Detection of Alcohol Intoxication Using Voice Features: A Controlled Laboratory Study

BRIAN SUFFOLETTO,[a,*] AYMAN ANWAR,[b] SEAN GLAISTER,[a] & ERVIN SEJDIC[b,c]

[a]*Department of Emergency Medicine, Stanford University, Palo Alto, California*

[b]*Department of Electrical & Computer Engineering, University of Toronto, Toronto, Ontario, Canada*

[c]*North York General Hospital, Toronto, Ontario, Canada*

ABSTRACT. Objective: Devices such as mobile phones and smart speakers could be useful to remotely identify voice alterations associated with alcohol intoxication that could be used to deliver just-in-time interventions, but data to support such approaches for the English language are lacking. In this controlled laboratory study, we compare how well English spectrographic voice features identify alcohol intoxication. **Method:** A total of 18 participants (72% male, ages 21–62 years) read a randomly assigned tongue twister before drinking and each hour for up to 7 hours after drinking a weight-based dose of alcohol. Vocal segments were cleaned and split into 1-second windows. We built support vector machine models for detecting alcohol intoxication, defined as breath alcohol concentration > .08%, comparing the baseline voice spectrographic signature to each subsequent timepoint and examined accuracy with 95% confidence intervals (CIs). **Results:** Alcohol intoxication was predicted with an accuracy of 98% (95% CI [97.1, 98.6]); mean sensitivity = .98; specificity = .97; positive predictive value = .97; and negative predictive value = .98. **Conclusions:** In this small, controlled laboratory study, voice spectrographic signatures collected from brief recorded English segments were useful in identifying alcohol intoxication. Larger studies using varied voice samples are needed to validate and expand models. (*J. Stud. Alcohol Drugs, 84,* 808–813, 2023)

DRINKING ALCOHOL TO INTOXICATION increases risks for numerous health and public safety problems (Alpert et al., 2022) including motor vehicle crashes (Naimi et al., 2018) and is responsible for an estimated 192 billion dollars in costs each year in the United States (Sacks et al., 2015). Mobile digital behavioral interventions reduce alcohol consumption in adults (Bendtsen et al., 2021; Kaner et al., 2017) and overcome barriers to in-person prevention and treatment (Venegas et al., 2021). Still, effects are small and the design is not yet optimized to provide just-in-time support (Nahum-Shani et al., 2017). Identifying individuals remotely who are intoxicated provides an opportunity to intervene in real time to mitigate further risks and subsequent harms.

Currently there are no commercially available tools to unobtrusively and effectively identify alcohol intoxication. Transdermal alcohol sensors (Russell et al., 2022) and portable breath alcohol meters (Norman et al., 2021) can accurately estimate blood alcohol content, but cost and availability-related barriers preclude widespread use (Piasecki, 2019). Self-report "drink counting" tools can accurately estimate blood alcohol content (Wray et al., 2014) but are too burdensome to be practically useful. Our group has shown that smartphone-based sensors can accurately detect alcohol intoxication remotely and that the most informative sensor features were related to time (i.e., day of week, time of day), movement (e.g., change in activities), device usage (e.g., screen duration), and communication (e.g., call duration) (Bae et al., 2018).

Separately, we have shown that smartphone-sensed gait (Suffoletto et al., 2018, 2020) can distinguish intoxicated from nondrinking states. Still, other measures of psychomotor function that are sensitive to acute alcohol consumption and can be readily measured remotely using existing technologies remain underdeveloped.

One measure of psychomotor function that could be useful in discriminating intoxicated from nonintoxicated states is the voice, the result of high-level sensory, cognitive, and motor processes requiring coordination of more than 100 muscles. There is a large and growing science of using voice to detect altered neurological "states" in individuals. For example, there are correlations found between prosodic, articulatory, and acoustic features of speech and clinical ratings of both depression and suicidality (Cummins et al., 2015). As well, phonation, articulation, and prosody abnormalities can detect untreated Parkinson's disease (Rusz et al., 2011). Preliminary research has also detected speech changes after administration of 3,4-methylenedioxymethamphetamine (MDMA; Agurto et al., 2020).

Although it has long been recognized that alcohol alters the acoustic-phonetic properties of speech (Johnson et al., 1990; Pisoni & Martin, 1989), human accuracy of classification is suboptimal (Baumeister & Schiel, 2013). Modern computing capabilities and signal analysis software help overcome the

Received: October 31, 2022. Revision: May 2, 2023.

This work was supported by National Institute on Alcohol Abuse and Alcoholism Grant R01AA023650.

*Correspondence may be sent to Brian Suffoletto at the Department of Emergency Medicine, Stanford University, 900 Welch Road, Suite 350, Palo Alto, CA 94304, or via email at: suffbp@stanford.edu.

doi:10.15288/jsad.22-00375

complexity of voice analysis. Specifically, methods to extract hierarchical acoustic features, perform iterative speaker normalization, and use machine learning techniques are all available to improve classification (Bone et al., 2014).

The widespread availability of audio samples through smartphones and smart speakers makes voice an attractive biomarker. For example, smartphones are near-ubiquitous (Pew Research Center, 2021), an estimated 35% of Americans own smart speakers (NPR, 2022), and a growing list of technologies incorporate voice commands (e.g., in-car systems). Most if not all of these systems contain audio recording, computing, and data transfer capabilities that could allow for remote and real-time inference of alcohol intoxication. Despite these capacities, however, no current commercially available voice-based recognition of alcohol intoxication exists.

In this proof-of-concept study, we test whether brief voice recordings and spectrographic (i.e., frequency-based) features can be useful to correctly discriminate alcohol-intoxicated from nonintoxicated states in adults. We chose to focus on frequency-based features given that prior research had shown that they are particularly good at discriminating levels of intoxication (Hollien et al., 2001). Findings could help inform new strategies for remote identification of intoxication for at-risk individuals to trigger just-in-time digital interventions to reduce health and public safety harms.

Method

Participants

From August to December 2018, we recruited 20 adults for a controlled laboratory study to test the effects of high doses of alcohol on blood and toxicology biomarkers. Participants were recruited via word of mouth and locally posted advertisements for a study to examine the effects of alcohol on the body. We conducted an initial screen by telephone to ensure individuals were at least 21 years old and consumed alcohol at least once per week. Consented participants then made appointments to come to the laboratory for one session that would last at least 7 hours and were instructed to abstain from consuming alcohol or using other psychoactive drugs during the 24 hours preceding the session. However, we did not confirm that participants had not used any psychoactive drugs on the day of the experiment. They were also told to fast and refrain from caffeine consumption at least 4 hours before the session. On the day of the session, participants were screened in person to verify age of at least 21 years using their driver license and a brief health survey. Individuals who reported any positive responses on the CAGE questionnaire (Liskow et al., 1995), hepatic/renal impairment, or peptic ulcer disease were excluded. Urine samples were also tested for pregnancy in female participants. Women who were pregnant or breastfeeding were excluded. For this study, two participants did not provide audio recordings because of technical difficulties; thus, 18 of 20 participants provided audio recordings and were included in our analyses.

Procedures

Participants presented to the Department of Emergency Medicine Applied Physiology Lab at the University of Pittsburgh at 8 A.M. After providing informed consent, participants completed a questionnaire including the 10-question Alcohol Use Disorders Identification Test (AUDIT; Saunders et al., 1993). Body weight and height were measured, and an intravenous line was placed to administer nausea medicine as needed (ondansetron 4 mg). Investigators prepared an ethanol oral dosing to achieve a goal peak breath alcohol concentration (BrAC) of > .20% using the Widmark formula as follows: 2 g/L × (0.7 L/kg [for men] or 0.6 L/kg [for women] × participant weight in kilograms) = dose of ethanol in grams/0.3156 g ethanol per milliliter = milliliter distilled spirits. Vodka was mixed with lime juice and simple syrup and administered according to standard procedures (Fillmore et al., 2000). Participants were given a maximum of 1 hour to finish alcohol consumption.

This high dose of alcohol was chosen given that the main aim of the original study was to identify toxicological biomarkers related to high doses of alcohol consumption. We attempted to minimize risk associated with high doses of alcohol by following the National Institute on Alcohol Abuse and Alcoholism (NIAAA) guidelines for alcohol administration (*NIAAA Guidance for Conducting Alcohol Administration Studies with Human Participants,* 2023), including not enrolling alcohol-naive participants or individuals with alcohol use disorder, pregnant women, or older adults. Participants were told the exact amount of alcohol they would be drinking, the expected BrAC values, and that they did not have to ingest the entire dose provided to them and could stop the administration of alcohol at any time if they felt uncomfortable or experienced any adverse effects. A study physician was available to assess risk and provide medical oversight throughout the study.

At baseline and each half-hour (for up to 7 hours), we measured BrAC (BACtrack S80; KHN Solutions, Inc., San Francisco, CA). Participants left the lab after 7 hours, when they could ambulate safely and had someone to drive them home. At baseline (before any alcohol consumption) and each hour for up to 7 hours, participants were asked to read a randomly allocated tongue twister while a smartphone was placed on the table 1 to 2 feet from their mouth to record the audio segment. No tongue twister was repeated within individuals. The order of tongue twisters was randomly generated for each participant. We used tongue twisters because they have been shown to be useful in inducing speech errors in healthy speakers (Goldstein et al., 2007) and identifying speech disorders (Kember et al., 2017).

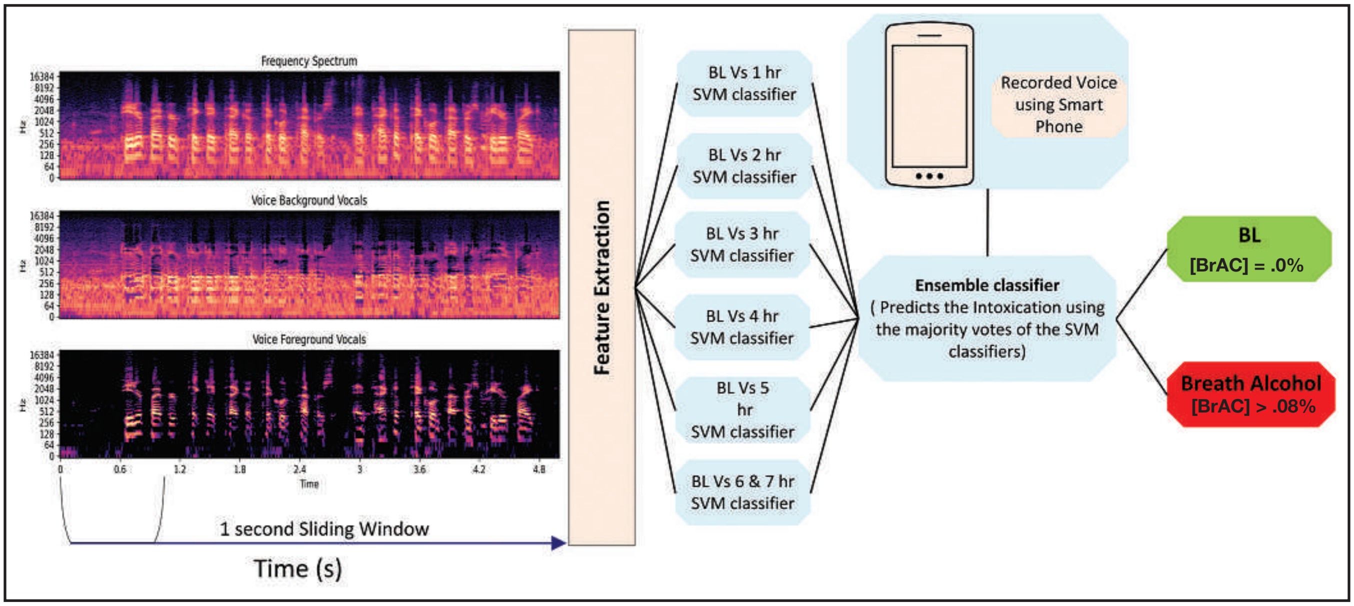

FIGURE 1. The classification scheme of the proposed algorithms. First, the voice records are separated into foreground and background vocals to remove noise and vocals in the background. Second, a sliding window of 1 second length slides over the voice foreground to extract the spectrogram features for different breath alcohol concentration (BrAC) levels. Then, the extracted features are fed into the corresponding support vector machine (SVM) model to separate the baseline (BL) level from the intoxicated vocals. Finally, the separated SVM predictions are fed to an ensemble classifier that predicts the presence of intoxication in the voice using a majority voting produced by the SVM classifiers.

Measures

Alcohol intoxication. We chose to use a threshold of BrAC > .08% as our classifier of alcohol intoxication because it has been used in prior studies of acute alcohol effects on psychomotor performance (Peacock et al., 2015) and represents the legal limit of blood alcohol in adult drivers in the United States.

Voice feature extraction. The spectrogram of each record showed some interference between speakers presented at the lab during the recording process. Hence, we first separated the vocals of the subject from the surrounding speakers by identifying regions of foreground and background sounds using nonlocal (i.e., Wiener) filtering (Jiang et al., 2002). After vocals separation, the time segments with no words were removed from the record and extracted foreground were used later in the segmentation and features extraction process. Afterward, we partitioned each record into nonoverlapping segments of 1 second length using the original sampling rate of the Voice Recorder App, which was 44,100 Hz.

Finally, a set of spectrogram features were then extracted from each segment. One feature was the Mel-frequency cepstral coefficient to describe the spectral envelope, which is generally found to outperform formant features across many speech-based classification tasks (Muda et al., 2010). Other features included the spectral centroid or the mean of the frequencies; spectral roll-off, defined as the energy of the spectrum (Yu et al., 2021); and spectral flatness or tonality coefficient (Dubnov, 2004). Finally, we examined spectral bandwidth and contrast (Klapuri et al., 2006). The extracted features from the spectrogram yielded a high number of features per frame, which increased the dimensionality of the data. Hence, we used principal component analysis to reduce the dimensionality of the data by computing the highest 50 principal components.

Analyses

Given that data are nested within individuals, and thus timepoints across individuals autocorrelated, we built support vector machine (SVM) models for detecting alcohol intoxication for each timepoint compared with the baseline when BrAC was 0%. A total of 18 participants were included in model training and testing. We used a leave-one-participant cross-validation in the training and testing process in which 17 subjects were used in the training in each run and 1 subject was used in the testing. We repeated that process for all subjects and calculated the average performance measures across all the runs. We present a Confusion Matrix showing the number of 1-second voice segments classified as true positives, true negatives, false positives, and false negatives for each timepoint comparison as well as for an ensemble classifier, which combines the predictions of the classification from the SVM models. For each new instance, the ensemble classifier selects the majority voting of all classifiers to predict the current instance class, as illustrated in Figure 1.

Table 1. Comparison and Confusion Matrix based on 1-second voice samples

Variable	BrAC: *M* (range)	TP	TN	FP	FN
BL vs. 1 hour	0 vs. .19 (.15 to .29)	168	255	8	3
BL vs. 2 hours	0 vs. .19 (.12 to .25)	176	236	5	3
BL vs. 3 hours	0 vs. .18 (.11 to .26)	178	193	8	2
BL vs. 4 hours	0 vs. .17 (.09 to .32)	182	139	1	5
BL vs. 5 hours	0 vs. .15 (.08 to .25)	179	149	2	4
BL vs. 6 & 7 hours	0 vs. .15 (.08 to .22)	184	114	1	2
Total		1,065	1,080	25	19

Notes: BrAC = breath alcohol concentration (gram%); TP = true positive; TN = true negative; FP = false positive; FN = false negative; BL = baseline.

Results

Participants

Mean age was 29 years (*SD* = 9.7), with a range of ages from 21 to 62 years. The majority (72%) of participants were men, and all participants were White and non-Hispanic. Mean AUDIT score was 5.8 (*SD* = 2.5), with six participants meeting criteria for risky drinking based on a score between 7 and 15. Mean weight was 76 kg (range: 51–102) and mean height 68 inches (range: 62–73). The BrAC was confirmed at 0% at baseline and increased above .08% in all participants by 1 hour. The mean and range of BrAC for each comparison are shown in Table 1.

Support vector machine models

Alcohol intoxication was predicted with an accuracy of 97.5% (95% CI [96.8, 98.2]); mean sensitivity = .98; specificity = .97; positive predictive value = .97; and negative predictive value = .98. Table 1 demonstrates the confusion metrics for each timepoint comparison.

Discussion

In this laboratory study, we built a model that correctly classifies voice spectrographic segments during states of alcohol intoxication from sober states with an accuracy of 98%. These findings are consistent with existing literature demonstrating the effect of alcohol intoxication on speech and voice (Bone et al., 2014; Fairbairn et al., 2015; Johnson et al., 1990; Pisoni & Martin, 1989). Our model outperformed the best-performing prior model using the only other known voice recording alcohol corpus we are aware of (i.e., German-language corpus from the Interspeech 2011 Speaker State Intoxication Sub-challenge [Schiel et al., 2012], which had an accuracy of 70% [Bone et al., 2014]). Our improved accuracy may be attributable to several potential causes. We used a standard set of tongue twisters given that they elicit diadochokinesis, or antagonistic syllable successions, which may function as a "phonetic stress test" for speech production and increase the sensitivity of models. Second, we used a standard set of tongue twisters as opposed to free speech samples, which reduces the variability between individuals and timepoints. Third, we used procedures to extract spectral or frequency-based features; therefore, our findings relate more to voice characteristics such as frequency and pitch as opposed to time-based features relating to phonemes and prosody, which may differ greatly between individuals.

Our study is limited in several key ways. Despite the inclusion of both men and women of varied ages and underlying alcohol use patterns, we studied a small sample of adults and thus were not able to externally validate our models. As well, we did not study non-White races or Hispanic ethnicity, which have been shown to be associated with different speech characteristics, including pitch (Andrianopoulos et al., 2001). We did not identify specific features of speech that have been shown to be sensitive to alcohol, such as speech volume (Fairbairn et al., 2015), errors (Pisoni & Martin, 1989), or deaffrication (Zihlmann, 2017). As well, speech samples collected in the real world may be shorter and not as sensitive to changes in cognitive or motor processes affected by alcohol. We also were not able to control for practice effects of completing tongue twisters but attempted to minimize this by using different tongue twisters for each timepoint and randomizing the order of phrases. We did not have voice samples when the BrAC was elevated above zero but less than .08 mg/dl; therefore, we cannot comment on whether voice signatures would be useful to detect lower risk drinking events. Finally, we were missing voice recordings of two participants, which could have influenced our models in unknown ways.

Several factors may impede real-world use of speech and voice signatures to identify alcohol intoxication states. First, environmental (i.e., background) noises and competing voices that exist in the real world, especially around drinking events, may require advanced pre-processing and filtering to be useful. Second, other factors that impair or alter speech, such as sleepiness or other substances (e.g., coffee, drugs), are likely to affect model accuracy in unpredictable ways. Third, it remains unknown whether individuals would perceive programs that process speech samples as intrusive; therefore, we do not know whether it would be an acceptable method to use in the real world. It is likely that optimized and acceptable programs would use other, less intrusive sensor-based features related to psychomotor function as a first-stage (i.e., highly sensitive) screen, and thus trigger voice analyses only as a second-stage (i.e., highly specific) validation.

We believe next steps to advance the science of using voice as a remote biomarker for acute alcohol intoxication should consider the following. There needs to be the funding and coordination of effort to create a large repository of high-quality audio samples across a range of adult ages, racial and ethnic subgroups, and blood or breath alcohol concentrations. There also need to be tools to allow researchers to easily and reliably extract phonetic and para-linguistic characteristics of speech from samples. There should also be serious consideration on relationship building with companies that already collect speech samples from smart speakers (e.g., Amazon via Alexa) to test models using real-world data. Finally, representative stakeholder groups should be queried on how they could see this technology being useful and acceptable. For example, if a smart speaker could identify alcohol intoxication, is it acceptable to prevent vehicle use?

In conclusion, we found in this proof-of-concept lab study that brief English speech samples are useful to classify alcohol-intoxicated states in adults. A much larger participant pool with more varied voice samples collected before and during the ascending and descending curves of alcohol intoxication is urgently needed to move the science of remote alcohol intoxication detection forward. As well, prospective studies testing algorithms using different sensor-based features related to psychomotor function are needed. Finally, studies to better understand the acceptability of different remote monitoring approaches are needed to ensure usability.

References

Agurto, C., Cecchi, G. A., Norel, R., Ostrand, R., Kirkpatrick, M., Baggott, M. J., Wardle, M. C., de Wit, H., & Bedi, G. (2020). Detection of acute 3,4-methylenedioxymethamphetamine (MDMA) effects across protocols using automated natural language processing. *Neuropsychopharmacology, 45,* 823–832. https://doi.org/10.1038/s41386-020-0620-4

Alpert, H. R., Slater, M. E., Yoon, Y.-H., Chen, C. M., Winstanley, N., & Esser, M. B. (2022). Alcohol consumption and 15 causes of fatal injuries: A systematic review and meta-analysis. *American Journal of Preventive Medicine, 63*(2), 286–300. https://doi.org/10.1016/j.amepre.2022.03.025

Andrianopoulos, A. V., Darrow, K. N., & Chen, J. (2001). Multimodal standardization of voice among four multicultural populations: Fundamental frequency and spectral characteristics. *Journal of Voice: Official Journal of the Voice Foundation, 15*(2), 194–219. https://doi.org/10.1016/S0892-1997(01)00021-2

Bae, S., Chung, T., Ferreira, D., Dey, A. K., & Suffoletto, B. (2018). Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. *Addictive Behaviors, 83,* 42–47. https://doi.org/10.1016/j.addbeh.2017.11.039

Baumeister, B., & Schiel, F. (2013, August). *Human perception of alcoholic intoxication in speech.* INTERSPEECH 2013, Lyons, France. https://www.isca-speech.org/archive/pdfs/interspeech_2013/baumeister13_interspeech.pdf

Bendtsen, M., McCambridge, J., Åsberg, K., & Bendtsen, P. (2021). Text messaging interventions for reducing alcohol consumption among risky drinkers: Systematic review and meta-analysis. *Addiction, 116*(5), 1021–1033. https://doi.org/10.1111/add.15294

Bone, D., Li, M., Black, M. P., & Narayanan, S. S. (2014). Intoxicated speech detection: A fusion framework with speaker-normalized hierarchical functionals and GMM supervectors. *Computer Speech & Language, 28*(2), 375–391. https://doi.org/10.1016/j.csl.2012.09.004

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication, 71,* 10–49. https://doi.org/10.1016/j.specom.2015.03.004

Dubnov, S. (2004). Generalization of spectral flatness measure for non-Gaussian linear processes. *IEEE Signal Processing Letters, 11*(8), 698–701. https://doi.org/10.1109/LSP.2004.831663

Fairbairn, C. E., Sayette, M. A., Amole, M. C., Dimoff, J. D., Cohn, J. F., & Girard, J. M. (2015). Speech volume indexes sex differences in the social-emotional effects of alcohol. *Experimental and Clinical Psychopharmacology, 23*(4), 255–264. https://doi.org/10.1037/pha0000021

Fillmore, M. T., Dixon, M. J., & Schweizer, T. A. (2000). Differential effects of alcohol on responses to negatively and positively primed stimuli. *Journal of Studies on Alcohol, 61*(6), 872–880. https://doi.org/10.15288/jsa.2000.61.872

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition, 103*(3), 386–412. https://doi.org/10.1016/j.cognition.2006.05.010

Hollien, H., DeJong, G., Martin, C. A., Schwartz, R., & Liljegren, K. (2001). Effects of ethanol intoxication on speech suprasegmentals. *Journal of the Acoustical Society of America, 110*(6), 3198–3206. https://doi.org/10.1121/1.1413751

Jiang, D.-N., Lu, L., Zhang, H.-J., Tao, J.-H., & Cai, L.-H. (2002). Music type classification by spectral contrast feature. *Proceedings. IEEE International Conference on Multimedia and Expo,* Lausanne, Switzerland, 113–116, vol.1. https://doi.org/10.1109/ICME.2002.1035731

Johnson, K., Pisoni, D. B., & Bernacki, R. H. (1990). Do voice recordings reveal whether a person is intoxicated? A case study. *Phonetica, 47*(3–4), 215–237. https://doi.org/10.1159/000261863

Kaner, E. F. S., Beyer, F. R., Garnett, C., Crane, D., Brown, J., Muirhead, C., Redmore, J., O'Donnell, A., Newham, J. J., de Vocht, F., Hickman, M., Brown, H., Maniatopoulos, G., & Michie, S. (2017). Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. *Cochrane Database of Systematic Reviews, Issue 9,* Article No. CD011479. https://doi.org/10.1002/14651858.CD011479.pub2

Kember, H., Connaghan, K., & Patel, R. (2017). Inducing speech errors in dysarthria using tongue twisters. *International Journal of Language & Communication Disorders, 52*(4), 469–478. https://doi.org/10.1111/1460-6984.12285

Klapuri, A. P., Eronen, A. J., & Astola, J. T. (2006). Analysis of the meter of acoustic musical signals. *IEEE Transactions on Audio, Speech, and Language Processing, 14*(1), 342–355. https://doi.org/10.1109/TSA.2005.854090

Liskow, B., Campbell, J., Nickel, E. J., & Powell, B. J. (1995). Validity of the CAGE questionnaire in screening for alcohol dependence in a walk-in (triage) clinic. *Journal of Studies on Alcohol, 56*(3), 277–281. https://doi.org/10.15288/jsa.1995.56.277

Muda, L., Begam, M., & Elamvazuthi, I. (2010). Voice recognition algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW). *Journal of Computing, 2.* https://doi.org/10.48550/arXiv.1003.4083

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2017). Just-in-Time Adaptive Interventions

(JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine, 52*(6), 446–462. https://doi.org/10.1007/s12160-016-9830-8

Naimi, T. S., Xuan, Z., Sarda, V., Hadland, S. E., Lira, M. C., Swahn, M. H., Voas, R. B., & Heeren, T. C. (2018). Association of state alcohol policies with alcohol-related motor vehicle crash fatalities among US adults. *JAMA Internal Medicine, 178*(7), 894–901. https://doi.org/10.1001/jamainternmed.2018.1406

National Institute on Alcohol Abuse and Alcoholism. (n.d.). *NIAAA Guidance for Conducting Alcohol Administration Studies with Human Participants, 2023.* https://www.niaaa.nih.gov/research/guidelines-and-resources/guidance_conducting_alcohol_administration_studies_human_participants

Norman, T., Peacock, A., Ferguson, S. G., Kuntsche, E., & Bruno, R. (2021). Combining transdermal and breath alcohol assessments, real-time drink logs and retrospective self-reports to measure alcohol consumption and intoxication across a multi-day music festival. *Drug and Alcohol Review, 40*(7), 1112–1121. https://doi.org/10.1111/dar.13215

NPR. (2022, June 16). *NPR & Edison Research: Smart speaker ownership reaches 35% of Americans.* https://www.npr.org/about-npr/1105579648/npr-edison-research-smart-speaker-ownership-reaches-35-of-americans

Peacock, A., Cash, C., & Bruno, R. (2015). Cognitive impairment following consumption of alcohol with and without energy drinks. *Alcoholism: Clinical and Experimental Research, 39*(4), 733–742. https://doi.org/10.1111/acer.12680

Pew Research Center. (2021, April 7). *Mobile Fact Sheet. Internet, Science & Tech.* https://www.pewresearch.org/internet/fact-sheet/mobile/

Piasecki, T. M. (2019). Assessment of alcohol use in the natural environment. *Alcoholism: Clinical and Experimental Research, 43*(4), 564–577. https://doi.org/10.1111/acer.13975

Pisoni, D. B., & Martin, C. S. (1989). Effects of alcohol on the acoustic-phonetic properties of speech: Perceptual and acoustic analyses. *Alcoholism: Clinical and Experimental Research, 13*(4), 577–587. https://doi.org/10.1111/j.1530-0277.1989.tb00381.x

Russell, M. A., Turrisi, R. J., & Smyth, J. M. (2022). Transdermal sensor features correlate with ecological momentary assessment drinking reports and predict alcohol-related consequences in young adults' natural settings. *Alcoholism: Clinical and Experimental Research, 46*(1), 100–113. https://doi.org/10.1111/acer.14739

Rusz, J., Cmejla, R., Ruzickova, H., & Ruzicka, E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *Journal of the Acoustical Society of America, 129*(1), 350–367. https://doi.org/10.1121/1.3514381

Sacks, J. J., Gonzales, K. R., Bouchery, E. E., Tomedi, L. E., & Brewer, R. D. (2015). 2010 national and state costs of excessive alcohol consumption. *American Journal of Preventive Medicine, 49*(5), e73–e79. https://doi.org/10.1016/j.amepre.2015.05.031

Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption—II. *Addiction, 88*(6), 791–804. https://doi.org/10.1111/j.1360-0443.1993.tb02093.x

Schiel, F., Heinrich, C., & Barfüsser, S. (2012). Alcohol language corpus: The first public corpus of alcoholized German speech. *Language Resources and Evaluation, 46,* 503–521. https://doi.org/10.1007/s10579-011-9139-y

Suffoletto, B., Dasgupta, P., Uymatiao, R., Huber, J., Flickinger, K., & Sejdic, E. (2020). A preliminary study using smartphone accelerometers to sense gait impairments due to alcohol intoxication. *Journal of Studies on Alcohol and Drugs, 81*(4), 505–510. https://doi.org/10.15288/jsad.2020.81.505

Suffoletto, B., Gharani, P., Chung, T., & Karimi, H. (2018). Using phone sensors and an artificial neural network to detect gait changes during drinking episodes in the natural environment. *Gait & Posture, 60,* 116–121. https://doi.org/10.1016/j.gaitpost.2017.11.019

Venegas, A., Donato, S., Meredith, L. R., & Ray, L. A. (2021). Understanding low treatment seeking rates for alcohol use disorder: A narrative review of the literature and opportunities for improvement. *American Journal of Drug and Alcohol Abuse, 47*(6), 664–679. https://doi.org/10.1080/00952990.2021.1969658

Wray, T. B., Merrill, J. E., & Monti, P. M. (2014). Using ecological momentary assessment (EMA) to assess situation-level predictors of alcohol use and alcohol-related consequences. *Alcohol Research: Current Reviews, 36,* 19–27.

Yu, Y., You, Y., Yang, Z., Liu, G., Li, P., Yang, Z., & Shan, W. (2021). *Spectral roll-off points variations: Exploring useful information in feature maps by its variations* (arXiv:2102.00369). http://arxiv.org/abs/2102.00369

Zihlmann, U. (2017). The effects of real and placebo alcohol on deaffrication. *Interspeech,* 3882–3886. https://doi.org/10.21437/Interspeech.2017-1579