Voice Patterns in Schizophrenia: A systematic Review and Bayesian Meta-Analysis

Parola Alberto[a], Simonsen Arndis[b,c], Bliksted Vibeke[b,c], Fusaroli Riccardo[c,d]

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

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

[c] The Interacting Minds Center - School of Culture and Society, Aarhus University, Denmark

[d] Department of Linguistics, Semiotics and Cognitive Science – School of Communication and Culture, Aarhus University, Denmark

Correspondence concerning this article should be addressed to:

Alberto Parola, Dipartimento di Psicologia, Via Po 14, 10123 Turin (Italy).

Phone number: +390116703065

Fax number: +39 0118159039

E-mail: alparola@unito.it

Word counts abstract (max 250): 214

Word counts text body (max 5000): 4682

Abstract

Voice atypicalities have been a characteristic feature of schizophrenia (SCZ) since its first definitions. They are often associated with core negative symptoms such as flat affect and alogia, and with the social impairments seen in the disorder. This suggests that voice atypicalities may represent a marker of clinical features and social functioning in SCZ. We systematically reviewed and meta-analyzed the evidence for distinctive acoustic patterns in SCZ, as well as their relation to clinical features. We identified 46 articles, including 55 studies with a total of 1254 patients with SCZ and 699 healthy controls. Summary effect size (Hedges'g) estimates were calculated using multilevel Bayesian modeling. We identified weak atypicalities in pitch variability (g = - 0.55) related to flat affect, and stronger atypicalities in proportion of spoken time, speech rate, and pauses (g's between -0.75 and -1.89) related to alogia and flat affect. However, the effects were modest compared to perceptual and clinical judgments, and characterized by large heterogeneity between studies. Moderator analyses revealed that tasks with a more demanding cognitive and social component had significantly larger effects both in contrasting patients and controls and in assessing symptomatology. In conclusion, studies of acoustic patterns are a promising but, yet unsystematic avenue for establishing markers of SCZ. We outline recommendations towards more cumulative, open, and theory-driven research.

Keywords:

Acoustic analysis, social communication, machine learning, biomarker, negative symptoms, speech signal.

Introduction

Individuals with schizophrenia (SCZ) display atypical voice patterns, qualitatively described in terms of poverty of speech, increased pauses, distinctive tone and intensity of voice[1–7]. Voice atypicalities have been reported since the first definitions of the disorder[8,9], are used in the clinical assessment process, and assume an even stronger relevance in the light of growing findings associating voice patterns to cognitive function, emotional states, and social engagement[10–19].

Voice atypicalities may thus constitute a window into the underlying clinical and cognitive features of the disorder. Indeed, they have been associated with core negative symptoms of SCZ such as blunted affect (e.g. diminished emotional expression, lack of vocal intonation), and alogia (e.g. poverty of speech, latency of speech and blocking)[2,3,20–22]. Negative symptoms are included among the primary diagnostic criteria of SCZ (DSM-V), and are associated with early age of onset, poor social and functional outcome, reduced quality of life, and poor response to medication and treatment[23–26]. Vocal expression also reflects a key component of social communication, a domain frequently impaired in individuals with SCZ[27–32]. Difficulties in controlling voice to express affective and emotional contents or to mark relevant information may dramatically reduce the ability of these individuals to communicate effectively in social context. Impairments in social communication may in turn lead to experience of failure in social situations, and to perceive negative social judgments on the part of others, resulting in social withdrawal and further aggravating the social cognitive impairments[28,33–37]. Voice atypicalities may thus represent an important biometric index that parallels both clinical features and social cognitive functioning of individuals with SCZ over time. A better comprehension of voice abnormalities could provide tools to better assess the cognitive and social features of this heterogeneous disorder. However, despite the importance in studying vocal expression in schizophrenia, and the routine assessments performed using interview-based clinical rating scales, our understanding of voice abnormalities in schizophrenia is limited. Previous work on voice atypicalities can be organized into three categories: qualitative perceptual ratings, quantitative acoustic analyses, and multivariate machine learning (ML) investigations. Most previous studies employing qualitative ratings reported robust differences between patients with SCZ and healthy controls (HC) across several perceptual features of their voice[4,38,39]. However informative, qualitative rating scales have serious limitations. They rely on raters' expertise and intuition, thus lacking scalability to large corpora, and they display low sensitivity to complex and multivariate acoustic patterns and variations in context and time[2,13,40,41]. A different approach involves the use of automated analysis of speech to identify acoustic features of vocal production, arguably with a greater reliability, sensitivity and validity. However, such studies have so far reported smaller and seemingly more contradictory findings: some indicate slower speech[42] , more pronounced pauses[43–45] and reduced prosodic variability[21,44,46]; while others indicate no reliable acoustic differences between individuals with SCZ and HC[47–49]. A meta-analysis of 13 studies[39] suggests large differences between individuals with SCZ and HC on pause and speech duration, and more modest on intensity and pitch variability. However, the number of studies included in the meta-analysis was small compared to the currently available literature and, given the high heterogeneity of patients with SCZ, a more systematic review accounting for the potential sources of heterogeneity in the effects is required: individual differences (e.g. gender, age and education), contextual factors (e.g. type of task) and clinical features (e.g. symptomatology and medication). A few studies have adopted a more fine-grained perspective, and assessed the relationship between acoustic measures and clinical features with some promise; however, the findings are still sparse[3,40,47,50,51].

Finally, more recent studies have tried to capitalize on the technological advancements in speech signal processing, and the application of multivariate ML techniques to better capture the complex, multivariate and often non-linear nature of acoustic patterns[52,53] (see also the appendix to Fusaroli et al. (2017)[48] for an introduction to ML techniques in the context of voice analysis). These studies extract more nuanced acoustic measures, e.g. spectral and glottal features, and assess how accurately the diagnosis can be identified only relying on acoustic measures. The results are promising[16,17,19,43,44,54], but a complete and comparative overview of the findings in SCZ is currently missing. Crucially, the reliability of ML results has been shown to be strongly dependent on the availability of large datasets and the validation of the findings across datasets[55–59].

Despite the promise of acoustic markers of clinical features in schizophrenia, it is yet unclear how to quantify them, that is, which acoustic features we should focus on, and the evidence for their relation to specific clinical features of the disorder. The aim of the present study was to fill this gap by systematically reviewing and meta-analyzing the current state of evidence for acoustic atypicalities in SCZ as a whole as well as their relation to the specific clinical features. Further, we evaluated the size and availability of previous datasets, and the attitudes towards data sharing of the authors of the studies reviewed to assess whether a more cumulative science of voice atypicalities in SCZ can be attempted. Note that the aim of this meta-analysis is less to provide a more accurate estimation of the voice atypicalities in SCZ than it is to provide the basis for more effective future studies, by identifying current practices, issues and promising venues.

Methods

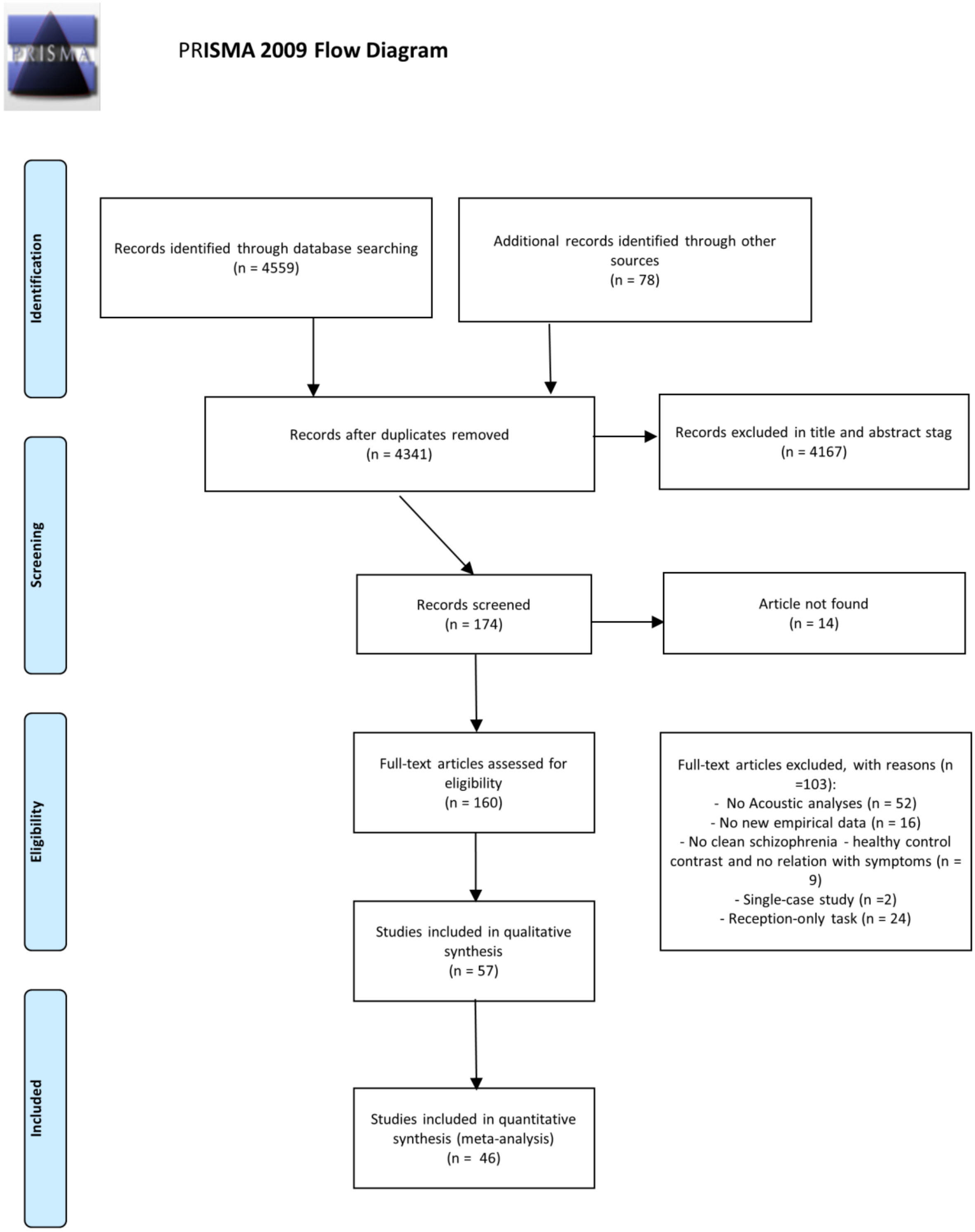
Inclusion criteria for literature search

We adopted the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Guidelines (PRISMA[60]) for transparent reporting of a systematic review. We pre-registered our protocol by specifying a priori the study rationale, eligibility criteria, search strategy, moderator variables, and statistical analyses (see https://bit.ly/2EEFeQZ). The literature search was conducted on Pubmed and Google Scholar, the latter including dissertations and unpublished manuscripts. The search terms used were (prosody OR inflection OR intensity OR pitch OR fundamental frequency OR speech rate OR voice quality OR acoustic OR intonation OR vocal) AND (schizo*). The search was conducted on August 21 2017, and updated on April 12 2018. We complemented the list by performing a backward and forward literature search: we screened the bibliography of the papers found and the papers citing them as identified by Google Scholar.

Articles were screened for eligibility by two authors (P.A and S.A). Study selection was conducted according to the following inclusion criteria: (a) empirical study, (b) quantification of acoustic features in the vocal production of participants with SCZ or schizoaffective disorder[1] (c) sample including at least two individuals with SCZ or schizoaffective disorder, (d) inclusion of a non-clinical comparison group, or an assessment of variation in acoustic features in relation to severity of clinical features. Clinical comparison groups (e.g. with depression) were excluded because the limited number of studies did not permit meta-analytic estimations. Fig. 1 shows the flow-diagram of study selection. We report the assessment of the risk of bias in the Supplementary Materials.

[1] We included schizotypy in literature search to better cover schizophrenia spectrum disorder. However, given schizotypy is only included in the schizophrenia spectrum in the ICD and is mentioned in the personality disorders in the DSM classification, we only included schizotypy in additional analysis in the supplementary material.

Figure 1


PRISMA 2009 Flow Diagram
Identification
Screening
Eligibility
Included
Records identified through database searching (n = 4559)
Additional records identified through other sources (n = 78)
Records after duplicates removed (n = 4341)
Records excluded in title and abstract stag (n = 4167)
Records screened (n = 174)
Article not found (n = 14)
Full-text articles assessed for eligibility (n = 160)
Full-text articles excluded, with reasons (n =103):
- No Acoustic analyses (n = 52)
- No new empirical data (n = 16)
- No clean schizophrenia - healthy control contrast and no relation with symptoms (n = 9)
- Single-case study (n =2)
- Reception-only task (n = 24)
Studies included in qualitative synthesis (n = 57)
Studies included in quantitative synthesis (meta-analysis) (n = 46)

Data extraction

For all the studies we reported the available clinical and demographic data, including pre-registered potential moderations. In particular we report: sample sizes, matching criteria, presence of a non-clinical control group, diagnosis, demographical variables (age, education, gender, language and ethnicity), clinical information (symptom clinical ratings, duration of illness, age of onset, hospitalization), level of intelligence (IQ), cognitive screening, medication. Further we extracted information about the speech production task, group-level acoustic estimates (mean and standard deviation), and correlation coefficients between acoustic measures and clinical ratings. We grouped speech production tasks into three categories: 1) *Constrained production* includes highly structured monological tasks such as reading aloud or repeating sequence of numbers. 2) *Free monological production* includes less constrained monological tasks such as description of pictures or videos, or providing narrative accounts (e.g. of a happy event, or of one's life). Compared to constrained production, free production is more challenging, as the linguistic materials are less pre-defined by the task. 3) *Social interaction* includes structured and semi-structured interviews, as well as spontaneous conversations. The production is dialogical and involves interpersonal factors and dynamics. Selected characteristics of included studies are available in Table 1.

Table 1. Selected characteristics of the studies included in the meta-analysis.

	Authors	Control group	Sample size and matching criteria	Clinical features	Task	Medication	Findings
1	Pinheiro et al. (2016)	YES	17 SZ 18 CT Age, sex, education, parental SES	PANNS, SANS, SAPS	Constrained production (reading single words)	NR	**Duration of utterance:** (words duration/ms): NS **Pitch mean** (Hz): NS **Intensity mean** (db): NS
2	Zhang et al. (2016)	YES	26 SZ 30 CT Age, sex, education	PANNS, SANS, CGI-S	Social interaction (phone conversation)	Not medicated	**Formants (F1, F2, F3, F4, F5, F6) (Hz/db):** NS **Formant bandwidth:** NS **Formants intensity variability (entropy):** NS **Spectral features: Mel-frequency cepstral coefficient (MFCC):** $p < .001$ (Lower in SZ) **Linear prediction coding (LPC): $p < .001$ (Higher in SZ)**
3	Bernardini et al. (2016)	NO	20 NA	PANNS, SANS	Spontaneous production (narrative)	NR	Correlations: **PANSS TOTAL: Percent time talking:** NS; **Pitch mean:** NS; **Pitch variability**: NS. SANS TOTAL: **Percent time talking**: NS; **Pitch mean**: NS; **Pitch variability**: NS. PANSS NEGATIVE: **Percent time talking**: NS; **Pitch mean**: NS; **Pitch variability**: NS. SANS FLAT AFFECT: **Percent time talking**: NS; **Pitch mean**: NS; **Pitch variability**: NS; SANS ALOGIA: **Percent time talking**: NS; **Pitch mean**: NS; **Pitch variability**: NS.
3	Bernardini et al. (2016)	NO	20 NA	PANNS, SANS	Spontaneous production (narrative)	NR	Correlations: **PANSS TOTAL: Percent time talking**: NS; **Pitch mean:** NS; **Pitch variability**: NS; SANS TOTAL: **Percent time talking**: NS; **Pitch mean**: NS; **Pitch variability**: NS. PANSS NEGATIVE: **Percent time talking**: NS; **Pitch mean**: $p = .04$ (Positive); **Pitch variability**: NS. SANS FLAT AFFECT: **Percent time talking**: NS; **Pitch mean**: NS; F0 SD: NS; SANS ALOGIA: **Percent time talking**: NS; **Pitch mean**: $p = .005$ (Positive); **Pitch variability**: $p < .05$ (Negative).
4	Martínez-Sánchez et al. (2015)	YES	45 SZ 35 CT Age, sex, education	BPRS	Constrained interaction (reading)	Medicated	**Pause percentage (>300ms):** $p < .001$ (Higher in SZ) **Pitch mean (Hz):** NS **Pitch variability (Hz):** NS **Intensity variability (Hz**): NS **Intensity mean (db):** $p < .001$ (Lower in SZ) Correlations: **BPRS TOTAL:** Proportion of pauses: NS; **Pitch mean**: NS; **Pitch variability**: NS; **Intensity mean**: NS. **BPRS NEGATIVE:** Proportion of pauses: NS; **Pitch mean**: NS; **Pitch variability**: NS; **Intensity mean**: NS **BPRS POSITIVE:** Proportion of pauses: $p = .021$ (Negative); **Pitch mean**: NS; **Pitch variability**: NS; **Intensity mean**: NS
6	Alpert et al. (2002)	NO	30 SZ NA	SANS	Social interaction (interview)	Medicated	Correlations: **SANS FLAT AFFECT**: **Percent time talking:** $p < .01$ (Negative); **Speech latency**: NS; **Pitch variability**: $p < .05$ (Negative); **Intensity variability**: NS; **SANS ALOGIA: Percent time talking**: NS; **Speech latency**: NS; **Pitch variability**: NS; **Intensity**

							variability: NS
7	Rapcan. et al. (2010)	YES	39 SZ 18 CT NR	SANS, BPRS	Constrained interaction (reading brief text)	Medicated	**Duration of utterance** (s.) p < .001 (Shorter in SZ) **Percent time talking:** NS **Duration of pauses** (s.): p = .011 **Percentage of silence:** p < .001 (Higher in SZ) **Number of pauses** (>250ms): p < .001 (Higher in SZ) **Pitch variability** (cv): NS **Intensity variability:** p < .04 (Higher in SZ) **Correlations: BPRS TOTAL**: **Duration of utterance**: p < .01 (Positive); **Proportion of silence**; p < .05 (Negative); **Duration of pauses**: NS; **Number of pauses**: NS; **Pitch variability**: NS; **Intensity variability**: p < .05 (Positive). **NEGATIVE SANS**: **Duration of utterance**: p < .05 (Positive); **Duration of pauses**: NS; **Proportion of silence**: p < .05 (Negative); **Number of pauses**: NS; **Pitch variability**: NS; **Intensity variability**: p < .05 (Positive).
9	Cannizzaro et al. (2005)	YES	13 SZ 6 CT NR	PANSS	Constrained (count) + Spontaneous (narrative elicitation)	Medicated	**Duration of pauses (>200msec): TASK 1 (Constrained): NS; TASK 2 (Free):** p < .001 (Higher in SZ). **Percentage of pauses: TASK 1:** NS; **TASK 2:** p <.001 (Higher in SZ) **Number of pauses: TASK 1:** NS; **TASK 2:** NS **Pause variability: TASK 1:** NS; **TASK 2:** p < .001 (Higher in SZ)
10	Graux et al. (2015)	YES	26 SZ 26 CT NR	NR	Constrained (read letter)	Medicated	**Pitch mean (Hz)**: p = .021 (Higher in SZ)
11	McGilloway et al. (2003)	YES	72 SZ 40 CT Age, sex	SANS	Constrained (read passage)	Medicated	**Duration of pauses**: NS **Intensity mean:** NS
12	Sison et al. (1996)	NO	24 SZ NA	SANS	Social interaction (interview)	Medicated	**Correlations: SANS FLAT AFFECT**: **Duration of utterance**: NS; **Duration of pauses**: p = .008 (Positive); **Pitch variability**: NS; **Intensity variability**: NS
13	Cohen et al. (2008)	YES	60 SZ 19 CT Age. sex and parental SEI.	SANS, SAPS	Spontaneous production (narrative)	Medicated	**Speech rate (words/sec):** NS **Pitch variability (Hz):** p = .041 (Lower in SZ) **Correlations: SANS FLAT AFFECT: Speech rate** p < .01 (Negative); **Pitch variability**: NS; **SANS ALOGIA: Speech rate**: p < .01 (Negative); **Pitch variability**: NS;
18	Cohen et al. (2013)	NO	26 SZ NA	BPRS	Spontaneous production (picture elicitation)	Medicated	**Correlations: BPRS TOTAL PSYCHO: Duration of pauses:** NS. **BPRS NEGATIVE PSYCHO: Duration of pauses**: NS **BPRS POSITIVE: Duration of pauses**: NS
20	Alpert et al. (1997)	YES	19 SZ 20 CT NR	SANS	Social interaction (interview) + Spontaneous production (monologue)	NR	**Duration of pauses:** p < .001 (Higher in SZ). **Response latency:** p < .001 (Higher in SZ).

21	Alpert et al. (2000)	YES	46 SZ 20 CT Age, education	SANS	Social interaction (semi-structured interview)	Medicated	**Speech rate** (words/s): p < .001 (Lower in SZ) **Percent time talking**: p = .012 (Lower in SZ) **Pitch variability** (SEMITONES): p = .001 (Lower in SZ) **Intensity variability** (db): NS <u>Correlations:</u> **SANS FLAT AFFECT: Percent time talking**: p < .01 (Negative); **Pitch variability**: p < .01 (Negative); **Intensity variability**: NS; **SANS ALOGIA: Percent time talking**: p < .01 (Negative); **Pitch variability**: NS; **Intensity variability**: NS
23	Covington et al. (2012)	NO	25 SZ	PANSS	Social interaction (PANNS interview)	Medicated	<u>Correlations</u>**: SANS TOTAL: Pitch mean: NS; Pitch variability:** NS; **PANSS NEGATIVE**: **Pitch mean**: NS; **Pitch variability**: NS; **PANSS POSITIVE**: **Pitch mean**: NS; **Pitch variability**: NS; **SANS FLAT AFFECT**: **Pitch mean**: NS; **Pitch variability**: NS; **SANS ALOGIA**: **Pitch mean**: NS; **Pitch variability**: NS;
24	Matsumoto et al. (2013)	YES	6 SZ 6 CT Age, education, IQ.	SANS, SAPS	Spontaneous production (picture elicitation)	Medicated	**Pause duration** (>250ms): NS. **Number of pauses**: p = .04 (Lower in SZ).
25	Alpert et al. (1994)	NO	17 SZ NA	SANS	Social interaction (semi-structured interview) + Spontaneous production (narrative elicitation)	Medicated	<u>Correlations</u>**: SANS ALOGIA: Speech rate:** p < .01 (Negative)**; Duration of pauses:** within-clauses: NS**;** between-clauses: p < .01 (Positive)**;** switching-clauses p < .01 (Positive)**;** **Number of pauses:** within-clauses: NS; between-clauses: NS; switching-clauses: NS; filled pause within-clauses: NS; filled pause between-clauses: NS
26	Kring et al. (1994).	NO	23 SZ NA	SANS, BPRS	Social interaction (semi-structured interview)	Unmedicated	<u>Correlations</u>**: BPRS TOTAL**: **Percent time talking**: NS; **BPRS POSITIVE: Percent time talking:** NS **BPRS NEGATIVE: Percent time talking:** NS **BPRS BLUNTED: Percent time talking:** NS **SANS TOTAL: Percent time talking**: NS
27	Pinheiro et al. (2017)	YES	15 SZ 16 CT Age. sex. and parental SES	PANNS, SANS, SAPS	Constrained interaction (reading single words)	Medicated	**Duration of utterance** (words/msec): p = .028 (Higher in SZ). **Pitch mean** (hz): NS **Intensity mean** (db): NS
28	Resnick et al. (1984)	YES	10 SZ 20 CT Age	NR	Social interaction (clinical interview)	Medicated	**Duration of pauses** (> 250ms): p = .013 (Higher in SZ). **Percent time talking**: NS
29	Mandal et al. (1990)	YES	40 SZ 60 CT NR	NR	Spontaneous production (facial expression picture elicitation)	Medicated	**Speech rate** (words/sec): p < .001 (Lower in SZ). **Percent time talking:** p < .001 (Lower in SZ). **Response latency:** p < .001 (Lower in SZ).
30	Tavano et al. (2008)	YES	37 SZ 37 CT Age and sex.	BPRS	Spontaneous production (narrative) + Social interaction interview)	Medicated	TASK 1 (Free): **Speech rate** (words/sec): p = .027(Lower in SZ). TASK 2 (Social): **Speech rate**: p < .001 (Lower in SZ).
31	Perlini et al.	YES	30 SZ	BPRS	Spontaneous production	Medicated	**Speech rate**: p = .009 (Lower in SZ).

	(2012)		30 CT Age, sex and education		(narrative) + Social interaction (interview)		
32	Rutter et al. (1997)	YES	12 SZ 12 CT Sex	NR	Social interaction (conversation)	NR	**Speech rate:** Task 1: NS**;** Task 2: NS; **Percent time talking:** Task 1: p = .019 (Lower in SZ); Tasks 2: NS
33	St-Hilaire et al. (2008)	YES	48 SZ 48 CT Age, sex, parental SES and ethnicity	BPRS	Social interaction (semi-structured interview)	Medicated	**Speech rate** (words/sec): p < .001 (Lower in SZ).
34	Shaw et al. (1999)	NO	30 SZ	SANS	Social interaction (interview)	Medicated	Correlations**: SANS FLAT AFFECT**: **Duration of pauses**: p < .01 (Positive); **Pitch variability**: NS
35	Docherty (2012)	YES	53 SZ 23 CT Age, sex, parent education, ethnicity	PANSS, PSYRATS	Social interaction (interview)	NR	**Speech rate**: NS
37	Rochester et al. (1977)	YES	40 SZ 20 CT Sex. age	NR	Social interaction (interviews)	Medicated	**Percent time talking: p = .001** (Lower in SZ). **Duration of pauses: p < .001** (Higher in SZ).
40	Compton et al. (2018)	YES	94 SZ 101 CT Age. ethnicity. race and marital status	PANSS, SANS, CAINS	Spontaneous production (narrative) + Constrained interaction (reading)	Medicated	**Pitch variability**: Task 1 (Free)**:** NS; Task 2 (Constrained): NS Correlations**: TOTAL SANS**: **Pitch variability**: P = .002 (Positive) **NEGATIVE PANSS**: **Pitch variability**: NS **SANS FLAT AFFECT**: **Pitch variability**: NS **SANS ALOGIA: Pitch variability**: p < .001 (Positive).
41	Salomé et al. (2002)	YES	10 SZ 10 CT Sex. and education	NR	Spontaneous production (narrative)	Medicated	**Speech rate**: p = .014 (Lower in SZ). **Number of pauses:** NS
43	Kliper et al. (2010)	YES	22 SZ 20 CT NR	SANS	Social interaction (interview) + Constrained interaction (reading)	NR	Correlations**: SANS TOTAL**: **Duration of utterance**: p < .001 (Negative); **Percent time talking**: p < .01 (Negative); **Duration of pauses**: p < .05 (Positive); **Intensity variability**: p < .01 (Negative)
44	Kliper. et al. (2015)	YES	22 SZ 20 CT Age. sex and education.	PANSS, SANS	Social interaction (clinical interviews)	Nr	**Duration of utterance**: p < .001 (Lower in SZ). **Percent time talking:** p < .001 (Lower in SZ). **Duration of pauses**: p < .001 (Higher in SZ). **Pitch variability**: p < .001 (Lower in SZ). **Intensity variability**: p < .001 (Higher in SZ). Correlations**: TOTAL PANSS**: **Duration of utterance**: NS; **Percent time talking**: NS; **Duration of pauses**: NS; **Pitch variability**: NS; **Intensity variability**: NS. **SANS TOTAL: Duration of utterance**: p = .04 (Negative); **Percent time talking**: p < .01 (Negative); **Duration of pauses**: p < .01 (Positive); **Pitch variability**: NS; **Intensity**

							variability: NS. **SANS FLAT AFFECT: Duration of utterance:** p < .01 (Negative); **Percent time talking**: p < .01 (Negative); **Duration of pauses**: p < .01 (Positive); **Pitch variability**: NS; **Intensity variability**: NS. **SANS ALOGIA: Duration of utterance:** p < .01 (Negative)**; Percent time talking**: p < .01 (Negative); **Duration of pauses**: p < .01 (Positive); **Pitch variability**: NS; **Intensity variability**: NS.
47	Ross et al. (2001)	YES	45 SZ 19 CT Not matched	SANS, SAPS, BPRS	Constrained production (repetition) + social interaction (interview)	Medication stabilized	TASK 1 (Constrained): **Pitch variability**: p < .0001 (Lower in SZ). TASK 2 (Social): **Pitch variability**: p < .0001 (Lower in SZ).
51	Püschel et al. (1998)	YES	45 SZ 45 CT Sex and age	SANS, PANSS, INSKA	Constrained production (counting and reading passage)	Medicated	**Duration of pauses: NS** **Number of pauses**: p = .0001 **Silence percentage:** p = .0002 **Duration of utterance**: p = .0001 **Total length of pauses** p = .0001 **Total length of utterances** p = .0001 **Pitch mean:** NS **Pitch variability:** p = .0137 **Intensity variability**: p = .0001
57	Meaux et al. (2018)	YES	36 SZ 25 CT Sex and education	BPRS	Spontaneous production (emotional picture elicitation)	NR	**Pitch variability**: Task 1 (Free): NS; Task 2: NS. **Intensity variability**: Task 1 (Free): NS; Task 2: NS. CORRELATION: **BPRS BLUNTED AFFECT: Pitch variability**: Task 1 (Free): NS; Task 2 (Free): NS; **Intensity variability**; Task 1 (Free): NS; Task 2 (Free): NS

Note: We included in the table those studies that reported: 1) descriptive statistics for SZ and HC groups, correlation coefficients, or statistical tests for these measures. P-values of statistical tests comparing individuals with SZ and HC and correlation coefficients have been extracted from original articles. When estimates for acoustic measures were reported for subgroups of patients, or for different task conditions within the same speech task, we averaged across them weighting the values by sample size, and we then computed independent samples t-test between individuals with SZ and HC groups. When in the original articles were provided estimates for acoustic measures but not p-value of the comparisons between groups, we computed independent samples t-test (or correlation coefficient) between individuals with SZ and HC control groups. When the authors provided us original data we recomputed independent samples t-tests (or correlation coefficients) using the original data. Clinical features: PANSS, The Positive And Negative Symptoms Scale; SANS, Scale for the Assessment of Negative Symptoms; SAPS, Scale for the Assessment of Positive symptoms. CGI-S*,* The Clinical Global Impression Scale; BPRS, Brief Psychiatric Rating Scale; PSYRATS, The Psychotic Symptom Rating Scales; CAINS, The Clinical Assessment Interview for Negative Symptoms; InSka, The Intentionality Scale.

When more than four studies reported statistical estimates for an acoustic measure, or correlation with symptomatology, we performed meta-analysis of the effects. When estimates for acoustic measures were reported for subgroups of patients, or for different task conditions within the same speech task, we averaged across them weighting the values by sample size.

In case of multivariate ML studies, we used a different focus. Multivariate ML approaches differ from the studies previously described in two main ways. While more traditional studies focus on a single feature at time, multivariate ML studies analyze multiple acoustic features simultaneously. While more traditional studies focus on best explaining all the current samples (minimizing within sample error), multivariate ML studies focus on generalizability of the results to new samples (minimizing out-of-sample error), e.g. by using validation and cross-validation techniques. In reviewing ML studies, we focused on reporting the algorithms adopted, the acoustic feature considered and the performance of the algorithms in either discriminating individuals with SCZ from HC with respect to the acoustic measures considered or predicting the severity of clinical features (e.g. negative symptoms) from acoustic measures (see Table S3 in appendix).

We contacted all authors to obtain missing group-level estimates and individual-level data. Statistics on authors' contact availability, propensity to respond and self-reported barriers to data sharing are also reported.

Statistical analysis

Meta-analyses were performed following well-established procedures[61–64] and complemented by a Bayesian framework[65,66]. To estimate the differences in vocal patterns between individuals with SCZ and HC we extracted the standardized mean difference (SMD; also known as Hedges' g). To estimate relations between vocal patterns and clinical features we extracted the raw correlation coefficient (Pearson's r). These effects were analyzed using 2-

level hierarchical Bayesian regression models to estimate the pooled effect sizes and corresponding credible (i.e., Bayesian confidence) intervals. The multilevel structure allowed us to explicitly model the heterogeneity (or σ^2) in the results of the studies analyzed. By including a random effect by study, we assumed that the variability in experimental design, acoustic analyses and population samples could generate heterogeneous findings and allowed the model to estimate such heterogeneity. We then measured and tested for heterogeneity of the studies using the Cochran's Q statistic[67], which reveals how much of the overall variance can be attributed to true between-study variance. To analyze the influence of potential moderators explaining between study heterogeneity, meta-regression models were applied separately. Note that only speech task presented enough data points to be analyzed as moderator. Other pre-registered moderators were not sufficiently reported and would have required access to individual level data for adequate treatment.

Priors were chosen to be only weakly informative so that their influence on the meta-analytic estimates were small, only discounting extreme values: a normal distribution centered at 0 (no effect), with a standard deviation of 0.5 for the overall effect, and a positive truncated normal distribution centered at 0, with a standard deviation of 0.5 for the heterogeneity of effects (standard deviation of random effects). We report 95% credible intervals (CIs), i.e. the intervals within which there is a 95% probability that the true value of the parameter (e.g. effect size) is contained, given the assumptions of the model. We provide evidence ratios (ER) and credibility scores. ERs quantify the evidence provided by the data in favor of the effect of diagnosis or of clinical feature (e.g. longer pauses in SCZ compared to HC) against the alternatives (e.g. same length or shorter pauses in schizophrenia). An ER equal to 3 indicates the hypothesis is 3 times more likely than the alternative. A credibility score indicates the percentage of posterior estimates falling above 0. Because Bayesian methods are less commonly used and understood, we also report p-values in order to reach a broader audience. Note that the p-values are calculated on the same 2-level hierarchical model as the

Bayesian inference, with the difference that p-value statistics rely on completely flat priors and assume Gaussian distributions for all estimated parameters.

To assess the potential role of speech production task in explaining the patterns observed, we compared the baseline model with a second multilevel Bayesian model including task as predictor of difference in vocal patterns. We used Leave-One-Out-Information-Criterion (LOOIC) and stacking weights indicating the probability that the model including task is better able to predict new data than baseline[68].

To explore the possibility of publication bias, potential for funnel plot asymmetry was examined visually and tested using the rank correlation test[69]. The raw data and analysis scripts are available at https://osf.io/qdkt4/. The supplementary materials report an additional analysis including schizotypy. All computation was done in R[70] relying on metafor, brms and Stan[64,71,72].

Results

3.1 Study selection

See Fig. 1 for full details on the selection. We were able to retrieve relevant statistical estimates from 46 articles (55 studies) from the texts or the authors. The meta-analysis included a total of 1254 patients (466 F) with SCZ and 699 controls (323 F). We contacted a total of 57 authors – including those of studies that were later deemed ineligible due to lack of statistical estimates – requesting additional information and individual level acoustic estimates for each participant: 40 (70.2%) responded and 10 (18%) provided at least some of the requested data. Chief reasons to decline sharing data were: i) effort required (n = 15, 50 %), ii) data loss (n = 14, 43.3% of respondents), iii) ethical concerns with data sharing (n = 3, 3 %), iv) skepticism towards quantitative meta-analyses (n =1, 3.3%). For full details on the email to the authors and their answers, see Supplementary Material.

3.2 Differences in acoustic patterns between individuals with schizophrenia and healthy controls

Detailed results are reported in Table 2. Hierarchical Bayesian meta-analyses revealed significant effects of diagnosis (in terms of Hedges' g) on pitch variability (-0.55, 95% CIs: -1.06, 0.09), proportion of spoken time (-1.26, 95% CIs: -2.26, 0.25), speech rate (-0.75, 95% CIs: -1.51, 0.04), and duration of pauses (1.89, 95% CIs: 0.72, 3.21), see Fig. 2. No significant effect was found for pitch mean (0.25, 95% CIs: -0.72, 1.30), intensity variability (0.739, 95% CIs: -2.01, 3.39), duration of utterance (-0.155, 95% CIs: -2.56, 2.26) and number of pauses (0.05, 95% CIs: -1.23, 1.13). We generally found high heterogeneity between studies, indicating a likely high diversity in samples and methods, and publication bias, indicating a tendency to publish only significant results, thus making the published literature not fully representative of the actual population of study (see Table 2).

Figure 2

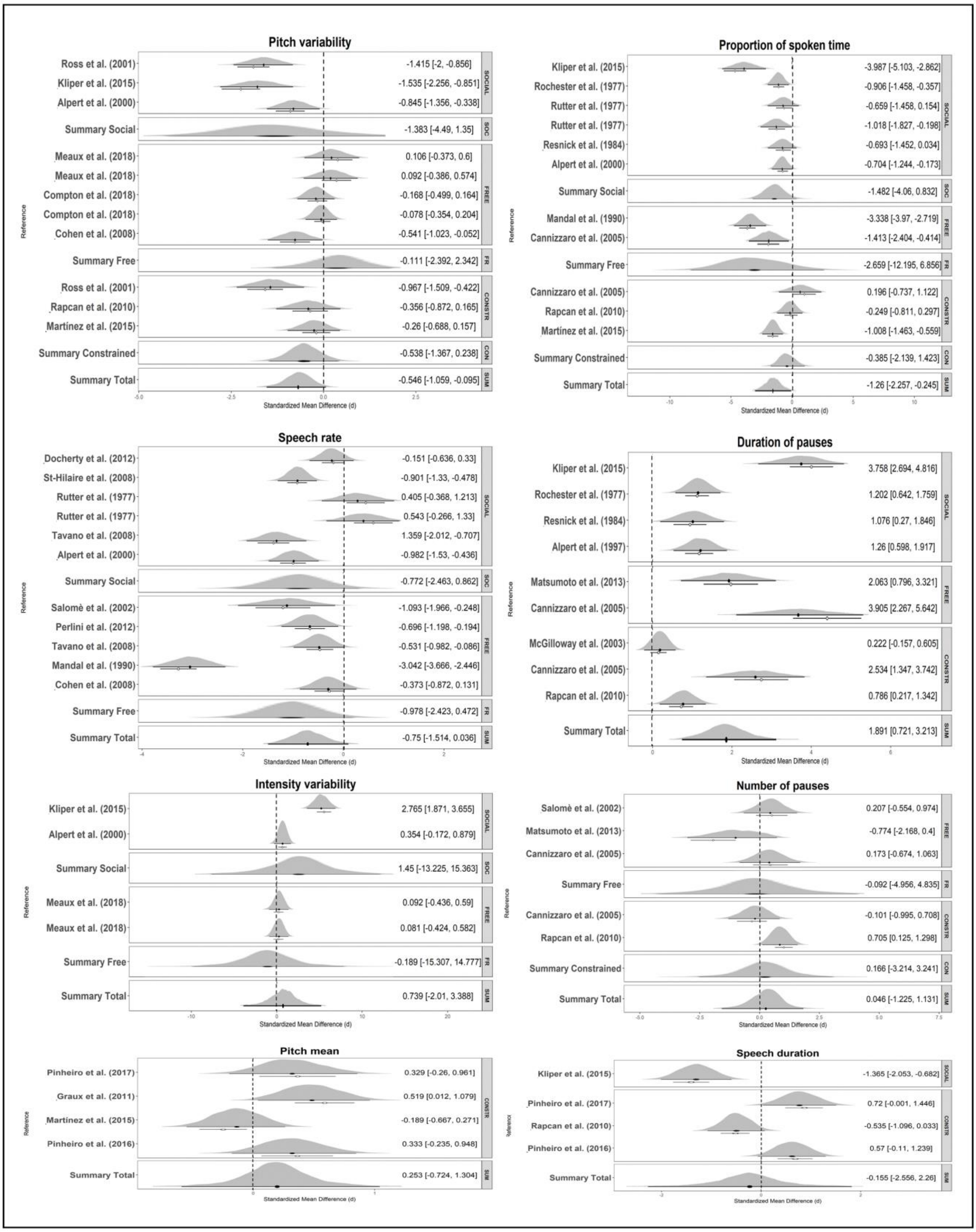

Pitch variability
Proportion of spoken time
Speech rate
Duration of pauses
Intensity variability
Number of pauses
Pitch mean
Speech duration
Reference
Standardized Mean Difference (d)

Table 2. Main Results of the meta-analysis for the effect of diagnosis on acoustic measures, and for the correlations between acoustic measures and symptoms ratings.

Effect of diagnosis on acoustic measures

Acoustic Features	**Participants (female) and median**	**Number of studies (articles)**	**Influential study**	**Estimates -Hedges'g [95% CI]**	**P - value**	**ER (Credibility)**	**Sigma squared [95% CI]**	**Q- stats (p-value)**	**Publication bias**
Pitch Mean	103 SZ (22 F) 95 CT (23 F)	4 (4)	Yes, Martinez Sanchez et al. (2015)	0.253 [-0.724, 1.304]	.273	3.467 (78%)	1.131 [0.005, 7.19]	8.282 (p = .041)	No, K = 0.0, p = 1.0
			Removing influential	0.505 [-0.859, 1.994]	.273	9.471 (90%)	2.143 [0.0, 16.934]		
Pitch variability	387 SZ (92F) 257 CT (106 F)	11 (8)	No	-0.546 [-1.059, -0.095]	.005	99.0 (99%)	0.566 [0.152, 1.64]	58.895 (p < .001)	Yes, K = -0.709, p = .002
Intensity variability	104 SZ (22 F) 65 CT (20 F)	4 (3)	Yes, Kliper et al. (2015)	0.739 [-2.01, 3.388]	.164	3.648 (78%)	7.888 [0.56, 47.934]	35.89 (p < .001)	No, K =1.0, p = .083
			Removing influential	0.152 [-1.005, 1.311]	.164	2.62 (72%)	1.932 [0.00, 13.486]		
Proportion of spoken time	267 SZ (106 F) 211 (98 F)	11 (9)	No	-1.26 [-2.257, -0.245]	.001	149.943 (99%)	2.538 [0.787, 7.224]	113.308 (p < .001)	No, K = -0.236 p= .359
Duration of utterance	93 SZ (30 F) 72 CT (30 F)	4 (4)	No	-0.155 [-2.556, 2.26]	.739	1.475 (60%)	6.045 [0.32, 37.49]	27.78 (p < .001)	No, K = 0.333, p = .75
Speech rate	336 SZ (111 F) 259 (107 F)	11 (9)	No	-0.75 [-1.514, 0.036]	.015	32.473 (97%)	1.447 [0.467, 3.915]	104.414 (p < .001)	No, K = -0.055, p = .879
Duration of pauses	221 SZ (128 F) 150 CT (92 F)	9 (8)	No	1.891 [0.721, 3.213]	< .001	234.294 (100%)	3.129 [0.754, 10.086]	75.624 (p < .001)	Yes, K = 0.667 p = .013
Number of pauses	68 SZ (23 F) 40 CT (13 F)	5 (4)	Yes, Matsumoto et al. (2013)	0.046 [-1.225, 1.131]	.782	1.321 (57%)	1.531 [0.017, 8.496]	11.61 (p = .02)	Yes, K = -1.0, p = .017
			Removing influential	0.355 [-0.991, 1.615]	.782	4.739 (83%)	1.481 [0.001, 10.692]		

Correlations between acoustic measures and clinical symptoms ratings

Acoustic Features	**Clinical features**	**Number of studies**	**Participants**	**Influential study**	**Pearson's r 95% CI**	**P - value**	**ER (credibility)**	**Sigma squared 95% CI**	**Q- stats**	**Publication bias**

Pitch Mean	Negative symptoms	7 (3)	107 SZ (33 F) 21	No	0.096 [-0.158, 0.346]	.136	4.198 (81%)	0.071 [0.00, 0.345]	9.976 (p = .126)	No, K = -0.619, p = .069
	Positive symptoms	4 (3)	107 (33 F) 21	No	-0.185 [-0.691, 0.316]	.04	6.89 (87%)	0.245 [0, 1.714]	3.586 (p = .31)	No, K = 0.333, p = .75
Pitch variability	General psychopatology	5 (4)	146 (48 F) 22	No	-0.091 [-0.34, 0.15]	.3	3.84 (79%)	0.057 [0, 0.354]	2.283 (p = .684)	No, K = 0, p =1.0
	Negative symptoms	11 (6)	261 (77 F) 22		-0.01 [-0.196, 0.144]	.836	1.117 (53%)	0.041 [0.002, 0.135]	19.292 (p = .037)	No, K = -0.2, p = 0.445
	Positive symptoms	4 (3)	107 (33 F) 21	Yes, Covington et al., (2012)	-0.027 [-0.686, 0.763]	.698	1.509 (60%)	0.525 [0.001, 4.294]	7.248 (p = .064)	No, K = 0.333, p = .75
				Removing influential	-0.05 [-0.715, 0.62]	.755	1.517 (60%)	0.422 [0.001, 2.944]		
	Alogia rating	9 (7)	313 (68 F) 26	No	-0.035 [-0.317, 0.22]	.465	1.5 (60%)	0.135 [0.032, 0.421]	45.478 (p < .001)	No, K = - 0.314, p = .246
	Flat affect rating	13 (10)	403 (81 F) 30	No	-0.106 [-0.262, -0.047]	.044	11.719 (92%)	0.053 [0.009, 0.153]	32.763 (p = .001)	No, K = -0.117 p = .582
Intensity variability	Flat affect rating	6 (5)	158 (22 F) 30	No	-0.005 [-0.324, 0.308]	.745	1.03 (51%)	0.117 [0.001, 0.658]	10.219 (p = .069)	No, K = -0.067, P = 1.0
Proportion of spoken time	General psychopatology	5(4)	124 (35 F) 22	Yes, Rapcan et al. (2010)	-0.026 [-0.53, 0.375]	.85	1.714 (63%)	0.268 [0.005, 1.411]	11.475 (p = .022)	No, K = -0.4, p = .483
				Removing influential	-0.069 [-0.536, 0.335]	.662	1.816 (65%)	0.235 [0.005, 1.335]		
	Negative psychopatology	9 (5)	146 (35 F) 22	No	-0.229 [-0.499, 0.035]	.198	23.29 (96%)	0.131 [0.027, 0.405]	35.506 (p < .001)	No, K = 0.333, p= .26
	Alogia rating	5 (4)	138 (23 F) 22	No	-0.413 [-0.723, -0.07]	< .001	58.259 (98%)	0.127 [0, 0.805]	7.344 (p = .119)	No, k = 0.333, p = .435
	Flat affect rating	6 (5)	161 (23 F) 22.5	No	-0.384 [-0.612, -0.082]	< .001	83.211 (99%)	0.08 [0, 0.456]	7.901 (p = .162)	Yes, K = .867, p = .017
Duration of pauses	Negative psychopatology	4 (4)	109 (30 F) 24	Yes, Rapcan et al. (2010)	0.302 [-0.199, 0.783]	.003	15.667 (94%)	0.246 [0, 1.754]	4.971 (p = .174)	No, K = 0.333, p = .75
				Removing influential	0.295 [-0.211, 0.757]	= .008	14.267 (93%)	0.37 [0, 2.129]		

Note: CI, credible interval; P values are 2-tailed; Evidence ratio (ER) quantify the evidence provided by the data in favor of the effect of associations between clinical features and acoustic measures (e.g. longer pauses associated to higher rating of alogia) against the alternatives (e.g. no association). An ER equal to 3 indicates the hypothesis is 3 times more likely than the alternative. A credibility score indicates the percentage of posterior estimates falling above 0.

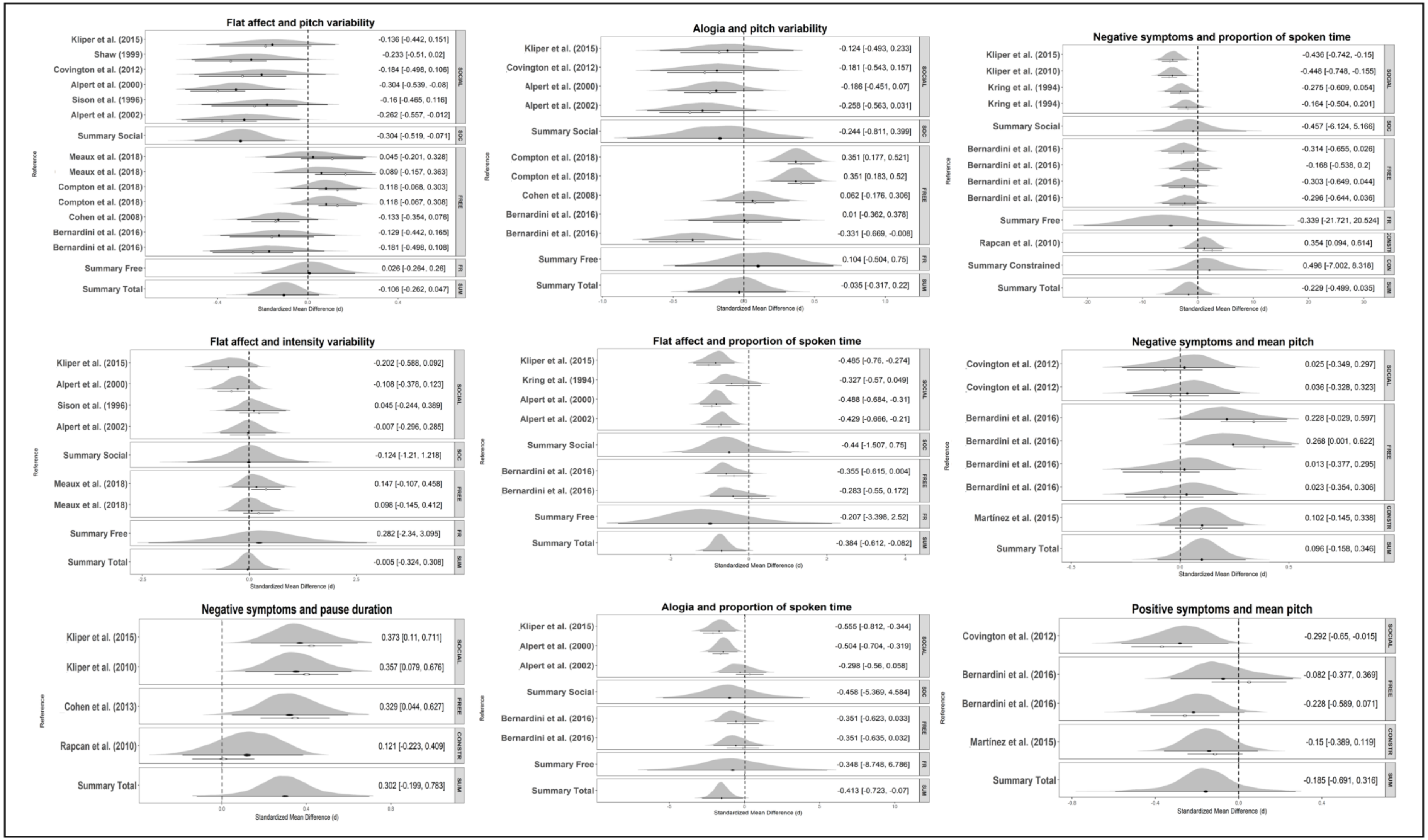
Moderator analysis

For detailed results, see Table S1 (in appendix). Adding the speech production task employed systematically increased the explained variability in SCZ atypicalities for pitch variability, proportion of spoken time, speech rate, number of pauses, duration of pauses and intensity variability (stacking weights: 100%). In general, we observe that dialogical and free speech show the biggest differences, while constrained monologue displays the smallest SCZ atypicalities in vocal patterns, except for pitch variability.

3.3 Correlation between acoustic measures and clinical ratings

For detailed results, see Table 2. Hierarchical Bayesian meta-analysis revealed significant overall correlation between flat affect and pitch variability (-0.11, 95% CIs: -0.26, 0.05) and proportion of spoken time (-0.38, 95% CIs: -0.61, -0.08), alogia and proportion of spoken time (-0.41, 95% CIs: -0.72, 0.07), positive symptoms and pitch mean (-0.19, 95% CIs: -0.69, 0.32), negative symptoms and pause duration (0.30, 95% CIs:-0.20, 0.78), see Fig. 3. No significant correlation was found between flat affect and intensity variability (-0.01, 95% CIs: -0.32, 0.31), alogia and pitch variability (-0.04, 95% CIs: -0.32, 0.22), general psychopathology and proportion of spoken time (-0.03, 95% CIs: -0.53, 0.375) and pitch variability (-0.09, 95% CIs:-0.34-, 0.15), positive symptoms and pitch variability (-0.03, 95% CIs: -0.68, .076), negative symptoms and pitch mean (0.01, 95% CIs:-0.16, 0.35), pitch variability (-0.01, 95% CIs:-0.20, 0.14), and proportion of spoken time (-0.23, 95% CIs: -0.50, 0.04) (see Table 2). We generally found high heterogeneity between studies, and publication bias (see Table 2).

Figure 3


Flat affect and pitch variability
Alogia and pitch variability
Negative symptoms and proportion of spoken time
Flat affect and intensity variability
Flat affect and proportion of spoken time
Negative symptoms and mean pitch
Negative symptoms and pause duration
Alogia and proportion of spoken time
Positive symptoms and mean pitch
Standardized Mean Difference (d)
Reference

Moderator analysis

For detailed results, see Table S2 (in appendix). Adding speech task to the model credibly improved it for correlations between pitch variability and positive symptom severity, negative symptom severity, alogia and flat affect, and between proportion of spoken time and total psychopathology, negative symptom severity, alogia and flat affect (stacking weight 100%). In general, we see that dialogic speech shows the strongest correlations with symptomatology, and constrained monological speech the weakest ones.

3.4. Multivariate machine learning (ML) studies

We found 4 ML articles fitting our criteria, all focused on identifying acoustic markers of the disorder[43,44,73,74] and 1 including the prediction of severity of clinical features from acoustic measures[73]. Three studies employed linear discriminant analysis (LDA) and one employed support vector machines to classify individuals with SZ vs. HC. All studies reported accuracy beyond 75% and up to 87.5%. All the results were cross-validated. Only one study[43] reported additional performance indices such as specificity, sensitivity, and area under the curve (AUC).

Only 1 study[73] attempted to predict the symptomatology (negative symptoms severity) from acoustic measures. The study relied on LDA and reported an accuracy of 78.6% in classifying individuals with SCZ with higher vs lower scores of negative symptoms (PANSS negative < 11 and SANS < 13), and 71.4 % accuracy in predicting a future (14 days after) measurement of negative symptoms.

Discussion

Overview

Early descriptions of schizophrenia point to atypical voice patterns and studies relying on perceptual judgments and clinical ratings of voice patterns have indeed found large differences between patients and controls[39]. This suggests the existence of acoustic markers of the disorder. We set out to systematically review and meta-analyze the literature on the topic to assess the evidence for atypical acoustic patterns as markers of the disorder and to better inform future research. We were able to analyze the aggregated data from 46 unique articles including 1212 individuals with SCZ and 699 HC. The univariate studies identified several null results, as well as weak atypicalities in pitch variability (perhaps in relation to flat affect), and stronger atypicalities in duration (possibly related to alogia and flat affect). The effect sizes suggest a within-sample discriminative accuracy between 66% and 80%, likely less if assessing new data. The multivariate ML studies paint a more promising picture, with overall out-of-sample accuracies between 76.5% and 87.5%. When assessing the relation between acoustic features and symptomatology, we found that specific symptoms that are more directly related to voice, e.g. in their description in clinical scales, yield slightly stronger results, with flat affect being related to speech variability and proportion of spoken time; and alogia being related to proportion of spoken time. Further, the results across all analyses suggest that dialogical productions, that is, tasks with a perhaps higher cognitive load and a more demanding social component, tend to involve larger effect sizes both in contrasting patients and controls and in assessing symptomatology. Free monological production follows and constrained production produces generally the smallest effects. Crucially, the studies analyzed mostly used widely different methods for sample selection, acoustic pre-preprocessing, feature extraction and selection. Indeed, we find large heterogeneity in the findings of the analyzed studies, and a large uncertainty in all our meta-analytic estimates.

What have we then learned? In line with a previous non-systematic meta-analysis (13 studies, Cohen et al 2014[39]), we do indeed find evidence for acoustic markers of schizophrenia, further supporting the relation between clinical features of SCZ and voice patterns. However, the effect sizes are too small for practical applications, not comparable to those of perceptual and clinical judgments, and in any case plagued by large between-studies variability. While good progress has been made in the field, the review highlights a number of issues to be overcome to more satisfactorily understand acoustic patterns in schizophrenia and their potential. In particular we identified the following obstacles to the scientific understanding of acoustic features in schizophrenia: i) small sample sizes in terms of both participants and repeated measures, ii) heterogeneous, not fully up-to-date and underspecified methods in data collection and analysis, leading to scarce comparability between studies; iii) very limited attempts at theory driven research directly tackling the mechanisms underlying atypical vocal patterns in schizophrenia. These are discussed below.

Sample size. Schizophrenia is a heterogeneous disorder, and indeed several studies attempted to more specifically investigate the relation of acoustic features with the symptomatology of the disorder. However, given the limited meta-analytic effect sizes and the awareness that replications tend to show a marked shrinkage of effect sizes[75], we need to move beyond small heterogeneous studies. The majority of the studies analyzed include between 20 and 30 patients, plausibly due to the difficulty in accessing clinical populations. However, an expected Cohen's d of 0.6 (pitch variability) would require at least 74 participants per group to reach a 95% power (calculations relying on G*Power[76]) at which effect size estimates are reliable[77]. If we considered the more conservative possibility of a smaller true effect size of 0.3, the required sample size would be 290 participants per group. While including as varied a sample as possible is an unavoidable concern, there are strategies to reduce the sample size needed. For instance, one could employ repeated measures, that is, collecting repeated voice samples over time. Using 10 repeated measures per participant brings the required sample from 290 participants per group to 82 (assuming that they are still representative of the full population). Repeated measures are also very useful to better understand the reliability of the acoustic patterns over re-testing and potentially across different contexts. In particular, we have seen that dialogical speech production tasks might yield stronger vocal differences, but without a controlled within-subject contrast it is difficult to assess whether this is due to the nature of the task or to other confounds in the sample and study design.

We had initially aimed to investigate the role of demographical (age, education, gender, language and ethnicity), cognitive and clinical features of the participants. However, we could not access sufficient information to perform these analyses, which would be best performed on individual-level data. Analysing how acoustic features vary with symptomatology and context of speech production can help uncover the mechanisms behind atypical vocal patterns and provide an additional insight into schizophrenia. Indeed, we observe that acoustic features are more strongly related to specific symptoms (alogia, flat affect) than to global scores of psychopathology.

Methods. We found that the field predominantly focuses on traditional acoustic features: pitch, intensity and duration measures. Even in these cases, the processing of the voice recordings and extraction of the features is poorly documented and arguably widely heterogeneous. Previous studies have found that different assumptions and settings in the feature extraction process might significantly affect the results (e.g. Kiss et al 2012[78] shows different results for different choice of ceiling in pitch extraction). Further, speech pathology and speech signal processing research has developed a wide array of acoustic features more directly relatable to production mechanisms like fine-grained muscle control, or clarity of articulation (for some examples see[79]), which are almost completely ignored in schizophrenia research. To overcome these barriers, we recommend the use of freely available open source software solutions providing standard procedures in the extraction of acoustic features and the documentation of the settings chosen[80,81]. Use of new features should be compared against this baseline to facilitate comparability between studies.

Further, the vast majority of the studies focused on one acoustic feature at a time failing to produce effects comparable to those found in perceptual judgment studies. This supports the idea that perception is a complex process, non-linearly combining multiple acoustic cues. Multivariate techniques may thus allow to better capture vocal atypicalities. Indeed, the four ML studies we were able to identify provide promising out-of-sample accuracies, indicating that voice of individuals with SCZ may contain enough information to reliable distinguish between the two populations. However, the almost complete lack of overlap in features and methods employed in these studies makes it hard to assess how reliable the findings are across samples and whether there are more promising features and algorithms we should focus on.

Theory-driven research. A common feature of many of the studies reviewed is the lack of theoretical background. For example, limited attention is paid to clinical features and their severity and the choice of the speech-production task and acoustic measures used is often under-motivated. On the contrary, by putting hypothesized mechanisms to the test, more theory-driven research on vocal production in schizophrenia would improve our understanding of the disorder itself. For instance, social cognitive impairments[82–84] would motivate hypotheses on prosodic patterns when speaking to an interlocutor, while lack of motivation and energy[85–87] would be reflected in a more general lack of articulatory clarity. By including different tasks with diverse cognitive and social constraints, it would be possible to produce more robust results not specifically bound to a specific context, and to investigate the mechanisms and contextual factors responsible for voice abnormalities.

Open Science. The recommendations to rely on large sample sizes, include individual differences, and cumulatively employ acoustic features from previous studies might seem too cumbersome, or even unreasonable, given the high costs of research, ethical and practical constraints in accessing clinical populations and proliferation of acoustic measures. This is why we recommend open science practices to be included already in the research design. Releasing in controlled and ethically sound ways one's datasets enables the construction of large collective samples and re-analysis of the data to replicate and extend previous findings. However, accessing previous datasets is currently unfeasible, due to lack of answers from corresponding authors, data loss and the practical and time-consuming hurdle of finding, preparing and sharing the data years after the study has been published. This suggests that planning data-sharing from the onset of the study is necessary to ensure a more open, collective and nuanced science of acoustic markers in schizophrenia, conscious of the individual differences and diverse symptomatology. Sharing identifiable (voice) data related to clinical populations requires serious ethical considerations and careful sharing systems, but there are available datasets of voice recordings in e.g. people with Parkinson's, bipolar disorder, depression and autism spectrum disorder[79,88–91], thus suggesting that these hurdles can be overcome. In line with these recommendations, all the data and the codes used in this manuscript are available at https://osf.io/qdkt4/.

Conclusion

We have systematically reviewed the evidence for acoustic markers of schizophrenia and its symptomatology, as well as the research practices employed. We did not find conclusive evidence for clear acoustic markers of schizophrenia, although pitch variability and duration are potential candidates. Multivariate studies are more promising, but their generalizability across samples could not be assessed. To advance the study of vocal markers of schizophrenia we outlined a series of recommendations towards more cumulative, open, and theory-driven research.

Reference List

1. Cohen AS, Mitchell KR, Docherty NM, Horan WP. Vocal expression in schizophrenia: Less than meets the ear. *J Abnorm Psychol*. 2016;125(2):299-309. doi:10.1037/abn0000136
2. Cohen AS, Najolia GM, Kim Y, Dinzeo TJ. On the boundaries of blunt affect/alogia across severe mental illness: Implications for Research Domain Criteria. *Schizophr Res*. 2012;140(1-3):41-45. doi:10.1016/j.schres.2012.07.001
3. Alpert M, Rosenberg SD, Pouget ER, Shaw RJ. Prosody and lexical accuracy in flat affect schizophrenia. *Psychiatry Res*. 2000;97(2-3):107-118. doi:10.1016/S0165-1781(00)00231-6
4. Hoekert M, Kahn RS, Pijnenborg M, Aleman A. Impaired recognition and expression of emotional prosody in schizophrenia: Review and meta-analysis. *Schizophr Res*. 2007;96(1-3):135-145. doi:10.1016/j.schres.2007.07.023
5. Andreasen NC, Hoffman R., Grove W. Mapping abnormalities in Language and cognition. In: Alpert A, ed. *Controversies in Schizophrenia*. New York, NY: Guilford Press; 1985:199-226.
6. Trémeau F, Malaspina D, Duval F, et al. Facial expressiveness in patients with schizophrenia compared to depressed patients and nonpatient comparison subjects. *Am J Psychiatry*. 2005;162(1):92-101. doi:10.1176/appi.ajp.162.1.92
7. Galynker I, Cohen L, Cai J. Negative Symptoms in Patients With Major Depressive Disorder: A Preliminary Report. *Neuropsychiatry Neuropsychol Behav Neurol*. 2000;13(3):171-176.
8. Bleuler E. *Dementia Praecox or the Group of Schizophrenias*. New York, NY: International University Press; 1911.
9. Kraepelin E. *Dementia Precox and Paraphrenia*. Edinburgh, UK: University of Edinburgh; 1919.
10. Wang K, An N, Li BN, Zhang Y, Li L. Speech Emotion Recognition Using Fourier

Parameters. *IEEE Trans Affect Comput*. 2015;6(1):69-75. doi:10.1109/TAFFC.2015.2392101

11. Trigeorgis G, Ringeval F, Brueckner R, et al. Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network. *2016 IEEE Int Conf Acoust Speech Signal Process*. 2016:5200-5204. doi:10.1109/ICASSP.2016.7472669
12. Yin B, Ruiz N, Chen F, Khawaja MA. Automatic cognitive load detection from speech features. *19th Australas Conf Comput Interact Entertain User Interfaces*. 2007;(November):249-255. doi:10.1145/1324892.1324946
13. Cohen AS, Elvevåg B. Automated computerized analysis of speech in psychiatric disorders. *Curr Opin Psychiatry*. 2014;27(3):203-209. doi:10.1097/YCO.0000000000000056
14. Pisanski K, Cartei V, McGettigan C, Raine J, Reby D. Voice Modulation: A Window into the Origins of Human Vocal Control? *Trends Cogn Sci*. 2016;20(4):304-318. doi:10.1016/j.tics.2016.01.002
15. Cohen AS, Hong SL. Understanding Constricted Affect in Schizotypy Through Computerized Prosodic Analysis. *J Pers Disord*. 2011;25(4):478-491. doi:10.1521/pedi.2011.25.4.478
16. Williams ME, Lee S, Bone D, et al. The Psychologist as an Interlocutor in Autism Spectrum Disorder Assessment: Insights From a Study of Spontaneous Prosody. *J Speech, Lang Hear Res*. 2014. doi:10.1044/2014_jslhr-s-13-0062
17. Tsanas A, Little MA, McSharry PE, Ramig LO. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson’s disease symptom severity. *J R Soc Interface*. 2011;8(59):842-855. doi:10.1098/rsif.2010.0456
18. Nevler N, Ash S, Jester C, Irwin DJ, Liberman M, Grossman M. Automatic measurement of prosody in behavioral variant FTD. *Neurology*. 2017;89(7):650-656.

doi:10.1212/WNL.0000000000004236

19. Faurholt-Jepsen M, Busk J, Frost M, et al. Voice analysis as an objective state marker in bipolar disorder. *Transl Psychiatry*. 2016;6:e856. doi:10.1038/tp.2016.123

20. Millan MJ, Fone K, Steckler T, Horan WP. Negative symptoms of schizophrenia: Clinical characteristics, pathophysiological substrates, experimental models and prospects for improved treatment. *Eur Neuropsychopharmacol*. 2014;24(5):645-692. doi:10.1016/j.euroneuro.2014.03.008

21. Ross ED, Orbelo DM, Cartwright J, et al. Affective-prosodic deficits in schizophrenia: comparison to patients with brain damage and relation to schizophrenic symptoms. *J Neurol Neurosurg Psychiatry*. 2001;70(5):597-604.

22. Andreasen NC. *Scale for the Assessment of Negative Symptom*. Iowa City: University of Iowa; 1984.

23. 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

24. 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

25. 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

26. 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

27. 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

28. 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

29. Colle L, Angeleri R, Vallana M, Sacco K, Bara BG, Bosco FM. Understanding the communicative impairments in schizophrenia: A preliminary study. *J Commun Disord*. 2013;46(3):294-308. doi:10.1016/j.jcomdis.2013.01.003

30. Brüne M, Bodenstein L. Proverb comprehension reconsidered—'theory of mind' and the pragmatic use of language in schizophrenia. *Schizophr Res*. 2005;75(2-3):233-239. doi:10.1016/j.schres.2004.11.006

31. Meilijson SR, Kasher A, Elizur A. Language performance in chronic schizophrenia: a pragmatic approach. *J Speech Lang Hear Res*. 2004;47(3):695-713. doi:10.1044/1092-4388(2004/053)

32. Champagne-Lavau M, Stip E. Pragmatic and executive dysfunction in schizophrenia. *J Neurolinguistics*. 2010;23(3):285-296. doi:10.1016/j.jneuroling.2009.08.009

33. Del-Monte J, Raffard S, Salesse RN, et al. Nonverbal expressive behaviour in schizophrenia and social phobia. *Psychiatry Res*. 2013;210(1):29-35. doi:10.1016/j.psychres.2013.05.034

34. Thoma P, Hennecke M, Mandok T, et al. Proverb comprehension impairments in schizophrenia are related to executive dysfunction. *Psychiatry Res*. 2009;170(2-3):132-139. doi:10.1016/j.psychres.2009.01.026

35. Bowie CR, Harvey PD. Communication abnormalities predict functional outcomes in

chronic schizophrenia: Differential associations with social and adaptive functions. *Schizophr Res*. 2008;103(1-3):240-247. doi:10.1016/j.schres.2008.05.006

36. Tan EJ, Thomas N, Rossell SL. Speech disturbances and quality of life in schizophrenia: Differential impacts on functioning and life satisfaction. *Compr Psychiatry*. 2014;55(3):693-698. doi:10.1016/j.comppsych.2013.10.016

37. Sparks A, McDonald S, Lino B, O'Donnell M, Green MJ. Social cognition, empathy and functional outcome in schizophrenia. *Schizophr Res*. 2010;122(1-3):172-178. doi:10.1016/j.schres.2010.06.011

38. Mueser KT, Sayers SL, Schooler NR, Mance RM, Haas GL. A multisite investigation of the reliability of the scale for the assessment of negative symptoms. *Am J Psychiatry*. 1994. doi:10.1176/ajp.151.10.1453

39. 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

40. Cohen AS, Mitchell KR, Docherty NM, Horan WP. Vocal expression in schizophrenia: Less than meets the ear. *J Abnorm Psychol*. 2016;125(2):299-309. doi:10.1037/abn0000136

41. Alpert M, Pouget ER, Silva R. Cues to the assessment of affects and moods: Speech fluency and pausing. *Psychopharmacol Bull*. 1995;31(2):421-424. doi:10.1016/j.joms.2006.09.018

42. Tavano A, Sponda S, Fabbro F, et al. Specific linguistic and pragmatic deficits in Italian patients with schizophrenia. *Schizophr Res*. 2008;102(1-3):53-62. doi:10.1016/j.schres.2008.02.008

43. Rapcan V, D'Arcy S, Yeap S, Afzal N, Thakore J, Reilly RB. Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia. *Med Eng Phys*. 2010;32(9):1074-1079. doi:10.1016/j.medengphy.2010.07.013

44. Martínez-Sánchez F, Muela-Martínez JA, Cortés-Soto P, et al. Can the Acoustic Analysis of Expressive Prosody Discriminate Schizophrenia? *Span J Psychol*. 2015;18(2015):E86. doi:10.1017/sjp.2015.85

45. Cannizzaro MS, Cohen H, Rappard F, Snyder PJ. Bradyphrenia and bradykinesia both contribute to altered speech in schizophrenia: A quantitative acoustic study. *Cogn Behav Neurol*. 2005;18(4):206-210. doi:10.1097/01.wnn.0000185278.21352.e5

46. 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

47. Meaux LT, Mitchell KR, Cohen AS. Blunted vocal affect and expression is not associated with schizophrenia: A computerized acoustic analysis of speech under ambiguous conditions. *Compr Psychiatry*. 2018;83:84-88. doi:10.1016/j.comppsych.2018.03.009

48. Fusaroli R, Lambrechts A, Bang D, Bowler DM, Gaigg SB. “Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis.” *Autism Res*. 2017;10(3). doi:10.1002/aur.1678

49. Docherty NM. Missing referents, psychotic symptoms, and discriminating the internal from the externalized. *J Abnorm Psychol*. 2012;121(2):416-423. doi:10.1037/a0026348

50. Zhang J, Pan Z, Gui C, Zhu J, Cui D. Clinical investigation of speech signal features among patients with schizophrenia. *Shanghai Arch Psychiatry*. 2016;28(2):95-102. doi:10.11919/j.issn.1002-0829.216025

51. 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

52. Bone D, Lee C-C, Chaspari T, Gibson J, Narayanan S. Signal Processing and Machine Learning for Mental Health Research and Clinical Applications [Perspectives]. *IEEE*

Signal Process Mag. 2017;34(5):196-195. doi:10.1109/MSP.2017.2718581

53. Huys QJM, Maia T V., Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci*. 2016;19(3):404-413. doi:10.1038/nn.4238
54. Bonneh YS, Levanon Y, Dean-Pardo O, Lossos L, Adini Y. Abnormal Speech Spectrum and Increased Pitch Variability in Young Autistic Children. *Front Hum Neurosci*. 2011;4(January):1-7. doi:10.3389/fnhum.2010.00237
55. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*.; 2013. doi:10.1007/978-1-4614-7138-7
56. Van Der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: A simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol*. 2014;14(1):1-13. doi:10.1186/1471-2288-14-137
57. Foody G. Impacts of Sample Design for Validation Data on the Accuracy of Feedforward Neural Network Classification. *Appl Sci*. 2017;7(9):888. doi:10.3390/app7090888
58. Bone D, Bishop SL, Black MP, Goodwin MS, Lord C, Narayanan SS. Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion. *J Child Psychol Psychiatry*. 2016;57(8):927-937. doi:10.1111/jcpp.12559
59. Chekroud A. T107. Why Validation Matters: A Demonstration Predicting Antipsychotic Response Using 5 Rcts. *Schizophr Bull*. 2018;44(Suppl 1(S157).
60. Stewart LA, Clarke M, Rovers M, et al. Preferred reporting items for a systematic review and meta-analysis of individual participant data: The PRISMA-IPD statement. *JAMA - J Am Med Assoc*. 2015. doi:10.1001/jama.2015.3656
61. Doebler P, Holling H. Meta-Analysis of Diagnostic Accuracy with mada. *R Packag*. 2015:1-15.

62. Field AP, Gillett R. How to do a meta-analysis. *Br J Math Stat Psychol*. 2010;63(3):665-694.

63. Quintana DS. From pre-registration to publication: A non-technical primer for conducting a meta-analysis to synthesize correlational data. *Front Psychol*. 2015;6(OCT):1-9. doi:10.3389/fpsyg.2015.01549

64. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36(3):1-18.

65. Williams DR, Rast P, Bürkner PC. Bayesian meta-analysis with weakly informative prior distributions. *PsyArXiv*. 2018. doi:10.17605/OSF.IO/7TBRM

66. Bürkner PC, Williams DR, Simmons TC, et al. Intranasal Oxytocin May Improve High-Level Social Cognition in Schizophrenia, but Not Social Cognition or Neurocognition in General: A Multilevel Bayesian Meta-analysis. *Schizophr Bull*. 2017;43(6):1291-1303. doi:10.1093/schbul/sbx053

67. Cochran WG. The Combination of Estimates from Different Experiments. *Biometrics*. 1954;10(1):101. doi:10.2307/3001666

68. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput*. 2017;27(5):1413-1432. doi:10.1007/s11222-016-9696-4

69. Begg CB, Mazumdar M. Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics*. 1994;50(4):1088. doi:10.2307/2533446

70. Core R Team. R: A language and environment for statistical computing. *R Found Stat Comput Vienna, Austria*. 2013.

71. Burkner PC. brms : An R Package for Bayesian Multilevel Models using Stan. *J Stat Softw*. 2013;80(Plummer):1-28. doi:10.18637/jss.v080.i01

72. Carpenter B, Guo J, Hoffman MD, et al. Stan : A Probabilistic Programming Language . *J Stat Softw*. 2017. doi:10.18637/jss.v076.i01

73. Püschel J, Stassen HH, Bomben G, Scharfetter C, Hell D. Speaking behavior and speech sound characteristics in acute schizophrenia. *J Psychiatr Res*. 1998. doi:10.1016/S0022-3956(98)00046-6

74. Kliper R, Portuguese S, Weinshall D. Prosodic analysis of speech and the underlying mental state. *Commun Comput Inf Sci*. 2016;604:52-62. doi:10.1007/978-3-319-32270-4_6

75. Klein RA, Vianello M, Hasselman F, et al. Many Labs 2: Investigating variation in replicability across samples and settings. *Adv Methods Pract Psychol Sci*. 2018;1(4):443-490.

76. Faul F, Erdfelder E, Lang A, Buchner A. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav Res Methods*. 2009. doi:10.3758/BRM.41.4.1149

77. Schönbrodt FD, Perugini M. At what sample size do correlations stabilize? *J Res Pers*. 2013. doi:10.1016/j.jrp.2013.05.009

78. Kiss G, van Santen JPH, Prud'Hommeaux ET, Black LM. Quantitative Analysis of Pitch in Speech of Children with Neurodevelopmental Disorders. *Interspeech*. 2012:1343-1346. http://20.210-193-52.unknown.qala.com.sg/archive/archive_papers/interspeech_2012/i12_1343.pdf.

79. Cummins N, Sethu V, Epps J, Schnieder S, Krajewski J. Analysis of acoustic space variability in speech affected by depression. *Speech Commun*. 2015;75:27-49. doi:10.1016/j.specom.2015.09.003

80. Degottex G, Kane J, Drugman T, Raitio T, Scherer S. COVAREP - A collaborative voice analysis repository for speech technologies. In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. ; 2014:960-964. doi:10.1109/ICASSP.2014.6853739

81. Eyben F, Wöllmer M, Schuller B. Opensmile: the munich versatile and fast open-

source audio feature extractor. *Proc ACM Multimed*. 2010:1459-1462. doi:10.1145/1873951.1874246

82. Green MF, Horan WP, Lee J. Social cognition in schizophrenia. *Nat Rev Neurosci*. 2015. doi:10.1038/nrn4005
83. Penn DL, Sanna LJ, Roberts DL. Social cognition in schizophrenia: an overview. *Schizophr Bull*. 2008;34(3):408-411. doi:10.1093/schbul/sbn014
84. Sergi MJ, Rassovsky Y, Widmark C, et al. Social cognition in schizophrenia: Relationships with neurocognition and negative symptoms. *Schizophr Res*. 2007. doi:10.1016/j.schres.2006.09.028
85. Trémeau F, Goldman J, Antonius D, Javitt DC. Inpatients with schizophrenia report impaired situational motivation but intact global and social motivation. *Psychiatry Res*. 2013. doi:10.1016/j.psychres.2013.05.031
86. Lysaker P, Bell M. Work and Meaning: Disturbance of Volition and Vocational Dysfunction in Schizophrenia. *Psychiatry*. 1995;58(4):392-400. doi:10.1080/00332747.1995.11024743
87. Frith CD. *The Cognitive Neuropsychology of Schizophrenia*. Hove: Lawrence Erlbaum; 1992.
88. Gratch J, Artstein R, Lucas G, et al. The Distress Analysis Interview Corpus of human and computer interviews. In: *LREC*. ; 2014.
89. Ambite JL, Tallis M, Alpert K, et al. SchizConnect: Virtual Data integration in neuroimaging. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. ; 2015. doi:10.1007/978-3-319-21843-4_4
90. Schuller B, Steidl S, Batliner A, et al. The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism. In: *Proceedings of the Annual Conference of the International Speech Communication Association,*

INTERSPEECH. ; 2013.

91. Tsanas A, Zañartu M, Little MA, Fox C, Ramig LO, Clifford GD. Robust fundamental frequency estimation in sustained vowels: Detailed algorithmic comparisons and information fusion with adaptive Kalman filtering. *J Acoust Soc Am*. 2014;135(5):2885-2901. doi:10.1121/1.4870484

Figure Legends

Figure 1. Flow chart showing the literature search and study selection process in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines.

Figure 2. Forest plots of effect sizes (Hedges'g) for all the acoustic measures. The x-axis report effect sizes (black dot, positive values indicate that individuals with SCZ are higher on that acoustic measures, while negative values the opposite), posterior distribution (density plot) and original data point (white dot) for each study. The y-axis indicates the studies for which statistical estimates have been provided.

The dotted vertical line indicates the null hypothesis (no difference between the populations). The studies are grouped by the speech task used to collect voice recordings (Constr = constrained monological, Free = free monological, Social = social interaction). When adding speech task credibly improved the model, we reported below each specific task group the summary effect size for that group. Filled diamonds represent summary effect sizes.

Figure 3. Forest plots of effect size (Pearson's r) for the correlations between clinical symptoms and acoustic measures. The x-axis report effect sizes (black dot, positive values indicate a positive relation between acoustic measures and clinical symptoms rating, e.g. increased pause duration associated with increased rating of alogia, while negative values the opposite), posterior distribution (density plot) and original data point (white dot) for each study. The y-axis indicates the studies for which statistical estimates have been provided.

The dotted vertical line indicates the null hypothesis (no difference between the populations). The studies are grouped as indicated in Figure 2