AUGUST 03 2023

Neural networks’ posterior probability as measure of effects of alcohol on speech FREE

Ratree Wayland ; Kevin Tang; Fenqi Wang ; Sophia Vellozzi ; Rahul Sengupta

Proc. Mtgs. Acoust 51, 060001 (2023)
https://doi.org/10.1121/2.0001764

ASA
LEARN MORE
Advance your science and career as a member of the
Acoustical Society of America

Proceedings of Meetings on Acoustics

Volume 51 http://acousticalsociety.org/

184th Meeting of the Acoustical Society of America

Chicago, Illinois

8-12 May 2023

Speech Communication: Paper 4aSC22

Neural networks' posterior probability as measure of effects of alcohol on speech

Ratree Wayland
Department of Linguistics, University of Florida, Gainesville, FL, 32611-5454, USA; ratree@ufl.edu

Kevin Tang
Department of English Languages and Linguistics, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, GERMANY; Kevin.Tang@hhu.ed

Fenqi Wang
Department of Linguistics, University of Florida, Gainesville, FL, 32611-5454, USA; fenqi@ufl.edu

Sophia Vellozzi and Rahul Sengupta
Department of Computer & Information Science & Engineering, University of Florida, Gainesville, FL, USA; s.vellozzi@ufl.edu; rahulseng@ufl.edu

Alcohol, a progressive central nervous system depressant, has been found to negatively affect not only cognitive functions but also the production of speech—a complex motor activity requiring a high degree of coordination. In this study, we estimate the degrees of deaffrication, spirantization, and retracted place of articulation for /t/, /d/, /s/, /ʃ /, /tʃ /, and /ʤ/ in a corpus of speech affected by alcohol. These estimations are based on posterior probabilities calculated by recurrent neural networks known as Phonet, which are trained to recognize anterior, continuant, and strident phonological features. The results obtained revealed both categorical and gradient errors in intoxicated speech, indicating the reliability of Phonet in quantifying fine-grained errors.

Published by the Acoustical Society of America

https://doi.org/10.1121/2.0001764
Proceedings of Meetings on Acoustics, Vol. 51, 060001 (2023)

1. INTRODUCTION

Alcohol, a progressive central nervous system depressant, has been been shown to have detrimental effects not only on cognitive functions, but also on the production of speech. Speech is a complex motor activity that requires a high degree of coordination. According to Johnson, Pisoni and Bernacki (1990),[8] the complexity of speech can be attributed to precise intergestural coordination and fine motor control involved in moving the articulators to target positions for different speech sounds. For instance, the relative timing of gestures performed by the vocal folds and the tip of the tongue, known as voice onset time (VOT), differentiates between /t/ and /d/ in English. In word onset position, the vocal fold vibration for /d/ starts slightly before or approximately simultaneously with the release of oral stop closure. Even a slight mistiming of these two gestures could result in a perceptually different consonant. In addition, the position of the tongue relative to the front teeth and the length of the constriction at the roof of the mouth distinguish /s/ from /ʃ/.[18] An /s/ is produced when the tongue tip is located near the front teeth, and the constriction at the roof of the mouth is relatively short (2.5 cm). However, the resulting sound will resemble /ʃ/ more if the constriction is slightly longer or wider or if the tongue tip is positioned slightly further back in the mouth.

In English, segmental errors observed in intoxicated speech include deaffrication (more stop-like production) of the sounds /tʃ/ and /dʒ/, spirantization (fricative-like production) of stop consonants, and palatalization (place retraction) of the alveolar fricative /s/ (produced as /ʃ/). For example, Pisoni and Martin (1989)[15] found that English speakers were unable to achieve a complete closure either before affricates or stops when intoxicated, with more pronounced and consistent effects observed for the voiced compared to the voiceless affricate. The alcohol-induced palatalization of /s/ in native speakers of English was first reported by Lester and Skousen (1974).[12] Cases of [s] as [ʃ] productions in intoxicated speech were mentioned by Chin & Pisoni (1997).[1] In 1991, Tanford and colleagues[19] reported a palatalized [s] pronunciation of the word Exxon by Captain Hazelwood, the commander of the oil tanker Exxon Valdez involved in the catastrophic oil spill in Prince William Sound, indicating possible alcohol impairment at the time of the accident. Finally, Johnson et al. (1993)[7] investigated the phonetic contexts in which [s] becomes [ʃ] and found that palatalization of [s] occurred only when it was followed by a voiceless affricate (e.g., as in *postulate*, *posturing* and *postulatable*).

However, previous studies have primarily focused on categorical errors in intoxicated speech, using measures such as perceptual judgment and phonetic transcription. These methods may be prone to perceptual bias, potentially leading to the oversight or imperceptibility of sub-contrastive or gradient errors occurring at a more subtle level beneath individual segments or features.[3] While the acoustic measurements have the potential to overcome perceptual biases, most studies have predominantly reported categorical (i.e., [s] is mispronounced as [ʃ]) rather than examining gradient phonetic errors (i.e., degrees of [ʃ]-like pronunciation) in intoxicated speech.

The goal of this study is to examine both gradient and categorical errors in a corpus of English intoxicated speech, using a neural network model known as Phonet. This approach draws inspiration from computational methods that utilize forced alignment to measure surface-level gradient phonetic variation. The Phonet model quantifies the gradient phonetic variation of deaffrication, spirantization, and place retraction by analyzing the posterior probability of relevant phonological features. These features, including [anterior], [continuant], and [strident], capture information about the relative location of the oral constriction, continuity of oral airflow, and intensity of high-frequency frication noise, respectively. In this study, a categorical error is defined as a sign shift of the phonological features, such as a transition from [+anterior] to [-anterior]. On the other hand, the gradiency of an error is reflected in the posterior probability values associated with a specific phonological feature.

A. PHONOLOGICAL FEATURES AND ALCOHOL-IMPAIRED SPEECH ERRORS

Phonemes are classified into classes based on their shared phonetic features. (See e.g., 6 for guides to phonological features). One common distinction is between [+consonantal] and [-consonantal] phonemes. Consonantal phonemes, such as stops, fricatives, affricates, nasals, and liquids, involve constriction of the articulators in the vocal tract and are classified as [+consonantal]. On the other hand, vowel and glide phonemes are typically classified as [-consonantal] because they do not involve the same level of constriction. Another important phonological feature is [syllabic]. [+syllabic] phonemes are the most sonorous segments and typically occupy the nucleus position of a syllable. Vowels and syllabic consonants /ɹ̩, l̩, n̩, m̩/ etc., are classified as [+syllabic] while other consonants including glides are classified as [-syllabic]. However, in the context of speech errors observed in intoxicated speech, the three relevant phonological features are [anterior], [strident], and [continuant]. These features capture specific aspects of the articulatory characteristics involved in the production of certain sounds during intoxication. The classification of phonemes based on these features allows for the analysis and understanding of the speech errors that occur under the influence of alcohol.

The [anterior] feature describes relative location of the tongue in the vocal tract. [+anterior] consonants are produced with an oral constriction near the front of the mouth, typically before the alveolar ridge. This category includes consonants produced with the involvement of the lips (labial consonants), the teeth (dental consonants), and the alveolar ridge (alveolar consonants). In contrast, [-anterior] consonants are produced with an oral obstruction occurring behind the the alveolar ridge.[4,5,9]

The [strident] feature characterizes consonants produced with high-frequency turbulent airflow. All affricates and fricatives produced at the labiodental, alveolar, palato-alveolar, retroflex, and uvular places of articulation are [+strident]. On the other hand, consonants that are classified as [-strident] are those that do not exhibit high-frequency frication noise such as stops, nasals, liquids and non-sibilant fricatives [f, v, θ, ð].

The [continuant] feature describes partial occlusion of the air passage. [+continuant] phonemes are produced with incomplete closure between articulators allowing continuous oral airflow. Fricatives, liquids, glides, and vowels are [+continuant] while stops and affricates are [-continuant]. Nasals are considered [-continuant] by some (because of airflow blockage through the oral cavity), but [+continuant] by others (because of continuous airflow through the nasal cavity). In this study, we specified them as [-continuant], treating them as consonants with a complete occlusion of the oral airflow.

B. PHONET

Phonet[22] is a bi-directional recurrent neural network model that is designed to recognize and classify phonemes into different phonological classes based on their phonological features. The model is trained using a segmentally-aligned acoustic corpus, obtained through forced alignment techniques. The input to Phonet consists of log-energy values distributed across triangular Mel filters. These values are computed from 25-ms windowed frames of each 0.5-second chunk of the input signal. By analyzing these acoustic features, the model learns to predict the posterior probabilities of relevant phonological features for the target segments (see 22 for details). Weighted categorical cross-entropy loss function was used. The weight factors for each class are based on the percentage of samples from the training set, that belong to each class. Adam optimizer[11] was used to train the model. Dropout and batch normalization layers were used to improve the generalization of the networks. The training lasted 81 epochs, with early stopping enabled (with a patience of 15 epochs). For more detail about the model and related procedures, see 22 and the publicly-available code at https://github.com/jcvasquezc/phonet. Once trained, posterior probabilities for relevant phonological features of the target segments can be computed by the model. Phonet has been found to be highly accurate in estimating degree of lenition in Spanish[20,21,23,24] and modelling the speech impairments of patients diagnosed with Parkinson's disease.[22]

2. THIS STUDY

This study quantifies the degrees of deaffrication, spirantization and place retraction of the affricates /tʃ/, /dʒ/; the stops /t/, /d/ and the fricatives /s/, /ʃ/ in a corpus of intoxicated English speech, using posterior probabilities of the [anterior], [strident] and [continuant] phonological features computed by Phonet.

A. METHODS

I. Materials

The target consonants for this study are English stops /t, d/ (Ns = 2,042 and 965), affricates /tʃ, dʒ/ (Ns = 129 and 145), and fricatives /s, ʃ/ (Ns = 1,308 and 84) from a corpus of intoxicated speech.[20]

II. Stimulus Recording Procedure

The corpus comprises recordings of four female native speakers of British English reading a naturally spoken dialogue, without using an animated or acting voice. The original text of the dialogue (based on 17), was edited to ensure gender and emotional neutrality, absence of overly long turns, and representation of the English phonemic inventory (available at 20). Two recordings, one in a sober state and one in a drunk state, were obtained from each participant on different days, with a time interval of 1-2 months. The recordings were conducted in a sound-attenuated room at a sampling rate of 44.1 kHz and 16-bit amplitude resolution in stereo, which were later converted to mono using Audacity. Participants were instructed not to eat, drink, or use mouthwash for two hours before each session and to refrain from smoking for at least half an hour before each recording session. To ensure absence of alcohol in their system, the participants' blood alcohol concentration (BAC) was measured using a breathalyzer [AlcoMate (Macomb Township. MI) Premium AL-7000] at the beginning of the sober session. The intoxicated recording session commenced once the speaker's BAC reached 0.12% after consumption of vodka or rum, mixed with juices.

III. Stimulus Pre-processing

The recordings were segmented into individual utterances and these individual utterances were then manually annotated to identify any disfluencies. Utterances containing disfluencies, accounting for 8.5% of the data, were excluded from further analysis. The remaining disfluency-free utterances were aligned using the Montreal Forced Aligner (version: 2.0),[13] utilizing the pretrained English model provided by the aligner.

IV. Phonet Training Procedure

A subset of the cleaned portion of 360 hours of Librispeech,[14] a large corpus of English audiobooks, was selected and used as a representative English speech sample. The corpus was then force-aligned using the Montreal Forced Aligner (version: 2.0).[13] The phone set was set to IPA. Other parameters were set at their default values. Model training with Phonet was performed on an NVIDIA GeForce RTX 3090 GPU using the Keras[2] library. The corpus was randomly split into a train subset (80%) and a test subset (20%) using the Python (Version 3.9) scikit-learn library [14]. Twenty-one Phonet models were trained for 20 phonological classes (consonantal, syllabic, voicing, labial, coronal, dorsal, lateral, nasal, rhotic, anterior, continuant, sonorant, strident, diphthong, high, low, back, round, stress, tense), and pause.

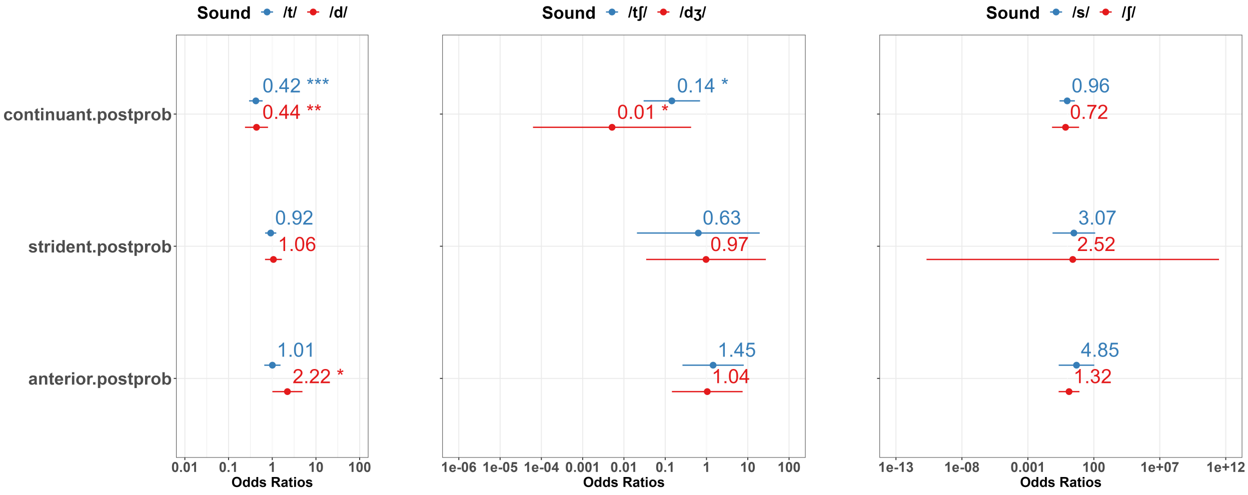
Only the test data was used in the internal evaluation of the posterior probabilities generated by the Phonet model. The model exhibited high accuracy as indicated by the unweighted average recall (UAR) ranging from from 91% (coronal) to 98% (pause). Importantly, the UARs for the anterior, continuant and strident features were found to be 93%, 92% and 97%, respectively. Subsequently, the trained model was applied to word tokens from our force-aligned intoxicated speech corpus, specifically targeting /t, d, tʃ, dʒ, s, ʃ/. The predictions were computed for 10-ms frames. For a token containing multiple frames, the average prediction from the middle frame(s) was taken as its prediction. The obtained anterior, continuant, and strident posterior probabilities for each target consonant were then employed for statistical analyses.

V. Statistical Analyses

The statistical analyses were performed using the `lme4` package in R.[16] Binary categorical variables were contrasted coded using (-0.5, 0.5). Speaker and word were included as random variables in the analyses. Two complementary analyses were performed. First, to examine if the posterior probabilities could predict the drinking status, a binary logistic regression analysis was performed with the three posterior probabilities (anterior, continuant and strident) as predictors and drinking status (sober and intoxicated) as the dependent variable, using the glmer function. A contrastive or categorical error was hypothesized if a feature emerged as a significant predictor. Second, to assess the gradiency of an error, a linear regression model (lmer) was performed, with drinking status as predictor and the three posterior probability values as dependent variables. An increase or decrease in the posterior probability of a feature would indicate the degree of error gradiency. In both analyses, the “drunk” status was set as the reference level. Based on previous studies,[8, 12] it was expected that the intoxicated condition would exhibit higher continuant and strident posterior probabilities for /t, d/ and /tʃ, dʒ/, while a lower anterior probability was expected for the fricative /s/ compared to the sober condition .

B. RESULTS

Figure 1a, 1b and 1c present the results of the binary logistic regressions with anterior, continuant and strident posterior probabilities as predictors and drinking status as the categorical, binary dependent variable.



(a) *(b)* *(c)*

Figure 1: Coefficients of the linear regression models for target sounds. An odds ratio $>$ 0.5 indicates an increased likelihood of being in a sober state, while an odds ratio $<$ 0.5 indicates a decreased likelihood of being in a sober state. *, **, *** indicate significant effects at p $>$ .05, .01 and .001, respectively.

From Figure 1, the continuant posterior probability is the only significant predictor of drinking status for

/t/ (Figure 1a), /tʃ/ and /dʒ/ (Figure 1b). The odd ratios suggest that as the continuant posterior probability increases, the likelihood of the speakers being sober decreases (odds ratios $<$0.5). For /d/, both continuant and anterior probabilities are the significant predictors. Similar to /t/, /tʃ/ and /dʒ/, for /d/, the likelihood of the speakers being sober decreases as the continuant posterior probability increases. However, as the anterior probability of /d/ increases, the likelihood of the sober status increases (odds ratios $>$0.5). No significant predictor was found for /s/ [odds ratios = 0.96-4.85, p$>$.05] nor /ʃ/ [odds ratios = 0.72-2.52, p$>$0.05] (Figure 1c). These results suggested that, under intoxication, categorical errors (i.e., [-continuant] $>$ [+continuant]) occurred for /t/, /d/, /tʃ/ and /dʒ/. For /d/, a categorical shift from [+anterior] $>$ [-anterior] also occurred. On the other hand, no categorical error was detected for /s/ or /ʃ/.

Tables 1, 2 and 3 summarize the results of the second analyses, the linear mixed-effect regression models with drinking status as predictors (reference level = drunk) and posterior probabilities of the three phonological features as the dependent variables.

Table 1: Summary of the linear regression models for anterior probability.

Predictor	Consonant	β	*P* value
Anterior	**/t/**	**0.03**	**0.003**
	/d/	**0.05**	**<0.001**
	/ʧ/	-0.00	0.960
	/ʤ/	-0.02	0.639
	/s/	**0.01**	**0.012**
	/ʃ/	0.02	0.709

From Table 1, we observe that the sober speech is predicted to have a significantly higher anterior probability for /t/, /d/ and /s/ [βs= 0.03, 0.05, 0.01; ps=0.003,$<$.001, =0.012.] compared to the drunken speech. However, there is no significant change in anteriority between the two speech conditions for /tʃ/, /dʒ/, and /ʃ/ [βs =-0.00, -0.02, -0.02; ps= $>$.05]. Although not statistically significant, the β values suggested that sober /tʃ/ and sober /dʒ/ are less anterior than drunk /tʃ/ and drunk /dʒ/ while sober /ʃ/ is more anterior than drunk /ʃ/. Overall, these results indicate that a) tongue tip location for /t/, /d/ and /s/ is significantly more front when sober than when intoxicated, and b) the shift in place of articulation is significantly greater for /t/, /d/ and /s/ than for /tʃ/, /dʒ/, and /ʃ/ under intoxication.

Table 2: Summary of the linear regression models for continuant probability.

Predictor	Consonant	β	*P* value
Continuant	**/t/**	**-0.10**	**<0.001**
	/d/	**-0.06**	**<0.001**
	/ʧ/	**-0.13**	**0.004**
	/ʤ/	**-0.07**	**0.004**
	/s/	0.01	0.152
	/ʃ/	-0.02	0.648

For the continuant probability (Table 2), significantly lower values are predicted for /t/, /d/, /tʃ/ and /dʒ/ [βs=-0.10, -0.06, -0.13, -0.07; ps$<$.01], indicating reduced continuancy under the sober condition compared to the drunk condition. However, no significant differences are observed for /s/ [β = 0.01, p=0.152] or /ʃ/ [β = -0.02, p = 0.648]. Although not statistically significant, the positive β value for /s/ suggests increased continuancy when sober, while the negative β for /ʃ/ suggests decreased continuancy under the sober condition. These results suggest that the oral constriction for the stops /t, d/, as well as the affricates /tʃ/ and /dʒ/, becomes significantly less complete under intoxication. In contrast, the size of the oral constriction remains unchanged for the /s/ and /ʃ/ when the speakers become intoxicated.

Table 3: Summary of the linear regression models for strident probability.

Predictor	Consonant	β	*P* value
Strident	**/t/**	**-0.06**	**<0.001**
	/d/	-0.02	0.266
	/ʧ/	-0.03	0.223
	/ʤ/	0.01	0.816
	/s/	**0.01**	**0.018**
	/ʃ/	0.00	0.781

Finally, Table 3 presents the results for the strident probability. It shows that the strident probability for /t/ is significantly lower under the sober condition, while the opposite is true for /s/. No significant difference [p>.05] is observed for the remaining consonants. Although not statistically significant, the β values for /d/ [-0.02] and /tʃ/ [-0.03] are negative, indicating reduced stridency in their production under the sober condition compared to when drunk. In contrast, the β value for /dʒ/ [0.01] is positive, indicating increased stridency, while the β value for /ʃ/ [0.00] is equal to 0, indicating no change in stridency for /ʃ/ when sober. These results suggest that /t/ is significantly less strident (produced with less turbulent noise) under the sober condition. On the other hand, drunk /s/ is less strident than its sober version. However, minimal change in the degree of stridency is observed for /d/, /tʃ/ and /dʒ/ and no change is predicted for /ʃ/.

3. DISCUSSION

In this study, a new computational approach, Phonet, was applied to a corpus of intoxicated English speech to detect categorical and gradient alcohol-induced speech error. The target consonants are stops /t, d/, affricates /tʃ, dʒ/ and fricatives /s, ʃ/. The error types examined are deaffrication, spirantization and retracted place of articulation. The gradient nature of these errors are estimated from posterior probabilities of three phonological features, [anterior], [continuant], [strident] computed by Phonet.

The results of the binary logistic regression models indicated that [continuant] was a significant predictor of drinking state for /t/, /tʃ/ and /dʒ/ while both [continuant] and [anterior] emerged as significant predictors for /d/. These findings suggested that the size of the oral constriction for these four consonants significantly widens from a sober to a drunk state. If a sign shift of a binary feature is responsible for its significant predictive power, then, these results could be interpreted to suggest that categorical errors ([-continuant] > [+continuant]) occur for /t/, /d/, /tʃ/ and /dʒ/ under intoxication. Following the same line of reasoning, a categorical shift from [+anterior] > [-anterior] also occurs for /d/ in the drunken state, indicating that,in addition to a significant degree of oral constriction widening, a simultaneous and significant amount of place retraction takes place for this consonant. The fact that /tʃ/ and /dʒ/ are [-anterior] may explain why they do not undergo further place retraction in the drunken state. Neutralization (loss of contrastivity) in anteriority between /t/ and /tʃ/ could account for why /t/ does not undergo place retraction, as both would become [-anterior] if /t/ were to retract. These results imply that articulatory planning may remain intact, but fine-grained motor control is partially lost under intoxication.

It was hypothesized that categorical shift in place of articulation would occur at least for /s/ (i.e., [+anterior] > [-anterior]).[8,12] However, no categorical error was detected for either /s/ or /ʃ/, at least not at the tested BAC level. It is possible that this error only emerges at a higher BAC level. Additionally, it is also worth noting that /s/ and /ʃ/ are both [+continuant, + strident]. It is possible that these shared and redundant features "add additional motoric instructions to enhance the saliency of the jeopardized features",[10] namely [anterior] in this case. However, the fact that this error has been previously attested suggests a limit to this enhancement effect. Further studies are needed to shed light on the relationship between error type and intoxication level.

Gradient errors are examined in the linear regression analyses. It was found that a shift in place of articulation was significantly greater for /t/, /d/ and /s/ than for /tʃ/, /dʒ/, and /ʃ/ under intoxication. These results suggest that the degree of place retraction in intoxicated speech may be constrained by the existing place feature of the affected consonants. Consonants with a [+anterior] feature exhibit a greater degree of place shift than consonants with a [-anterior] feature. A similar constraint also seems to apply to the [continuant] feature, resulting in gradiency along this dimension. Specifically, the posterior probability of continuant significantly increased for the [-continuant] consonants, /t/, /d/, /tʃ/, and /dʒ/, but not for the [+continuant] consonants, /s/ and /ʃ/. A higher degree of continuance (greater oral aperture) would lead to a reduction in the intensity of frication noise (i.e., stridency).

Furthermore, gradient errors in stridency were also found. Sober /t/ was found to be significantly less strident than drunk /t/. In contrast, the change in stridency was relatively small for /d/, /tʃ/ and /dʒ/. Additionally, drunk /s/ was significantly less strident than its sober counterpart, indicating a further widening of the oral aperture leading to a loss in stridency. However, no significant change in stridency was observed for /ʃ/. This result suggests that while the oral aperture could further widen for [-anterior, +continuant, +strident], /s/, it did not occur for [+anterior, +continuant, +strident], /ʃ/. Whether the oral constriction could be further widened for /ʃ/ at a higher level of intoxication remains to be further investigated.

4. CONCLUSION

Phonet successfully identified both categorical and gradient errors in intoxicated speech, demonstrating its reliability in quantify fine-grained errors. Nonetheless, our findings need to be confirmed with more subjects (male and female), and can be extended to languages with different contrastive phonological features,[20] and compared with speech by clinical populations, such as Parkinson's disease.

ACKNOWLEDGMENTS

This research was funded by NSF-National Science Foundation (SenSE). Award No. 2037266 – SenSE.

REFERENCES

[1] Steven Chin and David Pisoni, *Alcohol and speech*, Brill, 1997.

[2] F Chollet, *Keras.[online] available at: https://github. com/fchollet/keras*, Accessed **12** (2015), no. 01, 2021.

[3] Stefan A Frisch and Richard Wright, *The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue*, Journal of Phonetics **30** (2002), no. 2, 139–162.

[4] Tracy Alan Hall, *The phonology of coronals*, The Phonology of Coronals (1997), 1–186.

[5] ______, *Segmental features*, The Cambridge handbook of phonology **1118** (2007), 311–334.

[6] Bruce Hayes, *Introductory phonology*, vol. 7, John Wiley & Sons, 2008.

[7] K Johnson, MH Southwood, AM Schmidt, CM Mouli, AT Holmes, AA Armstrong, and AS WIlson, *A physiological study of the effects of alcohol on speech and voice*, 22nd annual Symposium on the Care of the Professional Voice at the Voice Foundation, 1993.

[8] Keith Johnson, David B Pisoni, and Robert H Bernacki, *Do voice recordings reveal whether a person is intoxicated? a case study*, Phonetica **47** (1990), no. 3-4, 215–237.

[9] Patricia A Keating, *Coronal places of articulation*, The special status of coronals: Internal and external evidence, Elsevier, 1991, pp. 29–48.

[10] Samuel Jay Keyser and Kenneth Noble Stevens, *Enhancement and overlap in the speech chain*, Language (2006), 33–63.

[11] Diederik P Kingma and Jimmy Ba, *Adam: A method for stochastic optimization*, 2014.

[12] Leland Lester and Royal Skousen, *The phonology of drunkenness*, Papers from the parasession on natural phonology (1974), 233–239.

[13] Michael McAuliffe, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger, *Montreal Forced Aligner: Trainable text-speech alignment using Kaldi.*, Interspeech, vol. 2017, 2017, pp. 498–502.

[14] Vassil Panayotov, Guoguo Chen, Daniel Povey, and Sanjeev Khudanpur, *Librispeech: an asr corpus based on public domain audio books*, 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, 2015, pp. 5206–5210.

[15] David B Pisoni and Christopher S Martin, *Effects of alcohol on the acoustic-phonetic properties of speech: perceptual and acoustic analyses*, Alcoholism: Clinical and Experimental Research **13** (1989), no. 4, 577–587.

[16] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022.

[17] Neil Simon, *The collected plays of neil simon. volume 2, a plume book*, 1986.

[18] Joanne D Subtelny, N Oya, and JD Subtelny, *Cineradiographic study of sibilants*, Folia Phoniatrica et Logopaedica **24** (1972), no. 1, 30–50.

[19] J Alexander Tanford, David B Pisoni, and Keith Johnson, *Novel scientific evidence of intoxication: Acoustic analysis of voice recordings from the exxon valdez*, The Journal of criminal law & criminology **82** (1991), no. 3, 579.

[20] Kevin Tang, Charles B Chang, Sam Green, Kai Xin Bao, Michael Hindley, Young Shin Kim, and Andrew Nevins, *Intoxication and pitch control in tonal and non-tonal language speakers*, JASA Express Letters **2** (2022), no. 6, 065202.

[21] Kevin Tang, Ratree Wayland, Fenqi Wang, Sophia Vellozzi, Rahul Sengupta, and Lori Altmann, *From sonority hierarchy to posterior probability as a measure of lenition: The case of Spanish stops*, The Journal of the Acoustical Society of America **153** (2023), no. 2, 1191–1203.

[22] Juan Camilo Vásquez-Correa, Philipp Klumpp, Juan Rafael Orozco-Arroyave, and Elmar Nöth, *Phonet: A tool based on gated recurrent neural networks to extract phonological posteriors from speech.*, Interspeech, 2019, pp. 549–553.

[23] Ratree Wayland, Kevin Tang, Fenqi Wang, Sophia Vellozzi, and Rahul Sengupta, *Quantitative acoustic versus deep learning metrics of lenition*, Languages **8** (2023), no. 2, 98.

[24] Ratree Wayland, Kevin Tang, Fenqi Wang, Sophia Vellozzi, Rahul Sengupta, and Lori Altman, *Lenition measures: Neural networks' posterior probability versus acoustic cues*, The Journal of the Acoustical Society of America **152** (2022), no. 4, A59–A59.