SIZE OF VELOPHARYNGEAL OPENING AND NASALITY MEASUREMENTS FROM ACOUSTIC FEATURES

Jahurul Islam* **and Bryan Gick†**
University of British Columbia, Vancouver, BC, Canada

1 Introduction

The study of nasality in speech has long been focused on developing methods to obtain accurate measures from acoustic signals [1, 2]. However, the effectiveness of these acoustic measures has been a subject of debate, primarily due to their limited generalizability across different speakers [3]. As a result, researchers have sought more direct measures of nasality, such as recordings of aerodynamic data in the form of nasal airflow [4, 5]. Nasal airflow (nasalance) data obtained through nasal airflow measurements is considered the gold standard due to its reliability for inter-speaker comparability. In recent years, there has been a growing interest in utilizing machine learning models to improve the prediction of nasality from acoustic features (NAF). Studies [6] have demonstrated that a principal component analysis (PCA)-based regression approach can yield remarkably similar results to nasalance data. Additionally, XGBoost learning algorithm has also been effectively used to obtain NAF from a wide range of acoustic correlates [7].

Parallel to these efforts, researchers have also explored techniques for measuring the actual size of the velopharyngeal opening (VPO) in nasal sounds. Recent studies [8] have investigated variations in the size of the VPO across nasal segments and languages, shedding light on the interplay between nasality and VPO dimensions.

Given the impressive predictive capabilities of NAF with respect to nasalance, it is worth exploring whether these acoustic measures can also be used to predict the actual size of the velopharyngeal opening (VPO) or the dynamic changes of VPO over time. Investigating the relationship between NAF and VPO can provide valuable insights into whether nasality in speech is more associated with the amount of air passing through the velopharyngeal port or with the size of the VPO. Understanding this relationship can contribute to a deeper understanding of the mechanisms underlying nasality in speech production.

2 Method

2.1 VPO data

For our investigation, we utilized sentence-level speech samples produced by four Canadian English speakers, obtained from the Université Laval X-ray videofluorography database [9]. To assess the size of the velopharyngeal opening (VPO), we extracted frames at a rate of 30 frames per second from the X-ray video files. These frames were processed using ImageJ software [10], where we counted the number of black pixels along a diagonal line approximating the movement of the velum. This pixel count served as a proxy for the size of the VPO. Further details regarding this method can be found in [8]. Figure 1 illustrates a sample frame that displays the size of the VPO during a nasal segment.

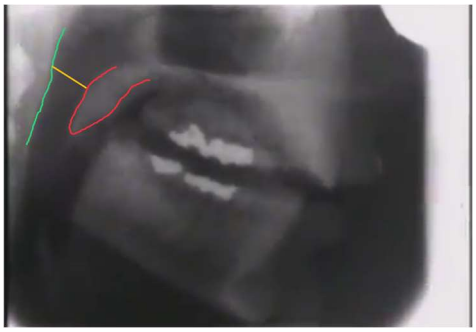

Figure 1: VPO measurement (red = velum outline, green = velopharyngeal wall, yellow = VPO)

2.2 NAF data

To obtain an acoustic measure of nasality, we followed the methods employed by Carignan and colleagues [7] and implemented a machine learning model based on the XGBoost algorithm. A total of 72 acoustic features were extracted from the audio signals. To measure 18 of these features, we used the Nasality Automeasure Praat [11] script developed by Will Styler [3]. These features included the frequencies, amplitudes, and bandwidths of F1-F3, P0 and P1 amplitude, P0 prominence, A1-P0 and A1-P1, along with their formant-compensated analogs. Additionally, A3-P0, H1-H2, and the first four spectral moments (center of gravity, variance, skew, and kurtosis) were measured. Furthermore, we extracted 14 Mel-frequency cepstral coefficients (MFCCs) using the *tuneR* R package. Lastly, delta coefficients were computed for all 36 features, resulting in a total of 72 features used for training the XGBoost model. These acoustic features were extracted at 11 evenly-spaced time points within a token. A token was included in the dataset if it belonged to one of the following phonological environments: CVC, CV#, VNC, CNV, CVN, VNV, VC#, VN#, NVN, or NVC. Out of the ten frames of measurements for each token, frames adjacent to a nasal segment were labeled as "nasal," while frames adjacent to an oral segment were labeled as "oral."

To construct the gradient-boosted decision tree models, we employed the R [12] package *XGBoost* (version 1.7.5.1). Following Carignan [6, 7], we trained the NAF model using oral and nasal observations, with oral labeled as 0 and nasal as 1, specifying the model to minimize linear regression error. The model generated predicted values on a scale of 0 to 1, indicating the degree of nasality.

* jahurul.islam741@gmail.com
† gick@mail.ubc.ca

3 Results

Figure 1 presents peak NAF and VPO values during nasal and oral segments. As the figure indicates, NAF values reveal similar patterns as the VPO values do; the NAF values for nasals are larger than the ones for oral; in fact, NAF values appear to distinguish nasals from orals in a more extreme way (i.e., the difference is larger).

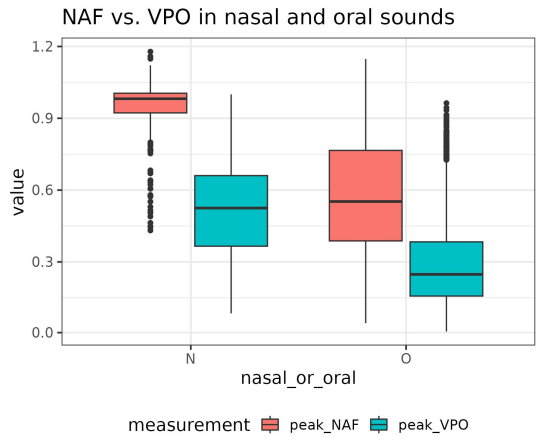

Figure 2: VPO vs. NAF in nasal and oral segments

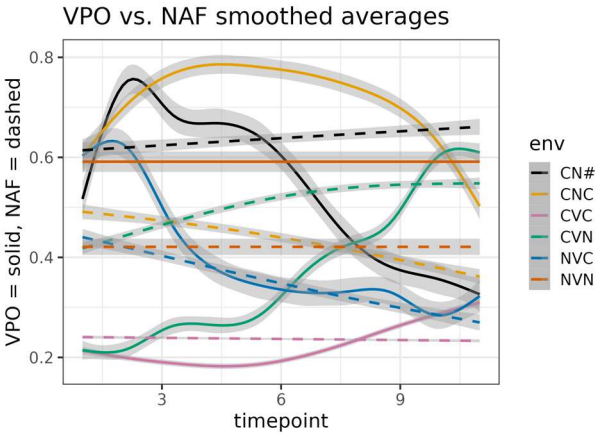

Figure 3: VPO vs. NAF for the duration of segments

To investigate the correspondence between VPO and NAF at a finer scale, Figure 3 presents VPO and NAF values across the duration of segments grouped according to their phonological environments (arbitrarily chosen from the dataset). The lines in the figure represent the smoothed average obtained via GAM method; the error bars represent 68% (1 SD) confidence intervals.

As Figure 3 reveals, NAF has a high degree of correspondence in revealing the overall changes to the degree of nasality over time. E.g., both VPO and NAF start low and end higher in CVN while both start high and lower in NVC, which completely aligns with the position of the nasal segment in the sequence. In CN#, however, the ending looks dramatically different; the VPO remains high while NAF goes low. This too is completely expected since the sequence is prepausal where the velum is lower for an inter-utterance rest position and there is no acoustic speech signal being produced (hence no nasal features).

4 Conclusion

The results of our study indicate a positive alignment between nasality measurements derived from acoustic signals and the VPO data. This finding suggests that NAF measurements can be a decent predictor of the dynamic changes in VPO across a segment. Having differently scaled NAF values, however, may not be directly correlated with the actual size of the VPO in terms of numeric values. The results also indicate that nasality is similarly associated with NAF and VPO in most environments.

Acknowledgments

This project has been supported by NSERC.

References

[1] Chen, M. Y. (1997). Acoustic correlates of English and French nasalized vowels. *JASA*, *102*(4).

[2] Pruthi, T., & Espy-Wilson, C. Y. (2007). Acoustic parameters for the automatic detection of vowel nasalization. In *Eighth Annual Conf. of the International Speech Communication Association*.

[3] Styler, W. (2017). On the acoustical features of vowel nasality in English and French. *JASA*, *142*(4).

[4] Cler, G., Lien, Y., Braden, M., … & Stepp, C. E. (2016). Objective measure of nasal air emission using nasal accelerometry. *J. Speech, Language, and Hearing Research*, *59*(5), 1018-1024.

[5] Carignan, C. (2018). Using ultrasound and nasalance to separate oral and nasal contributions to formant frequencies of nasalized vowels. *JASA*, *143*(5).

[6] Carignan, C. (2021). A practical method of estimating the time-varying degree of vowel nasalization from acoustic features. *JASA,* 149(2), 911-922.

[7] Carignan, C., Chen, J., Harvey, M., Stockigt, C., Simpson, J., & Strangways, S. (2023). An investigation of the dynamics of vowel nasalization in Arabana using machine learning of acoustic features. *Laboratory Phonology*, 14(1), 1-31.

[8] de Boer, G., Islam, J., Purnomo, C., Wu, L., & Gick, B. (2023). Revisiting the nasal continuum hypothesis: A study of French nasals in continuous speech. *J. Phon.*, *98*, 101244.

[9] Munhall, K. G., Vatikiotis-Bateson, E., & Tohkura, Y. (1995). X-ray film database for speech research. *JASA, 98*(2), 1222-1224.

[10] Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. doi:10.1038/nmeth.2089

[11] Boersma, Paul & Weenink, David (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.10.

[12] R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.