AN ACOUSTIC ANALYSIS OF CANNABIS-INTOXICATED SPEECH

Arian Shamei[*1] **and Sonya Bird**[†2]
[1]Department of Linguistics, University of British Columbia
[2]Department of Linguistics, University of Victoria

1 Introduction

Acoustic analysis has proven to be an effective method for discriminating the speech features of countless mental and physiological states such as alcohol intoxication [1], Schizophrenia [2], and heart disease [3]. All of these states result in alterations to the muscle-control underlying speech, which can be detected in the acoustic signal. The goal of this study was to determine if it is similarly possible to detect cannabis intoxication (CI) through acoustic analysis of speech.

2 Method

2.1 Participants

Eight medicinal users of cannabis (4 male, 4 female) aged 21-25 provided voice recordings before and after the consumption of medicinal cannabis. All participants were undergraduate students at the University of Victoria who used Cannabis to treat a variety of ailments. As all participants were medicinal users, all had previous experience consuming cannabis, although the duration and frequency of use varied greatly between participants.

2.2 Procedure and materials

The experiment was conducted utilizing a *pretest (PT) and retest (RT)* format, which involved completing the same task before (PT) and after (RT) cannabis consumption. Participants abstained from cannabis use for twenty-four hours prior to completing the PT. Following consumption, participants waited 30 minutes before completing the RT. In all cases, cannabis was consumed via inhalation, however the exact quantity and potency of cannabis consumed by each participant was not controlled; each participant provided their own cannabis and consumed the amount they normally would in accordance with their prescription and daily routine.

The task itself included three components used to measure the acoustic parameters of interest: a reading passage for measuring pitch trajectory, a list of words contrasting voiced~voiceless stops (e.g. *pit ~ bit*) to measure voice onset time (VOT), and a second list of words contrasting seven sustained English vowels to measure voice quality.

Pitch range and trajectory

Prosodic trajectories - pitch in particular - have been useful in distinguishing alterations to executive functions such as attention and planning in speaker states [2]. It stands to reason pitch trajectory may also distinguish CI. To test this, participants were asked to read the entire passage of *The North Wind and Sun* – a short, emotionally neutral text often used in phonetic research. In each condition (PT and RT), participants provided two readings of the story after familiarizing themselves with the text. Approximately twenty seconds of uninterrupted speech from the second repetition of the passage was selected for analysis, for each participant in each condition, by excluding the first and last stanza of the story. Two participant's data had to be excluded from the pitch trajectory analysis on the grounds that no fluent readings could be obtained in the RT, although PT readings were unaffected.

Intersyllabic and intrasyllabic pitch within the reading passage was measured using a similar methodology to that employed by [2] for their assessment of schizophrenic speech, namely their application of the Prosogram [4] for forensic assessments of speaker state: automatic syllable segmentation was implemented with a sampling rate of 0.005/s and an adaptive glissando threshold of [0.16-0.32/T2, DG = 30, dmin = 0.05], allowing rapid collection of F0 (pitch) data and cross-conditional comparisons of pitch trajectory.

Voice Onset Time

VOT provides an indirect measure of motor coordination, since it is a function of precise articulatory timing between the larynx and articulators in the vocal tract. To track the effect of CI on VOT, participants were asked to produce two repetitions of 20 minimal pairs (80 tokens), contrasting only in voicing of the first stop, eg : *pit ~ bit, dire ~ tire*. VOT was calculated by measuring the time in milliseconds (ms) between the release burst of the plosive and the onset of regular voicing in the following vowel. If the onset of voicing commenced after the release burst, the VOT was recorded as positive, but if voicing commenced before the release burst, the value was recorded as negative. For each speaker, the mean VOT of individual target phonemes - /b, p, t, d/ - was calculated to determine the standard deviation of VOT (σVOT). Phoneme-specific σVOTs were combined to calculate the mean σVOT of voiced phonemes (/b, d/) separately from voiceless phonemes (/p, t/), producing measurements of σVOT [+voice] and [-voice].

Vocal Quality

Measures of vocal quality have been used to generate insight about laryngeal physiological states such as muscle tension [3]. The goal of the *sustained vowel production task* (SVPT) was to allow measurements of laryngeal functioning in a non-

[*]arian@alumni.ubc.ca
[†]sbird@uvic.ca

speech setting where voluntary control of the larynx can be observed unimpeded. Participants were asked to produce the following words three times while sustaining the vowel for roughly five seconds: *see, saw, sue, sit, so, sat, say*. For each participant, the final two iterations (14 tokens) were selected for analysis in each condition. Jitter and Shimmer (local) were measured automatically with a Praat script, omitting the initial and final 25% of each token.

3 Results

3.1 Pitch range and trajectories

Results of the F0 analysis as described in 2.2.1 are provided in Table 1 below. All measurements are provided in semi-tones. Total pitch range and intrasyllabic trajectory were significantly reduced in RT compared to PT; intersyllabic trajectory and phonation trajectory were not significantly different in PT vs. RT.

Table 1: Pitch range and trajectories in semi-tones

Range P=.04		**Intra** P=.01		**Inter** P=.21		**Phon** P=.06	
PT	RT	PT	RT	PT	RT	PT	RT
11.9	9.5	15.2	12.1	18.3	18.9	16.4	14.9

Figure 1 below provides comparative prosograms of the PT (top) and RT (bottom). Note that intersyllabic trajectory is reduced in RT, but intersyllabic fluctuations remained consistent between conditions.

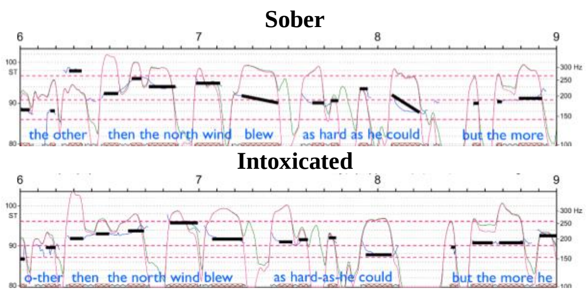

Figure 1: Comparative prosograms of PT (top) & RT (bottom)

3.2 Voice Onset Time

Table 2 below provides the mean σVOT measurements for both voiced (left) and voiceless (right) phonemes. The σVOT for voiced phonemes was significantly in the RT compared to the PT, meaning VOT was more variable in RT. Voiceless phonemes were comparable in the PT and RT.

Table 2: Standard deviation of voice onset time

σVOT [+voice] P<.005		**σVOT [-voice]** P=.12	
PT	RT	PT	RT
20ms	36.2ms	18.1ms	15.9ms

Figure 2 below illustrates prevoicing of onsets observed in the RT (right) but not in the PT (left). English stops are not normally prevoiced, but sometimes were in the RT, resulting higher σVOT [+voice] measurements in the RT condition.

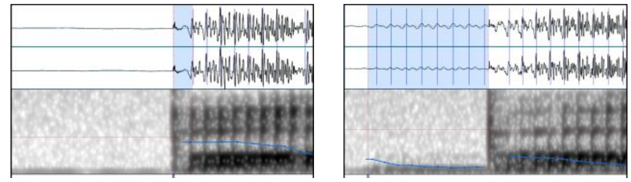

Figure 2: prevoicing in RT token (right) absent in PT(left)

3.3 Vocal Quality

Table 2 below provides the mean measurements of vocal quality for all participants as described in 2.2.2. Shimmer, but not jitter, was significantly lower in the RT to than PT.

Table 3: Jitter and Shimmer

Jitter P=.18		**Shimmer** P=.04	
PT	RT	PT	RT
0.38%	0.34%	3.7%	3.1%

4 Discussion

Analysis of pitch trajectories, VOT, and vocal quality indicate that CI affects speech stream acoustics in significant ways. The range of F0 following intoxication is significantly reduced, and these reductions match previous observations of schizophrenic speech [2], which reflect impairments to executive functions and processing load. Measurements of VOT demonstrate significant impairments to motor timing between the larynx and lips (/b, p/) and tongue (/d, t/). Measurements of vocal quality indicate significant reductions to shimmer which suggest greater flaccidity of the vocal folds and impairments to muscle tone [3]. Future work should investigate the utility of automated acoustic detection of CI.

Acknowledgments

This research was conducted on the unceded and ancestral territory of the WSANEC, Lkwungen, and Wyomilth people. We also wish to thank all the participants, whose volunteerism made this research possible.

References

[1] Wang, W. Y., Biadsy, F., Rosenberg, A., & Hirschberg, J. (2013). Automatic detection of speaker state: Lexical, prosodic, and phonetic approaches to level-of-interest and intoxication classification. *Computer Speech & Language*, *27*(1), 168-189.

[2] Martínez-Sánchez, F., Muela-Martínez, J. A., Cortés-Soto, P., Meilán, J. J. G., Ferrándiz, J. A. V., Caparrós, A. E., & Valverde, I. M. P. (2015). Can the acoustic analysis of expressive prosody discriminate schizophrenia?. *The Spanish journal of psychology*, *18*.

[3] Murton, O. M., Hillman, R. E., Mehta, D. D., Semigran, M., Daher, M., Cunningham, T., ... & Ausiello, D. (2017). Acoustic speech analysis of patients with decompensated heart failure: A pilot study. *The Journal of the Acoustical Society of America*, *142*(4), EL401-EL407.

[4] Mertens, P. (2004). The prosogram: Semi-automatic transcription of prosody based on a tonal perception model. In *Speech Prosody 2004, International Conference*.