

# Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia

Viliam Rapcan<sup>a,✉</sup>, Shona D'Arcy<sup>a</sup>, Sherlyn Yeap<sup>b</sup>, Natasha Afzal<sup>c</sup>, Jogin Thakore<sup>d</sup>, Richard B. Reilly<sup>a, d, e</sup>

<sup>a</sup> Trinity Centre for Bioengineering, Trinity College Dublin, Dublin 2, Ireland

<sup>b</sup> Our Lady's Hospital, Navan, Ireland

<sup>c</sup> Fraser Health Care, Vancouver, Canada

<sup>d</sup> The Cognitive Neurophysiology Laboratory, St. Vincent's Hospital, Fairview, Dublin, Ireland

<sup>e</sup> Trinity Institute of Neuroscience, Trinity College Dublin, Dublin 2, Ireland

## Abstract

Currently, there are no established objective biomarkers for the diagnosis or monitoring of schizophrenia. It has been previously reported that there are notable qualitative differences in the speech of schizophrenics. The objective of this study was to determine whether a quantitative acoustic and temporal analysis of speech may be a potential biomarker for schizophrenia.

In this study, 39 schizophrenic patients and 18 controls were digitally recorded reading aloud an emotionally neutral text passage from a children's story. Temporal, energy and vocal pitch features were automatically extracted from the recordings. A classifier based on linear discriminant analysis was employed to differentiate between controls and schizophrenic subjects.

Processing the recordings with the algorithm developed demonstrated that it is possible to differentiate schizophrenic patients and controls with a classification accuracy of 79.4% (specificity = 83.6%, sensitivity = 75.2%) based on speech pause related parameters extracted from recordings carried out in standard office (non-studio) environments.

Acoustic and temporal analysis of speech may represent a potential tool for the objective analysis in schizophrenia.

## 1. Introduction

Schizophrenia is a serious mental illness characterized by a fundamental disturbance in perception, thought and communication [1]. Characteristic symptoms of schizophrenia are categorized into positive and negative symptoms. Positive symptoms, such as delusions and hallucinations are those that appear to reflect an excess or distortion of normal functions [2]. Negative symptoms, such as loss of interest or asociality, are those that appear to reflect a diminution or loss of normal functions [3–5]. Disorganized speech, a positive symptom is characterized by tangential, loosely associated, or incoherent speech severe enough to substantially impair effective communication. Alogia, or poverty of speech, is a negative symptom characterized by lessening of speech fluency and productivity, thought to reflect slowing or blocked thoughts, and often manifested as short, 'empty' replies to questions. Another negative symptom, affective flattening, is the reduction in the range and intensity of emotional expression, including facial expression, voice tone, eye contact, and body language [6].

From a linguistic perspective, the duration of pauses and hesitations have been found to correlate strongly with the clinician's impressions of the patient's flat affect and alogia [7]. Alpert [8] has shown that patients with this flat affect spoke with less inflection and have used less of their available time to describe recent experiences, i.e. were less fluent. Docherty et al. [9] have demonstrated that sustained attention impairment and impaired sequencing abilities (i.e. adequate words and phrases ordering for the communication of intended meanings), which are often found in schizophrenia, are highly predictive of communication failures related to language structure. Furthermore, Wisniecki et al. [10] have shown, using a simple counting and picture description task, that average pause length was indicative of motor retardation in those with negative symptom schizophrenia compared to controls. A number of different groups have used analysis of speech patterns so as to quantify the differences in length of pauses between those with mental illness and controls. These studies demonstrated that the length of pauses of the schizophrenic patients was longer in comparison with control subjects [11–13].

It has been demonstrated using MRI imaging, that structural changes can be found in the brains of those with schizophrenia [14–17] and that cognitive impairment affects up to 73% of schizophrenic patients [18]. Rapoport et al. [19] reported a

✉ Corresponding author. Tel.: +353 18964214; fax: +353 16795554.

E-mail addresses: rapcanv@tcd.ie, viliam@rapcan.eu (V. Rapcan).

Table 1  
Patient profile, medication and psychiatric rating.

	Female		Male	
	Mean	SD	Mean	SD
Number of subjects	12	–	27	–
Meds (CPZ) (mg)	818	347	708.4	346
In/Out Patient	2 in/10 out	–	8 in/19 out	–
BPRS	34.9	7.7	37.8	12.7
SANS	17.6	17.5	21.7	17.9
LOI (years)	18.4	12.4	12.6	10.4
Age (years)	49.2	15.6	39.2	11.5

BPRS is the Brief Psychiatric Rating Scale, SANS is the Scale for Assessment of Negative Symptoms, CPZ is chlorpromazine equivalents in mg, and the LOI the Length of Illness in years.

7% reduction of thalamus volume per year, in adolescents with schizophrenia. A similar volume reduction of the putamen and thalamus were reported in patients with probable Alzheimer’s disease [20]. It was found that the decrease in volume of these deep grey matter structures correlated linearly with impaired global cognitive performance. The putamen, as a part of the striatum in the basal nuclei, connects with many other structures and pathways in the brain, forming a series of complex circuits between the cerebral cortex, basal nuclei and thalamus. One of the major circuits in this network is the motor circuit [21]. Significant reduction in grey matter in frontal areas in adolescent schizophrenic patients over a 3–5 year period was reported by Rapoport [22]. The frontal lobes contain major motor control and speech production areas, including the primary motor cortex (Brodmann’s area 4 (BA-4)), the premotor area (BA-6) and Broca’s area (BA-44, BA-45), which are responsible for sequencing and controlling the motor movements required for the production of speech [21].

Two symptoms of schizophrenia in particular are associated with speech; alogia, which may reflect slowing or blocked thoughts, and cognitive impairment. The study presented here investigated temporal speech parameters such as the number, length of pauses and utterances as indicators of these symptoms. Affective flattening as a symptom of schizophrenia and structural changes in the brain of schizophrenics may be indicative of changes in functioning of articulators, hence this study investigated vocal pitch and energy parameters.

The aim of this study was to determine whether the speech characteristics of those with schizophrenia were significantly different from healthy controls, specifically to establish which components of speech can serve as a basis for objective analysis in schizophrenia.

## 2. Methods

### 2.1. Subjects

Informed consent was obtained from 39 (12 females) patients who fulfilled the Diagnostic and Statistical Manual (DSM) of mental disorders IV criteria [23] using the Structured Clinical Interview for DSM [24] (mean ( ±SD) age=42.3 ± 13.5 years) from the St. Vincent’s Hospital Fairview catchment area in Dublin. The mean duration of illness (time from diagnosis) was 14.4 ( ± 11.2) years.

All patients were physically healthy and were not accepted onto the study if they had a co-morbid DSM IV [23] psychiatric diagnosis including alcohol or illicit drug abuse. The severity of illness was rated in all of the recruited patients by one of us (SY), using the Brief Psychiatric Rating Scale (BPRS) [25] and the Scale for Assessment of Negative Symptoms (SANS) [26]. Patients were being prescribed a variety of different antipsychotic medications that have been converted to chlorpromazine (mg) equivalents in Table 1.

Control subjects comprised of 18 subjects (10 females) with a mean ( ±SD) age of 40.5 ( ± 12.9) years. All subjects were physically healthy and had no personal or family history of psychiatric illness and were recruited from within the local community. None of the control subjects was taking any form of prescribed or over the counter medication.

All schizophrenic and control subjects were literate. The study had approval from the Ethics Committee of St. Vincent’s Hospital Fairview. After complete description of the study to the subjects, written informed consent was obtained.

### 2.2. Data acquisition and statistical analysis

The system implementation is comprised of three main parts—data acquisition, feature extraction and classification (see Fig. 1).

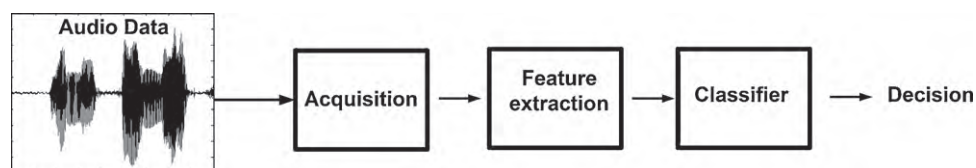
All audio files were recorded in a quiet room on a minidisc recorder (SONY MZ-B10) with direct digital 16-bit sampling and at a sampling rate of 22 kHz. Although every effort was taken to keep a constant distance during acquisition between the speaker and microphone, variation in acoustic amplitude exists between speakers due to individual variation in speaking volume.

An emotionally neutral text was selected for the recording and was found to take, on average, three minutes to read aloud. This passage was specifically chosen for its verbal and semantic simplicity and has been employed in similar studies in schizophrenia [27]. A brief extract of this passage from a children’s story is given below.

#### Text passage

“The thing which attracted her most, however, was the waving and roaring of the three old fir trees on these windy days. She would run away repeatedly from whatever she might be doing, to listen to them, for nothing seemed so strange and wonderful to her as the deep mysterious sound in the tops of the trees. .... ”

All speech signals were inspected visually before data analysis. Pauses longer than 250 ms were removed at the beginning and the end of each recording. This was carried out to ensure that analysis is performed from the moment the person starts to speak until the person stops speaking (otherwise the Number of Pauses would be increased in error by two, having a corresponding incorrect increase in duration of pauses). All audio files were high-pass filtered at 80 Hz with 7th order type II Chebyshev filter (roll-off of 55 dB/octave) prior to processing in order to remove the ambient noise that exists in the lower frequency bands.



### 2.3. Feature extraction

The basis of the feature extraction process was to identify from the audio signal what sections are speech and what are non-speech. Non-speech includes silence, external noises and breaths. In addition to this, a Speech/Non-speech threshold was automatically estimated for each audio file. The audio file was segmented into 50 ms windows and the energy calculated using Eq. (1). It was empirically estimated that approximately 15–20% of the entire recording was silence. Fifteen percent (value selected by iterative process) of the segments with the lowest energy were selected and maximum absolute amplitude value in each segment was calculated. Subsequently the average value from these maximal values was calculated. This value was used as the Speech/Non-speech threshold, where segments below this value were considered non-speech and segments above this value were considered speech.

The following features were extracted for each file:

1. Number of Pauses
2. Mean Pause Duration
3. Proportion of Silence
4. Mean Utterance Duration
5. Total Recording Time
6. Total Length of Pauses
7. Total Length of Utterances
8. Relative Variation in Energy
9. Relative Variation in Vocal Pitch

#### 2.3.1. Temporal features

Processing of each audio file consisted of classifying successive segments as voiced or silent depending on whether their amplitude was above or below the Speech/Non-speech threshold discussed above. For non-speech segments, durations of 250 ms or greater were considered pauses [28], as anything shorter than this is considered a natural part of speech where segments of low energy are a result of transitions from one word to another. As the last step, speech segments shorter than 100 ms surrounded by pauses on both sides were transformed to silence, because these segments usually represent clicks, knocks on the microphone, coughs and other artefacts of noise and are incorrectly selected as speech segments.

#### 2.3.2. Energy measures

The energy of a discrete time signal [29] is defined as

$$E_x \stackrel{\text{def}}{=} \sum_{n=-\infty}^{\infty} |x(n)|^2$$

where  $x(n)$  is the speech signal.

“Mean Energy per Second” (MES) and “Standard Deviation of Mean Energy per Second” (SDMES) were estimated for each speech signal. In order to combat some of the variance in the speech signal, the “Relative Variation in Energy” (RVE) was calculated (Eq. (2)).

$$RVE = \frac{SDMES}{MES} \quad (2)$$

#### 2.3.3. Vocal pitch estimation

To calculate the vocal pitch, the Variable Length Average Magnitude Difference Function (VLAMDF) [30] was employed. This function provides efficient pitch estimation which can be implemented in real-time. The main principle of VLAMDF is to estimate the pitch by comparing the similarity between the original signal and its shifted version. The VLAMDF algorithm uses variable-length

speech samples and is defined as

$$E^{\text{VLAMDF}}(\tau) = \frac{1}{\tau} \sum_{n=0}^{\tau-1} |s(n) - s(n + \tau)| \quad (3)$$

where  $\tau = \tau_{\min}, \tau_{\min+1}, \dots, \tau_{\max}$  and  $s(n)$  is the speech sample sequence.

The pitch is estimated according to the minimum of the waveform and is obtained by the equation

$$T_p^{\text{VLAMDF}} = \arg \underset{\tau = \tau_{\min}}{\tau_{\max}} \text{MIN} [E^{\text{VLAMDF}}(\tau)] \quad (4)$$

For vocal pitch estimation, the speech recording was divided into 40 ms frames. As the pitch period of the human voice will not exceed 20 ms [30], the frame length of 40 ms was considered sufficiently long for vocal pitch estimation. The vocal pitch was estimated in voiced frames, whereas unvoiced and speech onset or offset frames were excluded from pitch estimation.

“Mean Vocal Pitch” (MVP) and “Standard Deviation of Mean Vocal Pitch” (SDMVP) were calculated. In a similar manner to the energy calculation in order to normalize for the inherent gender variation in pitch the “Relative Variation in Vocal Pitch” (RVVP) was calculated (Eq. (5)).

$$RVVP = \frac{SDMVP}{MVP} \quad (5)$$

### 2.4. Classification

Analysis of the histogram of all features extracted from all speech signals for schizophrenic and control group show an approximate Gaussian distribution. Therefore, a classifier based on Linear Discriminant Analysis (LDA) [31] was chosen to differentiate between control and schizophrenia audio samples based on these features. With the relatively small number of subjects available, cross-fold validation [32] was used to maximize training and determine classification accuracies. The variance of the performance estimates was decreased by averaging results from multiple runs of cross-validation where a different random split of the training data into folds is employed for each run. In this study 39 repetitions of 18-fold cross-validation were used to estimate classifier performance. For each run of cross-fold validation the number of normal and abnormal cases was equal.

Classifier performance was measured using sensitivity, specificity, positive predictivity, negative predictivity and the overall accuracy. These measures were calculated as per the definition of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) presented in Table 2.

Receiver operator characteristic curves (ROC curves) [33], which provide valuable information on the classifier’s ability to discriminate between two classes over the complete spectrum of decision thresholds, were calculated. The ROC curve is a graph of sensitivity vs. (100% – specificity), as the a-priori probabilities of the two classes are swept between zero and one. It provides information on clinical usefulness since it presents a trade-off in costs between false positives and false negatives and can be employed to decide the threshold for different clinical requirements e.g. screening vs. pre-surgical diagnosis. The area enclosed by the ROC plot is a metric

Table 2  
Definitions of true positives/negatives and false positives/negatives.

		True classification	
		Pathology	Normal
Predicted classification	Pathology	True positive (TP)	False positive (FP)
	Normal	False negative (FN)	True negative (TN)

Table 3  
Average feature values (and standard deviation) for schizophrenic and control subjects.

	Schizophrenics		Controls	
	Mean	SD	Mean	SD
Number of Pauses ( n )	68.72	19.77	54.05	8.94
Mean Pause Duration (s)	0.63	0.08	0.57	0.08
Proportion of Silence	0.19	0.04	0.16	0.02
Mean Utterance Duration (s)	2.69	0.63	3.02	0.48
Total Recording Time (s)	222.15	42.80	194.58	30.64
Total Length of Pauses (s)	42.80	12.42	30.56	6.10
Total Length of Utterances (s)	179.35	35.78	164.02	25.46
Relative Variation in Energy	0.67	0.21	0.58	0.13
Relative Variation in Pitch	0.22	0.05	0.24	0.08

against which other classifier's configurations and/or features can be compared.

### 3. Results

#### 3.1. Feature values

Table 3 displays the average feature values for the patient and control groups. A clear difference in these values can be observed, with some features showing larger differences. The Number of Pauses, Proportion of Silence and the Total Length of Pauses were found to increase by 27%, 23% and 40% respectively for the patient group compared to the control group. The variations in their distributions are clearly smaller for the control group.

To determine if these differences are statistically significant, a two-tailed, unequal sample sizes, unequal variance, Student's  $t$ -test was applied to the data. The following features were found to be statistically significant—Number of Pauses ( $t(55) = 3.87$ ,  $p < 0.0003$ ), Mean Pause Duration ( $t(35) = 2.62$ ,  $p < 0.01$ ), Proportion of Silence ( $t(54) = 4.81$ ,  $p < 0.00001$ ), Mean Utterance Duration ( $t(43) = 2.15$ ,  $p < 0.04$ ), Total Recording Time ( $t(45) = 2.77$ ,  $p < 0.008$ ), Total Length of Pauses ( $t(55) = 4.99$ ,  $p < 0.000007$ ), Relative Variation in Energy ( $t(49) = 2.12$ ,  $p < 0.04$ ). The only features not achieving statistical significance were the Total Length of Utterances ( $t(45) = 1.85$ ,  $p < 0.07$ ) and the Relative Variation in Vocal Pitch ( $t(24) = 0.88$ ,  $p < 0.4$ ).

#### 3.2. Classification results

To assess the ability of the LDA classifier to distinguish between schizophrenic patients and controls, the LDA classifier was trained using three feature sets. At first, the LDA classifier was trained on a combination of all the features listed in Table 3, and the performance of this classifier may be seen in Table 4, Feature Set 1.

The training and testing was repeated for each individual feature set and certain features were found to result in better classification accuracy than others. Predictably, the two feature sets which were found not to have significantly different data distributions resulted

Table 4  
Linear Discriminant Analysis (LDA) classifier performance metrics.

	Feature Set 1	Feature Set 2	Feature Set 3
Sensitivity (%)	72.64	75.21	72.36
Specificity (%)	78.63	83.62	85.47
Positive predictivity (%)	77.86	82.02	83.39
Negative predictivity (%)	74.24	77.40	75.75
Overall accuracy (%)	75.64	79.42	78.92
Area under the ROC curve	0.79	0.82	0.80

Feature Set 1—all features combined, Feature Set 2—Feature Set 1 minus Total Length of Utterances and Relative Variation in Vocal Pitch, Feature Set 3—Number of Pauses, Proportion of Silence, Total Recording Time and Total Length of Pauses.

in the poorest classification performance. The classification result using Total Utterance Duration and Mean Utterance Duration was 60% and 62% respectively compared to 75% and 77% classification for Number of Pauses and Proportion of Silence. The combined feature set was then reduced removing Total Length of Utterances and Relative Variation in Vocal Pitch. The results of the LDA training/testing may be seen in the Table 4, Feature Set 2. Further reduction in the features set removed the next poorest performing features; Mean Pause Duration and Mean Utterance Duration and Relative Variation in Energy (see Table 4, Feature Set 3).

The classification performance when training on the two reduced feature sets outperform the original classifier for all classifier metrics, however between the two reduced feature set the results are not as clear. While the Feature Set 2 (only omitting Total Utterance Duration and Relative Variation in Vocal Pitch) performs better than the Feature Set 3 (additionally missing Mean Pause Duration and Mean Utterance Duration and Relative Variation in Energy) for nearly all classification results except Specificity and Positive Predictability. The higher specificity suggests more controls being classified correctly while the slightly higher positive predictability suggests a higher likelihood of classifying a patient correctly.

#### 3.3. Correlation analysis

Table 5 lists the correlation coefficients between extracted features and clinical variables—dosages of medication, symptom scales (BPRS, SANS), and Length of Illness.

To determine if these correlation coefficients are statistically significant, a one-tailed Fisher's  $F$ -test was applied to the data. Each of the clinical variables has a correlation of statistical significance ( $p < 0.05$ ) with at least one of the temporal features. Highest correlation coefficients, which were statistically significant at  $p < 0.01$ , were between Length of Illness and Proportion of Silence ( $F(1,35) = 8.33$ ), Mean Utterance Duration ( $F(1,35) = 9.74$ ) and Relative Variation in Energy ( $F(1,35) = 8.74$ ), further between SANS and Total Length of Utterances ( $F(1,35) = 10.08$ ) and also between BPRS and Mean Utterance Duration ( $F(1,35) = 9.90$ ) (see Table 5).

### 4. Discussion

This article presents a study on discriminating between schizophrenic and healthy subjects based on temporal and acoustic analysis of speech signals. A feature extraction algorithm has been developed to reliably extract a range of temporal and acoustic features, some of which have been found to classify speakers into two states; schizophrenics and controls.

From classification results of two reduced feature sets it is clear that it is the pause related features that were most significant in differentiating between schizophrenics and controls. Classification was carried out on each feature set individually and it was the Number of Pauses, Proportion of Silence and Total Length of Pauses that had the highest overall accuracy for classification. On average, patients with schizophrenia tend to insert more pauses in their read speech (+27%, see Table 3), which causes the Proportion of Silence and Total Length of Pauses to be higher for patients with schizophrenia than control subjects. Mean Pause Duration is also slightly higher (+10%, see Table 3) for patients with schizophrenia. This result confirms our hypothesis that the speech characteristics of patients with schizophrenia contain more and longer pauses.

These differences in temporal features of speech may be related to the cognitive impairment which affects people with schizophrenia [18]. Cognitive impairment is associated with psychomotor retardation and has deteriorating effects on the production of speech. A study carried out by D'Arcy et al. [34] investigated tem-

Table 5

Correlation coefficients between features, symptom scales, dosages of medication and Length of Illness.

	Correl. with Meds	Correl. with BPRS	Correl. with SANS	Correl. with LOI
Number of Pauses	0.02	-0.19	-0.03	0.11
Mean Pause Duration	0.31	0.10	-0.01	0.04
Proportion of Silence	0.33 <sup>a</sup>	-0.34 <sup>a</sup>	-0.35 <sup>a</sup>	0.43 <sup>b</sup>
Mean Utterance Duration	-0.22	0.46 <sup>b</sup>	0.37 <sup>a</sup>	-0.46 <sup>b</sup>
Total Recording Time	-0.13	0.25	0.38 <sup>a</sup>	-0.28
Total Length of Pauses	0.18	-0.14	-0.02	0.11
Total Length of Utterances	-0.22	0.35 <sup>a</sup>	0.46 <sup>b</sup>	-0.38 <sup>a</sup>
Relative Variation in Energy	-0.34 <sup>a</sup>	0.37 <sup>a</sup>	0.37 <sup>a</sup>	-0.44 <sup>b</sup>
Relative Variation in Vocal Pitch	-0.15	0.07	0.02	-0.18

<sup>a</sup> Statistically significant at  $p = 0.05$ .<sup>b</sup> Statistically significant at  $p = 0.01$ .

poral features of read and spontaneous speech along with several standard clinical measures of cognitive function in elderly speakers. Profiles of pauses inserted during a reading speech task, similar to that presented in this study, proved effective when discriminating between older adults with poor cognitive function and healthy older adults. Saykin et al. [35] compared two schizophrenic patient groups—first episode patients and previously treated patients, with healthy control group. Patients groups had nearly identical profiles showing generalized impairment, particularly in verbal memory and learning attention-vigilance, and speeded visual-motor processing and attention.

In terms of the changes in the brain areas of schizophrenic patients and the negative symptom, affective flattening, which is associated with the speech of schizophrenics, it was hypothesized that a lack of variation in energy and vocal pitch would be observed in speech of schizophrenics. Of the remaining features; utterance related features and energy/pitch variation, it is the Relative Variation in Energy that had the highest success rate in classification. Given the variation in individual speaker volume, the absolute values for Mean Energy and Standard Deviation of Energy are not valid for inter-subject comparison. Relative measures of these features were calculated and Relative Variation in Energy results in a classification accuracy of 63.3%. The distribution of the Relative Variation in Vocal Pitch was found not to be statistically significant ( $p = 0.4$ ) and was not a useful measure in differentiating between the two groups. No significant difference was found between speech parameters of in-patients and out-patients.

The ability to discriminate between healthy controls and schizophrenic subjects of our study was lower than the discriminating performance of 95.2% shown in the study of Stassen et al. [27]. This may be a result of Stassen's training of the discriminant analysis (DA) classifier. The classifier was trained on recordings from 84 subjects (42 schizophrenics, 42 controls). The same 84 subjects were re-recorded 14 days later and the speech parameters from these recordings tested on the trained classifier. Stassen et al. also demonstrated that speech parameters for healthy subjects [36] and schizophrenic patients [27] remain stable over the period of 14 days and are highly correlated. Using the same subjects in the training and test samples, and taking into account the stability of speech parameters over time, an increased accuracy of the DA classifier may be expected. Also, Stassen et al. recorded data in an acoustically shielded room. The recordings reported in this study were carried out in a quiet room, more akin to what one might expect in real world clinical settings. Nevertheless, a good discrimination accuracy of 79.4% was achieved. Therefore, the results from the study presented here demonstrate that the patients' voice characteristics can differentiate from those of healthy subjects.

New features, such as the fundamental frequency amplitude and fundamental frequency contour, which when combined with the actual measured temporal, energy and pitch features, may lead to an increase in the system's performance. A larger database of speech samples from schizophrenic patients displaying a range of

symptoms, as well as samples from healthy controls would contribute to a more robust classification system and may possibly show better correlation between acoustic features and dosages of medication, symptom scales and/or Length of Illness.

Using speech for assessment of schizophrenia may enable development of remote and fully automated system. Studies by Moran et al. [37] and Wormald et al. [38] presented automatic assessment systems for monitoring vocal fold pathologies. In this way, high quality assessments may be carried out remotely, resulting in the use of available medical resources in an efficient and effective manner, as well as providing psychiatrists with up to date results with minimal processing.

Longitudinal studies to assess the changes in acoustic and temporal features with time are currently being investigated. Monitoring of the effects of medication on cognitive functioning of schizophrenic patients using regular voice recordings may provide the psychiatrist with valuable information about the suitability of the prescribed medication.

#### Acknowledgments

This work was supported by Enterprise Ireland, under grant number POC/2006/003. The authors thank Dr. Nils Penard for his input to the statistical analysis of the data. The authors would like to thank the volunteers at St. Vincent's Hospital, Fairview in Dublin for assisting this study.

#### Conflict of interest

All authors report no competing interests.

#### References

- [1] Bleuler E. *Dementia praecox, or the group of schizophrenias*. Translated by J. Zinkin (1911). New York: International University Press; 1950.
- [2] Peralta V, Manuel JC. Factor structure and clinical validity of competing models of positive symptoms in schizophrenia. *Biol Psychiatry* 1998;44:107–14.
- [3] Roy MA, DeVriendt X. Positive and negative symptoms in schizophrenia: a current overview. *Can J Psychiatry* 1994;39:407–14.
- [4] Crow TJ. Brain changes and negative symptoms in schizophrenia. *Psychopathology* 1995;28:18–21.
- [5] Blanchard JJ, Mueser KT, Bellack AS. Anhedonia, positive and negative affect, and social functioning in schizophrenia. *Schizophr Bull* 1998;24:413–24.
- [6] United States, Public Health Service, Office of the Surgeon General, Center for Mental Health Services, National Institute of Mental Health. *Mental health: a report of the surgeon general*. Rockville, MD/Pittsburgh, PA: Dept. of Health and Human Services, U.S. Public Health Service; 1999.
- [7] Alpert M, Kotsaftis A, Pouget ER. At issue: speech fluency and schizophrenic negative signs. *Schizophr Bull* 1997;23:171–7.
- [8] Alpert M. Prosody and lexical accuracy in flat affect schizophrenia. *Psychiatry Res* 2000;97:107–18.
- [9] Docherty NM, Strauss ME, Dinzeo TJ, St-Hilaire A. The cognitive origins of specific types of schizophrenic speech disturbances. *Am J Psychiatry* 2006;163:2111–8.
- [10] Wisniecki A, Cannizzaro M, Cohen H, Snyder P. Speech impairments in neurodegenerative diseases/psychiatric illnesses. In: *Encyclopedia of language & linguistics*. Elsevier; 2006, 758–764.

- [11] Alpert M, Rosen A, Welkowitz J, Sobin C, Borod JC. Vocal acoustic correlates of flat affect in schizophrenia. Similarity to Parkinson's disease and right hemisphere disease and contrast with depression. *Br J Psychiatry Suppl* 1989;51-6.
- [12] Greden JF, Albala AA, Smokler IA, Gardner R, Carroll BJ. Speech pause time: a marker of psychomotor retardation among endogenous depressives. *Biol Psychiatry* 1981;16:851-9.
- [13] Kring AM, Alpert M, Neale JM, Harvey PD. A multimethod, multichannel assessment of affective flattening in schizophrenia. *Psychiatry Res* 1994;54:211-22.
- [14] Shenton M, Dickey C, Frumin M, McCarley R. A review of MRI findings in schizophrenia. *Schizophr Res* 2001;49:1-52.
- [15] Harrison PJ, Lewis DA. Neuropathology of schizophrenia. In: Steven R, Hirsch DRW, editors. *Schizophrenia*. 2nd ed. 2007. p. 310-25.
- [16] Weinberger DR, Marengo S. Schizophrenia as a neurodevelopmental disorder. In: Steven R, Hirsch DRW, editors. *Schizophrenia*. 2nd ed. 2007. p. 326-48.
- [17] Rund BR. Is there a degenerative process going on in the brain of people with schizophrenia? *Front Hum Neurosci* 2009;3:36.
- [18] Palmer BW, Heaton RK, Paulsen JS, Kuck J, BraffD, Harris MJ, et al. Is it possible to be schizophrenic yet neuropsychologically normal? *Neuropsychology* 1997;11:437-46.
- [19] Rapoport JL, Giedd J, Kumra S, Jacobsen L, Smith A, Lee P, et al. Childhood-onset schizophrenia. Progressive ventricular change during adolescence. *Arch Gen Psychiatry* 1997;54:897-903.
- [20] de Jong LW, van der Hiele K, Veer IM, Houwing JJ, Westendorp RGJ, Bollen ELEM, et al. Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. *Brain* 2008;131:3277-85.
- [21] Ferrand CT. *Speech science: an integrated approach to theory and clinical practice*. Boston: Pearson/Allyn and Bacon; 2007.
- [22] Rapoport JL. Progressive cortical change during adolescence in childhood-onset schizophrenia: a longitudinal magnetic resonance imaging study. *Arch Gen Psychiatry* 1999;56:649-54.
- [23] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 4th ed. Washington: American Psychiatric Association; 1994.
- [24] Spitzer RL, Williams JB, Gibbon M, First MB. The Structured clinical interview for DSM-III-R (SCID). I. History, rationale, and description. *Arch Gen Psychiatry* 1992;49:624-9.
- [25] Overall JE, Gorham DR. The brief psychiatric rating scale. *Psychol Rep* 1962;10:799-812.
- [26] Andreasen NC. *The scale for the assessment of negative symptoms (SANS)*. Iowa City: University of Iowa; 1983.
- [27] Stassen HH, Albers M, Püschel J, Scharfetter C, Tewesmeier M, Woggon B. Speaking behavior and voice sound characteristics associated with negative schizophrenia. *J Psychiatr Res* 1995;29:277-96.
- [28] Eisler FG. *Psycholinguistics: experiments in spontaneous speech*. London/New York: Academic Press; 1968.
- [29] Deller JR, Proakis JG, Hansen JHL. *Discrete-time processing of speech signals*. Upper Saddle River, NJ: Prentice-Hall; 1993.
- [30] Xiao-Dan M, Jengshyang P, Sheng-He S. Efficient algorithms for speech pitch estimation. In: *Proceedings of the 2001 international symposium on intelligent multimedia, video and speech processing*. 2001. p. 421-4.
- [31] Duda R, Hart P, Stork D. In: Edition S, editor. *Pattern classification*. Wiley-Interscience; 2000.
- [32] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th international joint conference on Artificial intelligence*, vol. 2. 1995.
- [33] Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39:561-77.
- [34] D'Arcy S, Rapcan V, Penard N, Morris M, Robertson I, Reilly R. Speech as a means of monitoring cognitive function of elderly speakers. In: *Proceedings of the interspeech conference*. 2008.
- [35] Saykin AJ, Shtasel DL, Gur RE, Kester DB, Mozley LH, Stafiniak P, et al. Neuropsychological deficits in neuroleptic naive patients with first-episode schizophrenia. *Arch Gen Psychiatry* 1994;51:124-31.
- [36] Stassen HH. Affective state and voice: the specific properties of overtone distributions. *Methods Inf Med* 1991;30:44-52.
- [37] Moran RJ, Reilly RB, de Chazal P, Lacy PD. Telephony-based voice pathology assessment using automated speech analysis. *IEEE Trans Biomed Eng* 2006;53:468-77.
- [38] Wormald RN, Moran RJ, Reilly RB, Lacy PD. Performance of an automated, remote system to detect vocal fold paralysis. *Ann Otol Rhinol Laryngol* 2008;117:834-8.