Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 7298–7304
Marseille, 20-25 June 2022

A new European Portuguese corpus for the study of Psychosis through speech analysis

Maria Forjó[1], Daniel Neto[2], Alberto Abad[1], H. Sofia Pinto[1], Joaquim Gago[3]
[1]INESC-ID/Instituto Superior Técnico, University of Lisbon, Portugal
[2]Serviço de Saúde da Região Autónoma da Madeira, Portugal
[3]Nova Medical School/Centro Hospitalar de Lisboa Ocidental, Lisboa, Portugal
maria.beato.forjo@tecnico.ulisboa.pt, danielcarvalhoneto@gmail.com, alberto.abad@inesc-id.pt,
sofia@inesc-id.pt, joaquimgago@me.com

Abstract

Psychosis is a clinical syndrome characterized by the presence of symptoms such as hallucinations, thought disorder and disorganized speech. Several studies have used machine learning, combined with speech and natural language processing methods to aid in the diagnosis process of this disease. This paper describes the creation of the first European Portuguese corpus for the identification of the presence of speech characteristics of psychosis, which contains samples of 92 participants, 56 controls and 36 individuals diagnosed with psychosis and medicated. The corpus was used in a set of experiments that allowed identifying the most promising feature set to perform the classification: the combination of acoustic and speech metric features. Several classifiers were implemented to study which ones entailed the best performance depending on the task and feature set. The most promising results obtained for the entire corpus were achieved when identifying individuals with a Multi-Layer Perceptron classifier and reached an 87.5% accuracy. Focusing on the gender dependent results, the overall best results were 90.9% and 82.9% accuracy, for female and male subjects respectively. Lastly, the experiments performed lead us to conjecture that spontaneous speech presents more identifiable characteristics than read speech to differentiate healthy and patients diagnosed with psychosis.

Keywords: Psychosis, Schizophrenia, Speech Analysis, Machine Learning

1. Introduction

A study conducted in 2017 found that around 729 million people in the world live with a mental health disorder[1]. These disorders are complex and can take many forms, such as depressive disorder, anxiety disorders, eating disorders and schizophrenia, among others.

The use of computational resources to assist during the diagnosis, study and control of these diseases has been developed throughout the years for many of them, like depressive disorder (Cohn et al., 2009). Studies have demonstrated that the use of a computational approach could be a valuable asset to psychiatrists, in particular when focused on characteristics such as facial expressions, speech characteristics and the semantics of speech (Chaika, 1982; Alpert et al., 1997; Yang et al., 2013; Gosztolya et al., 2020).

Psychosis is a clinical syndrome characterized by the presence of symptoms such as hallucinations, delusions, thought disorders and disorganized speech (Gaebel and Zielasek, 2015). This disorder is characterized by the individual's distortion and misconception of the surrounding environment, being also commonly associated with feelings of emotional distress and fear. This clinical syndrome can be a consequence of mental illnesses like schizophrenia, physical diseases, substance abuse or stressful life events.

Psychotic episodes are time intervals during which symptoms of psychosis affect the day-to-day life of an individual. The characteristics of these episodes differ depending on the person and episode, but the need for treatment is vital, otherwise, the psychosis can carry on indefinitely.

The disorganized speech symptom of psychosis is often characterized by patients expressing themselves through incoherent, and often irrelevant discourses, that contain a range of traits that are indispensable during the diagnosis of the disease.

It is crucial that individuals with psychosis are diagnosed as early as possible, since the treatment offers the best chance to recover, and the prognosis is affected by the delay in starting the medication. It is also important to notice that since the first episode of psychosis is most likely to occur during the early stages of an individual's life when one is developing one's personality, forming new relationships and shaping what their future holds, the treatment allows them to proceed with their development and enables a healthy future. This not only benefits the individual but also decreases the negative impact that psychosis has on the family and reduces associated problems such as depression and substance abuse.

As explained previously, some of the most common symptoms showed by individuals with psychosis are related to speech. These patients develop specific speech traits that can be recognized by psychiatrists in the diagnosis process. As shown by Rapcan et al. (2010),

[1]https://ourworldindata.org/mental-health, accessed on September 20th 2021

a quantitative acoustic and temporal analysis of the speech is a biomarker for the speech characteristics of psychotic patients, making speech a biomarker for psychosis. Some speech traits continue to be present even when patients are medicated and stabilized. Therefore a computational approach could be a valuable asset to the psychiatrist as one of the inputs to be used during both diagnosis and follow up of the patient after the diagnosis process, by signalling the presence of speech characteristics identified in patients with psychosis. This could provide additional information to support doctors in the diagnosis and follow up and ultimately help patients lead healthy and prosperous lives.
The main long term goal of this project is to develop a tool that supports psychiatrists in the diagnosis and follow up of patients diagnosed with psychosis in Portugal. At this initial stage our goal is to develop a tool that identifies recordings from diagnosed patients. To this end, the first European Portuguese corpus of speech recordings from individuals diagnosed with psychosis and healthy controls was created. The corpus collects recordings from a total of 92 subjects, 56 from individuals without a known diagnosis of psychosis, to serve as controls; and 36 from individuals diagnosed and medicated for psychosis that are being followed in mental health units. Some preliminary classification experiments were conducted with the collected data using different classification models trained on top of various feature sets, including speech metrics and acoustic-prosodic features. The results obtained allowed identifying the combination of speech metrics features as the most promising characteristics to focus on when identifying the presence of psychotic speech characteristics. Overall, an 87.5% accuracy was obtained when identifying individuals among the entire corpus, while a 90.9% and 82.9% accuracy was obtained for the female and male corpus sub-sets, respectively.

2. Related Work

In order to recognize if someone has schizophrenia through speech, previous studies demonstrate that the task conducted by the participant is highly dependent on the speech's characteristics which the study focuses on. The most common method to evaluate how connected, consistent and coherent a speech is, is to make the participants speak fluently and in a natural way.
There are two types of studies related to this: linguistic characteristics focused studies and paralinguistic characteristics focused studies.
A linguistic approach was explored by Elvevåg et al. (2010). In the study conducted, the participants were asked to perform a verbal fluency task and then an open-ended interview which allowed to verify if patients with schizophrenia produce fewer words and less complex sentences than healthy individuals. To this end, the speeches of the participants were analysed for different characteristics. The results allowed to observe that best classification results are obtained based on statistical-based semantic features, achieving accuracy results over 75% depending on the specific questions asked.
A paralinguistic approach was developed by Gosztolya et al. (2020), where the use of temporal parameters from the speech along with machine learning was applied to distinguish patients suffering from schizophrenia and bipolar disorder and their sub-groups. To study the speech of the participants, specific temporal parameters focused on the amount of hesitation in the speech were calculated. The achieved results detect differences in prosody, lack of tone, and inflexion in schizophrenic patients, and patients with Formal Thought Disorder made fewer filled pauses than controls. The accuracy of recognizing schizophrenia patients from bipolar disorder patients was higher than 80%, but the distinction between types of schizophrenia was around 60%.
Knowing that psychosis affects a patient's ability to speak fluently and consistently, it is possible to recognize that some of the best results come from the evaluation of samples from tasks that require the participants to speak as naturally as possible. However, another common and successful approach is through a Verbal Fluency task, which is used to evaluate the semantic coherence between the words produced by the participants. This method is appropriate because it does not require an entire speech to evaluate coherence, being, therefore, less computationally complex.

3. Corpus Design and Collection

3.1. Protocol definition

The protocol for data collection was defined based on the intersection of the tasks presented in the literature that showed the greatest potential, and endorsed the best results, with the first-hand experience of professionals who deal with psychosis patients on a daily basis. The protocol consists of a set of clear directions that takes about 20 minutes for each participant and is conducted by an assistant. This protocol was approved by the Ethics Committees from Instituto Superior Técnico, Casa de Saúde São João de Deus and Centro Hospitalar Lisboa Ocidental.

3.2. Speech production tasks

The data collected for each participant consists of 7 different tasks, each chosen for not requiring any personal information or sensitive topics from the participant. Each task allows the study of different linguistic and paralinguistic characteristics, by inciting the participant to speak naturally, to enumerate categories or to read a text.
The speech production tasks included in the protocol are:

1. Verbal Fluency Task - Words starting with "P";
2. Verbal Fluency Task - Animals;
3. Reading Task - "The North Wind and the Sun";

4. Storytelling Task - "The three little pigs";
5. Affective Image Description Task - Positive Image;
6. Affective Image Description Task - Neutral Image;
7. Affective Image Description Task - Negative Image.

3.2.1. Verbal Fluency Tasks

A verbal fluency test is a short test to assess a participant's verbal functioning (Shao et al., 2014). It usually involves two tasks: a category fluency task, also mentioned as semantic fluency task and a letter fluency task, also mentioned as phonemic fluency task. During these tasks, the participants are given 1 minute to produce as many unique words as they can remember within a category (in category fluency tasks) or starting with a given letter (in letter fluency tasks). The participant's score in each task is the number of unique correct words produced.

In the protocol, the incorporated verbal fluency tasks are the first and second tasks. With these tasks both the semantic and phonemic fluency of the participant can be studied:

- To study the phonemic fluency - Words starting with "P";
- To study the semantic category - Animals.

Expected results

From the literature that was studied and the knowledge from the psychiatrists in our team, one expects that, overall, the control participants would produce more unique items and have more filled pauses than the psychotic participants (Holshausen et al., 2014). Regarding differences between both tasks, it was anticipated that both controls and psychotic participants would show more ease with the category Animals, than the letter "P". Lastly, a difference in the number of words produced between people with different education levels was also expected for both controls and psychotic participants.

3.2.2. Reading Task

The reading task was included in the protocol aiming at a deeper focus of the speech characteristics not related to coherence and meaning of the content spoken by the participant. Having all the participants read the same text allows the study of the differences in characteristics such as number of pauses and speaking rate. Regarding the choice of text, there was a need to choose one that was not too complicated and therefore would not be easily affected by the education level. As a result, the text "The North Wind and the Sun" was chosen for being phonetically balanced, its simplicity, not being too long and not having much punctuation.

Expected results

It was expected that the readings of the fable "The North Wind and the Sun" would show that controls make more frequent pauses, and show a greater vocal range compared to a more monotone reading by the psychotic participants (Rapcan et al., 2010).

3.2.3. Storytelling Task

In contrast to the reading task, the inclusion of a storytelling task in the protocol allows for a deeper study of the coherence and content of a participants' speech, by studying the story told and how it was told. This task brings out a more natural and fluent speech and helps to investigate the ability of a psychotic participant to tell a story and follow a line of thought. The story chosen to be told by the participants was "The three little pigs", since it is a very simple and well-known story. In the event of a participant not knowing the story, the person conducting the protocol read it and then asked the participant to tell it back.

Expected results

It was anticipated that storytelling would allow the detection of differences in the coherence of the stories that were told. In this sense, psychotic participants were expected to tell more confusing stories than control subjects (Elvevåg et al., 2007). This task was also expected to allow the study of the prosody characteristics of each participant in a more natural and unscripted task.

3.2.4. Affective Image Description Tasks

The final set of tasks in the protocol requires the participants to look at three affective images and describe what they see, or make up a story.

By including three different images, a positive, a neutral and a negative one, it is possible to study how a participant reacts to each one of them, and the ability to recognize emotions.

Expected results

This task was expected to show that psychotic participants have more difficulties recognizing the positive and negative emotions and therefore in coherently describing those two images (Mota et al., 2017). Additionally, it was also anticipated that the study of the prosody of the voice of a participant would depict an overall more monotone voice when analysing psychotic participants' samples.

3.3. Corpus Characterization

The corpus contains samples of 92 recordings of the protocol which include the participation of 56 individuals without a known diagnosis of psychosis, designated control sub-set from now on, and 36 individuals diagnosed with psychosis, designated patients sub-set from now on. Table 1 summarizes the basic demographics information of the participants included in the corpus. The information collected was: gender, age and education level (number of years of education, considering *University degree* or higher equivalent to 15 years). In terms of gender distribution, the corpus contains 45 female participants and 47 male participants. Regarding age and education, the patients participants were in average older and less educated than the control ones.

Regarding the recording set-up and the acoustic conditions during the recording sessions, all participants were recorded with a Zoom H4N Pro recorder using two microphones: 1 lapel microphone and 1 microphone attached to the recording device at $\sim$1 meter distance. The recording sessions were conducted in different rooms with distinct acoustics properties and background noises, resembling realistic conditions of medical appointments.

Lastly, since the corpus was acquired in pandemic times, some of the recordings were performed with masks and others without.

3.3.1. Controls sub-set

The control participants sub-set of the study is composed of samples from individuals who have not had a psychotic episode. This sub-set contains samples of 56 recordings of participants doing the protocol, of which 31 are females and

	Controls			Patients		
	male	fem.	Σ	male	fem.	Σ
#participants	25	31	56	22	14	36
age (mean ± sd.)	33.8 ± 13.1	37.4 ± 15.0	35.8 ± 14.3	52.3 ± 15	54.3 ± 11.6	53.1 ± 14.5
education (mean ± sd.)	14.5 ± 1.4	14.1 ± 1.9	14.3 ± 1.7	7.7 ± 3.8	10.2 ± 3.8	8.7 ± 4.0

Table 1: Demographic characterization of the corpus.

25 are males. All participants that accepted our invitation were included.

3.3.2. Patients sub-set

The psychotic patients sub-set of the study is composed of samples from individuals who have had a psychotic episode and are being followed in Centro Hospitalar Lisboa Ocidental (CHLO) or Casa de Saúde São João de Deus (CSSJD). This sub-set contains samples of 36 recordings of participants doing the protocol, of which 14 are females and 22 are males. All patients that volunteered were included in the study.
In addition to the demographic information gathered, the following information about the subjects was also obtained: total BPRS value (Brief Psychiatric Rating Scale - used to measure psychiatric symptoms[2]) and the number of years since the patient has been diagnosed. These indicators were gathered to allow a posterior study on how they influence the results obtained.

3.4. Qualitative analysis of the samples

3.4.1. Verbal Fluency Tasks

Control Participants

The control participants engaged in this task without hesitations, and used the entire minute. Depending on the education level, it was possible to recognize some differences in the words they remembered, their complexity and the line of thought they were following. Some were going through similar words from a phonetics point of view, others decided to go through different environments and natural habitats and remember the words starting with "P" and animals present there, and others just said words that randomly came to their minds. The *Words starting with "P"* sub-task was shown to be more difficult, and the participants overall enumerated fewer exemplars in this sub-task than in the *Animals* sub-task.

Patient Participants

Most participants diagnosed with psychosis when asked to perform these tasks uttered a reduced number of examples compared to the controls. The control participants said a mean average of 23.79 words per minute, while patients said a mean average of 12.69 words per minute. These values were established counting the different number of items produced by each individual, not counting number, gender or diminutive variations of an animal. The *Words starting with "P"* sub-task proved itself to be the most difficult of the two, during which a lot of participants showed themselves impatient and uneasy. Regarding the *Animals* sub-task, it depended greatly on the individual, since some just said a handful of animals and decided to stop, in a similar way to what was observed for *Words starting with "P"* sub-task, while others showed interest in thinking and trying to remember more animals, not giving up. It was possible to recognize differences based on the education level of the participant, regarding the number of words they remembered, their complexity, and their ability to keep thinking even though they did not remember more words to enumerate.
Overall, diagnosed psychotic participants enumerated more words in the *Animals* sub-task, and participants with higher education levels tended to say more exemplars than others in both sub-tasks. Regarding the recognition of a line of thought in the answers given, most participants just tried to remember words that fit the category and did not follow an identifiable line of thought.

3.4.2. Reading Task

Control Participants

The control participants read the "The North Wind and the Sun" text without showing much difficulty, regardless of the education level. This proves that the text is adequate and is not evaluating the reading abilities of each individual, but can focus on the speech and how someone reads the text. Moreover, different individuals made different pauses, but as a whole, no individual read the entirety of the text without any pauses.

Patient Participants

Patients read the "The North Wind and the Sun" text showing some difficulties, mostly those with lower educational levels. The reading times for the text by these individuals was considerably lower than the reading times by the controls. Overall diagnosed psychotic participants read the entire text without pausing, compared to a more prosodically rich reading by the controls.

3.4.3. Storytelling Task

Control Participants

The way in which the participants tell the "The three little pigs" story varies for numerous reasons. The first observation was that some people do not remember the story, even though they were told the story numerous times in their childhood. Secondly, the details with which a participant tells the story sometimes is not directly correlated to remembering the story but to their personality. Lastly, some participants tell the story with voiceovers, this is thought to be a direct consequence of life experiences: people who have told the story to children tell it more child-oriented and with a more engaging voice prosody.

Patients Participants

Diagnosed psychotic participants showed themselves nervous when asked to tell the story "The three little pigs". After asking to hear the story first, they would then tell a summary of it, and most did not focus on anything specifically when telling it and were very nervous to remember it. In some

[2]https://www.psychiatrictimes.com/view/bprs-brief-psychiatric-rating-scale, accessed on October 2nd

cases, the storytelling was not concluded. It was also clear that the participants who remembered the story told it in a way that reflected their personality and life experiences: less talkative people told shorter versions of the story and people who had child-oriented professions told the story with a more engaging voice prosody.

3.4.4. Affective Image Description Task

Control Participants

From the descriptions given by the participants of the images in the protocol, it was possible to recognize the differences in the personalities and life experiences of the participants. Firstly, the study showed that individuals who were more talkative explored the images at a greater extent, while others, in comparison, would only mention a unique word. Secondly, in the first image, although all participants easily recognized the positive image, younger people tended to mention getting together with their friends and having fun. On the contrary, older people associated the image more to relax and being at ease. Regarding the second image, the neutral one, not much difference was noticed between the answers. The only variation may be associated with whether the participant is a student or employed: students recognized comfort the most, while employed individuals recognized rest. Lastly, in the negative image, all participants acknowledged the negativity of the image, but the most prominent difference was that individuals with more life experience, also mentioned that it could as well be someone thinking and reflecting about something that is troubling or has been going through their mind.

Patients Participants

Similarly to the control participants, the descriptions given by the subjects diagnosed with psychosis allowed the observation of personality traits and life experiences that were influencing how they interpreted each image. The first observation regarding the first image was that most of these participants, although able to recognize freedom in the image, did not make reference to happiness itself, only feelings that alluded to it. Regarding the second image, most of the participants limited their answers to describing the elements they saw in the image, not alluding to any feeling. Moreover, some of the participants showed difficulty in describing the negative image and recognizing the sadness in it, only mentioning the surroundings of the woman, and saying she was reflecting. Overall, it was possible to notice that in comparison to the control participants, patients showed more difficulty in recognizing the positive and negative emotions, and provided shorter and faster descriptions of the images.

4. Automatic Recognition of Psychotic Characteristics

Preliminary classification experiments have been conducted to discriminate the presence of speech characteristics common in individuals diagnosed with psychosis using the newly collected corpus. To the best of our knowledge, this type of approach has not been carried out for European Portuguese, and the ones conducted in other languages did not consider speech production tasks corresponding to different types of speech. Moreover, not only the acoustic and prosodic characteristics are examined, but also speech metrics based on the content. To this end, both the audio sample itself and the corresponding audio transcription are needed.

Acoustic characteristics were analysed using the GeMAPS feature set (Eyben et al., 2016). The GeMAPS feature set used is a standard acoustic-prosodic parameter set with 62 features, commonly used in various areas of automatic voice analysis and paralinguistics. On the other hand, automatically generated transcriptions –using INESC-ID's transcriber TRIBUS (Carvalho and Abad, 2021) – were used to compute the *speaking rate* and *articulation rate* of each audio sample. The speaking rate is measured by the total number of words divided by the number of minutes the speech took, and the articulation rate is measured by the total number of words divided by the number of minutes the speech took minus the time spent pausing. In order to obtain these metrics, timestamps of the beginning and end of each word are needed. We considered several transcribers for this task and evaluated their recognition performance in the "Vento Norte" corpus (a pilot corpus created with control participants reading the "The North Wind and the Sun" text). Overall, INESC-ID's TRIBUS transcriber was the one with the best results. While not considered in this work, transcriptions will also allow studying the coherence and content of a participant's speech in future studies.

In terms of classifiers, we have compared the following: a Support Vector Machine, with variations in the hyperparameters (SVM); a Multi-Layer Perceptron (MLP); a Random Forest (RF); a k-Nearest Neighbors with variations in k (kNN); and a Naïve Bayes classifier (NB). Regarding the selection of the most adequate hyperparameters for the SVM, we ran a grid search and tested the following parameters:

- kernel: linear, polynomial, sigmoid and radial basis function.
- C: 1e-6,1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1, 10, 100, 1000 and 10000.
- γ (for the rbf kernel): 1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1, auto and scale.
- degree (for the polynomial kernel): 2,3,4 and 5.

These classifiers were trained and evaluated following a leave-one-speaker-out cross-validation strategy.

The classification performance was assessed in terms of accuracy and F-measure.

The experiments reported in the remaining of this section focused on the analysis of the different speech tasks –including the importance of read versus spontaneous speech–, the comparison of acoustic and speech metrics features and the influence of gender in the results.

4.1. Analysis of speech production tasks

The first set of experiments focuses on the analysis of the importance of the different speech production tasks. To this end, we have considered 4 tasks that correspond to the tasks performed by the participants of each category, that is, *Reading*, *Storytelling*, *Image Description* and *Verbal Fluency*. For each task, all classifiers were trained and evaluated considering three different feature sets: Acoustic Features only (A), Speech Metrics only (S) and Acoustic Features + Speech Metrics (A + S).

Table 2 shows the results obtained in each task for the best classifiers provided with each feature set. As it can be seen, the task that obtained the best results was the Storytelling task. In this case, these results were obtained with a Multi-Layer Perceptron classifier trained with both the acoustic and

speech metrics features combined. Further analysing the results of Table 2, it is possible to observe the following:

- The results obtained with the Verbal Fluency task are considerably lower than those obtained from the other tasks;
- The results obtained with the Storytelling task are in general slightly better than those obtained from the other tasks;
- Comparing the results obtained for the Storytelling task and the Reading task, it is possible to notice that the Storytelling task achieves better results. This can be an indicator that spontaneous speech contains more information than read speech to correctly identify patients suffering psychosis.

Classifier	Feature Set	Accuracy	F-measure
Reading task			
SVM	Acoustic	0.837	0.789
NB	Speech Metrics	0.791	0.725
SVM	Acoustic & Speech Metrics	**0.846**	**0.8**
Storytelling task			
SVM	Acoustic	0.841	0.781
MLP	Speech Metrics	0.840	0.767
MLP	Acoustic & Speech Metrics	**0.875**	**0.819**
Image Description task			
SVM	Acoustic	**0.849**	**0.806**
SVM	Speech Metrics	0.652	0.579
MLP	Acoustic & Speech Metrics	0.826	0.778
Verbal Fluency task			
RF	Acoustic	**0.774**	**0.696**
NB	Speech Metrics	0.597	0.274
RF	Acoustic & Speech Metrics	0.760	0.656

Table 2: Best results obtained for each task with each feature set.

4.2. Analysis of acoustic features and speech metrics

Table 2 allows identifying that the classifiers based on acoustic features achieve in general better results than those based only on the speech metrics. This was expected since the speech metrics used in this study are only two (speaking rate and articulation rate) compared to the 62 features that comprise the GeMaps set. Nevertheless, speech metrics seem to encode important cues given the very remarkable results obtained in some tasks with only 2 features (i.e. reading and storytelling tasks). Thus, we consider the use of other similar features based on timing and transcription information worth to be considered in future studies.

Regarding the combination of both acoustic features and speech metrics, it is possible to notice that this always results in improved performance in all tasks, specially in the Storytelling task. As such, we can state that the two types of features encode complementary useful information for the identification of psychotic speech.

As mentioned above, the speech metrics were calculated based on the transcriptions provided by the TRIBUS transcriber. To study the influence that the transcription errors had on the classification performance of the systems based on speech metrics, we computed new speech metrics for the reading task using the timing information obtained with forced alignment. The results obtained showed that, for the reading task, the transcriber did not negatively influence the classification performance. For the other tasks, given that the corpus does not contain manual transcriptions, informal validation of the automatic transcriptions seem to show that these are not as precise as for the reading task. As such, one can expect a more negative influence of these automatic transcriptions on the quality of the speech metrics for the other tasks.

4.3. Analysis of gender-dependent results

In order to analyse possible gender differences, we carried out experiments in which the speech classifiers were separately trained and evaluated for each gender.

Table 3 shows the best results obtained in the male and female experiments for each task. The best results are obtained for the female corpus sub-set for the Storytelling task with a MLP trained with the combination of acoustic and speech metrics features. In the case of male subjects, the best results are obtained in the Image description task, also with a MLP trained with all features. With the exception of the Storytelling task, in all the other tasks male results are considerably better than the female ones in terms of F-measure. Overall, in this corpus, it seems that the speech of male individuals allows a more accurate distinction between controls and patients than the female speech. However, this may be a consequence of particularities of the data set and the distribution of classes.

Classifier	Feature Set	Sub-set	Accuracy	F-measure
Reading task				
SVM	A	Female	0.750	0.522
NB	S	Male	**0.760**	**0.756**
Storytelling task				
MLP	A + S	Female	**0.909**	**0.833**
MLP	A + S	Male	0.837	0.799
Image Description task				
MLP	A + S	Female	0.818	0.714
MLP	A + S	Male	**0.829**	**0.826**
Verbal Fluency task				
NB	A	Female	0.705	0.381
RF	A	Male	**0.702**	**0.663**

Table 3: Evaluation metrics for the corpus sub-sets for each task the classifier was provided.

5. Conclusions and Future Work

Psychosis is a clinical syndrome that affects millions around the world. To help these patients get access to the treatment they need without becoming targets of discrimination and live a normal life, the correct and early diagnosis is essential.

This paper presents the protocol for the creation of the first European Portuguese speech corpus for the study of psychosis, together with some preliminary experiments. The joint work with psychiatrists stressed the need for the creation of a corpus that allows the study not only of coherence and semantic features of speech, as targeted by previous research work in the area, but also of specific speech characteristics, such as speech metrics and acoustic features. Moreover, it was also clear that the protocol should involve different tasks –not requiring private information– and that should go beyond guided dialogues between the patient and specialized

mental health professionals. The remarkable preliminary results obtained in the experiments proves the ability of machine learning models to identify the presence of speech psychotic characteristics in the collected data.

In spite of the promising results, our study faced some limitations that need to be noticed and considered for future improvements: (1) data unbalance: the corpus does not include an equally balanced number of participants for all age groups and education levels; (2) lack of untreated patients: only diagnosed and treated patients were included due to the fact that there is no access to unmedicated patients in the Portuguese health-care context, which raises the question about the applicability of the developed models for early detection and the relation between medicated and unmedicated psychotic speech characteristics; (3) varying recording conditions: different room locations were used for the recordings; and (4) mask effect: the ongoing COVID-19 pandemic prevented the collection of all the recordings without masks.

Overall, we consider this study to be a trail-blazer to mental health classification in European Portuguese, that opens the door to the inclusion of computerized methods to support mental health professionals in the diagnosis and diagnosis, prescription and follow up of psychosis in Portugal. It also opens the door to many different future courses of action. These include not only the use of more linguistic features, such as word embeddings, and adding the Verbal Fluency score (number of different exemplars created) to the feature set, but also studying coherence of discourse and existence of specific terms like "Dreams" and "Voices", as described in the literature.

6. Acknowledgements

We thank Dra. Ana Moreira for her insight and support during the creation of the protocol and the recording process of the patients corpus subset at CHLO. We also thank the staff from Unidade de Saúde Mental de Oeiras - Centro Hospitalar Lisboa Ocidental for their help and support during the patients corpus recording process. This work was supported by national funds through *Fundação para a Ciência e a Tecnologia* (FCT) under project UIDB/50021/2020.

7. Bibliographical References

Alpert, M., Kotsaftis, A., and Pouget, E. (1997). Speech Fluency and Schizophrenic Negative Signs. *Schizophrenia bulletin*, 23:171–177.

Carvalho, C. and Abad, A. (2021). TRIBUS: An end-to-end automatic speech recognition system for European Portuguese. In *Proc. IberSPEECH 2021*, pages 185–189.

Chaika, E. (1982). At Issue: Thought Disorder or Speech Disorder in Schizophrenia? *Schizophrenia bulletin*, 8:587–594.

Cohn, J. F., Kruez, T. S., Matthews, I., Yang, Y., Nguyen, M. H., Padilla, M. T., Zhou, F., and De la Torre, F. (2009). Detecting depression from facial actions and vocal prosody. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–7.

Elvevåg, B., Foltz, P. W., Weinberger, D. R., and Goldberg, T. E. (2007). Quantifying incoherence in speech: an automated methodology and novel application to schizophrenia. *Schizophrenia Research*, 93(1-3):304–316.

Elvevåg, B., Foltz, P. W., Rosenstein, M., and Delisi, L. E. (2010). An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. *Journal of neurolinguistics*, 23(3):270–284.

Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., Narayanan, S. S., and Truong, K. P. (2016). The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. *IEEE Transactions on Affective Computing*, 7(2):190–202.

Gaebel, W. and Zielasek, J. (2015). Focus on psychosis. *Dialogues in clinical neuroscience*, 17(1):9–18.

Gosztolya, G., Bagi, A., Szalóki, S., Szendi, I., and Hoffmann, I. (2020). Making a Distinction between Schizophrenia and Bipolar Disorder Based on Temporal Parameters in Spontaneous Speech. *INTERSPEECH 2020*.

Holshausen, K., Harvey, P. D., Elvevåg, B., Foltz, P. W., and Bowie, C. R. (2014). Latent semantic variables are associated with formal thought disorder and adaptive behavior in older inpatients with schizophrenia. *Cortex*, 55:88–96.

Mota, N. B., Copelli, M., and Ribeiro, S. (2017). Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. *nature portfolio journal Schizophrenia*, 3(1):18.

Rapcan, V., D'Arcy, S., Yeap, S., Afzal, N., Thakore, J., and Reilly, R. B. (2010). Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia. *Medical Engineering & Physics*, 32(9):1074–1079.

Shao, Z., Janse, E., Visser, K., and Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, 5:772.

Yang, Y., Fairbairn, C., and Cohn, J. F. (2013). Detecting depression severity from vocal prosody. *IEEE Transactions on Affective Computing*, 4(2):142–150.