SPEAKER-DEPENDENT FEATURE GENERALIZABILITY FOR THE DETECTION OF ALCOHOL INTOXICATION

Xinglei Liu [*1], **Arian Shamei** [†1,2,3], **and Rima Seiilova** [‡1]
[1]Tenvos Incorporated, Sacramento, California, USA
[2]École de Technologie Supérieure, Montreal, Québec, Canada
[3]University of British Columbia, Vancouver, British Columbia, Canada

1 Introduction

Previous studies have investigated the impact of alcoholic intoxication on fundamental frequency and formants [1–3]. In this paper, alongside exploring F0 and formants, we aim to examine additional vowel and consonant features. Recognizing the considerable variability of acoustic features across speakers, we approach alcohol intoxication prediction in a speaker-dependent manner. Our objective is to determine whether certain features can generalize across speakers in predicting alcohol intoxication. To evaluate this, we calculate the average feature importance across speakers as a measure of the feature's overall generalizability.

2 Method

2.1 Feature Extraction

The Alcohol Language Corpus (ALC) provided Praat TextGrid files which include word and phoneme alignments. Using Parselmouth, a python interface for Praat [4, 5], we utilized the TextGrid files to get vowel and consonant intervals. Based on the intervals, for each speaker, we extracted 10 features from vowels (F0 - F3, jitter, shimmer, harmonics-to-noise ratio, vowel euclidean distance, duration and duration variability), 6 from consonants (spectral skewness and kurtosis, center of gravity, duration and duration variability, harmonics-to-noise ratio), along with the onset consonant-to-vowel ratio. We used the parselmouth library to compute F0 - F3, jitter, shimmer, harmonics-to-noise ratio, center of gravity, skewness, kurtosis. When calculating F0 - F3, jitter, shimmer, and harmonics-to-noise ratio, we set the pitch floor to 125 Hz and the pitch ceiling to 300 Hz for females, and to 85 Hz and 200 Hz, respectively, for males. The onset consonant-to-vowel ratio was computed using words with the first phoneme as a consonant and the second phoneme as a vowel. To account for individual and phonetic differences, we normalized F0 - F3 by individual speakers, while normalizing center of gravity, skewness, and kurtosis by the type of consonant, since these features exhibit significant variability across consonant types.

TABLE 1 – Number of participants and files used in this paper

	Participants	Intoxicated Files	Sober Files
Female	43	946	2120
Male	54	1254	2661

*. xinglei@tenvos.com
†. arian@tenvos.com
‡. rima@tenvos.com

2.2 Feature Importance

To compute feature importance, we used random forest models as speaker-dependent models. Given that we used 0.08 as the threshold to divide sober and intoxicated states, each speaker has on average 22 intoxicated recordings and 49 sober recordings, resulting in class imbalance. To address this issue, we have downsampled the sober class for each speaker before fitting the random forest model. Then we used 5-fold cross-validation to evaluate the models and calculate the average feature importance using the associated mean decrease in Gini impurity (GI).

3 Results

Among all speakers, spectral skewness and kurtosis of consonant features demonstrate the highest levels of generalizability, with mean Gini impurity (GI) values of .11 and .09, respectively. Figure 1 shows that for male models, F0 emerges as the second most generalizable feature; however, its generalizability is comparatively diminished in female models. Conversely, vowel duration and F2 with mean Gini impurity (GI) values of .04 and .03 respectively are consistently identified as the least generalizable features across both male and female models.

The mean F1 score across all speaker-dependent models is 69.0%, with 69.8% precision and 71.8% recall. As depicted in Table 2, male models outperform female models on average. This reflects that chosen features demonstrate better efficacy in classifying intoxication among male speakers. Notably, the F1 scores of the speaker-dependent random forests exhibit a wide range, spanning from 36% to 96%, suggesting that these features may not perform well for certain individuals despite their overall generalizability across the population.

TABLE 2 – Performance for male and female speaker-dependent random forest models

	Female	Male
Avg Precision	66.7%	72.2%
Avg Recall	68.6%	74.3%
Avg F1	65.8%	71.6%
Min F1	42.7%	35.9%
Max F1	95.6%	96.0%

4 Discussion and Conclusion

In this paper, we explored the generalizability of phonetic features across a diverse range of speakers in the context of al-

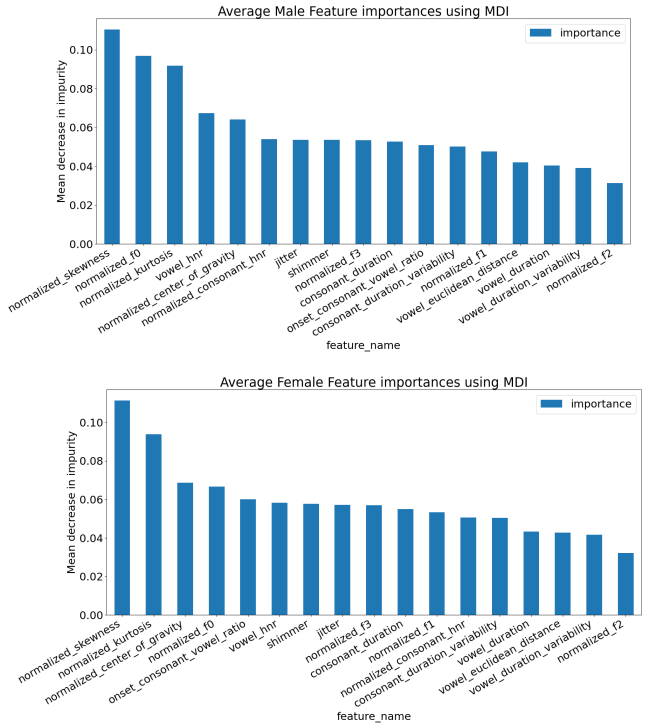

FIGURE 1 – Bar plots illustrating the mean decrease in Gini impurity for 17 features for both males and females. The x-axis represents the names of the features, while the y-axis denotes the mean decrease in Gini impurity.

coholic intoxication. By extracting 17 acoustic features from each audio file and training individual random forest models for each speaker, we assessed the average decrease in Gini impurity as a measure of generalizability. The performance of models varied across speakers and genders, but based on feature importance, consonant spectral skewness and kurtosis are the most generalizable features across all speakers. The high generalizability of spectral skewness and kurtosis across speakers likely reflects generalized changes to vocal tract physiology during intoxication, yet it remains unclear how these changes to vocal tract physiology interact with individual variability in speech motor control. Moving forward, further exploration of additional features will help us understand more about individual diversity and shared patterns in feature alterations following intoxication.

5 Acknowledgements and Disclosures

This research was funded by a NSF SBIR Phase 1 grant awarded to Tenvos Incorporated of Sacramento, California. Tenvos Incorporated has a financial interest in the development of acoustic speaker state detection systems. This research was conducted by the authors on behalf of Tenvos Incorporated.

References

[1] F. Schiel, Chr. Heinrich, and V. Neumeyer. Rhythm and formant features for automatic alcohol detection. In *Proceedings of the INTERSPEECH 2010 Conference*, pages 458–461, Chiba, Japan, 2010.

[2] B. Baumeister, Ch. Heinrich, and F. Schiel. The influence of alcoholic intoxication on the fundamental frequency of female and male speakers. *The Journal of the Acoustical Society of America*, 132(1) :442–451, 2012.

[3] Arian Shamei and Xinglei Liu. Expansion of vowel space following alcohol intoxication, Nov 2023.

[4] Yannick Jadoul, Bill Thompson, and Bart De Boer. Introducing parselmouth : A python interface to praat. *Journal of Phonetics*, 71 :1–15, 2018.

[5] Paul Boersma. Praat : doing phonetics by computer [computer program]. *http ://www. praat. org/*, 2011.

[6] F. Schiel, Chr. Heinrich, and S. Barfüßer. Alcohol language corpus. *Language Resources and Evaluation*, 46(3) :503–521, 2012.

[7] Leo Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. CRC Press, Boca Raton, 1984.