voicesense

CASE STUDY

BEER-YAAKOV MENTAL HEALTH

Prosodic Vocal Analysis and Mental Disorders

OBJECTIVES

The research examined the relationship between depression and acoustic vocal patterns (prosody), using the Voicesense mobile device for mental health monitoring.

The study was designed to support two objectives:

- Characterizing prosodic speech pattern differences between mental health patients and non-patients.
- Characterizing prosodic speech pattern changes over time among subjects inflicted by mental illness.

METHOD

Subjects: The subjects sample consisted of a depressed patient group, a schizophrenic patient group and non-patient group (control). The psychiatric subjects were acute and ambulatory patients of the Beer Yaakov-Ness Ziona mental health center in Israel. The non-patient control group subjects were recruited by a survey company for general vocal research purposes of Voicesense Ltd. A total number of 158 subjects were included in the study, 42 depressed patients (18 males and 24 females), 12 schizophrenic subjects (5 males and 7 females) and 104 non-patient subjects (53 males and 51 females).

The vocal samples of the study were mobile phone conversation recordings of the subjects. On average, each subject recorded 16.4 conversations.
Overall, the study consisted of 3509 recorded conversations, 1016 recordings of the depressed patient group, 489 recordings of schizophrenic group and 2004 recordings of the non-patient group. These 3509 recordings consisted the dataset sample of the study.

Tools:

The Psychiatric assessments included the following tools:

- Hamilton depression inventory
- Hamilton anxiety inventory
- PANNS – psychiatric state severity

The voice sampling was carried out using the Voicesense mobile software application for mental health monitoring. The subjects' voices were recorded while they were engaged in regular phone conversations through the application (with full consent and without identification).
At least two minutes of the subject's voice were collected in each call. The recorded audio was sent by the application to central processing in Voicesense secured cloud servers.

The voice analysis was carried out using Voicesense central vocal analysis software, focused on prosodic features of the speech. The analysis is language independent and content free (no understanding of what is being said). It calculates over 200 raw voice parameters per recording, consisting of a wide range of acoustic features. Thousands of datapoints are averaged, calibrated and normalized to reflect the individual's personal speech patterns in every given recording. These patterns were then analyzed using machine learning models (separate training and test sub-samples), in order to select and weight the vocal parameters that best correlate with Depression and Schizophrenia and reach stable and reliable predictive models. The software output were unified vocal-based scores of depression risk and schizophrenia risk per each recording.

Procedure:

- Phase A (acute psychiatric state)—the subjects were recruited when they arrived in acute mental health state to the hospital or clinic. They went through a psychiatric assessment and their vocal samples were collected while they were still in acute phase.
- Phase B (remission psychiatric state)—the subjects were sampled again after reported by health staff to be in remission (typically a few months after phase A, and not less than one month after phase A). They went through a second psychiatric assessment and if indeed remission was verified by the assessment then a second vocal sample set was collected.

The non-patient subjects' data was collected in one phase. The subjects downloaded the mobile software to their phones and their conversations were recorded within a two-week period.

Statistical analysis was then carried out to examine the study objectives. The vocal scores of the depressed and schizophrenic subjects were compared to the vocal scores of the non-patient groups. And the vocal scores of the subjects in acute phase were compared to the vocal scores of the subjects in the remission phase. Pearson correlation, Anova variance analysis and positive and negative predictive values analysis (confusion matrix) were used for the statistical analysis.

RESULTS

Depression

The differences in the vocal depression risk scores between depressed patients in acute state and between the non-patient group were highly significant. Results were significant both when the analysis used an average of all the voice recordings per subject and when only a single recording was used. (Statistical figures will be provided in the detailed summary).

The differences in the vocal depression risk scores between depressed patients in remission state and between the non-patient group were highly

Results were significant both when the analysis used an average of all the voice recordings per subject and when only a single recording was used. (Statistical figures will be provided in the detailed summary). The differences in the vocal depression risk scores between depressed subjects in acute state and between depressed subjects in remission were significant. (Statistical figures will be provided in the detailed summary).

Schizophrenia

The differences in the vocal schizophrenic risk scores between schizophrenic patients in acute state and between the non-patient group were highly significant. Results were significant both when the analysis used an average of all the voice recordings per subject and when only a single recording was used. (Statistical figures will be provided in the detailed summary). There weren't enough schizophrenic subjects in remission, so there are no comparisons between schizophrenic subjects in remission to the non-patient group or the schizophrenic in acute state. Predictive power of the Vocal Depression risk score
The predictive power of the vocal depression risk model, or its statistical fit to the Hamilton reference scores, was evaluated using binary confusion matrix analysis.

The Hamilton scores were grouped into two categories: 'Low Depression' and High Depression'. The vocal depression scores were also grouped into two categories: 'Low depression risk' and 'High depression risk'
As explained in the Method paragraph above, the Vocal depression model was developed using a training sub-sample (on which the model was trained) and a test sub-sample (on which the model was tested). Therefore, the predictive power of the model would be best evaluated by the confusion matrix results for the Test sub-sample. The differences between the Training and the Test sub-samples also give an indication of the expected stability of the model.

The confusion matrix results for accuracy, sensitivity and specificity were all between 70%-80%. (Statistical figures will be provided in the detailed summary).

CONCLUSIONS

The study offers support to its two main objectives:

The results demonstrated strong and significant differences between the vocal risk scores of depressed and schizophrenic subjects and between the equivalent vocal risk scores of the non-psychiatric subjects. These results suggest that vocal analysis can be used as a risk indicator for mental illness in the general population as a potential screening tool.
The results also demonstrated significant differences between subjects in acute and remission psychiatric states. These results suggest that vocal analysis can be used as an indicator for tracking changes over time among psychiatric patients and possibly offer early alerts of change in psychiatric state.
Other conclusions relate to the usability of the tool (significant results were obtained even with one audio sample per subject), as well as to the tool's accuracy, sensitivity and specificity indicators (will be elaborated with the detailed summary).