Speech markers of MDMA and Oxytocin ingestion

Presenter: Carla Agurto

Motivation

- **Correct assessment** of substance use disorders is essential for treatment planning and referral to adjunctive services.
- **Clinical reviews** are used for this purpose. However, they are inherently **subjective** and have limitations in terms of how comprehensively they can assess this condition.
- **Speech** data has the potential to become a powerful tool to provide **quantitative information** about emotion beyond that achieved by subjective assessments.
Protocol: All participants received, in randomized order, doses of placebo, MDMA at two different concentrations (0.75 mg/kg and 1.5 mg/kg), and Oxytocin.

Procedure:
• Participants were asked to perform two speech tasks (5-minute duration) in each session:
  ➢ Monologue (no listener was present)
  ➢ Description about important person in their life (listener was present and minimum participation was allowed)
• Recordings were manually transcribed.

Subjects: 31 subjects (12 F: 24.6 + 4.7 years, 19 M: 24.1 + 4.5 years)
Mel-Frequency cepstral coefficients (MFCC). MFCCs characterize the voice spectrum on a non-linear scale which is similar to pitch perception in the human auditory system.

\[
\text{Mel}(f) = 2595 \log \left( 1 + \frac{f}{700} \right)
\]

- Vowel space (e.g. distribution of formants which measure vocal tract resonances)
- Voice stability (e.g. jitter, shimmer)
- Noise measurements (e.g. harmonics to noise ratio)
- Spectral characterization (e.g. slope of frequency spectrum)
- Temporal features (e.g. articulation rate, pause duration distribution)

Comparison with American reference vowel space
Semantic Features

✓ Latent Semantic Analysis (LSA) was used to evaluate the content of speech.

✓ For our text corpus, we used TASA, a collection of educational materials compiled by Touchstone Applied Science Associates (7651 documents and 12 190 931 words, from a vocabulary of 77 998 distinct words).

✓ Transcripts were lemmatized using NLTK lemmatizer and only nouns were used for the analysis.

✓ Cosine distance between each word uttered in each session to the following words of interest was calculated.

affect, anxiety, compassion, confidence, disdain, emotion, empathy, fear, feeling, forgive, friend, happy, intimacy, love, pain, peace, rapport, sad, support, think, and talk.
Psycholinguistic Features

Lexical content:

- Content words, total words, empty words (e.g. thing), fillers (e.g. uh ), typetoken (unique words/total words)
- Brunet’s index: $W = NV^{-0.165}$
- Honore’s statistic: $R = 100 \times \log(N/(1-V_1/V))$
  
  where $N$ is the total text length, $V_1$ words uttered once, $V$ is total vocabulary used.

We used CPIDR software to calculate:

- Number of ideas found in the transcripts in each session
- Propositional density = numbers ideas/total words.

Parts of speech:

- Number of pronouns, nouns, verbs, determiners, indefinites, definite, and Is normalized by the total number of words
**Statistical Analysis**

We performed Wilcoxon signed rank test, a paired test, designed to evaluate the effects in the same subject. To correct for multiple comparisons, false discovery rate (FDR) correction at q<0.05 was performed.

**Partial Correlations**

Using the features with the lowest p-value, we analyze the partial correlations between them as a measure of causal associations.

\[ \rho_{XY \cdot Z} = \frac{\rho_{XY} - \rho_{XZ} \rho_{ZY}}{\sqrt{1 - \rho_{XZ}^2} \sqrt{1 - \rho_{ZY}^2}}. \]

**Classification**

We standardized the features (mean = 0 and standard deviation = 1), and performed a nested leave-subject-out cross validation approach using Linear SVM and Random Forest:

<table>
<thead>
<tr>
<th>Type of features</th>
<th>Classification tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acoustic Features</td>
<td>Placebo vs. MDMA low dose</td>
</tr>
<tr>
<td>Semantic Features</td>
<td>Placebo vs. MDMA medium dose</td>
</tr>
<tr>
<td>Psycholinguistic Features</td>
<td>Placebo vs. Oxytocin</td>
</tr>
<tr>
<td>All features</td>
<td>MDMA low dose vs. MDMA medium dose</td>
</tr>
<tr>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
## Results

### Statistical Analysis

- We observe a different behavior for monologues and description tasks.
- Psycholinguistic features (lexical content and number of ideas) show statistical significance for the description task.
- F2 helps distinguish OT from PBO. Previous approaches showed that positive valence (elation, pleasure, etc.) resulted in higher F2 values.

<table>
<thead>
<tr>
<th>Type of Feature</th>
<th>Monologues</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Feature Name (p-value)</td>
<td>Feature Name (p-value)</td>
</tr>
<tr>
<td></td>
<td>Acoustic</td>
<td>Semantic</td>
</tr>
<tr>
<td>MDMA 0.75 vs. PBO</td>
<td>Pitch\textsubscript{a} – 7.2E-4</td>
<td>Think – 4.1E-3*</td>
</tr>
<tr>
<td>MDMA 1.5 vs. PBO</td>
<td>Pitch\textsubscript{a} – 2.3E-3</td>
<td>Talk – 4.6E-3*</td>
</tr>
<tr>
<td>MDMA 0.75 vs. MDMA 1.5</td>
<td>MFCC #12\textsubscript{a} – 7.9E-3</td>
<td>Talk – 8.9E-3</td>
</tr>
<tr>
<td>Oxytocin vs. PBO</td>
<td>Pitch\textsubscript{a} – 7.4E-4</td>
<td>Support – 1.9E-1</td>
</tr>
<tr>
<td></td>
<td>MFCC #4\textsubscript{b} – 2.9E-2</td>
<td>Think – 2.1E-1</td>
</tr>
<tr>
<td></td>
<td>F2\textsubscript{c} – 5.2E-5*</td>
<td>Emotion – 2.1E-2</td>
</tr>
<tr>
<td></td>
<td>F2\textsubscript{b} – 8.4E-4*</td>
<td>Anxiety – 5.3E-2</td>
</tr>
</tbody>
</table>

- We observe a different behavior for monologues and description tasks.
- Psycholinguistic features (lexical content and number of ideas) show statistical significance for the description task.
- F2 helps distinguish OT from PBO. Previous approaches showed that positive valence (elation, pleasure, etc.) resulted in higher F2 values.
Partial Correlations (Description task)

- Stronger positive correlations are found in subjects under MDMA.
- Both MDMA doses share strong correlations that are not present in placebo.
- Love as a concept shares correlation with other concepts like think (OT) and talk (MDMA) but was absent in placebo.
- MDMA can make people feel anxious, even fearful that others want to harm them. This could explain why the concept of anxiety is positively correlated with pain.
Results

Classification Accuracy *(using Linear SVM and Random Forest)*

- Classification using acoustic features is better for the monologues than the description task.
- On the other hand, features obtained from transcripts (semantic, psycholinguistic) are more informative when the subjects are describing persons.

<table>
<thead>
<tr>
<th>Type of Feature</th>
<th>Monologues</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature Name (p-value)</td>
<td>Acoustic</td>
<td>Semantic</td>
</tr>
<tr>
<td>MDMA 0.75 vs. PBO</td>
<td>0.85*</td>
<td>0.71</td>
</tr>
<tr>
<td>MDMA 1.5 vs. PBO</td>
<td>0.71</td>
<td>0.68</td>
</tr>
<tr>
<td>MDMA 0.75 vs. MDMA 1.5</td>
<td>0.81*</td>
<td>0.52</td>
</tr>
<tr>
<td>Oxytocin vs. PBO</td>
<td>0.87</td>
<td>0.74</td>
</tr>
</tbody>
</table>

* Best performance was achieved with Random Forest.
Conclusions

• We present the first study that uses characteristics of speech to identify subjects that are under the influence of MDMA and oxytocin.

• The most relevant acoustic features correlate with positive valence, which supports previous research of drug effects using subjective analyses.

• For the description task, we observe that features that characterize language complexity (lexical content) are useful to discriminate placebo from MDMA.

• We observe that subjects behave differently when they are speaking to another person (description task) or not (monologue task). This can be seen in the difference in classification performance.

• Our results suggest that speech may be a potential future solution to replace subjective analyses to improve the reliability of the clinical studies for drug effects assessments.
Thank you for your attention