A new model for objective estimation of hypernasality from dysarthric speech

Michael Saxon[1], Julie Liss[2], Visar Berisha[1,2]

Ira A. Fulton Schools of Engineering[1], College of Health Solutions[2], Arizona State University

INTRODUCTION

Hypernasality is a common disordered speech symptom, characterized by excessive nasal resonance. It is caused by velopharyngeal port dysfunction, an inability to properly regulate airflow between the oral and nasal cavities. Such modulation requires intact muscle strength and precise motor control (Novotny et al., 2016), thus hypernasality is exhibited in a variety of neurological conditions; automated measures of hypernasality would thus prove valuable in neurological clinical settings.

The gold standard in hypernasality assessment is clinician opinion ratings (Kent, 1996). While averaging multiple subjective ratings improves validity, the practice is untenable in most clinical settings.

The acoustic correlates of hypernasality manifest variably, challenging development of objective measures. Broadly, hypernasality introduces a resonance in the lower frequencies (Kummer, 1996) for voiced sounds; for unvoiced sounds, hypernasality impacts articulatory precision (Woo, 2012).

We introduce and evaluate a new set of acoustic features that leverage the advantages of both approaches, following the intuition that increases in hypernasality result in two perceptible changes: unvoiced phonemes become less precise and voiced phonemes become nasalized.

METHODS

Our **Nasalization/Articulatory Precision** (NAP) features separately evaluate voiced and unvoiced phonemes using two acoustic models that are trained exclusively on larger corpora of healthy speech.

Nasalization Model. We train an acoustic model using a corpus of healthy, read speech (Panayotov et al., 2015). We separate all voiced phonemes into nasal and non-nasal classes, and train a Gaussian mixture model (GMM) to calculate the log-likelihood ratio that a phoneme belongs to the nasalization class over the non-nasalization class.

Articulatory Precision Model. The articulatory precision model computes the precision of unvoiced phonemes, similar to (Witt & Young, 2000), as implemented in (Tu et al., 2016). This yields a likelihood ratio, computed for every unvoiced sound, estimating the precision of each.

Model evaluation: Using a dysarthric speech corpus of 75 speakers (40 male) exhibiting varying levels of hypernasality (38 Parkinson's disease; 6 Huntington's disease; 16 ataxia; 15 amyotrophic lateral sclerosis). All read 5 sentences, for which hypernasality severity ratings (7-point scale) were provided by 14 speech language pathologists. Model predictions are compared against severity ratings.

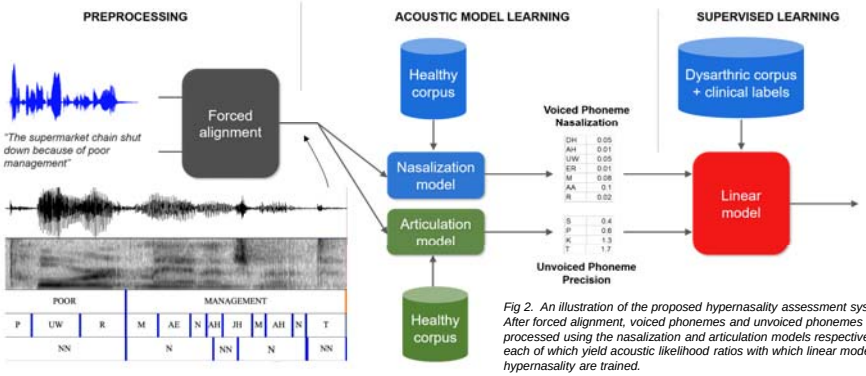

Fig 2. An illustration of the proposed hypernasality assessment system. After forced alignment, voiced phonemes and unvoiced phonemes are processed using the nasalization and articulation models respectively, each of which yield acoustic likelihood ratios with which linear models of hypernasality are trained.

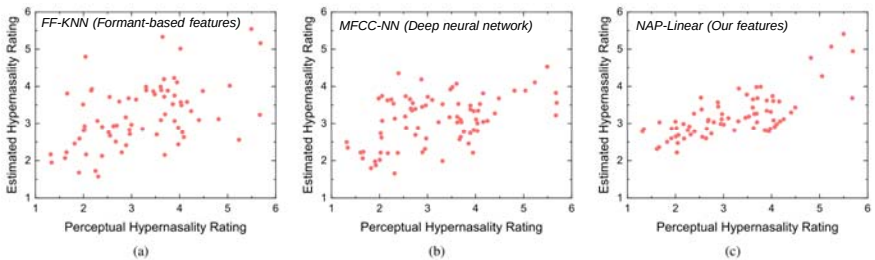

Fig 3. Cross validation results for models predicting clinician hypernasality rating from baseline formant features (FF-KNN) (a), a MFCCs with a neural network (MFCC-NN) (b), and our features (NAP-Linear) (c), demonstrating how NAP-Linear achieves both higher accuracy and better correlation than either of the baseline approaches.

Train on	LOSO		HD, PD, ALS		Ataxia, PD, ALS		Ataxia, HD, ALS		Ataxia, PD, HD	
Test on	LOSO		Ataxia		HD		PD		ALS	
	MAE	PCC	MAE	PCC	MAE	PCC	MAE	PCC	MAE	PCC
FF-Linear	0.871	0.180	0.823	0.042	0.666	-0.751	1.316	0.351	1.426	-0.425
FF-Additive	0.789	0.435	0.730	-0.123	0.693	-0.557	1.334	0.277	1.260	0.429
FF-KNN	0.754	0.481	0.781	0.333	0.567	0.381	1.218	0.402	1.227	-0.039
MFCC-NN	0.884	0.458	0.904	-0.120	0.429	0.568	0.800	0.457	1.233	0.315
NAP-Linear *ours*	**0.587**	**0.722**	**0.546**	**0.750**	**0.559**	**0.737**	**0.509**	**0.697**	**0.597**	**0.527**

Table 1. Evaluation comparison of the NAP features with existing approaches for predicting hypernasality. The input features are "FF," hand-engineered formant features, "MFCC," Mel frequency cepstral coefficients, and "NAP," our Nasalization/Articulatory Precision features. The classifiers include "Linear," simple linear regression, "Additive," additive forward regression, "KNN," K-nearest neighbor selection, and "NN," a neural network as defined in (Vikram et al., 2018). MAE represents mean absolute error, and PCC is the Pearson correlation coefficient between the predicted nasality scores and the true clinician-assessed nasality scores. LOSO denotes "leave one speaker out" cross validation.

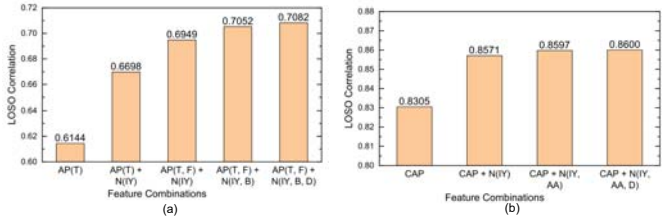

Fig 4. (Left) Additive regression results predicting clinician rating from (a) a subset of the NAP features (articulation feature for T, F and nasalization feature for IY, B, D) and (b) clinician-rated articulatory precision alongside a subset, in leave-one-speaker-out correlation (LOSO). (b) demonstrates that there are hypernasality-correlated qualities in the patient speech that go beyond general articulatory precision, that are captured by the N component of our NAP features.

We evaluate NAP as an input feature to a linear regression model, trained to predict speaker hypernasality against two others, the best hand-engineered formant features (Styler, 2015) and an MFCC-processing end-to-end neural model (Vikram et al., 2018). Two cross validation schemes, "leave one speaker out," (LOSO), and "leave one disease out," (LODO), were employed.

CONCLUSIONS

Results show that NAP features generalize even when training on hypernasal speech from one disease and evaluating on another disease, and are more predictive than both the neural models and hand-engineered models in both LOSO and LODO cross-validation (Table 1).

The NAP features achieve consistent performance across all LODO classes. This suggests that these features are a robust measure of hypernasality, relatively invariant to the disease-specific co-modulating variables that hinder the performance of other approaches.

The NAP model has limitations. Its reliance on aligned transcripts makes it only useful in a controlled clinical setting. Because there are no nasalized voiceless phonemes in English to train a nasalization model, we instead must use articulatory precision as a proxy for hypernasality in voiceless phonemes. Increased hypernasality typically implies reduced articulatory precision, but the converse is not necessarily true. As such, it is possible for speakers to exhibit reduced precision for other reasons than hypernasality.

Despite these limitations, NAP features show promise as a component in diagnostic hypernasality tracking tools, with better predictive performance and generalization than the state of the art.

REFERENCES

M. Novotny, J. Rusz, R. Cmejla, H. Ruzickova, J. Klempir, and E. Ruzicka, "Hypernasality associated with basal ganglia dysfunction: evidence from Parkinson's disease and Huntington's disease," PeerJ, vol. 4, p. 2530, 2016.

R. Kent, "Some limits to the auditory-perceptual assessment of speech and voice disorders,"American Journal of Speech Language Pathology, 1996.

A. Kummer and L. Lee, "Evaluation and Treatment of Resonance Disorders," Language, Speech, and Hearing in Schools, vol. 27, pp. 271–281, Jul 1996.

A. Woo, "Velopharyngeal dysfunction," Semin Plast Surg, vol. 26, no. 4, pp. 170–177, Nov 2012.

V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April 2015, pp. 5206–5210.

S. Witt and S. Young, "Phone-level pronunciation scoring and assessment for interactive language learning," Speech Commun., vol. 30, no. 2-3, pp. 95–108, Feb. 2000. [Online]. Available: http://dx.doi.org/10.1016/S0167-6393(99)00044-8

M. Tu, A. Grabek, J. Liss, and V. Berisha, "Investigating the role of L1 in automatic pronunciation evaluation of L2 speech," Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, vol. 2018-September, pp. 1636–1640, 1 2018.

C. M. Vikram, A. Tripathi, S. Kalita, and S. Prasanna, "Estimation of hypernasality scores from cleft lip and palate speech." in Proc. Interspeech 2018, 2018, pp. 1701-1705.

College of Health Solutions
Arizona State University