

PREDICTION OF NEGATIVE SYMPTOMS OF SCHIZOPHRENIA FROM EMOTION RELATED LOW-LEVEL SPEECH SIGNALS

Debsubhra Chakraborty[†], Zixu Yang*, Yasir Tahir[†], Tomasz Maszczyk[†], Justin Dauwels[†], Nadia Thalmann[†], Jianmin Zheng[†], Yogeswary Maniam*, Nur Amirah*, Bhing-Leet Tan* and Jimmy Lee*

[†]Nanyang Technological University, Singapore and *Institute of Mental Health, Singapore

Introduction

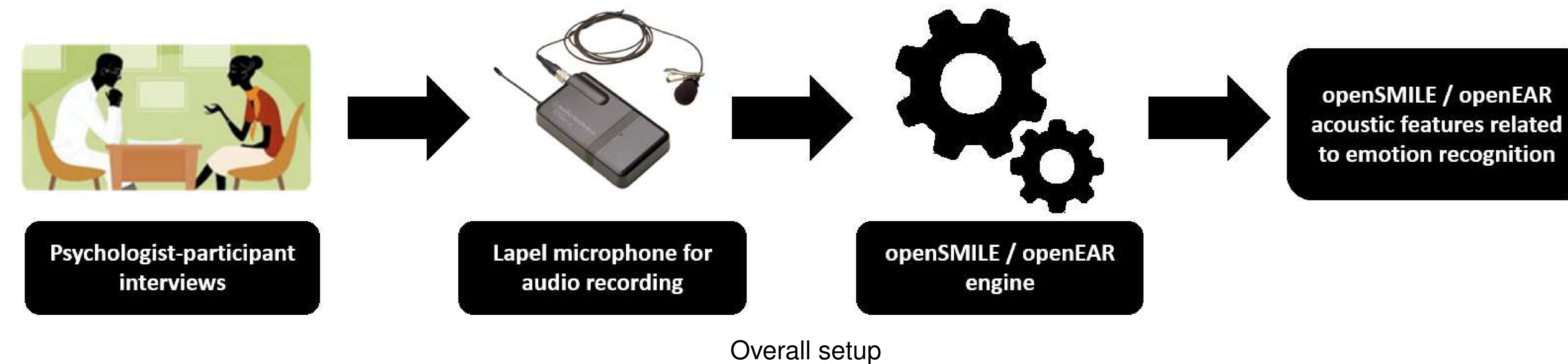
- Schizophrenia is a chronic mental disorder **affecting millions globally**.
- It is characterised by **positive and negative symptoms**.
- Negative symptoms are **difficult to detect** with few effective drug treatments.
- One of the key negative symptoms is the **lack of emotions in facial expressions and prosody**.
- Emotional **impairment in speech prosody** has been **studied lesser** compared to blunting of facial affect.
- Patients are unable to identify emotion from voice due to their **inability to process low-level acoustic features**, such as pitch, intensity, frequency etc. [1, 2].

Research Questions

- 1 Can the low-level acoustic signals be used to detect the severity of negative symptoms?
- 2 Can the same signals be used to differentiate between patients and healthy individuals?

		Patients (N = 52)	Controls (N = 26)
Age	Mean (years)	30.3	29.6
	Range (years)	20-46	19-47
Gender	Male	25	12
	Female	27	14
Ethnicity	Chinese	44	22
	Malay	5	3
	Indian	3	1
Education	University	7	4
	Diploma/ Vocational High School	28 17	15 7

Demographics data of participants.



Design of Experiment

- 78 total participants in the study, with 52 **Patients** and 26 **Healthy Controls**.
- The **audio and video** of the participants are **recorded** while they are **interviewed** by a trained psychometrician.
- The psychometrician rates the participants' behavior on the **NSA-16 scale** [3].
- The interview is semi-structured, with no time-limit for responses, and no role-playing.
- Average duration of interview is 26 minutes, amounting to **34 hours of audio** data analysed.

System Overview

- Portable H4n recorder and lapel microphones to record audio.
- Low-level audio descriptors related to emotion recognition from openSMILE [4].
- The descriptors are: **Intensity, Loudness, MFCC (12), Pitch, Probability of voicing, F₀ envelope, LSF (8), and ZCR**.
- 19 statistical measures on the descriptors and their delta coefficients take the total number of features to **988**.

Results

First, we present results of binary classification on the relevant NSA-16 items:

NSA-16 Items Classification

NSA-16 Item	Accuracy	Baseline Accuracy	Algorithm
Prolonged time to respond	82.69%	67.31%	kNN
Restricted speech quantity	82.69%	59.61%	Gaussian SVM
Impoverished speech content	80.77%	51.92%	Linear SVM
Emotion reduced range	61.54%	53.85%	kNN
Affect: Reduced modulation of intensity	78.85%	50.00%	Adaboosted DT
Reduced expressive gestures	84.62%	65.38%	Adaboosted DT

Now, the results for Patients v/s Controls classification:

Patient v/s Controls Classification

Features	Accuracy	Baseline Accuracy	Algorithm
openSMILE audio	79.49%	66.67%	Linear SVM
openSMILE audio + movement signals [5]	86.36%	65.15%	Linear SVM

All classifications were performed with appropriate feature-selection and leave-one-out cross-validation.

Discussion

- NSA-16 items related even to **speech and gestures** can be reliably **predicted**.
- Prosody dysfunction and reduced display of gestures often co-exist with emotional impairment in negative symptoms of schizophrenia.
- It also leads to high accuracy for **Patient v/s Control** classification.

Conclusion

- Inability to process emotions in speech is a salient feature of negative symptoms.
- We utilized this effect to predict the subjective ratings and differentiate between patient and healthy individuals.
- Our goal is to move towards **automated, objective assessment of mental health disorders**.
- Such systems can assist clinicians to screen and track individuals for mental illnesses.

Acknowledgements

- Singapore Ministry of Health's National Medical Research Council Center Grant.
- Nanyang Institute of Technology in Health and Medicine Grant.
- Being Together Center, Institute for Media Innovation.
- RRIS Rehabilitation Research Grant.
- Interdisciplinary Graduate School.

References

- [1] Rinat Gold, Pamela Butler, Nadine Revheim, David I Leitman, John A Hansen, Ruben C Gur, Joshua T Kantrowitz, Petri Laukka, Patrik N Juslin, Gail S Sillipo, et al. Auditory emotion recognition impairments in schizophrenia: relationship to acoustic features and cognition. *American Journal of Psychiatry*, 169(4):424-432, 2012.
- [2] David I Leitman, Petri Laukka, Patrik N Juslin, Erica Saccente, Pamela Butler, and Daniel C Javitt. Getting the cue: sensory contributions to auditory emotion recognition impairments in schizophrenia. *Schizophrenia bulletin*, 36(3):545-556, 2008.
- [3] Nancy C Andreasen. Negative symptoms in schizophrenia: definition and reliability. *Archives of General Psychiatry*, 39(7):784-788, 1982.
- [4] Florian Eyben, Martin Wöllmer, and Björn Schuller. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1459-1462. ACM, 2010.
- [5] Debsubhra Chakraborty, Yasir Tahir, Zixu Yang, Tomasz Maszczyk, Justin Dauwels, Daniel Thalmann, Nadia Magnenat Thalmann, Bhing-Leet Tan, and Jimmy Lee. Assessment and prediction of negative symptoms of schizophrenia from rgb+ d movement signals. 2017.