, which are among the primary diagnostic criteria and prognostic indicators of the disorder (e.g. response to treatment and reduced likelihood of remission[4–7]). Vocal behavior may constitute a window into the underlying social and cognitive features of the disorder[8]. For example, the social and cognitive impairments frequently reported in schizophrenia[9–11] may be reflected in difficulties in speech fluency (e.g. increased pauses), or in controlling the voice to express affective and emotional contents and to mark relevant information[12,13]. In other words, not only can the quantitative analysis of vocal behavior scaffold the current evaluation of negative symptoms[14–19], but it could also offer a more fine-grained perspective on its clinical, social and cognitive dimensions, e.g., social and cognitive functioning over time[8,18,20,21] .

However, while extensive literature exists on vocal atypicalities in schizophrenia, with studies spanning back to the 60's, the findings are often contradictory and difficult to interpret. A recent meta-analysis[3] identified weak atypicalities in pitch variability - potentially related to flat affect- and stronger atypicalities in proportion of spoken time, speech rate, and pauses - potentially related to alogia and flat affect. The effects had large heterogeneity, and were modest compared to clinical judgments of vocal atypicalities[22]. The studies were noted

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

to have small sample sizes, high heterogeneity in methods and features analyzed, with little to no attention to the heterogeneity of the disorder[23–25] and to the replicability and generalizability of previous results on diverse samples[20,25–27]. Further, voice quality features - highlighted as important by speech pathologists and speech processing research[28,29]- had been largely neglected.

To assess the robustness and potential clinical impact of biobehavioral vocal markers of schizophrenia, we need to understand under which conditions they vary and under which they can be relied upon. We need to map which variations in clinical and socio-cognitive features might underlie voice atypicalities, and how stable they are across languages, time and recordings. For example, although vocal behavior has been shown to be influenced by linguistic and cultural factors, all studies of vocal markers of schizophrenia have investigated single monolingual samples. We need to assess how voice atypicalities relate to the development of the disorder, for instance, whether they are a long-term consequence of chronicity, already present at illness onset, and vary along with symptom severity, thus being potentially useful for tracking the development of the disorder and monitoring the symptomatology over time[30–32]. Another crucial issue is how antipsychotic drugs relate to these atypicalities and impact our ability to use them as biobehavioral markers. For example, effects of antipsychotic medication have been hypothesized to affect language in different ways, such as causing extrapyramidal motor symptoms or increasing negative symptoms by blocking dopamine receptors[33].

In other words, there is a need for a more rigorous cumulative scientific approach to understand vocal and prosodic atypicalities in schizophrenia: the synthesis and integration of data across studies and laboratories, in order to assess which patterns might generalize across contexts and samples, identify possible sources of variation and improve estimations. In this work, we provide the first steps towards such an approach.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

First, we made use of the recommendations developed in the previous meta-analysis[3] and collected a large dataset of multiple audio-recordings of patients with schizophrenia and controls in four different languages and three language families (Germanic, Mandarin-Chinese, Japonic). Such a setup provides a stronger basis for estimating the robustness of vocal atypicalities, within and between subjects and samples, as well as their variability. Further, this design allows us to explicitly compare results across different languages for the first time in schizophrenia, thus accounting for the natural differences in vocal patterns across languages.

Second, we provided a more systematic investigation of the acoustic features potentially associated with schizophrenia. While a previous meta-analysis[3] identified relevant estimates for a limited set of features (n = 8), pertaining to pitch and rhythm, recent investigations into Parkinson's disease and depression[28,34] suggest that voice quality and articulatory features might have higher discriminative power compared to traditional acoustic features. In other words, they might be particularly involved in the mechanisms underlying the disorders. We therefore extended the acoustic features investigated to include them.

Third, we cumulatively build on previous findings by explicitly (but critically) including the meta-analytic findings in the statistical Bayesian analysis of the current study. This practice - referred to as "informed priors" or "posterior passing"[35] - allows us to directly estimate how well our results match previous findings and in which ways they deviate, but also potentially increases the precision of our estimates.

Fourth, we include a more comprehensive assessment of the patients' symptomatology and clinical profile. Specifically, we model the relationships between acoustic features and pharmacotherapy, relevant clinical aspects, demographic and social features, relationships that have rarely been jointly investigated in previous studies[36].

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Finally, we rely on an open methodology: not only do we carefully describe the methodology used and test the robustness of the results to variations in the methodology; but we also use open-source software, extracting the features in a reproducible manner with openly available scripts.

By providing an initial consolidation and test of acoustic atypicalities in schizophrenia that systematically builds on and extends the previous literature, we aim to set the foundations for more critical theory development, more cumulative and holistic approaches to the understanding of the underlying mechanisms and potentially the development of applications that constructively support clinical practices.

Methods

Participants

We collected a Danish (DK), German (GE), Chinese (CH), and Japanese (JP) cross-linguistic dataset involving 231 participants with schizophrenia (105 DK, 61GE, 51CH, 14JP) and 238 matched controls (HC) (116DK, 62GE, 43CH, 17JP). The samples for the present study were collected in separate studies assessing mentalizing ability in patients with schizophrenia and healthy controls. Information on demographics, IQ, psychopathology, and social functioning is summarized in **Table 1**. Detailed information on each study is reported in the **Supplementary Material (SM) - S1**.

Table 1. Demographic and clinical characteristics of patients with schizophrenia and healthy controls (HC).

Corpus	Danish		German		Chinese		Japanese	
Diagnosis	SCZ N = 105	HC N =116	SCZ N= 61	HC N=62	SCZ N =51	HC N =43	SCZ N= 14	HC N =17
N. of recording	N = 900	N = 989	N=612	N= 609	N = 401	N = 340	N= 121	N= 144
Age	26.5 (8.82)	26.4 (8.96)	31.7 (9.92)	33.2 (8.79)	27.2 (7.25)	29.7 (8.72)	28,6 (6.71)	41.9 (13.7)

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Education	12.9 (2.74)	14.9 (2.61)	12.1 (1.48)	12.3 (1.11)	12.7 (2.69)	14.1 (2.37)	11.9 (1.37)	14.7 (2.93)
Sex (n. of females and %)	45 (43%)	50 (43%)	22 (36%)	24 (38%)	23 (45%)	19 (44%)	4 (29%)	11 (65%)
Verbal IQ	89.13 (18.7)	102.12 (16.1)	112.0 (15.6)	116.5 (16.9)	96.3 (16.6)	100.27 (14.7)	NA	NA
SANS total	9.66 (4.41)	NA	NA	NA	7.61 (2.99)	NA	NA	NA
SAPS total	10.32 (4.91)	NA	NA	NA	7.23 (4.79)	NA	NA	NA
PANSS total	NA	NA	54.65 (10.63)	NA	75.79 (10.44)	NA	58.86 (15.25)	NA
PANSS negative	NA	NA	14.61 (4.19)	NA	20.09 (5.24)	NA	15.28 (3.95)	NA
PANSS positive	NA	NA	11.31 (3.13)	NA	18.51 (4.37)	NA	12.86 (4.10)	NA
Illness duration (months)	8.87 (6.68)	NA	54.38 (83.24)	NA	63.13 (68.87)	NA	83.86 (91.33)	NA
PSP	56.01 (15.70)	86.46 (6.72)	NA	NA	54.09 (8.02)	NA	NA	NA

Note: The table displays means and standard deviations of demographic (age, education and sex) and clinical information. Clinical symptoms were measured using the Scale for the Assessment of Negative Symptoms (SANS)[37], the Scale for the Assessment of Positive Symptoms (SAPS)[38], and the Positive and Negative Syndrome Scale (PANSS)[39]. Social functioning was measured using the Personal and Social Performance scale (PSP)[40]. NA = data not available; SCZ = patients with schizophrenia, HC = healthy controls.

Voice recordings

The dataset included a total of 2.034 recordings of individuals with schizophrenia (mean recording length = 17.3 seconds (sec), sd = 15.6sec) and 2.082 recordings of control participants (recording length = 18.5sec, sd = 12.7sec). Voice recordings were collected using the Animated Triangles Task (ATT)[41,42]. The task is generally used to measure theory of mind (ToM) and involves twelve video clips representing an interaction between animated geometrical shapes (triangles). The participants were asked to provide an interpretation of what was going on in each animation and their answers were audio-recorded. A detailed

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

description of the task and its validity for assessing speech production is included in **SM2**. Recording setting and experimental procedure are described in detail in **SM3**, and were kept constant across the different sites**.** All recordings were manually pre-processed to remove background noise and interviewer speech when present, and to ensure that all recordings analyzed had adequate audio quality. A full description of the process and of the extracted acoustic features is available in **SM3**.

Statistical Modeling

Analysis of effect of diagnosis on acoustic features

To estimate the differences between individuals with schizophrenia and HC in the different acoustic measures, we used Bayesian multilevel Gaussian regression models on the current data with each acoustic feature as outcome, and diagnosis (schizophrenia vs. HC) and language (DK, GE, CH, JP) as predictors. Within the same model, we separately assessed the effect of diagnosis for each language, and modeled varying effects of participants, i.e., intercepts and slopes, separately for each group and language. For each acoustic feature, we built a first model with weakly informative priors, i.e., expectations of no effects of diagnosis, thus conservatively regularizing the model parameters, reducing overfitting and leading to improved predictions[43]. We then built a second model with an informed prior (when available), that is meta-analytic effect size (ES), and compared results across the two models. We aimed to assess whether the effects of diagnosis are robust across changes of priors, and whether the skeptical and informed priors led to more robust inference, that is, in lower estimated out-of-sample error - measured in terms of Leave-One-Out based stacking weights[44]. To evaluate the potential role of biological sex (male vs. female), age and level of intelligence, we built additional models, one per each moderator interacting with group separately in the four languages. We then reported the model estimates for the interaction, including credible (i.e., Bayesian confidence) intervals (CIs) and evidence ratios (ERs), i.e.,

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

evidence in favor of the effect observed against alternative hypotheses (see **SM4**). When ER was weak (below 10, that is, less than ten times as much evidence for the effect as for alternative hypotheses), we also calculated the ER in favor of the null hypothesis. Further details are presented in the **SM4**. In addition to the more traditional acoustic features included in the previous meta-analysis, we extracted 24 novel voice quality acoustic features including both spectral and glottal properties of voice[28], for a total of 32 including those of rhythm. Median and interquartile range (IQR) were calculated for each of these measures (see **SM3**). We then estimated the differences between individuals with schizophrenia and HC for all measures (see **SM5**). We report additional analyses in **SM7** to assess the robustness of the findings: we repeat all analyses on audio segments of 6 sec. to control for recording length. The results generally support our main findings and we report here only qualitative divergences.

The relationship between acoustic features and clinical features, pharmacotherapy and social functioning.

To assess the relationship between acoustic features and clinical ratings, we built Bayesian multilevel regression models with each acoustic feature as outcome, and clinical features (one at a time) as ordinal predictors. We separately assessed the relationship between the different acoustic features and clinical ratings for each language, and modeled varying effects of participants, i.e., intercepts and slopes, separately for each language. This analysis was performed on the schizophrenia group only (see **SM4** for more details).

To assess the effect of medication, patients were divided into two categories based on the mechanism of action of antipsychotic medication[45], namely patients with (1) low D2R occupancy, i.e., Clozapine, Olanzapine, Paliperidone, Quetiapine or (2) high D2R occupancy, i.e., Aripiprazole, Amisulpride, Risperidone, Ziprasidone, Sertindole (see SM6). Antipsychotic dose was converted to chlorpromazine (CPZ) equivalents[46]. We used Bayesian

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

multilevel regression models, as described above, with each acoustic feature as outcome, and antipsychotic medication (high D2R, low D2R) and language (DK, CH, GE) as predictors. For each acoustic feature we built a first model with weakly informative priors, that is expectations of no effects of medication, and we then built a second model with informed prior (relying on DeBoer et al. 2020[33], see **SM6**). Our main aim was to assess the generalizability of previous findings on the effect of medication on speech production in schizophrenia, more than evaluating the causal pathways of medication on speech production, which would have required a larger sample and more detailed information on potential confounders (see discussion section). To assess the role of drug dosage, duration of illness (DUI) and social functioning (PSP scale), we built Bayesian multilevel regression models with each acoustic feature as outcome, and chlorpromazine equivalents, DUI and PSP scale score (one at a time) as predictors. We also compared acoustic patterns between patients with first-episode (FES) schizophrenia and chronic patients. We used Bayesian multilevel regression models with each acoustic feature as outcome, illness onset (FES patients, chronic patients and healthy controls) and language as predictors (see **SM6** for more details).

The code used for the analysis and the extracted features are openly available (see **SM8**).

Results

Effect of Diagnosis

The detailed results and comparisons to meta-analytic effects are reported in **Table 2** and **Figure 1**. We only partially replicated previous meta-analytic findings: only reduced pitch variability was found across all datasets, and with smaller effect sizes. Duration atypicalities (reduced speech rate, increased pause duration) were replicated only in the German and Danish corpora, and with smaller effect sizes. We also identified a new potential marker across languages: longer utterance duration. In agreement with the inconsistent replications, meta-analytically informed models were more robust and generalizable to new data (LOO

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

weights for skeptic models above .75) only in about half of the models, indicating that meta-analytic findings were not fully representative of the current samples.

Biological sex, age and level of intelligence of the participants also affected the group differences, although inconsistently across languages.

Our results are generally confirmed when only the 6-second segments of the audio recordings are analyzed, although with some differences: the results of the robustness analysis (see **SM7**) tend to be more consistent with the results of the meta-analysis[3] (e.g., reduced speech rate also in German, reduced speech percentage in Danish, no evidence of reduced pause duration in Chinese and Japanese). Interestingly, the robustness analysis also showed that patients with schizophrenia were more likely to give very short responses to the task than controls (total recording duration < 6sec, see **SM7**).

Table 2. Estimated standardized mean difference (HC- schizophrenia) for the eight acoustic measures present in the meta-analysis, as estimated separately by the meta-analytically informed and the skeptical models

Acoustic Features	Group (HC – SCZ)	Group * Biological sex (M-F)	Age	IQ
Pitch Median	MA priors = 0.25 (-0.72, 1.30)			
Skeptical DK	0.05 (-0.16 0.26) ER = 1.94 ER01 = 3.53	0.1 (-0.1 0.29) ER = 3.77	**0.02 (0 0.03) ER = 32.33**	0 (-0.01 0) ER = 2.09 ER01 = 119.29
Skeptical GE	-0.06 (-0.29 0.17) ER = 1.98 ER01 = 3.09	0.04 (-0.3 0.38) ER = 1.43 ER01 = 3.26	0 (-0.02 0.01) ER = 2 ER01 = 51.01	0 (-0.01 0.01) ER = 1.11 ER01 = 87.94
Skeptical CH	**0.25 (-0.05 0.55) ER = 11.05**	**0.28 (-0.01 0.57) ER = 16.76**	**0.02 (0 0.04) ER = 45.3**	0.01 (0 0.02) ER = 4.11
Skeptical JP	**-0.57 (-0.98 -0.13) ER = 58.88**	**0.63 (0.18 1.08) ER = 77.12**	-0.01 (-0.08 0.04) ER = 1.04 ER01 = 11.37	-0.01 (-0.86 0.82) ER = 1.01 ER01 = 0.93
Informed DK	0.07 (-0.15 0.28) ER = 2.3 ER01 = 3.35	NA	NA	NA
Informed GE	-0.04 (-0.26 0.18) ER = 1.58 ER01 = 3.97	NA	NA	NA
Informed CH	**0.28 (-0.02 0.57) ER = 15**	NA	NA	NA
Informed JP	**-0.44 (-0.86 0) ER = 19.2**	NA	NA	NA
Stacking weight	**Informed Model 1.00**			
Pitch IQR	MA priors = **-0.55 (-1.06, 0.09)**			
Skeptical DK	**-0.19 (-0.32 -0.05) ER = 62.29**	0.16 (-0.1 0.41) ER = 5.29	0.01 (0 0.02) ER = 9	0 (-0.01 0) ER = 3.76
Skeptical GE	**-0.29 (-0.46 -0.11) ER = 343.83**	-0.03 (-0.34 0.28) ER = 1.29 ER01 = 3.72	0 (-0.01 0.02) ER = 1.83 ER01 = 59.83	0 (-0.01 0.01) ER = 1.46 ER01 = 128.11
Skeptical CH	**-0.2 (-0.43 0.04) ER = 11.02**	0.03 (-0.4 0.46) ER = 1.25 ER01 = 1.33	0.01 (-0.01 0.03) ER = 5.49	0 (-0.01 0) ER = 3.22

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Skeptical JP	**-0.5 (-0.85 -0.13) ER = 64.79**	0.09 (-0.59 0.77) ER = 1.43 ER01 = 1.75	0.01 (-0.02 0.03) ER = 1.92 ER01 = 27.15	0 (-0.82 0.83) ER = 1.01 ER01 = 0.97
Informed DK	**-0.24 (-0.37 -0.1) ER = 832.33**	NA	NA	NA
Informed GE	**-0.35 (-0.52 -0.19) ER = 9999**	NA	NA	NA
Informed CH	**-0.31 (-0.52 -0.1) ER = 134.14**	NA	NA	NA
Informed JP	**-0.57 (-0.84 -0.3) ER = 2499**	NA	NA	NA
Stacking weight	**Informed Model 1.00**			
Speech Rate	MA priors -0.75 (-1.51, 0.04)			
Skeptical DK	**-0.37 (-0.54 -0.22) ER = 9999**	0.22 (-0.09 0.53) ER = 6.99	0 (-0.02 0.01) ER = 1.21 ER01 = 57.1	0 (-0.01 0.01) ER = 1.98 ER01 = 109.93
Skeptical GE	**-0.12 (-0.35 0.1) ER = 4.46**	0.29 (-0.17 0.74) ER = 5.49	**0.01 (0 0.03) ER = 11.95**	0 (-0.01 0.01) ER = 1.09 ER01 = 105.22
Skeptical CH	0.07 (-0.18 0.32) ER = 2.09 ER01 = 2.94	-0.1 (-0.56 0.36) ER = 1.8 ER01 = 2.94	0.01 (-0.01 0.03) ER = 3.14	0 (0 0.01) ER = 5.52
Skeptical JP	0.13 (-0.28 0.55) ER = 2.43 ER01 = 1.68	-0.39 (-1.12 0.34) ER = 4.43	0.02 (-0.01 0.06) ER = 5.08	0 (-0.82 0.83) ER = 1.01 ER01 = 1
Informed DK	**-0.41 (-0.57 -0.25) ER = 9999**	NA	NA	NA
Informed GE	**-0.21 (-0.43 0) ER = 19.58**	NA	NA	NA
Informed CH	-0.04 (-0.28 0.19) ER = 1.62 ER01 = 19.09	NA	NA	NA
Informed JP	-0.19 (-0.58 0.19) ER = 3.77	NA	NA	NA
Stacking weight	**Skeptic Model 1.00**			
Speech Percentage	MA priors -1.26 (-2.26, 0.25)			
Skeptical DK	-0.02 (-0.19 0.15) ER = 1.4 ER01 = 4.7	-0.04 (-0.36 0.28) ER = 1.35 ER01 = 3.67	0 (-0.01 0.02) ER = 1.69 ER01 = 46.55	0 (-0.01 0.01) ER = 1.51 ER01 = 114.89
Skeptical GE	0.07 (-0.18 0.3) ER = 2.24 ER01 = 2.94	0.25 (-0.23 0.73) ER = 3.99	**0.02 (0 0.03) ER = 15.08**	0 (0 0.01) ER = 7.16
Skeptical CH	**0.43 (0.17 0.68) ER = 332.33**	-0.25 (-0.75 0.24) ER = 4.05	**0.02 (0 0.04) ER = 10.64**	**0.01 (0 0.02) ER = 13.86**
Skeptical JP	**0.34 (-0.07 0.75) ER = 10.82**	-0.22 (-0.95 0.48) ER = 2.28 ER01 = 1.47	**0.04 (0.01 0.07) ER = 43.05**	-0.01 (-0.82 0.8) ER = 1.02 ER01 = 1.02
Informed DK	-0.09 (-0.26 0.09) ER = 3.87	NA	NA	NA
Informed GE	-0.06 (-0.3 0.18) ER = 1.86 ER01 = 109.19	NA	NA	NA
Informed CH	**0.28 (0.02 0.53) ER = 24.91**	NA	NA	NA
Informed JP	-0.03 (-0.46 0.39) ER = 1.13 ER01 = 62.41	NA	NA	NA
Stacking weight	**Skeptic Model 0.50**			
Pause Number	MA priors 0.05 (-1.23, 1.13)			
Skeptical DK	**-0.18 (-0.31 -0.05) ER = 73.63**	0.34 (0.07 0.6) ER = 57.14	-0.01 (-0.02 0.01) ER = 5.33	0 (0 0.01) ER = 3.27
Skeptical GE	**-0.22 (-0.42 -0.02) ER = 32.22**	0.15 (-0.24 0.55) ER = 2.78 ER01 = 2.43	0.01 (0 0.03) ER = 7.36	0 (-0.01 0.01) ER = 2.06 ER01 = 99.91
Skeptical CH	-0.17 (-0.4 0.06) ER = 8.16	-0.34 (-0.77 0.1) ER = 8.99	0 (-0.02 0.02) ER = 1.23 ER01 = 41.58	0 (-0.01 0) ER = 5.31
Skeptical JP	-0.06 (-0.44 0.32) ER = 1.58 ER01 = 2.15	-0.28 (-0.97 0.43) ER = 2.94 ER01 = 1.35	**0.03 (0 0.06) ER = 20.1**	0 (-0.82 0.81) ER = 1.02 ER01 = 1.03
Informed DK	**-0.18 (-0.31 -0.04) ER = 64.36**	NA	NA	NA

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Informed GE	**-0.21 (-0.41 -0.02) ER = 25.18**	NA	NA	NA
Informed CH	-0.16 (-0.4 0.08) ER = 6.82	NA	NA	NA
Informed JP	-0.05 (-0.42 0.32) ER = 1.38 ER01 = 2.24	NA	NA	NA
Stacking weight	**Skeptic Model 1.00**			
Pause duration	MA priors 1.89 (0.72, 3.21)			
Skeptical DK	**0.23 (0.09 0.38) ER = 180.82**	-0.13 (-0.41 0.14) ER = 3.48	0 (-0.02 0.01) ER = 1.93 ER01 = 53.4	0 (-0.01 0.01) ER = 1.34 ER01 = 103.28
Skeptical GE	**0.17 (-0.03 0.37) ER = 12.04**	-0.17 (-0.59 0.26) ER = 2.91 ER01 = 2.18	-0.01 (-0.03 0) ER = 9.73	0 (-0.01 0) ER = 1.97 ER01 = 110.6
Skeptical CH	**-0.19 (-0.42 0.03) ER = 11.5**	0.27 (-0.14 0.69) ER = 6.09	-0.02 (-0.04 0) ER = 8.67	0 (-0.01 0) ER = 2.71 ER01 = 85.99
Skeptical JP	**-0.3 (-0.65 0.07) ER = 10.52**	0.28 (-0.4 0.95) ER = 3.05	**-0.04 (-0.07 0) ER = 30.65**	0 (-0.82 0.81) ER = 1.02 ER01 = 1.02
Informed DK	**0.3 (0.16 0.44) ER = 4999**	NA	NA	NA
Informed GE	**0.31 (0.11 0.51) ER = 177.57**	NA	NA	NA
Informed CH	-0.02 (-0.24 0.21) ER = 1.36 ER01 = 4640.1	NA	NA	NA
Informed JP	0.17 (-0.21 0.58) ER = 3.08	NA	NA	NA
Stacking weight	**Informed Model 0.77**			
Utterance Duration	MA priors -0.13 (-1.571 1.342)			
Skeptical DK	**0.17 (0.02 0.33) ER = 29.12**	**-0.34 (-0.67 -0.02) ER = 23.57**	0.01 (-0.01 0.02) ER = 2.86 ER01 = 43.25	**-0.01 (-0.01 0) ER = 10.66**
Skeptical GE	**0.37 (0.15 0.58) ER = 332.33**	0.1 (-0.35 0.55) ER = 1.77 ER01 = 2.42	**0.01 (0 0.03) ER = 12.79**	0 (0 0.01) ER = 3.5
Skeptical CH	**0.21 (-0.03 0.44) ER = 12.39**	-0.17 (-0.56 0.23) ER = 3.29	0 (-0.01 0.01) ER = 2 ER01 = 64.92	0 (0 0.01) ER = 3.2
Skeptical JP	**0.81 (0.45 1.15) ER = 1427.57**	-0.05 (-0.84 0.72) ER = 1.13 ER01 = 1.52	**0.04 (0.01 0.07) ER = 42.29**	-0.01 (-0.84 0.82) ER = 1.01 ER01 = 0.96
Informed DK	**0.18 (0.02 0.34) ER = 27.65**	NA	NA	NA
Informed GE	**0.37 (0.16 0.58) ER = 475.19**	NA	NA	NA
Informed CH	**0.21 (-0.02 0.44) ER = 13.99**	NA	NA	NA
Informed JP	**0.89 (0.51 1.25) ER = 4999**	NA	NA	NA
Stacking weight	**Skeptic Model 1.00**			
Number of Utterance				
Skeptical DK	**-0.21 (-0.35 -0.07) ER = 171.41**	**0.37 (0.1 0.64) ER = 76.52**	-0.01 (-0.02 0.01) ER = 4.31	0 (0 0.01) ER = 3.39
Skeptical GE	**-0.23 (-0.43 -0.03) ER = 29.21**	0.18 (-0.22 0.58) ER = 3.37	0.01 (0 0.03) ER = 8.36	0 (0 0.01) ER = 2.45 ER01 = 97.14
Skeptical CH	-0.14 (-0.37 0.1) ER = 4.93	**-0.37 (-0.81 0.05) ER = 12.12**	0 (-0.02 0.02) ER = 1 ER01 = 41.2	0 (-0.01 0) ER = 5.51
Skeptical JP	-0.1 (-0.48 0.27) ER = 2.12 ER01 = 1.99	-0.34 (-1.02 0.33) ER = 3.78	**0.03 (0 0.06) ER = 11.5**	0 (-0.83 0.83) ER = 0.98 ER01 = 1.07
Total Duration				
Skeptical DK	-0.03 (-0.22 0.15) ER = 1.64 ER01 = 4.24	0.19 (-0.17 0.56) ER = 4.16	0 (-0.02 0.02) ER = 1.02 ER01 = 48.82	0 (0 0.01) ER = 2.19 ER01 = 141.8
Skeptical GE	0.01 (-0.19 0.22) ER = 1.17 ER01 = 3.93	0.14 (-0.27 0.53) ER = 2.56 ER01 = 2.56	0 (-0.01 0.02) ER = 1.92 ER01 = 45.72	0 (-0.01 0.01) ER = 1.2 ER01 = 110.36

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

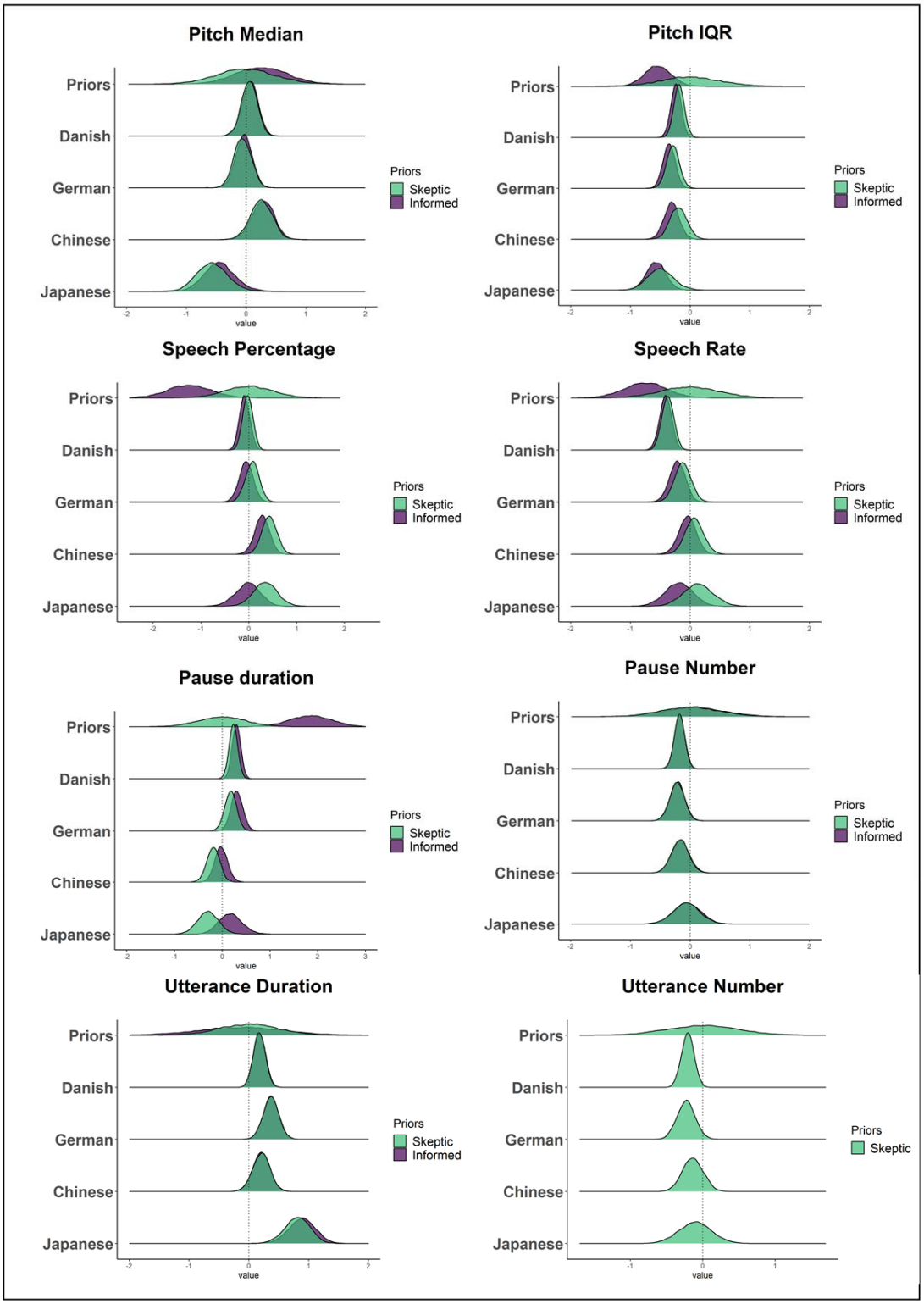
Skeptical CH	**-0.21 (-0.43 0.02) ER = 13.06**	0.54 (0.11 0.97) ER = 50.02	0 (-0.02 0.02) ER = 1.19 ER01 = 41.85	0 (-0.01 0.01) ER = 1.18 ER01 = 100.72
Skeptical JP	-0.27 (-0.63 0.12) ER = 7.35	-0.08 (-0.72 0.55) ER = 1.42 ER01 = 1.75	**-0.03 (-0.06 0) ER = 17.66**	0 (-0.82 0.85) ER = 0.99 ER01 = 1.02

Note: The first column reports the main effect of the diagnostic group (across sex and age), respectively from the meta-analysis (MA), for the skeptical Danish (DK), German (GE), Chinese (CH) and Japanese (JP) models, and for the informed ones. The second column indicates the interaction between the effect of the diagnostic group and biological sex (male–female), that is, the difference in effect of group between the male and the female participants. The third column reports the interaction between the effect of diagnostic group and age, that is, the change in effect size as age increases by 1 standard deviation (SD). The fourth column reports the interaction between the effect of diagnostic group and level of intelligence (IQ), the change in effect size as IQ increases by 1 SD. ER indicates the evidence ratio for the difference, ER01 the evidence ratio for the null effect. Bold text indicates findings for which there is more than anecdotal evidence (evidence ratio above 10). Abbreviations: SCZ, schizophrenia; HC, healthy controls. NA = not applicable.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



1

2

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Novel Acoustic features

The detailed results are reported in **Table S5_A**. Generally, we found some evidence for reduced formants median frequency and formants variability, and for increased median relative amplitude (H1H2) and reduced H1H2 variability. However, these findings are small and not robust across languages. Results for the other novel features are very uncertain or inconsistent across languages. As for more traditional features, we found that sex, age, and level of intelligence again affect several differences.

Effect of Symptoms

Detailed results and comparison with meta-analytic findings are reported in **Table 3**. Clinical features generally correlated with acoustic features, and these associations were in line with meta-analytic priors. However, we did not find reliable and robust associations across all languages. The associations are generally stronger for temporal-duration measures (lower speech percentage, increased pause duration and reduced speech rate are associated with higher flat affect, alogia and negative symptom severity), whereas pitch measures (median and IQR) are more weakly associated with clinical ratings. Most of the correlations are between small and moderate (5-16% of explained variance), and vary across languages and rating scales.

Table 3. Estimated standardized relation between acoustic and clinical features. ER indicates the evidence ratio for the difference, ER01 the evidence ratio for the null effect.

Rating scales	SANS	SAPS	SANS - Alogia	SANS - Flat affect
Pitch Median				
MA priors	0.096 [-0.158, 0.346]	-0.185 [-0.691, 0.316]		
Skeptical DK	-0.01 (-0.19 0.15) ER = 1.22 ER01 = 6.52	0.1 (-0.05 0.25) ER = 6.87	0.11 (-0.05 0.27) ER = 6.77	-0.04 (-0.19 0.11) ER = 2.01 ER01 = 6.14
Skeptical CH	-0.01 (-0.35 0.33) ER = 1.13 ER01 = 3.12	-0.07 (-0.3 0.14) ER = 2.34 ER01 = 4.07	-0.03 (-0.47 0.4) ER = 1.21 ER01 = 2.41	-0.15 (-0.45 0.15) ER = 4.39
Informed DK	-0.02 (-0.19 0.16) ER = 1.32 ER01 = 3.89	0.1 (-0.04 0.25) ER = 8.39		

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Informed CH	0.04 (-0.26 0.34) ER = 1.54 ER01 = 2.16	-0.08 (-0.3 0.13) ER = 2.89 ER01 = 4.52		
Stacking weight	**Informed Model 1.00**	**Informed Model 0.56**		
Pitch IQR				
MA priors	-0.01 [-0.196, 0.144]	-0.027 [-0.686, 0.763]	-0.035 [-0.317, 0.22]	-0.106 [-0.262, -0.047]
Skeptical DK	0.01 (-0.02 0.05) ER = 2.91 ER01 = 24.69	-0.01 (-0.03 0.02) ER = 1.64 ER01 = 33.01	0.01 (-0.02 0.04) ER = 1.69 ER01 = 31.04	0 (-0.03 0.02) ER = 1.4 ER01 = 39.65
Skeptical CH	-0.06 (-0.25 0.1) ER = 2.48 ER01 = 5.84	-0.03 (-0.14 0.07) ER = 2.21 ER01 = 8.3	**-0.32 (-0.6 -0.1) ER = 124**	**-0.26 (-0.42 -0.11) ER = 599**
Informed DK	0.01 (-0.02 0.05) ER = 2.95 ER01 = 11.54	-0.01 (-0.03 0.02) ER = 1.62 ER01 = 51.25	0.01 (-0.02 0.04) ER = 1.66 ER01 = 20.87	0 (-0.03 0.02) ER = 1.45 ER01 = 16.74
Informed CH	-0.05 (-0.23 0.11) ER = 2.15 ER01 = 2.47	-0.03 (-0.14 0.07) ER = 2.24 ER01 = 11.95	**-0.3 (-0.53 -0.1) ER = 141.86**	**-0.24 (-0.39 -0.1) ER = 749**
Stacking weight	**Informed Model 1.00**	**Skeptic Model 1.00**	**Skeptic Model 1.00**	**Informed Model 1.00**
Speech percentage				
MA priors	-0.229 [-0.499, 0.035]		-0.413 [-0.723, -0.07]	-0.1 (-0.18 -0.01) ER = 34.29
Skeptical DK	**-0.12 (-0.23 -0.02) ER = 47.78**	0.04 (-0.05 0.13) ER = 3.37	-0.03 (-0.12 0.06) ER = 2.17 ER01 = 9.57	**-0.09 (-0.17 0) ER = 23.49**
Skeptical CH	-0.02 (-0.3 0.24) ER = 1.15 ER01 = 3.88	0.07 (-0.1 0.24) ER = 3.17	-0.11 (-0.46 0.21) ER = 2.48 ER01 = 2.76	-0.06 (-0.3 0.17) ER = 2.03 ER01 = 4.01
Informed DK	**-0.13 (-0.24 -0.02) ER = 51.63**	NA	-0.03 (-0.13 0.06) ER = 2.55 ER01 = 10.71	**-0.09 (-0.18 -0.01) ER = 31.97**
Informed CH	-0.05 (-0.33 0.2) ER = 1.62 ER01 = 3.16	NA	-0.18 (-0.54 0.14) ER = 4.28	-0.11 (-0.34 0.12) ER = 3.46
Stacking weight	**Informed Model 1.00**		**Informed Model 0.88**	**Informed Model 1.00**
Speech rate				
Skeptical DK	**-0.18 (-0.28 -0.1) ER = 1999**	-0.02 (-0.09 0.06) ER = 1.91 ER01 = 13.81	**-0.13 (-0.22 -0.05) ER = 499**	**-0.11 (-0.18 -0.04) ER = 332.33**
Skeptical CH	-0.15 (-0.38 0.05) ER = 8.09	0.05 (-0.09 0.19) ER = 2.87 ER01 = 6.31	**-0.22 (-0.54 0.06) ER = 9.64**	**-0.17 (-0.37 0.01) ER = 14.62**
Pause duration				
MA priors	0.302 [-0.199, 0.783]			
Skeptical DK	**0.06 (0.02 0.11) ER = 186.5**	0 (-0.03 0.04) ER = 1.16 ER01 = 29.07	**0.04 (0 0.08) ER = 24.1**	**0.03 (0 0.07) ER = 21.99**
Skeptical CH	-0.03 (-0.22 0.16) ER = 1.52 ER01 = 5.35	-0.02 (-0.14 0.11) ER = 1.43 ER01 = 8.44	-0.02 (-0.29 0.25) ER = 1.17 ER01 = 4.15	0.01 (-0.17 0.2) ER = 1.21 ER01 = 5.81
Informed DK	**0.07 (0.02 0.11) ER = 221.22**	NA	NA	NA
Informed CH	-0.03 (-0.23 0.18) ER = 1.42 ER01 = 6.85	NA	NA	NA
Stacking weight	**Informed Model 1.00**			
Number of pauses				

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Skeptical DK	**-0.06 (-0.13 0.01) ER = 10.11**	0.02 (-0.04 0.08) ER = 2.1 ER01 = 15.8	-0.03 (-0.09 0.03) ER = 3.6	**-0.05 (-0.11 0) ER = 15.81**	
Skeptical CH	0.08 (-0.14 0.3) ER = 2.86 ER01 = 3.72	**-0.15 (-0.3 -0.01) ER = 23.19**	0.19 (-0.05 0.46) ER = 9.71	0.08 (-0.08 0.26) ER = 3.91	
Duration of utterance					
Skeptical DK	**-0.07 (-0.13 -0.01) ER = 67.97**	**-0.05 (-0.09 0) ER = 16.7**	**-0.08 (-0.14 -0.03) ER = 374**	-0.03 (-0.08 0.02) ER = 6.33	
Skeptical CH	-0.1 (-0.28 0.06) ER = 5.97	**0.1 (0 0.21) ER = 19.76**	**-0.25 (-0.51 -0.04) ER = 35.14**	**-0.18 (-0.34 -0.04) ER = 49.85**	
Number of utterances					
Skeptical DK	**-0.06 (-0.13 0.01) ER = 13.89**	0.01 (-0.05 0.07) ER = 1.43 ER01 = 18.21	-0.04 (-0.1 0.02) ER = 6.63	**-0.06 (-0.11 0) ER = 20.82**	
Skeptical CH	0.06 (-0.15 0.28) ER = 2.24 ER01 = 4.43	**-0.15 (-0.3 -0.02) ER = 27.99**	0.15 (-0.09 0.41) ER = 5.74	0.06 (-0.11 0.24) ER = 2.71 ER01 = 4.99	
Rating scales	**PANSS Total**	**PANSS Negative**	**PANSS Positive**	**PANSS Conversation**	**PANSS Blunted affect**
Pitch Median					
MA priors		0.096 [-0.158, 0.346]	-0.185 [-0.691, 0.316]		
Skeptical CH	-0.3 (-0.79 0.16) ER = 5.97	-0.14 (-0.46 0.17) ER = 3.31	-0.24 (-0.6 0.11) ER = 6.73	0 (-0.25 0.26) ER = 1.1 ER01 = 3.97	-0.07 (-0.37 0.22) ER = 1.85 ER01 = 3.26
Skeptical JP	-0.43 (-1.03 0.18) ER = 7	-0.4 (-1.01 0.23) ER = 6.36	**-0.51 (-1.11 0.08) ER = 11.66**	NA	NA
Skeptical GE	-0.1 (-0.48 0.28) ER = 1.92 ER01 = 2.45	-0.03 (-0.34 0.27) ER = 1.29 ER01 = 3.41	**-0.34 (-0.71 -0.01) ER = 21.56**	-0.12 (-0.3 0.06) ER = 6.52	0.26 (-0.14 0.73) ER = 5.87
Informed CH	NA	-0.1 (-0.4 0.2) ER = 2.32 ER01 = 1.83	-0.26 (-0.62 0.1) ER = 7.72	NA	NA
Informed JP	NA	-0.18 (-0.66 0.31) ER = 2.66 ER01 = 1.06	**-0.63 (-1.28 0.03) ER = 16.7**	NA	NA
Informed GE	NA	-0.02 (-0.3 0.25) ER = 1.18 ER01 = 2.37	**-0.36 (-0.74 -0.03) ER = 25.79**	NA	NA
		Skeptic Model 1.00	**Skeptic Model 1.00**		
Pitch IQR					
MA priors	-0.091 [-0.34, 0.15]	-0.01 [-0.196, 0.144]	-0.027 [-0.686, 0.763]	0 (-0.25 0.26) ER = 1.1 ER01 = 3.97	-0.07 (-0.37 0.22) ER = 1.85 ER01 = 3.26
Skeptical CH	0.23 (-0.11 0.56) ER = 6.95	-0.15 (-0.35 0.05) ER = 8.45	**0.4 (0.24 0.57) ER = Inf**	-0.1 (-0.24 0.03) ER = 9.45	-0.07 (-0.26 0.11) ER = 2.7 ER01 = 5.13
Skeptical JP	0.04 (-0.43 0.53) ER = 1.26 ER01 = 2.21	-0.08 (-0.61 0.45) ER = 1.5 ER01 = 1.84	0.06 (-0.44 0.56) ER = 1.44 ER01 = 2.01	NA	NA
Skeptical GE	**-0.16 (-0.33 0.01) ER = 16.39**	**-0.27 (-0.44 -0.14) ER = 1999**	-0.01 (-0.17 0.16) ER = 1.29 ER01 = 5.93	-0.04 (-0.13 0.04) ER = 4.15	-0.1 (-0.34 0.08) ER = 4.43
Informed CH	0.18 (-0.14 0.48) ER = 5.4	-0.13 (-0.3 0.05) ER = 7.08	**0.41 (0.25 0.59) ER = 5999**	-0.1 (-0.24 0.03) ER = 8.85	-0.08 (-0.25 0.08) ER = 3.62
Informed JP	0 (-0.43 0.43) ER = 1.05 ER01 = 1.71	-0.05 (-0.39 0.29) ER = 1.42 ER01 = 1.18	0.06 (-0.5 0.63) ER = 1.38 ER01 = 2.72	NA	NA
Informed GE	**-0.15 (-0.32 0.01) ER = 15.57**	**-0.25 (-0.39 -0.12) ER = 1199**	-0.02 (-0.19 0.15) ER = 1.36 ER01 = 8.98	-0.05 (-0.13 0.03) ER = 4.63	-0.1 (-0.29 0.06) ER = 5.3
	Informed Model 1.00	**Skeptic Model 1.00**	**Skeptic Model 1.00**	**Skeptic Model 1.00**	**Informed Model 1.00**

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Speech percentage					
MA priors		0.302 [-0.199, 0.783]			
Skeptical CH	0.03 (-0.45 0.5) ER = 1.18 ER01 = 2.06	**-0.31 (-0.62 0) ER = 18.05**	0.14 (-0.23 0.5) ER = 2.97 ER01 = 2.22	-0.15 (-0.38 0.07) ER = 7.02	-0.2 (-0.49 0.08) ER = 7.6
Skeptical JP	0.24 (-0.25 0.73) ER = 4.03	0.14 (-0.42 0.67) ER = 2.01 ER01 = 1.72	0.19 (-0.36 0.72) ER = 2.64 ER01 = 1.55	NA	NA
Skeptical GE	**-0.26 (-0.57 0.04) ER = 11.88**	-0.12 (-0.36 0.12) ER = 3.93	-0.11 (-0.38 0.17) ER = 2.91 ER01 = 3.11	**-0.17 (-0.32 -0.03) ER = 36.97**	-0.17 (-0.5 0.13) ER = 5.63
Informed CH	0.03 (-0.45 0.53) ER = 1.18 ER01 = 2.11	**-0.32 (-0.63 -0.02) ER = 22.35**	NA	**-0.19 (-0.42 0.03) ER = 11.88**	**-0.26 (-0.48 -0.05) ER = 51.63**
Informed JP	0.23 (-0.29 0.74) ER = 3.56	0 (-0.51 0.48) ER = 0.97 ER01 = 1.75	NA	NA	NA
Informed GE	**-0.26 (-0.57 0.04) ER = 11.42**	-0.14 (-0.38 0.08) ER = 5.51	NA	**-0.19 (-0.34 -0.04) ER = 56.14**	**-0.23 (-0.51 -0.03) ER = 34.71**
	Informed Model 1.00	**Informed Model 1.00**		**Skeptic Model 1.00**	**Informed Model 1.00**
Speech Rate					
Skeptical CH	**-0.38 (-0.76 -0.01) ER = 20.28**	-0.16 (-0.41 0.08) ER = 6.48	-0.01 (-0.28 0.27) ER = 1.13 ER01 = 3.78	-0.13 (-0.32 0.06) ER = 6.8	-0.11 (-0.35 0.11) ER = 3.74
Skeptical JP	0.26 (-0.28 0.8) ER = 3.98	0.15 (-0.46 0.74) ER = 2.05 ER01 = 1.44	0.3 (-0.28 0.84) ER = 4.45	NA	NA
Skeptical GE	**-0.39 (-0.69 -0.1) ER = 61.5**	**-0.24 (-0.47 -0.02) ER = 23.9**	-0.12 (-0.4 0.15) ER = 3.22	**-0.18 (-0.34 -0.04) ER = 51.63**	-0.21 (-0.55 0.08) ER = 8.01
Mean Pause Duration					
MA priors		0.302 [-0.199, 0.783]			
Skeptical CH	0.14 (-0.24 0.49) ER = 2.78 ER01 = 2.15	**0.38 (0.22 0.54) ER = 1199**	-0.12 (-0.39 0.15) ER = 3.39	**0.23 (0.1 0.36) ER = 351.94**	**0.26 (0.11 0.43) ER = 351.94**
Skeptical JP	**-0.5 (-0.96 -0.01) ER = 20.98**	-0.42 (-0.96 0.13) ER = 8.93	**-0.49 (-1 0.02) ER = 16.05**		
Skeptical GE	**0.11 (0 0.23) ER = 18.54**	**0.07 (-0.02 0.16) ER = 10.45**	0.02 (-0.09 0.13) ER = 1.85 ER01 = 9	**0.11 (0.04 0.17) ER = 299**	**0.11 (-0.01 0.28) ER = 15.48**
Informed CH	NA	**0.39 (0.24 0.55) ER = 5999**	NA	NA	NA
Informed JP	NA	-0.37 (-0.95 0.21) ER = 6.18	NA	NA	NA
Informed GE	NA	**0.07 (-0.01 0.16) ER = 10.83**	NA	NA	NA
		Informed Model 1.00			
Pause Number					
Skeptical CH	**-0.51 (-0.94 -0.08) ER = 33.09**	**-0.32 (-0.62 -0.04) ER = 29.15**	0.02 (-0.32 0.37) ER = 1.15 ER01 = 2.78	**-0.3 (-0.5 -0.09) ER = 92.75**	**-0.25 (-0.52 0) ER = 19.91**
Skeptical JP	0.24 (-0.26 0.73) ER = 3.94	0.14 (-0.41 0.67) ER = 2.07 ER01 = 1.79	0.33 (-0.2 0.81) ER = 6.23	NA	NA
Skeptical GE	**-0.2 (-0.39 -0.03) ER = 32.9**	**-0.12 (-0.26 0.02) ER = 12.95**	-0.02 (-0.19 0.14) ER = 1.52 ER01 = 5.83	**-0.14 (-0.23 -0.06) ER = 271.73**	**-0.15 (-0.39 0.03) ER = 11.35**
Utterance duration					
Skeptical CH	**0.3 (0 0.59) ER = 18.05**	**0.2 (0.02 0.38) ER = 29.61**	0.1 (-0.12 0.32) ER = 3.63	0.09 (-0.03 0.22) ER = 8.51	**0.19 (0.06 0.33) ER = 60.22**
Skeptical JP	0.06 (-0.44 0.54) ER = 1.5 ER01 = 1.96	-0.01 (-0.53 0.49) ER = 1.04 ER01 = 2.05	0.11 (-0.4 0.6) ER = 1.94 ER01 = 1.87	NA	NA

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Skeptical GE	**-0.19 (-0.33 -0.07) ER = 110.11**	**-0.17 (-0.29 -0.07) ER = 544.45**	-0.07 (-0.21 0.06) ER = 4.38	**-0.07 (-0.14 0) ER = 18.29**	-0.07 (-0.27 0.1) ER = 3.06
Utterance Number					
Skeptical CH	**-0.49 (-0.9 -0.07) ER = 34.71**	**-0.34 (-0.62 -0.07) ER = 40.1**	0.03 (-0.3 0.35) ER = 1.25 ER01 = 3.21	**-0.3 (-0.5 -0.11) ER = 145.34**	**-0.24 (-0.5 -0.01) ER = 21.47**
Skeptical JP	0.25 (-0.25 0.73) ER = 4.16	0.12 (-0.42 0.65) ER = 1.83 ER01 = 1.68	0.35 (-0.17 0.83) ER = 7.15	NA	NA
Skeptical GE	**-0.22 (-0.41 -0.03) ER = 39**	**-0.14 (-0.29 0) ER = 20.66**	-0.03 (-0.21 0.14) ER = 1.68 ER01 = 5.22	**-0.15 (-0.25 -0.07) ER = 314.79**	**-0.15 (-0.38 0.04) ER = 9.95**

Note: Abbreviations: ER, evidence ratio for the difference; ER01, evidence ratio for the null effect. The Scale for the Assessment of Negative Symptoms (SANS), the Scale for the Assessment of Positive Symptoms (SAPS), and the Positive and Negative Syndrome Scale (PANSS). Danish (DK), German (GE), Chinese (CH) and Japanese (JP) models. NA = not applicable.

Medication, duration of illness and social functioning (PSP)

We found that medication was related to vocal patterns, but inconsistently across languages. The more widespread patterns were that patients who use high D2R occupancy drugs, compared to patients who use low D2R occupancy drugs, show reduced pitch variability, higher number of pauses, longer utterance duration and higher speech rate. Further, higher drug dosage (CPZ equivalents) was associated with lower pitch median, lower speech rate, increased pause and utterance duration and reduced total number of words (see **Table S6_B**). We generally found no difference between patients with FES and patients with chronic schizophrenia, and only weak evidence supporting a role of illness duration: acoustic atypicalities are present already at disease onset and not just associated with a longer and more severe course of the disorder (see **Table S6_C**). Finally, we found that increased speech percentage and reduced pause duration were associated with higher scores on PSP, although only in the Danish corpus (see **Table S6_D**).

Discussion

In the present paper, we aimed at developing a critical, cumulative scientific approach to the understanding of vocal and prosodic atypicalities in schizophrenia. Relying on a previous meta-analysis[3] of the field, we systematically assessed the generalizability of established and

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

novel acoustic markers on a new large cross-linguistic dataset. We also assessed whether explicitly incorporating previous findings as informed priors would increase the generalizability of the results and provide additional insights on heterogeneity of the findings compared to previous literature.

Is there a universal generalizable acoustic profile of schizophrenia?

Our study assessed the generalizability of findings across a heterogeneous dataset (4 different languages). In other words, we assessed whether previous findings would be shown in a new study applying analogous experimental and/or statistical procedures, i.e., replication[47], "to populations with for instance a different language, age distribution, or other demographic and clinical characteristics", i.e., generalization[48]. We only found a minimal generalizable acoustic profile of schizophrenia: reduced pitch variability and increased utterance duration, albeit with modest effect sizes. Given the heterogeneity of previous studies and uncertainty about publication bias reported in a previous meta-analysis[3], even these minimal cross-linguistically generalized findings are far from trivial.

One possible mechanism for the generalizable acoustic profile is a relation to negative symptoms, emotional and effort-related ones. Reduced pitch variability is related to monotone speech and flat affect[49–51], and increased utterance duration to lower energy and increased vocal effort[52,53]. Further, the most promising of the novel features we investigated (e.g., reduced formant frequency and increased H1H2, albeit not consistent across all languages) also fit this explanation[54–56]. Reduced formant frequencies have been found to be associated with clinical ratings of blunted affect and alogia in schizophrenia[55], and a decrease in articulatory effort in patients with depression[28]. Furthermore, many studies in patients with depression have shown that an increase in H1H2 can be considered one of the acoustic indicators of breathiness and associated with psychomotor retardation[57]. However, we should

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

be skeptical of any simplistic explanation as yet, since we did not find a cross-linguistically robust association between these acoustic features and clinical ratings of negative symptoms. In particular, we argue that heterogeneity of the studies and samples has not been insufficiently accounted for, so far.

Source of heterogeneity in the voice profiles of patients with schizophrenia

The second crucial contribution of this study is highlighting the importance and complexity of clinical, socio-demographic, contextual (e.g., speech task) and linguistic differences in assessing vocal markers of schizophrenia.

Clinical heterogeneity

Individuals with schizophrenia present wildly heterogeneous constellations of clinical features: symptoms, onset and duration of the disorder, as well as medications[24,27]. Moreover, some of the clinical ratings are based on the perceptual assessment of speech features, and accordingly, we should expect acoustic heterogeneity co-varying with clinical heterogeneity[36,54,55]. Indeed, we found associations between acoustic parameters and clinical ratings, with duration aspects being more closely related to clinical ratings, in line with meta-analytic findings[3]. Lower proportion of speech and reduced speech rate, as well as longer pause duration were generally associated with higher ratings of negative symptoms and, in particular, alogia and flat affect.

However, the acoustic atypicalities were generally smaller than differences in clinical ratings[33,43], and their relation to clinical ratings was often inconsistent across languages and rating scales. This might have several explanations. The first is that the acoustic features analyzed are only a subset of those actually used by clinicians to produce clinical ratings of alogia and blunted affect[15]. Different approaches using larger sets of acoustic features and machine learning techniques could be required to better characterize the acoustic markers of clinical ratings[54–56]. Second, the divergence between acoustic features and clinical ratings can

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

be explained by the fact that these two signals have different temporal (i.e., precision with respect to time) and spatial (i.e., precision with respect to environmental changes) resolution, thus expecting high convergence between the two may not be realistic[49]. Third, different clinical scales might not be fully overlapping in the symptoms they include and in their definition of the symptoms[58,59], and linguistic and cultural differences may also affect the frequency, expression and rating of symptoms, thus generating several inconsistencies[60,61].

Different clinical profiles also often imply different medication profiles, and different medications can differentially impact vocal production[41]. Indeed, D2R drug occupancy and medication dosage were shown to relate to acoustic patterns, albeit inconsistently across languages. This suggests that the field needs a more fine-grained assessment of the different medications involved in schizophrenia, its comorbidities and their impact on voice.

Finally, we found no differences between patients with FES and chronic patients, and a limited role of illness duration on acoustic profiles: vocal atypicalities are already present at the onset of the disease and not only associated with a longer and more severe course of the disorder, thus having the potential for tracking its development and monitoring the symptomatology over time.

Socio-demographic heterogeneity

Another crucial source of heterogeneity in acoustic patterns is socio-demographic heterogeneity. Indeed, we found several reliable effects of sex and age on vocal atypicalities. However, the picture is currently very sparse, and more than identifying systematic effects, we rather recommend socio-demographic variables be taken explicitly into account. Promising progress has been made in normative modeling[62,63], which relies on large samples to provide expectations, accounting for clinical and socio-demographic features, and assess individual deviations from such expectations. This approach is particularly relevant in light of recent evidence showing how computational speech analysis may be prone to serious bias

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

determined by socio-demographic factors, such as racial identity[64], as it may help to identify such potential bias. An alternative approach is to collect larger samples and use propensity scores[65,66] to better match patients and controls and account for potential confounders. These approaches may also help to tackle the problem of selection bias, as most previous studies have used convenience samples with relatively homogeneous clinical and sociodemographic features (e.g., chronic or FES patients), that may not be fully representative of the schizophrenia spectrum and its heterogeneity.

Linguistic heterogeneity

Not least, albeit previously neglected, linguistic and cultural differences can also play a role. We modeled each language separately to account for the heterogeneity in the different corpora, thus allowing us to compare results across languages. In general, we found that atypical voice patterns were more similar within Germanic (Danish and German) and non-Germanic (Japanese and Chinese) language families. The most prominent results of the previous meta-analysis[3] (reduced speech percentage and speech rate, longer pause duration) were not consistently found across all languages in our study. One possible source of this discrepancy is the amount of noise and bias present in previous studies. Meta-analyses are often afflicted by publication bias, and heterogeneous quality in the estimates being collated, which often results in consistent discrepancies with multi-lab replications[67]. However, a complementary explanation is that quantitative vocal aspects may vary with the languages being studied, and that vocal atypicalities reported to be abnormal in schizophrenia may not be universally expressed in the same way. For instance, prosodic use of speech rate and pauses to create emphasis may differ across languages. Thus, it is possible that decreased prosodic emphasis is a universal feature of schizophrenia, but the acoustic ways in which this can be measured vary across languages. Vocal patterns must also be considered in relation to

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

their cultural and linguistic context[8,68], and the generalizability across social and linguistic groups should be systematically tested in future studies.

Acoustic heterogeneity and speech task

The use of a common standardized task (i.e., the ATT[41]) allowed us to compare voice patterns across the different corpora. This is particularly important considering that a previous meta-analysis[3] has shown that varying speech tasks may yield different outcomes, and recent studies further support this finding[55,51]. Further, our task included repeated measurements and allowed us to model intra-speaker variability, which previous studies found to be relevant[69–71]. Even by controlling for intra-speaker variability and keeping the task constant, we found large and important differences in voice patterns between the corpora. Further, we noticed that the patients produced very short responses (< 6 sec) much more frequently than controls, and when excluding these trials, we observed that results were more in line with meta-analytic findings[3]. This suggests that acoustic patterns may be task-related (with the task also interacting with language) and not robust across different tasks.

Limitations and future perspectives

The large differences in terms of the clinical and socio-demographic features within our corpus is a limitation of the present study. For example, the sample size differs across languages, and the samples in the different languages are not exactly matched in terms of the relevant clinical features, such as medication or symptom profiles. Thus, larger sample sizes might be more representative than smaller ones, and partially different subpopulations might be investigated across languages. Further, even if we kept the task and the experimental procedure constant across sites, a more controlled acoustic setting, with high-quality headset microphones placed at a constant distance from the mouth, would have allowed to further reduce potential differences across the different recording settings. This variability, albeit limited, between our corpora in terms of these features, may have contributed to the

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

differences in the main results; however, it also provided an opportunity to assess the generalizability of results across more varied conditions and thus provided more generalizable results. Moreover, although we collected a larger multilingual sample compared to previous studies[3,72], an ideal sample systematically representative of the schizophrenic spectrum and its clinical and sociodemographic variability would be even larger[63]. Future efforts should thus be directed to build a large open cross-linguistic corpus of speech recordings of patients with schizophrenia able to capture linguistic, cultural, socio-economic and clinical variability. This multilingual corpus may represent an ideal benchmark dataset for testing the reliability and generalizability (e.g., out of sample predictability) of voice analysis results in schizophrenia[73], and the necessary ground for assessing its clinical applicability[74]. Not least, future studies should focus more on cross-diagnostic comparisons aimed at capturing symptom dimensions which extend over a single disorder[75,76], and implement longitudinal designs able to test more complex hypothesis on the interaction between antipsychotic medication type and dosage, clinical (e.g., illness severity and duration) and sociodemographic (e.g., sex differences[77]) characteristics, and speech production.

Another limitation is that we focused on single features as markers of schizophrenia, and did not vary the speech task. For example, the specific social context in which the speech task takes place, the social actors involved in it, and the communicative goal to be fulfilled can influence speech production and thus acoustic patterns. Speaking to a superior vs a peer, having a formal interview vs. an informal chat, all involve different prosodic patterns. Individual participants might perceive the experiment context differently from each other, some as more formal than others; and this is further complicated by cultural factors affecting how interactions with researchers are perceived and dealt with, and therefore the acoustic patterns. Further, linguistic and vocal patterns are inherently multidimensional, with different acoustic features interacting with each other, and cross-linguistic variations potentially

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

affecting these interactions. Looking at single features could thus be reductive: future studies should focus more on examining patterns of shared variance across features[78-80], their relations with different speech tasks (in terms of social and cognitive demands)[78,79], and with linguistic variability[8,80]. However, this would require testing more fine-grained hypotheses on mechanisms relying on formal linguistic theory[68] and on psychopathological functioning theory.

Conclusion

Overall, we found scarce evidence for a universal, distinctive vocal pattern that characterizes schizophrenia: vocal patterns are highly heterogeneous with different sources of heterogeneity interacting with each other at different levels. These results raise some questions on generalizability of previous findings and on the possibility to cumulatively build on them[81]. However, they also indicate where future attempts may be directed: a larger shared multilingual corpus representative of the heterogeneity of the schizophrenic spectrum, a more explicit focus on multidimensional acoustic patterns and their relations with speech tasks, and a self-correcting cumulative approach. These are the necessary conditions to develop effective clinical applications in the near future that can target different ranges of patients and address the issue of potential bias.

Acknowledgments

A.P is supported by a Marie Skłodowska-Curie Actions – H2020-MSCA-IF-2018 grant (ID: 832518, Project: MOVES). A.S is supported by the Carlsberg Foundation. K. K has been supported by the Japan Society for the Promotion of Science (JSPS) (PE 07550). The project has been supported by seed funding from the Interacting Minds Center, Aarhus University.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Bibliography

1. Bleuler E, Aschaffenburg G. *Handbuch Der Psychiatrie*. (Deuticke, ed.).; 1911.
2. Kraepelin E. *Dementia Precox and Paraphrenia*. University of Edinburgh; 1919.
3. Parola A, Simonsen A, Bliksted V, Fusaroli R. Voice patterns in schizophrenia: A systematic review and Bayesian meta-analysis. *Schizophr Res*. Published online 2020. doi:10.1016/j.schres.2019.11.031
4. Couture SM, Granholm EL, Fish SC. A path model investigation of neurocognition, theory of mind, social competence, negative symptoms and real-world functioning in schizophrenia. *Schizophr Res*. 2011;125(2-3):152-160. doi:10.1016/j.schres.2010.09.020
5. Rabinowitz J, Levine SZ, Garibaldi G, Bugarski-Kirola D, Berardo CG, Kapur S. Negative symptoms have greater impact on functioning than positive symptoms in schizophrenia: Analysis of CATIE data. *Schizophr Res*. 2012;137(1-3):147-150. doi:10.1016/j.schres.2012.01.015
6. Häfner H, Löffler W, Maurer K, Hambrecht M, An Der Heiden W. Depression, negative symptoms, social stagnation and social decline in the early course of schizophrenia. *Acta Psychiatr Scand*. 1999;100(2):105-118. doi:10.1111/j.1600-0447.1999.tb10831.x
7. Tandon R, Keshavan MS, Nasrallah H a. Schizophrenia, “just the facts” what we know in 2008. 2. Epidemiology and etiology. *Schizophr Res*. 2008;102(1-3):1-18. doi:10.1016/j.schres.2008.04.011

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

8. Palaniyappan L. More than a biomarker: could language be a biosocial marker of psychosis? *npj Schizophr 2021 71*. 2021;7(1):1-5. doi:10.1038/s41537-021-00172-1

9. Parola A, Berardinelli L, Bosco FM. Cognitive abilities and theory of mind in explaining communicative-pragmatic disorders in patients with schizophrenia. *Psychiatry Res*. 2018;260:144-151. doi:10.1016/j.psychres.2017.11.051

10. Bambini V, Arcara G, Bechi M, Buonocore M, Cavallaro R, Bosia M. The communicative impairment as a core feature of schizophrenia: Frequency of pragmatic deficit, cognitive substrates, and relation with quality of life. *Compr Psychiatry*. 2016;71:106-120. doi:10.1016/j.comppsych.2016.08.012

11. Bliksted V, Fagerlund B, Weed E, Frith C, Videbech P. Social cognition and neurocognitive deficits in first-episode schizophrenia. *Schizophr Res*. 2014;153(1-3):9-17. doi:10.1016/j.schres.2014.01.010

12. Cohen AS, Dinzeo TJ, Donovan NJ, Brown CE, Morrison SC. Vocal acoustic analysis as a biometric indicator of information processing: Implications for neurological and psychiatric disorders. *Psychiatry Res*. 2015;226(1):235-241. doi:10.1016/j.psychres.2014.12.054

13. Cohen AS, McGovern JE, Dinzeo TJ, Covington MA. Speech deficits in serious mental illness: A cognitive resource issue? *Schizophr Res*. 2014;160(1-3):173-179. doi:10.1016/j.schres.2014.10.032

14. Insel TR. Digital Phenotyping: Technology for a New Science of Behavior. *JAMA*. 2017;318(13):1215-1216. doi:10.1001/JAMA.2017.11295

15. Chandler C, Foltz PW, Cohen AS, et al. Machine learning for ambulatory applications of neuropsychological testing. *Intell Med*. 2020;1-2:100006. doi:10.1016/J.IBMED.2020.100006

16. Cohen AS, Cox CR, Masucci MD, et al. Digital Phenotyping Using Multimodal Data.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Curr Behav Neurosci Reports. 2020;7(4):212-220. doi:10.1007/S40473-020-00215-4/FIGURES/3

17. Ben-Zeev D, Brian R, Wang R, et al. CrossCheck: Integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse. *Psychiatr Rehabil J*. 2017;40(3):266-275. doi:10.1037/prj0000243

18. Corcoran CM, Cecchi GA. Using Language Processing and Speech Analysis for the Identification of Psychosis and Other Disorders. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2020;5(8):770-779. doi:10.1016/j.bpsc.2020.06.004

19. Galatzer-Levy IR, Abbas A, Koesmahargyo V, et al. Facial and vocal markers of schizophrenia measured using remote smartphone assessments. *medRxiv*. 2020;1(646). doi:10.1101/2020.12.02.20219741

20. Hitczenko K, Mittal VA, Goldrick M. Understanding Language Abnormalities and Associated Clinical Markers in Psychosis: The Promise of Computational Methods. *Schizophr Bull*. 2021;47(2):344-362. doi:10.1093/SCHBUL/SBAA141

21. Corcoran R, Frith CD. Autobiographical memory and theory of mind: Evidence of a relationship in schizophrenia. *Psychol Med*. 2003;33(5):897-905. doi:10.1017/S0033291703007529

22. Cohen AS, Mitchell KR, Elvevåg B. What do we really know about blunted vocal affect and alogia? A meta-analysis of objective assessments. *Schizophr Res*. 2014;159(2-3):533-538. doi:10.1016/j.schres.2014.09.013

23. Schnack HG. Improving individual predictions: Machine learning approaches for detecting and attacking heterogeneity in schizophrenia (and other psychiatric diseases). *Schizophr Res*. 2019;214:34-42. doi:10.1016/J.SCHRES.2017.10.023

24. Gratton C, Mittal VA. Embracing the Complexity of Heterogeneity in Schizophrenia: A New Perspective From Latent Clinical-Anatomical Dimensions. *Schizophr Bull*.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2020;46(6):1337-1338. doi:10.1093/schbul/sbaa122

25. Honnorat N, Dong A, Meisenzahl-Lechner E, Koutsouleris N, Davatzikos C. Neuroanatomical heterogeneity of schizophrenia revealed by semi-supervised machine learning methods. *Schizophr Res*. 2019;214:43-50. doi:10.1016/J.SCHRES.2017.12.008

26. Fisher AJ, Medaglia JD, Jeronimus BF. Lack of group-to-individual generalizability is a threat to human subjects research. *Proc Natl Acad Sci U S A*. 2018;115(27):E6106-E6115. doi:10.1073/PNAS.1711978115/-/DCSUPPLEMENTAL

27. Dickinson D, Pratt DN, Giangrande EJ, et al. Attacking Heterogeneity in Schizophrenia by Deriving Clinical Subgroups From Widely Available Symptom Data. *Schizophr Bull*. 2018;44(1):101-113. doi:10.1093/SCHBUL/SBX039

28. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. *Speech Commun*. 2015;71:10-49. doi:10.1016/J.SPECOM.2015.03.004

29. Arora S, Baghai-Ravary L, Tsanas A. Developing a large scale population screening tool for the assessment of Parkinson's disease using telephone-quality voice. *J Acoust Soc Am*. 2019;(145(5)):2871-2884. doi:10.1121/1.5100272

30. Ben-Zeev D, Buck B, Kopelovich S, Meller S. A technology-assisted life of recovery from psychosis. *npj Schizophr 2019 51*. 2019;5(1):1-4. doi:10.1038/s41537-019-0083-y

31. Arevian AC, Bone D, Malandrakis N, et al. Clinical state tracking in serious mental illness through computational analysis of speech. *PLoS One*. 2020;15(1):e0225695. doi:10.1371/JOURNAL.PONE.0225695

32. Tan EJ, Meyer D, Neill E, Rossell SL. Investigating the diagnostic utility of speech patterns in schizophrenia and their symptom associations. *Schizophr Res*. 2021;238:91-

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

98. doi:10.1016/J.SCHRES.2021.10.003

33. de Boer JN, Voppel AE, Brederoo SG, Wijnen FNK, Sommer IEC. Language disturbances in schizophrenia: the relation with antipsychotic medication. *npj Schizophr*. 2020;6(1):1-9. doi:10.1038/s41537-020-00114-3

34. Moro-Velazquez L, Gomez-Garcia JA, Arias-Londoño JD, Dehak N, Godino-Llorente JI. Advances in Parkinson's Disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects. *Biomed Signal Process Control*. 2021;66:102418. doi:10.1016/J.BSPC.2021.102418

35. Brand CO, Ounsley JP, Van der Post DJ, Morgan TJH. Cumulative Science via Bayesian Posterior Passing. *Meta-Psychology*. 2019;3. doi:10.15626/mp.2017.840

36. Oomen PP, de Boer JN, Brederoo SG, et al. Characterizing speech heterogeneity in schizophrenia-spectrum disorders. *J Psychopathol Clin Sci*. 2022;131(2):172-181. doi:10.1037/ABN0000736

37. Andreasen NC. *Scale for the Assessment of Negative Symptom*. University of Iowa; 1984.

38. *Scale for the Assessment of Positive Symptoms (SAPS)*. University of Iowa; 1984.

39. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13(1):261-276. doi:10.1093/schbul/13.2.261

40. Nasrallah H, Morosini P, Gagnon DD. Reliability, validity and ability to detect change of the Personal and Social Performance scale in patients with stable schizophrenia. *Psychiatry Res*. 2008;161(2):213-224. doi:10.1016/j.psychres.2007.11.012

41. Abell F, Happé F, Frith U. Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cogn Dev*. Published online 2000. doi:10.1016/S0885-2014(00)00014-9

42. Castelli F, Happé F, Frith U, Frith C. Movement and Mind: A Functional Imaging

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Study of Perception and Interpretation of Complex Intentional Movement Patterns. *Neuroimage*. 2000;12(3):314-325. doi:10.1006/NIMG.2000.0612

43. Gelman A, Vehtari A, Simpson D, et al. Bayesian Workflow. *arXiv Prepr arXiv201101808*. Published online November 3, 2020. doi:10.48550/arxiv.2011.01808

44. Yao Y, Vehtari A, Simpson D, Gelman A. Using Stacking to Average Bayesian Predictive Distributions (with Discussion). *https://doi.org/101214/17-BA1091*. 2018;13(3):917-1007. doi:10.1214/17-BA1091

45. Aringhieri S, Carli M, Kolachalam S, et al. Molecular targets of atypical antipsychotics: From mechanism of action to clinical differences. *Pharmacol Ther*. 2018;192:20-41. doi:10.1016/J.PHARMTHERA.2018.06.012

46. Leucht S, Samara M, Heres S, Patel MX, Woods SW, Davis JM. Dose Equivalents for Second-Generation Antipsychotics: The Minimum Effective Dose Method. *Schizophr Bull*. 2014;40(2):314-326. doi:10.1093/SCHBUL/SBU001

47. Goodman SN, Fanelli D, Ioannidis JPA. What does research reproducibility mean? *Sci Transl Med*. Published online June 1, 2016:96-102. doi:10.1126/SCITRANSLMED.AAF5027

48. Vandenbroucke JP, Von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLOS Med*. 2007;4(10):e297. doi:10.1371/JOURNAL.PMED.0040297

49. Cohen AS, Schwartz E, Le TP, et al. Digital phenotyping of negative symptoms: the relationship to clinician ratings. *Schizophr Bull*. 2021;47(1):44-53. doi:10.1093/schbul/sbaa065

50. 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(5):347-353. doi:10.1016/S0022-3956(02)00016-X

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

51. Compton MT, Lunden A, Cleary SD, et al. The aprosody of schizophrenia: Computationally derived acoustic phonetic underpinnings of monotone speech. *Schizophr Res*. 2018;197:392-399. doi:10.1016/j.schres.2018.01.007

52. Beechey T, Buchholz JM, Keidser G. Measuring communication difficulty through effortful speech production during conversation. *Speech Commun*. 2018;100:18-29. doi:10.1016/J.SPECOM.2018.04.007

53. Traunmüller H, Eriksson A. Acoustic effects of variation in vocal effort by men, women, and children. *J Acoust Soc Am*. 2000;107(6):3438. doi:10.1121/1.429414

54. De Boer JN, Voppel AE, Brederoo SG, et al. Acoustic speech markers for schizophrenia-spectrum disorders: a diagnostic and symptom-recognition tool. *Psychol Med*. Published online 2021:1-11. doi:10.1017/S0033291721002804

55. Cohen AS, Cox CR, Le TP, et al. Using machine learning of computerized vocal expression to measure blunted vocal affect and alogia. *npj Schizophr 2020 61*. 2020;6(1):1-9. doi:10.1038/s41537-020-00115-2

56. Tahir Y, Yang Z, Id DC, et al. Non-verbal speech cues as objective measures for negative symptoms in patients with schizophrenia. Published online 2019:1-17.

57. Asiaee M, Vahedian-azimi A, Atashi SS, Keramatfar A, Nourbakhsh M. Voice Quality Evaluation in Patients With COVID-19: An Acoustic Analysis. *J Voice*. Published online October 1, 2020. doi:10.1016/J.JVOICE.2020.09.024

58. Fried EI. The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *J Affect Disord*. 2017;208:191-197. doi:10.1016/J.JAD.2016.10.019

59. Micoulaud-Franchi J, Quiles C, Batail J, L'encephale CL-, 2018 undefined. Making psychiatric semiology great again: A semiologic, not nosologic challenge. *Elsevier*. doi:10.1016/j.encep.2018.01.007

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

60. Khan A, Yavorsky C, Liechti S, et al. A rasch model to test the cross-cultural validity in the positive and negative syndrome scale (PANSS) across six geo-cultural groups. *BMC Psychol 2013 11*. 2013;1(1):1-18. doi:10.1186/2050-7283-1-5

61. Aggarwal NK, Tao H, Xu K, Stefanovics E, Zhening L, Rosenheck RA. Comparing the PANSS in Chinese and American inpatients: Cross-cultural psychiatric analyses of instrument translation and implementation. *Schizophr Res*. 2011;132(2-3):146-152. doi:10.1016/J.SCHRES.2011.08.003

62. Marquand AF, Kia SM, Zabihi M, Wolfers T, Buitelaar JK, Beckmann CF. Conceptualizing mental disorders as deviations from normative functioning. *Mol Psychiatry 2019 2410*. 2019;24(10):1415-1424. doi:10.1038/s41380-019-0441-1

63. Marquand AF, Rezek I, Buitelaar J, Beckmann CF. Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. *Biol Psychiatry*. 2016;80(7):552-561. doi:10.1016/J.BIOPSYCH.2015.12.023

64. Hitczenko K, Cowan HR, Goldrick M, Mittal VA. Racial and Ethnic Biases in Computational Approaches to Psychopathology. *Schizophr Bull*. 2022;48(2):285-288. doi:10.1093/SCHBUL/SBAB131

65. Ali MS, Prieto-Alhambra D, Lopes LC, et al. Propensity score methods in health technology assessment: Principles, extended applications, and recent advances. *Front Pharmacol*. 2019;10:973. doi:10.3389/FPHAR.2019.00973/BIBTEX

66. Gooden TE, Gardner M, Wang J, et al. The risk of mental illness in people living with HIV in the UK: a propensity score-matched cohort study. *Lancet HIV*. 2022;9(3):e172-e181. doi:10.1016/S2352-3018(21)00319-2

67. Kvarven A, Strømland E, Johannesson M. Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nat Hum Behav*. 2020;4(4):423-434. doi:10.1038/s41562-019-0787-z

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

68. Çokal D, Zimmerer V, Turkington D, et al. Disturbing the rhythm of thought: Speech pausing patterns in schizophrenia, with and without formal thought disorder. *PLoS One*. 2019;14(5):1-14. doi:10.1371/journal.pone.0217404

69. Dellwo V, Leemann A, Kolly M-J. Rhythmic variability between speakers: Articulatory, prosodic, and linguistic factors. *J Acoust Soc Am*. 2015;137(3):1513. doi:10.1121/1.4906837

70. Kanber E, Lavan N, McGettigan C. Highly Accurate and Robust Identity Perception From Personally Familiar Voices. *J Exp Psychol Gen*. Published online 2021. doi:10.1037/XGE0001112

71. Kreiman J, Park SJ, Keating PA, Alwan A. The Relationship Between Acoustic and Perceived Intraspeaker Variability in Voice Quality. In: *In Sixteenth Annual Conference of the International Speech Communication Association.* ; 2015.

72. Cohen AS, Mitchell KR, Elvevåg B. What do we really know about blunted vocal affect and alogia? A meta-analysis of objective assessments. *Schizophr Res*. 2014;159(2-3):533-538. doi:10.1016/j.schres.2014.09.013

73. Rocca R, Yarkoni T. Putting Psychology to the Test: Rethinking Model Evaluation Through Benchmarking and Prediction: *https://doi.org/101177/25152459211026864*. 2021;4(3). doi:10.1177/25152459211026864

74. Loth E, Ahmad J, Chatham C, et al. The meaning of significant mean group differences for biomarker discovery. *PLOS Comput Biol*. 2021;17(11):e1009477. doi:10.1371/JOURNAL.PCBI.1009477

75. Stein F, Buckenmayer E, Brosch K, et al. Dimensions of Formal Thought Disorder and Their Relation to Gray- and White Matter Brain Structure in Affective and Psychotic Disorders. *Schizophr Bull*. 2022;48(4):902-911. doi:10.1093/SCHBUL/SBAC002

76. Tang SX, Hänsel K, Cong Y, et al. Latent Factors of Language Disturbance and

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Relationships to Quantitative Speech Features. *medRxiv*. Published online April 1, 2022:2022.03.31.22273263. doi:10.1101/2022.03.31.22273263

77. Brand BA, Haveman YRA, De Beer F, De Boer JN, Dazzan P, Sommer IEC. Antipsychotic medication for women with schizophrenia spectrum disorders. *Psychol Med*. 2022;52(4):649-663. doi:10.1017/S0033291721004591

78. Parola A, Gabbatore I, Berardinelli L, Salvini R, Bosco FM. Multimodal assessment of communicative-pragmatic features in schizophrenia: a machine learning approach. *npj Schizophr 2021 71*. 2021;7(1):1-9. doi:10.1038/s41537-021-00153-4

79. Parola A, Salvini R, Gabbatore I, Colle L, Berardinelli L, Bosco FM. Pragmatics, Theory of Mind and executive functions in schizophrenia: Disentangling the puzzle using machine learning. *PLoS One*. Published online 2020. doi:10.1371/journal.pone.0229603

80. Lau JCY, Patel S, Kang X, et al. Cross-linguistic patterns of speech prosodic differences in autism: A machine learning study. Pegoraro C, ed. *PLoS One*. 2022;17(6):e0269637. doi:10.1371/JOURNAL.PONE.0269637

81. Parola A, Lin JM, Simonsen A, et al. Speech disturbances in schizophrenia: Assessing cross-linguistic generalizability of NLP automated measures of coherence. *Schizophr Res*. Published online August 1, 2022. doi:10.1016/J.SCHRES.2022.07.002

medRxiv preprint doi: https://doi.org/10.1101/2022.04.03.22273354; this version posted August 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure Legends

Figure 1. Comparing meta-analysis, skeptical expectations and results. Each panel presents a separate acoustic measure, with the x-axis corresponding to standardized mean differences (schizophrenia - HC) equivalent to Hedges' g, with estimates above 0 indicating higher scores for patients with schizophrenia. The overlap between the skeptical and meta-analytic posterior distributions suggests no real advantage in using meta-analytic informed priors compared to skeptic ones.