

# Can the Acoustic Analysis of Expressive Prosody Discriminate Schizophrenia?

Francisco Martínez-Sánchez<sup>1</sup>, José Antonio Muela-Martínez<sup>2</sup>, Pedro Cortés-Soto<sup>2</sup>,  
Juan José García Meilán<sup>3</sup>, Juan Antonio Vera Ferrándiz<sup>1</sup>, Amaro Egea Caparrós<sup>1</sup> and  
Isabel María Pujante Valverde<sup>1</sup>

<sup>1</sup> Universidad de Murcia (Spain)

<sup>2</sup> Universidad de Jaén (Spain)

<sup>3</sup> Universidad de Salamanca (Spain)

**Abstract.** Emotional states, attitudes and intentions are often conveyed by modulations in the tone of voice. Impaired recognition of emotions from a tone of voice (receptive prosody) has been described as characteristic symptoms of schizophrenia. However, the ability to express non-verbal information in speech (expressive prosody) has been understudied. This paper describes a useful technique for quantifying the degree of expressive prosody deficits in schizophrenia, using a semi-automatic method, and evaluates this method's ability to discriminate between patient and control groups. Forty-five medicated patients with a diagnosis of schizophrenia were matched with thirty-five healthy comparison subjects. Production of expressive prosodic speech was analyzed using variation in fundamental frequency ( $F_0$ ) measures on an emotionally neutral reading task. Results revealed that patients with schizophrenia exhibited significantly more pauses ( $p < .001$ ), were slower ( $p < .001$ ), and showed less pitch variability in speech ( $p < .05$ ) and fewer variations in syllable timing ( $p < .001$ ) than control subjects. These features have been associated with «flat» speech prosody. Signal processing algorithms applied to speech were shown to be capable of discriminating between patients and controls with an accuracy of 93.8%. These speech parameters may have a diagnostic and prognosis value and therefore could be used as a dependent measure in clinical trials.

Received 27 September 2014; Revised 13 January 2015; Accepted 4 March 2015

**Keywords:** Schizophrenia, flat affect, expressive prosody, speech analysis.

The difficulty in identifying and expressing emotional states is one of the major symptoms of schizophrenia as well as a predictor of social adjustment (Brazo, Beaucousin, Lecardeur, Razafimandimby, & Dollfus, 2014; Hooker & Park, 2002; Kee, Green, Mintz, & Brekke, 2003); This “flat” affective state is present in at least 66% of schizophrenics (Trémeau et al., 2005).

The difficulty to identify the emotional state of others in facial expression (Kohler, Walker, Martin, Healey, & Moberg, 2010) and in the tone of their voice (Hoekert, Kahn, Pijnenborg, & Aleman, 2007) has been well established as one of the main predictors of deterioration in all phases of the disorder: from the first episode (Horan et al., 2012), during its chronicity (Green et al., 2012), and even in its high-risk states (Allot et al., 2014).

These difficulties are part of the alterations in social cognition (described as the ability to construct representations of the relationship between oneself and others,

and to use them flexibly in behavior regulation, Adolphs, 2001), along with social perception social, theory of mind, social knowledge and attributional style, which have been consistently linked to schizophrenia (Couture, Penn, & Roberts, 2006; Green & Horan, 2010; Kring & Ellis, 2013; Penn, Sanna, & Roberts, 2008).

Similarly, the emotional expressive ability is also impaired (Cohen, Kim, & Najolia, 2013), although there is evidence (Kring & Moran, 2008) indicating the dissociation between expressiveness and emotional experience, since flat affect does not necessarily lead to a reduction in emotional experience.

Prosody has been studied to a lesser extent than the ability to identify and express emotions facially. In speech, not only the changes in the melody produced by variations in the frequency of opening and closing of the vocal cords are perceived, but also the changes of rhythm, speed, intonation, pauses, intensity and other spectral alterations that are perceived by the listener as melodic variations, and interpreted subjectively as paralinguistic signals, essential for the understanding and interpretation of the utterance and the identification of the emotional and motivational state of the speaker (Patel, 2008).

---

Correspondence concerning this article should be addressed to Francisco Martínez-Sánchez. Dpto. de Psicología Básica y Metodología. Facultad de Psicología. Universidad de Murcia. Campus Universitario de Espinardo. 30100. Murcia (Spain).

E-mail: franms@um.es

Receptive prosody studies have shown that patients exposed to voice samples with different emotions and questioned about the emotion they have heard show a marked difficulty to identify such emotions. An extensive meta-analysis (Hoekert et al., 2007) comprising twenty articles with a total of 663 patients, has described a significant effect size ( $d = -1.24$ ) when comparing the performance of schizophrenic participants and controls.

Regarding expressive prosody, the studies that have collected voice samples and have asked participants to encode different emotional states have concluded that there is a marked difficulty in expressing emotions verbally (Putnam & Kring, 2007). Hoekert et al.'s (2007) meta-analysis has also concluded there was a significant effect size ( $d = -1.11$ ) between the eleven studies reviewed with data from 186 patients.

Most of the studies on expressive prosody have used tasks in which participants were asked to encode various emotional states perceived in voice (Hoekert et al., 2007); fewer studies have asked their participants to spontaneously narrate sad, happy and anger events they have experienced (Alpert, Rosenberg, Pouget, & Shaw, 2000; Shaw et al., 1999). These methods have, in our view, reduced the ecological validity of the study of expressive prosody, since high intensity discrete emotional states (anger, sadness, fear, etc.) are rarely expressed in colloquial language in the manner described in these studies. In everyday social interaction, the affective state, intentions and aptitude are constantly and inexorably present in the tone of voice, though less explicitly. An alternative method is used by Cohen et al. (Cohen, Iglesias, & Minor, 2009; Cohen & Hong, 2011), consisting of analyzing the prosody of verbal responses triggered by emotional stimuli presentations, using the "International Affective Picture System" (Lang, Bradley, & Cuthbert, 2005).

In contrast to these methods, only three studies have used emotionally neutral readings (Cohen, Alpert, Nienow, Dinzeo, & Docherty, 2008; Dickey et al., 2012; Leentjens, Wielaert, Harskamp, & Wilmink, 1998). To evaluate the non-emotional expressive prosody in schizophrenia, this paper presents a potentially useful technique, using an emotionally neutral text that has been proven useful in quantifying prosody in other disorders (Martínez-Sánchez, Meilán, Pérez, Carro, & Arana, 2012). The procedure consists of the semi-automatic analysis of the variations in the trajectory of pitch and height perception of the fundamental frequency (of the vocal cords' vibrations) of the vocalic syllable nuclei, as this is the point of greatest loudness, using a purely acoustic base to extract the harmonic peak without the need of phonetic segmentation. The data derived from the behavior of the  $F_0$  in the intensity of vocalic segments yield a complete melodic pattern of

the speaker that shows significant changes in tone, both upstream (prosodic peaks) and downstream (prosodic valleys), within the syllabic nucleus, as well as between different nuclei (see Annex I for a description of the prosodic parameters used).

This procedure has many advantages over previously used prosodic analyses, since it increases the reliability and validity of results, speeds up the production of prosodic parameters and minimizes the influence of the coding skills of the subject, as it uses an emotionally neutral text. It also increases its ecological validity, as it is closer to the colloquial language used in everyday interactions. Finally, the procedure does not require phonetic segmentation, which virtually eliminates any errors the experimenter could commit in the process of quantification, as well as any differences in estimation between various experimenters.

In the present work, this procedure is used in order to objectively quantify the deficits in expressive prosody in schizophrenia, as well as to assess its discriminatory power between groups. It is hypothesized that the group of schizophrenia patients will show a significantly flatter prosodic profile, characterized by less variability in the dynamics and the path of the vocalic nuclei and in voice intensity as well as an increase in the number of pauses than those obtained by the control group.

## Method

### Statistical design

A cross-sectional, analytical, observational and retrospective design was used.

### Participants

A sample of 80 participants, divided into two groups, was recruited: 45 patients diagnosed with schizophrenia and 35 asymptomatic controls.

The group of patients ( $M_{\text{age}} = 39.49$ ,  $SD = 10.89$ ; 71.1% male) were recruited from various Mental Health units of the Andalusian Health Service of the province of Jaen (Therapeutic and Rehabilitation Community of the Jaen area and the Mental Health Clinics of Martos and Andujar). All participants were evaluated using the clinical version of the Structured Clinical Interview (First, Spitzer, Gibbon, & Williams, 1996), following the criteria established in the *DSM-IV-TR* (2000). The average duration of the disorder, from its initial diagnosis, was 21.17 years ( $SD = 5.65$ ), the mean number of relapses was 3.47 ( $SD = 1.89$ ) and the mean time elapsed since the last relapse was 45.88 days ( $SD = 25.67$ ). The average dose of chlorpromazine equivalent units was 669.88 mg / day ( $SD = 559.31$ ).

Meanwhile, the control group ( $M_{age} = 35.34$ ,  $SD = 10.48$ ; 62.9% male) was matched with the patient sample for age, sex and educational level, and was extracted from the same social environment as the patient sample. They had no history of mental or neurological disorders or drug or alcohol abuse, which were considered exclusion criteria.

### Materials and Procedure

Brief Psychiatric Rating Scale (BPRS; Lukoff, Nuechterlein, & Ventura, 1986) in the Spanish validation by Peralta and Cuesta (1994). The 0–5 points response range was used as it increases the inter-rater reliability (Bech, Larsen, & Andersen, 1988). It is composed of 18 items, to be administered by an experienced therapist after a semi-structured interview (15–25 minutes in length). Each item is scored using a Likert-type scale with 5 levels of intensity, where 0 represents the absence of the symptom and 4 represents extreme gravity.

In order to record speech, a professional Foxtek FR-2LE recorder was used, with a resolution of 24 bits and a 48 kHz sampling rate, using a cardioid AKG D3700S microphone. Samples were edited using the acoustic voice analysis 5.1.42 Praat program (Boersma & Weenink, 2013). Annex I contains the definition of the various parameters used.

The study was conducted between June and December of 2013. All participants were adequately informed in order to sign their consent according to the protocols of the Bioethics Committees of the participating institutions. This study complied with the ethical principles of the Declaration of Helsinki for medical research involving human subjects. The procedure was performed in a single session, initially collecting socio-demographic information (age, marital status, etc.) and clinical data (age at onset of symptoms and diagnosis, duration of illness, number of admissions, time since last admission, etc.) as well as administering the BPRS. The doses of chlorpromazine equivalent units ingested by patients were also registered. Subsequently, speech recordings were made.

The task entailed reading the first paragraph (405 syllables) of "Don Quijote" by Miguel de Cervantes. The recordings were performed in a silent room (but not acoustically isolated), placing the microphone 8 cm away and at a 45° angle from the participant's mouth in order to prevent any aerodynamic noise.

To quantify prosodic patterns, the automatic prosodic transcription of the recordings was performed using the algorithms implemented by Mertens (2004) on the Praat program (Boersma & Weenink, 2013). The estimation of the prosodic speech profile was performed analyzing the variations in the height trajectory and pitch perception (prosodic peaks and valleys) of the  $F_0$  of the

vocalic syllable nuclei that contain voice signals, on a peak intensity delimited -3dB and -9dB to left and right, respectively, in order to represent the melodic movements perceived by the human ear. The value of the left limit (-3dB) eliminates most of the microprosodic disturbances and stylizes the beginning of the syllable, while the right (-9dB) limit preserves the variations in tone of accented vowels.

In this paper, a detection range for the  $F_0$  of 65–650 Hz was established for 0.005s windows; the following threshold intensity was established for the automatic segmentation in the stylization of the algorithm:  $Glissando = 0.32/T^2$ ,  $DG = 30$ ,  $dmin = 0.05$ . To determine the presence of a vowel, a  $0.32/T^2$  semitone threshold was allocated, where  $T$  is the duration of the vowel in seconds. If the tone's exchange rate is higher than the threshold defined by the perceptual values of voice detection, a value proportional to the *glissando* threshold (continuous slippage of the melodic line in the same syllable) was assigned, whereas if the value is lower than the threshold, the same value as the median of the voice sample analyzed was assigned. It should be noted that while the standard psychoacoustic threshold for isolated vowels is  $G = 0.16/T^2$ , voice flow is rarely linear during natural speech, hence, the value assigned in this present study has been shown to more adequately model voice variations, especially in automatic transcription.

### Data analyses

The IBM SPSS (version 21) statistical package was used. The Student- $t$  test for independent samples was used to define the differential prosodic profile between groups and Pearson correlations were used to assess the relationship between variables. Finally, a discriminant analysis was performed in order to assess the ability of prosodic variables to classify subjects into both groups.

### Results

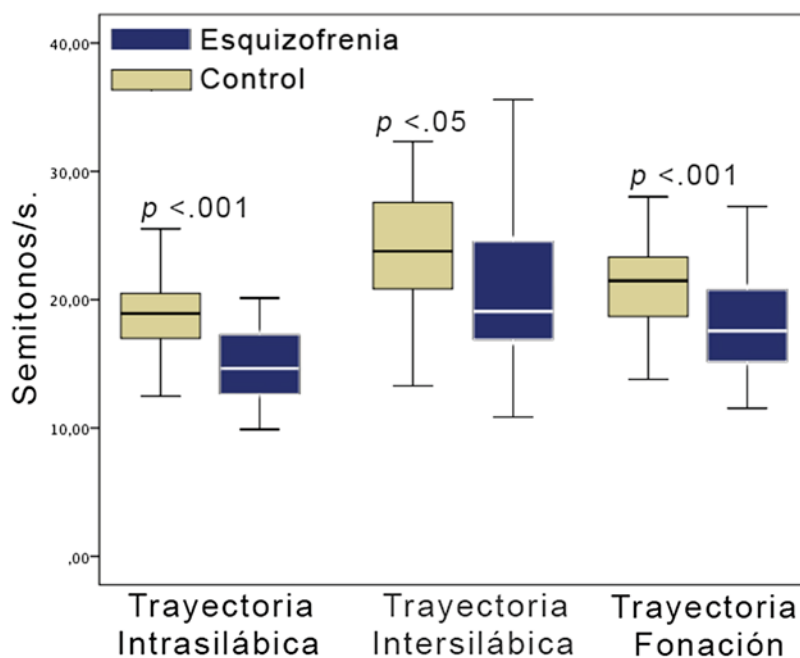
Different statistical tests were conducted to assess the absence of statistically significant intergroup differences among the sociodemographic variables. No differences for the variable "age" ( $t_{79} = -1.76$ ;  $p = .091$ ), educational level (measured in months of schooling;  $t_{79} = 1.74$ ;  $p = .085$ ), or for the distribution of sex per group ( $\chi^2 = .611$ ;  $p = .434$ ) were observed.

The mean comparisons show the existence of significant intergroup differences in all the prosodic variables, but not for those dependent on the frequency ( $F_0$ ) (Table 1). The main prosodic parameters (valleys, prosodic dynamics, inter- and intra-syllabic and phonation trajectories) of the Schizophrenia patient group yielded significantly lower levels than those obtained by the control group (Figure 1), showing in general,

**Table 1.** Descriptive statistics and mean comparison tests of the analyzed prosodic parameters

VV. DD.	Group		<i>t</i>	<i>p</i>
	Control (SD)	Schizophr. (SD)		
Task duration (s)	83.85 (15.65)	124.12 (43.15)	-5.25	.001
Intensity (dB)	73.79 (1.40)	71.08 (1.50)	8.22	.001
Pause rate $\geq$ 300 ms (%)	28.58 (5.98)	37.64 (10.69)	-4.49	.001
Mean F <sub>0</sub> (Hz)	156.57 (43.89)	143.02 (40.33)	1.43	.156
F <sub>0</sub> SD (Hz)	29.67 (8.32)	27.50 (10.36)	1.00	.303
F <sub>0</sub> Range (ST)	11.00 (2.90)	11.27 (3.84)	-3.43	.733
Syllabic dynamics (%)	4.31 (3.02)	2.80 (2.50)	2.44	.017
Prosodic peaks (%)	1.09 (1.10)	.76 (1.19)	1.26	.206
Prosodic valleys (%)	3.23 (2.22)	2.05 (1.85)	2.59	.011
Intra-syllabic Trajectory (ST/s)	18.51 (2.87)	14.11 (2.68)	7.05	.001
Inter- syllabic Trajectory (ST/s)	24.26 (4.89)	20.86 (7.20)	2.39	.014
Phonation Trajectory (ST/s)	21.32 (3.54)	17.26 (3.97)	4.75	.001

Note: dB = Decibels; Hz = Hertz; F<sub>0</sub> = Fundamental frequency; ST/s = Semitones/second.

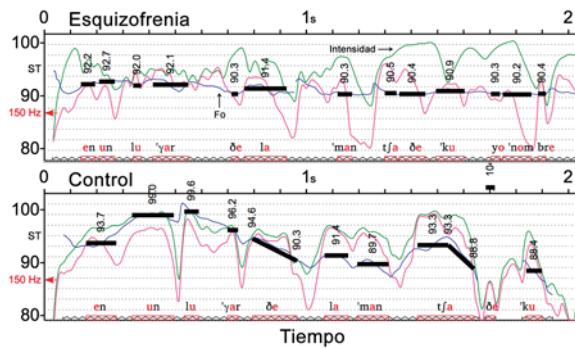
**Figure 1.** Prosodic trajectories of syllable nuclei for both groups.

a sparsely prosodic and melodically flatter speech than that of controls (Figure 2). Moreover, they spent more time performing the task, made a greater proportion of pauses and exhibited a significantly lower voice intensity, accentuating their perception of dysprosody.

As expected, education level significantly affected the performance of the task in the control group. For example, in the schizophrenia group, the higher the educational level, the lower the time spent on reading the text and the lower the proportion of pauses performed. It should be noted that for a test to be used for screening, it must be scarcely sensitive to the effects

of education in the pathological group; the poor correlation between prosodic variables and educational level in the patient group for this task can be appreciated in Table 2.

The years of chronicity of the disorder is the clinical variable most strongly associated with prosodic variables in the schizophrenia patients group; the greater the number of years elapsed since diagnosis of the disorder, the smaller the intrasyllabic ( $r = -.377$ ;  $p = .028$ ), and phonation ( $r = -.422$ ;  $p = .013$ ) trajectories were. Similarly, the time elapsed since the last relapse correlated with the phonation trajectory ( $r = .404$ ;  $p = .018$ ).



**Figure 2.** Example of prosogram from a patient with schizophrenia and from a control participant.

**Table 2.** Correlations between prosodic variables and months of schooling

	Control	Schizophrenia
	Education	Education
Task duration	-.506+	-.353
% Pauses	-.567+	-.305
Syllabic dynamics	-.475+	-.323
Prosodic peaks	-.561+	-.289
Prosodic valleys	-.371	-.259
Intra-syllabic Trajectory	-.083	-.208
Inter-syllabic Trajectory	-.173	.027
Phonation Trajectory	-.134	-.019

Note: +  $p < .006$  (Bonferroni's correction).

Drug treatment did not induce significant changes in prosody, although a trend was observed which suggests that the higher the dose of chlorpromazine was, the less time was spent on the reading task ( $r = -.280$ ;  $p = .063$ ) and fewer pauses were made ( $r = -.251$ ;  $p = .091$ ). Moreover, the BPRS scores were not significantly correlated to the prosodic variables, although the positive symptoms scale score negatively correlated with the intensity of voice ( $r = -.346$ ;  $p = .021$ ). The scale's item that evaluates blunted affect did not correlate significantly with any prosodic parameter, obtaining the highest degree of correlation with the "Intrasyllabic trajectory" variable and with blunted affect ( $-.219$ ;  $p = .118$ ). No significant correlations were found between the total scores of the BPRS scale and the prosodic parameters evaluated, or between these parameters and the scores obtained in the positive and negative symptoms subscales.

Finally, a discriminant analysis was performed, in order to assess the ability of the prosodic parameters studied to distinguish subjects in both groups. The canonical discriminant function explained 100% of the variance (canonical correlation = .828;  $\lambda_{14} = .314$ ;  $\chi^2_{14} = 82.16$ ;  $p < .001$ ), with Intensity, Intrasyllabic trajectory,

Total length, Phonation trajectory and Pause rate being the variables with the greatest discriminating power (with standardized canonical discriminant function coefficients of .630, .541,  $-.403$ , .365 y  $-.344$  respectively). Conversely, the dependent variables of the fundamental frequency ( $M F_0 = -.003$ ;  $F_0$  Range =  $-.026$ ;  $F_0$  SD = .077) were those that showed less discriminatory power. Overall, the discriminant analysis allowed for the correct classification of 93.8% of the original data (Table 3); in the cross-validation analysis, 87.5% of cases were correctly classified.

## Discussion

The objective of this research was to study the differential expressive prosodic patterns in a group of schizophrenic patients and in a group of asymptomatic subjects using an unemotional text as the reading task. The results show the existence of marked intergroup differences, with the clinical group exhibiting a slow and low-intensity speech, with many pauses. Moreover, this group's dysprosody is characterized by scarce tone changes, both within the syllable nucleus, and between adjacent syllabic nuclei. All these characteristics yield an "emotionally flattened" speech (Treméau, 2006).

These results concur with those obtained in several studies. Firstly, the low dynamics of both the inter- and intra-syllabic nuclei observed in our results, in spite of not have been studied before, is the result of the scarce changes in  $F_0$ . This is consistent with previous studies (Alpert et al., 2000) reporting that the schizophrenic speech is characterized by few inflections of speech. Secondly, the large number of pauses has also been previously identified as a characteristic of the disease (Alpert, Kotsaftis, & Pouget, 1997; Clemmer, 1980; Cohen & Elvevåg, 2014; Cohen, Mitchell, & Elvevåg, 2014). They can be attributed to the reduction in verbal fluency and the discrimination of phonemes, processes that are altered in schizophrenia (Johnson-Selfridge & Zalewski, 2001; Kugler & Caudrey, 1983). Thirdly, the voice signal is less intense (Pascual, Solé, Castellón, Abadía, & Tejedor, 2005), which also accentuates the perception of dysprosodic speech. Finally, these subjects take longer to complete the task, evidencing the limited cognitive resources available to perform a complex task such as reading (Cohen, McGovern, Dinzeo, & Covington, 2014; Melinder & Barch, 2003).

The dependent parameters of the fundamental frequency (mean, standard deviation and range) showed no differences between groups. This is not unexpected, as even though these measures contribute to variations in tone, they are microprosodic measures, as opposed to the suprasegmental parameters studied, which have been shown to be significant intergroup discriminators.



**Table 3.** Percentage of correctly classified cases in the discriminant analysis

	Prognosticated group membership	
	Control	Schizophrenia
Control	32 (91.4 %)	3 (8.6 %)
Schizophrenia	2 (4.4 %)	43 (95.6 %)

Note: 93.8 % of the original cases were classified correctly.

The observed differences in the variations in the syllabic dynamics of the vocalic nucleus are especially noteworthy, as they show a differential prosodic profile in both groups regarding their melodic slopes. Thus, although the percentage of upward vocalic nuclei (expressed by prosodic peaks) did not differ between the groups, the downward vocalic nuclei (prosodic valleys) were significantly lower in the group with schizophrenia. In the Spanish language, declarative intonation is typically downward, fully explaining the above-mentioned, while the interrogative intonation is, however, generally upward, presenting an incomplete utterance (Cantero, 2003). As the text used in the task was a declarative text, the clinical group's intonation was clearly inadequate, a fact that the listener may perceived as discordant with the meaning of the sentence being read.

The clinical variables have proven to be insignificant in expressive prosody. Our results show a correlation between the years of chronicity of the disorder and the impairment of intersyllabic and phonation trajectories. Although there is a lack of data from other investigations to compare, it is known that the ability to identify facial and prosodic expressions reduces with the duration of illness (Kucharska-Pietura, David, Masiak, & Phillips, 2005; Silver, Goodman, Knoll, Isakov, & Modai, 2005; Silver, Shlomo, Turner, & Gur, 2002). Although drug treatment does not induce changes in prosody, beyond increasing the duration of the task and the number of pauses, Hoekert et al.'s (2007) meta-analysis does not verify the existence of any significant relationships between these variables.

The relationship between the increase in positive symptoms and the reduction in the intensity of the voice is paradoxical, because one would expect the negative symptoms to be related to a lower intensity of voice (considering it an indicator of flat prosody), as reported by Cohen and Hong (2011), although their results did not reach statistical significance, while positive symptoms would correlate to an increase in verbal fluency. It is known that positive symptoms correlate with the difficulty in discriminating the fundamental frequency changes in sentences with an emotional

content (Matsumoto et al., 2006). Similarly, patients who experience hallucinations and delusions show alterations when identifying changes in the tone of voice (Johns et al., 2001).

Finally, the scores of the BPRS subscales have not yielded any significant correlations with prosody. These scales' limitations are well-known (Cohen & Elvevåg, 2014; Nicholson, Chapman, & Neufeld, 1995). Even though they are able to identify differences of up to six standard deviations when comparing negative symptoms of patients and controls (Emmerson et al., 2009), they are relatively insensitive to changes in the patient's condition and they induce response biases that make it difficult for even trained evaluators to notice specific aspects of behavior related to alogia and blunted affect within the patient's speech (Alpert, Shaw, Pouget, & Lim, 2002). On the other hand, our results coincide with those reported by Cohen et al. (2013), who found no relationship between the symptoms of a schizophrenic group and variables related to the expressive prosody (number of pauses and variability of  $F_0$ ).

While the slow implementation of the task (duration and rate of pauses) is conditioned by educational level in both groups (though to a greater extent in the clinical group), the variables related to the trajectory of the syllabic nuclei exhibit no relationship to these variables in any of the two groups, showing that the prosody results are scarcely sensitive to the cultural level, coinciding with the results reported by Leentjens et al. (1998).

The used procedure has proven to be, in our view, a valid and reliable alternative to accurately record non-emotional expressive prosody in colloquial speech, as its results can relate to patients' everyday interactions with their environment. The small number of existing studies using non-emotional stimuli to assess the expressive prosody in schizophrenia is surprising (Cohen et al., 2008; Dickey et al., 2012; Leentjens et al., 1998).

Furthermore, the procedure provides numerous advantages to those traditionally employed. It allows to objectively quantify the degree of dysprosody in a fast and non-intrusive way, without causing discomfort to the patient. Additionally, the acoustic analyses are highly sensitive to changes in the voice; therefore, their use is potentially useful in the study of the evolution of the disorder and in evaluating drug treatments. The high discriminating ability (93.8%) is higher than that achieved by Kliper et al. (80.95%; Kliper, Vaizman, Weinshall, & Portuguese, 2010).

This study has several limitations. Firstly, patients were medicated, making it impossible to assess the differential effect of drug treatment on prosody; however, the correlations obtained seem to rule out any connection. Secondly, although several variables that can modulate the results (age, education level, etc.) have been evaluated, it is necessary to expand the number of

variables that can influence the results. Thirdly, the possible existence of differences in prosody dependent on the demand of the task should be researched, since the cognitive demand required while reading a neutral text is very different from that required by speech in response to stimuli of an emotional character. The differences in the demands imposed by the tasks can potentially explain the disparity in the results obtained in various investigations, since cognitive resources particularly determine speech production in patients with schizophrenia.

Future studies, with an extended sample, may shed more data on the utility of this procedure and its clinical implications, especially in longitudinal studies. The obtained results support the use of the procedure; however the refinement of acoustic analysis algorithms should be sought, in order to achieve higher levels of discriminatory power between groups.

## References

- Adolphs R.** (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology*, *11*, 231–239. [http://dx.doi.org/10.1016/S0959-4388\(00\)00202-6](http://dx.doi.org/10.1016/S0959-4388(00)00202-6)
- Allott K. A., Schäfer M. R., Thompson A., Nelson B., Bendall S., Bartholomeusz C. F., ... Amminger G. P.** (2014). Emotion recognition as a predictor of transition to a psychotic disorder in ultra-high risk participants. *Schizophrenia Research*, *153*, 25–31. <http://dx.doi.org/10.1016/j.schres.2014.01.037>
- Alpert M., Kotsaftis A., & Pouget E. R.** (1997). At issue: Speech fluency and schizophrenic negative signs. *Schizophrenia Bulletin*, *23*, 171–177.
- Alpert M., Rosenberg S. D., Pouget E. R., & Shaw R. J.** (2000). Prosody and lexical accuracy in flat affect schizophrenia. *Psychiatry Research*, *97*, 107–118. [http://dx.doi.org/10.1016/S0165-1781\(00\)231-6](http://dx.doi.org/10.1016/S0165-1781(00)231-6)
- Alpert M., Shaw R. J., Pouget E. R., & Lim K. O.** (2002). A comparison of clinical ratings with vocal acoustic measures of flat affect and alogia. *Journal of Psychiatry Research*, *36*, 347–353. [http://dx.doi.org/10.1016/S0022-3956\(02\)00016-X](http://dx.doi.org/10.1016/S0022-3956(02)00016-X)
- Bech P., Larsen J. K., & Andersen J.** (1988). The BPRS: Psychometric Developments. *Psychopharmacology Bulletin*, *24*(1), 117–121.
- Boersma P., & Weenink D.** (2013). *Praat: Doing phonetics by computer (5.1.42 version)*. [Computer software]. Amsterdam, The Netherlands: University of Amsterdam. Retrieved from <http://www.praat.org/>
- Brazo P., Beaucousin V., Lecardeur L., Razafimandimby A., Dollfus S.** (2014). Social cognition in schizophrenic patients: The effect of semantic content and emotional prosody in the comprehension of emotional discourse. *Frontiers in Psychiatry*, *5*, 120. <http://dx.doi.org/10.3389/fpsy.2014.00120>
- Cantero F. J.** (2003). Fonética y didáctica de la pronunciación [Phonetics and Didactics of pronunciation]. In A. Mendoza (Ed.), *Didáctica de la Lengua y la Literatura para Primaria* [Didactics of the language and literature for elementary]. (pp. 1–24). Madrid, Spain: Pearson.
- Clemmer E. J.** (1980). Psycholinguistic aspects of pauses and temporal patterns in schizophrenic speech. *The Journal of Psycholinguistic Research*, *9*, 161–185. <http://dx.doi.org/10.1007/BF01067469>
- Cohen A. S., & Elvevåg B.** (2014). Automated computerized analysis of speech in psychiatric disorders. *Current Opinion in Psychiatry*, *27*, 203–209. <http://dx.doi.org/10.1097/YCO.0000000000000056>
- Cohen A. S., & Hong S. L.** (2011). Understanding constricted affect in schizotypy through computerized prosodic analysis. *Journal of Personality Disorders*, *25*, 478–491. <http://dx.doi.org/10.1521/pedi.2011.25.4.478>
- Cohen A. S., Alpert M., Nienow T. M., Dinzeo T. J., & Docherty N. M.** (2008). Computerized measurement of negative symptoms in schizophrenia. *Journal of Psychiatry Research*, *42*, 827–836. <http://dx.doi.org/10.1016/j.jpsychires.2007.08.008>
- Cohen A. S., Iglesias B., & Minor K. S.** (2009). The neurocognitive underpinnings of diminished expressivity in schizotypy: What the voice reveals. *Schizophrenia Research*, *109*, 38–45. <http://dx.doi.org/10.1016/j.schres.2009.01.010>
- Cohen A. S., Kim Y., & Najolia G. M.** (2013). Psychiatric symptom versus neurocognitive correlates of diminished expressivity in schizophrenia and mood disorders. *Schizophrenia Research*, *146*, 249–253. <http://dx.doi.org/10.1016/j.schres.2013.02.002>
- Cohen A. S., McGovern J. E., Dinzeo T. J., & Covington M. A.** (2014). Speech deficits in serious mental illness: A cognitive resource issue?. *Schizophrenia Research*, *160*, 173–179. <http://dx.doi.org/10.1016/j.schres.2014.10.032>
- Cohen A. S., Mitchell K. R., & Elvevåg B.** (2014). What do we really know about blunted vocal affect and alogia? A meta-analysis of objective assessments. *Schizophrenia Research*, *159*, 533–538. <http://dx.doi.org/10.1016/j.schres.2014.09.013>
- Couture S. M., Penn D. L., & Roberts D. L.** (2006). The functional significance of social cognition in schizophrenia: A review. *Schizophrenia Bulletin*, *32*, S44–S63. <http://dx.doi.org/10.1093/schbul/sbl029>
- Dickey C. C., Vu M. A. T., Voglmaier M. M., Niznikiewicz M. A., McCarley R. W., & Panych L. P.** (2012). Prosodic abnormalities in schizotypal personality disorder. *Schizophrenia Research*, *142*, 20–30. <http://dx.doi.org/10.1016/j.schres.2012.09.006>
- Emmerson L. C., Ben-Zeev D., Granholm E., Tiffany M., Golshan S., & Jeste D. V.** (2009). Prevalence and longitudinal stability of negative symptoms in healthy participants. *International Journal of Geriatric Psychiatry*, *24*, 1438–1444. <http://dx.doi.org/10.1002/gps.2284>
- First M. B., Spitzer R. L., Gibbon M., & Williams J. B. W.** (1996). *Structured Clinical Interview for DSM-IV Axis I Disorders—Clinician Version (SCID-CV)*. New York, NY: New York State Psychiatric Institute, Biometrics Research Department.
- Green M. F., & Horan W. P.** (2010). Social cognition in Schizophrenia. *Current Directions in Psychological Science*, *19*, 243–248. <http://dx.doi.org/10.1177/0963721410377600>
- Green M. F., Bearden C. E., Cannon T. D., Fiske A. P., Helleman G. S., Horan W. P., ... Nuechterlein K. H.** (2012). Social cognition in schizophrenia, Part 1: Performance

- across phases of illness. *Schizophrenia Bulletin*, 38, 854–864. <http://dx.doi.org/10.1093/schbul/sbq171>
- Hoekert M., Kahn R. S., Pijnenborg M., & Aleman A.** (2007). Impaired recognition and expression of emotional prosody in schizophrenia: Review and meta-analysis. *Schizophrenia Research*, 96, 135–145. <http://dx.doi.org/10.1016/j.schres.2007.07.023>
- Hooker C., & Park S.** (2002). Emotion processing and its relationship to social functioning in schizophrenia patients. *Psychiatry Research*, 112, 41–50. [http://dx.doi.org/10.1016/S0165-1781\(02\)00177-4](http://dx.doi.org/10.1016/S0165-1781(02)00177-4)
- Horan W. P., Green M. F., DeGroot M., Fiske A., Helleman G. S., Kee K., ... Nuechterlein K. H.** (2012). Social cognition in schizophrenia, Part 2: 12-month stability and prediction of functional outcome in first-episode patients. *Schizophrenia Bulletin*, 38, 865–872. <http://dx.doi.org/10.1093/schbul/sbr001>
- Johns L. C., Roseell S., Frith C., Ahmad F., Hemsley D., Kuipers E., & McGuire P. K.** (2001). Verbal self-monitoring and auditory verbal hallucinations in patients with schizophrenia. *Psychological Medicine*, 31, 705–715. <http://dx.doi.org/10.1017/S0033291701003774>
- Johnson-Selfridge M., & Zalewski C.** (2001). Moderator variables of executive functioning in schizophrenia: Meta-analytic findings. *Schizophrenia Bulletin*, 27, 305–316. <http://dx.doi.org/10.1093/oxfordjournals.schbul.a006876>
- Kee K. S., Green M. F., Mintz J., & Brekke J. S.** (2003). Is emotion processing a predictor of functional outcome in schizophrenia? *Schizophrenia Bulletin*, 29, 487–497. <http://dx.doi.org/10.1093/oxfordjournals.schbul.a007021>
- Kliker R., Vaizman Y., Weinsall D., & Portuguese S.** (2010). *Evidence for depression and schizophrenia in speech prosody*. Athens, Greece: Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics.
- Kohler C. G., Walker J. B., Martin E. A., Healey K. M., & Moberg P. J.** (2010). Facial emotion perception in schizophrenia: A meta-analytic review. *Schizophrenia Bulletin*, 36, 1009–1019. <http://dx.doi.org/10.1093/schbul/sbn192>
- Kring A. M., & Ellis O.** (2013). Emotion deficits in people with schizophrenia. *Annual Review of Clinical Psychology*, 9, 409–433. <http://dx.doi.org/10.1146/annurev-clinpsy-050212-185538>
- Kring A. M., & Moran E. K.** (2008). Emotional response deficits in schizophrenia: Insights from affective science. *Schizophrenia Bulletin*, 34, 819–834. <http://dx.doi.org/10.1093/schbul/sbn071>
- Kucharska-Pietura K., David A. S., Masiak M., & Phillips M. L.** (2005). Perception of facial and vocal affect by people with schizophrenia in early and late stages of illness. *The British Journal of Psychiatry*, 187, 523–528. <http://dx.doi.org/10.1192/bjp.187.6.523>
- Kugler B. T., & Caudrey D. J.** (1983). Phoneme discrimination in schizophrenia. *The British Journal of Psychiatry*, 142, 53–59. <http://dx.doi.org/10.1192/bjp.142.1.53>
- Lang P. J., Bradley M. M., & Cuthbert B. N.** (2005). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-6*. Gainesville, FL: University of Florida.
- Leentjens A. F. G., Wielaert S. M., Harskamp F. V., & Wilmsink F. W.** (1998). Disturbances of affective prosody in patients with Schizophrenia; a cross sectional study. *The Journal of Neurology, Neurosurgery, & Psychiatry*, 64, 375–378. <http://dx.doi.org/10.1136/jnnp.64.3.375>
- Lukoff D., Nuechterlein K. H., & Ventura J.** (1986). Manual for the Expanded Brief Psychiatric Rating Scale (BPRS). *Schizophrenia Bulletin*, 12, 594–602.
- Martínez-Sánchez F., Meilán J. J., Pérez E., Carro J., & Arana J. M.** (2012). Patrones prosódicos en sujetos con la Enfermedad de Alzheimer [Prosodic profiles of patients suffering from Alzheimer's Disease]. *Psicothema*, 24(1), 16–21.
- Matsumoto K., Samson G. T., O'Daly O. D., Tracy D. K., Patel A. D., & Shergill S. S.** (2006). Impaired prosodic discrimination in patients with schizophrenia. *British Journal of Psychiatry*, 89, 180–181.
- Melinder M. R. D., & Barch D. M.** (2003). The influence of a working memory load manipulation on language production in schizophrenia. *Schizophrenia Bulletin*, 29, 473–485. <http://dx.doi.org/10.1093/oxfordjournals.schbul.a007020>
- Mertens P.** (2004). Le prosogramme: Une transcription semi-automatique de la prosodie [The prosogram: Semi-automatic transcription of prosody]. *Cahiers de l'Institut de Linguistique de Louvain*, 30, 7–25. <http://dx.doi.org/10.2143/CILL.30.1.519212>
- Michelas A., Faget C., Portes C., Lienhart A. S., Boyer L., Lançon C., & Champagne-Lavau M.** (2014). Do patients with schizophrenia use prosody to encode contrastive discourse status? *Frontiers in Psychology*, 5, 755. <http://dx.doi.org/10.3389/fpsyg.2014.00755>
- Nicholson I. R., Chapman J. E., & Neufeld R. W. J.** (1995). Variability in BPRS definitions of positive and negative symptoms. *Schizophrenia Research*, 17, 177–185. [http://dx.doi.org/10.1016/0920-9964\(94\)00088-P](http://dx.doi.org/10.1016/0920-9964(94)00088-P)
- Pascual S., Solé B., Castellón J. J., Abadía M. J., & Tejedor M. J.** (2005). Prosodia afectiva y reconocimiento facial y verbal de la emoción en la esquizofrenia [Expressive prosody and facial and verbal emotional recognition in schizophrenia]. *Revista de Psiquiatría de la Facultad de Medicina de Barcelona*, 32, 179–183.
- Patel A. D.** (2008). *Music, language and the brain*. Oxford, UK: Oxford University.
- Penn D. L., Sanna L. J., & Roberts D. L.** (2008). Social perception in schizophrenia: An overview. *Schizophrenia Bulletin*, 34, 408–411.
- Peralta V., & Cuesta M. J.** (1994). Validación de la escala de los síndromes positivo y negativo (PANSS) en una muestra de esquizofrénicos españoles [Standardization of the Positive and Negative Syndrome Scale (PANSS) in a sample of Spanish schizophrenic patients]. *Actas Luso-Españolas de Neurología y Psiquiatría*, 22, 171–177.
- Putnam K. M., & Kring A. M.** (2007). Accuracy and intensity of posed emotional expressions in unmedicated schizophrenia patients: Vocal and facial channels. *Psychiatry Research* 151, 67–76. <http://dx.doi.org/10.1016/j.psychres.2006.09.010>
- Shaw R. J., Dong M., Lim K. O., Faustman W. O., Pouget E. R., & Alpert M.** (1999). The relationship between affect expression and affect recognition in schizophrenia.



*Schizophrenia Research*, 37, 245–250. [http://dx.doi.org/10.1016/S0920-9964\(98\)00172-8](http://dx.doi.org/10.1016/S0920-9964(98)00172-8)

**Silver H., Goodman C., Knoll G., Isakov V., & Modai L.** (2005). Schizophrenia patients with a history of severe violence differ from nonviolent schizophrenia patients in perception of emotions but not cognitive function. *Journal of Clinical Psychiatry*, 66, 300–308. <http://dx.doi.org/10.4088/JCP.v66n0305>

**Silver H., Shlomo N., Turner T., & Gur R. C.** (2002). Perception of happy and sad facial expressions in chronic schizophrenia: Evidence for two evaluative systems.

*Schizophrenia Research*, 55, 171–177. [http://dx.doi.org/10.1016/S0920-9964\(01\)00208-0](http://dx.doi.org/10.1016/S0920-9964(01)00208-0)

**Trémeau F. A.** (2006). Review of emotion deficits in schizophrenia. *Dialogues in Clinical Neuroscience*, 8, 59–70.

**Trémeau F., Malaspina D., Duval F., Correa H., Hager-Budny M., Coin-Bariou L., ... Gorman J. M.** (2005). Facial expressiveness in patients with schizophrenia compared to depressed patients and nonpatient comparison subjects. *The American Journal of Psychiatry*, 162(1), 92–101. <http://dx.doi.org/10.1176/appi.ajp.162.1.92>

## Annex I. Definition of the used parameters

Measure	Description
Duration	Total duration of the reading task, including pauses (s.)
Intensity	Voice intensity (dB)
% Pauses	Rate / Percentage of inter-syllabic pauses with a duration $\geq 300$ ms relative to the total time of phonation (%)
$M F_0$	Mean value of the Fundamental frequency ( $F_0$ ) (Hz)
$F_0 SD$	Standard Deviation of the $F_0$ (Hz)
$F_0$ Range	Syllabic nuclei amplitude (2%-98%) (ST)
Syllabic dynamics	Percentage of syllables with important changes in tone [abs (changes) $\geq 4ST$ ] (%)
Prosodic peaks	Percentage of syllables with upward tone changes ( $\geq 4ST$ ) (%)
Prosodic valleys	Percentage of syllables with downward tone changes ( $\leq -4ST$ ) (%)
Intra-syllabic Trajectory	Tone changes (abs. values) of the syllabic nuclei / duration (ST/s)
Inter-syllabic Trajectory	Tone changes (abs. values) of the syllabic nuclei, with the exception of pauses / duration (ST/s)
Phonation Trajectory	Sum of the inter- and intra-syllabic / phonation time (ST/s)