34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Normalization Strategies for Nasal Acceleration to Assess Velopharyngeal Function

Elias B. Thorp, Boris T. Virnik, Cara E. Stepp

Abstract— **Velopharyngeal function is essential for intelligible speech production, but can often be impaired. Current clinical care could be improved with the use of reliable and objective methods of assessment appropriate for home use. This paper explores the use of a combined nasal acceleration and acoustic sensor to assess velopharyngeal function. Speech production data in nasalized and non-nasalized contexts is recorded from N=6 healthy participants and three normalization strategies are assessed. Normalizing data to maximally nasal productions results in a reduction of between-speaker variability. Using a filtered speech signal can reduce the effects of intra-speaker variability caused be differences in loudness. The normalization strategies pursued show high discriminability of nasalization in vowels with an inexpensive sensor appropriate for home use.**

I. Introduction

The velopharyngeal (VP) port is the small opening in the back of the mouth that connects the nasal cavity to the pharynx. During speech, this port is opened and closed by the velum to control the amount of acoustic energy allowed to exit the nose. In English only the three “nasal” phonemes /m/, /n/, and /ŋ/ (e.g., “dim”, “din”, and “ding”) require an open VP port. Velopharyngeal dysfunction (VPD) is characterized by the inability to properly close the VP port, resulting in speech that is hypernasal, with too much acoustic energy escaping from the nose [1]. VPD can lead to a dramatic decrease in speech intelligibility [2, 3], which can be physically limiting, socially isolating, and can negatively impact childhood development [4, 5]. This disorder is relatively common, with roughly 1 of every 100 children at risk [6-8].

Current rehabilitation methods and objective assessment techniques for VPD are limited. The most common device for objective VP assessment is the nasometer. This device consists of a baffle pressed against the upper lip of the user that separates the acoustic energy from the nose with the energy from the mouth [9]. Microphones record the nasal and oral acoustic output and take the ratio of these quantitates providing an acoustic correlate of perceived nasality [10], which has been shown to be an effective rehabilitation tool [11]. Nasometers, however, are bulky, cumbersome and not designed for at home use. Previous studies employing a flexible endoscope to provide visual feedback of the VP port to patients produced positive initial results [12], but the use of this technique is limited due to the invasiveness of the procedure.

The ratio of nasal acceleration to total acoustic output (NATAR) has been shown to correlate well with listener judgments of nasality [13]. However, this measure does not provide an estimate of VP function appropriate for comparisons within participants or across groups due to anatomical variability across participants as well as sound pressure level differences across vowels [14]. Inherent differences in sound pressure level as a function of vowel cause problems with nasometer recordings as well, with significant differences in nasalance estimates found as a function of vowel [15].

The purpose of this study is to investigate several normalization strategies with the goal of providing a robust estimate of VP function with minimal variability. To lower the variability across participants, the control signal will be normalized over the maximum achievable control signal by each participant, which will be realized by asking them to produce a loud /n/ sound. In addition, we attempt to reduce within-participant variability caused by the inherent differences in vowel sound pressure level by using filtered acoustic output. Specifically, these inherent differences are most prevalent near the second speech formant (upper harmonics) [16], thus we will use a low-pass filtered acoustic signal rather than the entire acoustic signal to reduce these differences in vowel level. A reliable non-invasive assessment of VP function has utility for improving clinical assessment of VPD. In addition, a sensor and measure appropriate for home use could be used to design novel rehabilitation methods to speed recovery of patients with VPD.

II. Methods

A. Hardware Design

This experiment was performed using a specially designed sensor consisting of a wide-band accelerometer (BU Series Knowles Acoustics) and a standard headset with microphone (PC131, Sennheiser). The accelerometer was attached to the outside of the nose of each participant using medical grade double-sided tape. The accelerometer measures the vibrations of the nasal cavity transferred to the skin of the nose. The accelerometer was connected to a simple circuit for power consisting of a single AAA battery connected in series with a 5.6 KΩ resistor. The headset

*Research supported by Boston University.

E. B. Thorp is with the Department of Biomedical Engineering, Boston University, Boston, MA 02215 USA (email: ebthorp@bu.edu).

B. T. Virnik is with the Department of Biomedical Engineering, Boston University, Boston, MA 02215 USA (email: virnikb@bu.edu).

C. E. Stepp is with the Departments of Speech, Language, & Hearing Sciences and Biomedical Engineering, Boston University, Boston, MA 02215 USA (phone: 617-353-7487; fax: 617-353-5074; e-mail: cstepp@bu.edu).

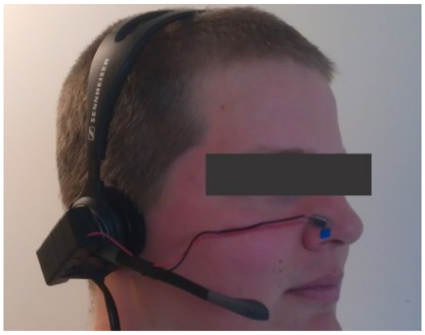

Fig. 1. Acoustic-Accelerometer Sensor as applied to a participant.

microphone measures the combined acoustic output from both the nasal cavity and oral cavity.

The output connector of the combined acoustic-accelerometer is a standard 1/8 inch stereo jack which can be plugged into any computer or recording devices with a standard stereo line-in port. For the experiment below, all data was sampled at 44100 Hz.

B. *Experimental Design*

This experiment was performed on six healthy volunteers with no history of speech or hearing disorders. The participants consisted of three males and three females with an average age of 21 years (STD = 0.5 years). All participants were native speakers of American English. Participants were tested during a single 15-min session and were compensated $5 for their participation. Participants completed written consent in compliance with the Boston University Institutional Review Board.

Participants were asked to read a list of 18 tokens in a comfortable but clear voice while wearing the specially designed sensor. Stimuli consisted of nine nasal and nine non-nasal syllable-pairs matched by place of articulation (e.g., "bob" was matched with "mom"; see Table I). In American English, vowels produced in a nasal context (near nasal consonants) are produced in a nasalized manner. Sensor data was digitally acquired using Praat and a standard computer line-in. For all syllable-pairs, further analysis was performed only on the vowel, which was extracted manually using Praat.

C. *Normalization Strategies*

The raw nasal acceleration provides sensitive information about VP function, but is susceptible to between-participant differences in anatomy and within-participant differences in overall loudness and effort. Three normalization strategies were explored in the current study. The first was the NATAR, the ratio of the nasal acceleration to the total acoustic output, which has previously been shown to correlate with listener perceptions of nasality [13]. For all normalization strategies, the accelerometer output was high pass filtered to eliminate noise using a second order Butterworth filter with a cutoff frequency of 80Hz. The NATAR was calculated by dividing the root-mean-square (rms) of the nasal acceleration by the rms of the total acoustic output as measured by the microphone (eq. 1).

$$\text{NATAR} = \frac{\text{rms(Nasal Acceleration)}}{\text{rms(Total Acoustic Output)}} \qquad (1)$$

The second normalization strategy was the normalized nasal acceleration to acoustic ratio (nNATAR). The nNATAR was calculated by taking the NATAR and normalizing it by dividing by the maximum possible NATAR which was defined as occurring during production of a loud /n/ sound (eq. 2).

$$\text{nNATAR} = \frac{NATAR}{NATAR\ during /n/} \qquad (2)$$

The final normalization strategy used a filtered acoustic output to provide lower variability with-in participants and across vowels. This normalization strategy was the normalized nasal acceleration to low passed acoustic ratio (nNALPAR). For the nNALPAR, the total acoustic output was low pass filtered with a cutoff frequency of 250 Hz. The nNALPAR is thus the rms of the nasal acceleration divided by the filtered total acoustic output, normalized over the nNALPAR calculated during the production of a loud /n/ (eq. 3 and 4).

$$\text{NALPAR} = \frac{\text{rms(Nasal Acceleration)}}{\text{rms(Filtered Acoustic Output)}} \qquad (3)$$

$$\text{nNALPAR} = \frac{\text{NALPAR}}{\text{NALPAR during /n/}} \qquad (4)$$

D. *Analysis*

The three normalization strategies were applied offline for each participant's productions using MATLAB (Mathworks, Natick MA). Statistical analysis was performed using Minitab Statistical Software (Minitab Inc., State College, PA). A two factor repeated measures analysis of variance (ANOVA) was performed to assess the effects of nasality (nasal versus non-nasal) and syllable-pair as well as the interaction between the two. *Post hoc* two-sided Tukey's Simultaneous tests were performed as appropriate. Statistical analyses were performed using an alpha level of 0.05 for significance.

In addition, effect size was used to quantitatively measure the efficacy of each normalization strategy. The effect size was calculated as the difference between the mean across nasalized vowels and the mean control signal across non-nasalized vowels divided by the combined standard deviation of the two (eq. 5).

$$\text{Effect Size} = \frac{\text{mean(nasal)} - \text{mean(non-nasal)}}{\text{combined std}} \qquad (5)$$

The effect size for each control signal was calculated for each participant as well as all for the participants' data combined.

Sensitivity and specificity analysis was performed on the data collected by defining a nasality threshold as the level of each normalized signal that discriminates nasal and non-

TABLE I
SYLLABLE PAIRS

Nasal	knee	man	mean	min	mom	moo	nan	nom	noon
Non-Nasal	dee	bad	bead	bid	bob	boo	dad	dob	dude

TABLE II
ANOVA RESULTS

Normalization Strategy	Factor	DF	F	p
NATAR	Nasality	1	65.0	<0.001
	Syllable-Pair	8	3.9	<0.001
	Nasality × Syllable-Pair	8	1.0	0.458
nNATAR	Nasality	1	201.4	<0.001
	Syllable-Pair	8	14.1	<0.001
	Nasality × Syllable-Pair	8	3.8	0.001
nNALPAR	Nasality	1	352.3	<0.001
	Syllable-Pair	8	5.7	<0.001
	Nasality × Syllable-Pair	8	2.1	0.04

nasal productions. Based on this nasality threshold, measured signals that fall above the threshold were marked a "nasal" and those falling below the threshold as "non-nasal." At each nasality threshold, sensitivity (the ratio of true positives to the sum of true positives and false negatives) and specificity (the ratio of true negatives to the sum of true negatives and false positives) were calculated. The sensitivity and specificity were plotted to produce a receiver operating characteristic (ROC curve) to show the discrimination performance of the three normalization strategies. Based on the ROC, the AUC (area under the curve) was determined using numerical integration (trapezoidal rule) for each normalization strategy. Positive likelihood ratios (PLR) were calculated as the sensitivity / (1 – specificity).

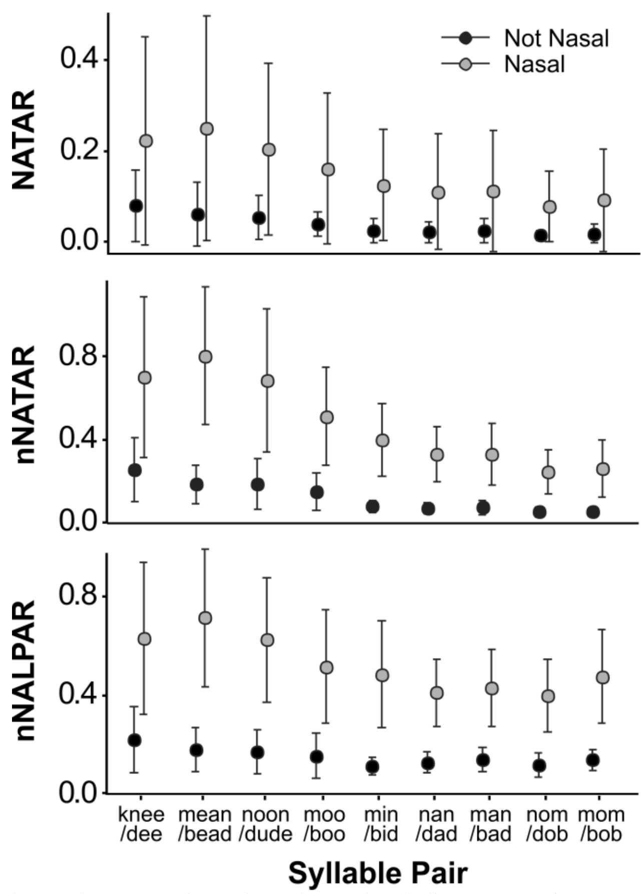

Fig. 2. Primary Experimental Results. Markers indicate means +/- SE. ANOVA found a significant effect of both nasality and syllable-pair for all three normalization strategies.

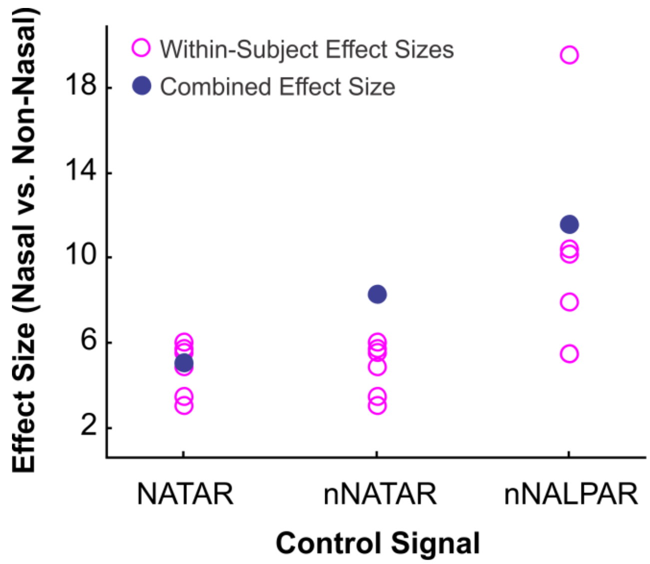

Fig. 3. Effect Sizes. The effect size between nasalized and non-nasalized productions is shown for each participant individually (open circles) and over all participants (closed circles) for each of the three normalization strategies.

III. RESULTS

All three normalization strategies showed significant effects of nasality as well as syllable-pair (see results of ANOVAs in Table II). Both nNATAR and nNALPAR also showed a significant effect of the interaction between nasality and syllable-pair. Figure 2 shows plots of the primary experimental findings. Post-hoc Tukey's tests on nasality showed that all 3 normalization strategies resulted in values that were significantly increased during nasalized syllables relative to non-nasalized syllables ($p < 0.05$).

Effect sizes between nasalized and non-nasalized productions are shown in Fig. 3. Individual within-participant effect sizes show no improvement between NATAR and nNATAR, although the combined effect size (comparison using mixed participant data) shows a 63% increase from 5.1 to 8.3. Use of nNALPAR shows an improvement both in individual within-participant effect sizes as well as the combined effect size. The effect size between nasalized and non-nasalized productions using nNALPAR is 11.6, a 127% increase from the effect size using NATAR.

Results of ROC analysis are shown in Fig. 4 and shows the sensitivity and specificity of the three normalization strategies in discriminating nasalized from non-nasalized vowels. All three normalization strategies performed relatively well when compared with chance detection rate (dashed gray line). NATAR had an AUC = 0.83 and a maximum PLR = 15 occurring at a threshold of 0.19. nNATAR had an AUC = 0.92 and a maximum PLR = 26 occurring at a threshold of 0.41. Finally, nNALPAR had an AUC = 0.97 and a maximum PLR = 40 occurring at a threshold of 0.33. Overall, nNATAR outperformed NATAR, and nNALPAR outperformed both NATAR and nNATAR. These results match well with the effect-size analysis.

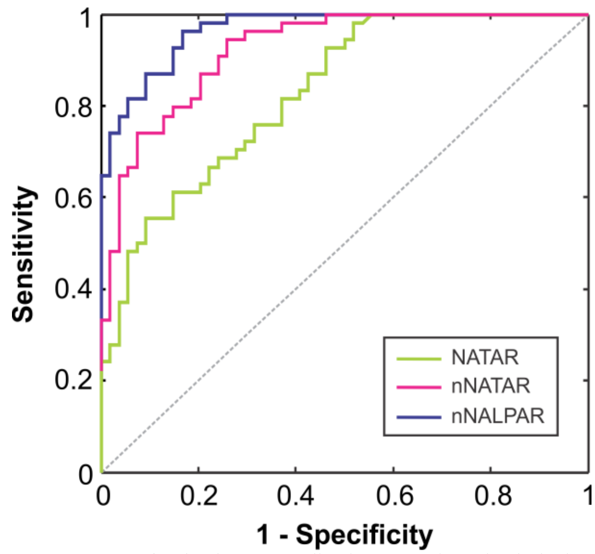

Fig. 4. ROC curve for the three normalization strategies. The dashed gray line shows chance performance.

IV. Discussion

A. Using maximal nasalization

Based on the results seen here, normalizing data to maximally nasal productions results in a reduction of between-speaker variability. This is evidenced by the effects sizes shown in Fig. 3. Although individual within-participant effect sizes do not show improvement between NATAR and nNATAR, the combined effect size shows a large increase. Although these results seem promising, normalization from a single maximal measure can also add variance in cases of unreliable maximal nasality elicitation. In our future work we will assess the reliability of maximal measures across participants and across time.

B. Using a filtered speech signal

By using a low-pass filtered speech signal, the effects of intra-speaker variability caused by differences in inherent vowel sound pressure level can be reduced. From Fig. 2. the increase in separation between nasalized and non-nasaslized production is particularly apparent. The normalization strategies pursued in nNATAR and nNALPAR increase the separation between nasal and non-nasal while decreasing intra-speaker variability.

In our current work, nNALPAR was calculated by low pass filtering the total acoustic output with a cutoff frequency of 250 Hz. The goal was to select a cutoff frequency that would confine the resultant signal to energy from the fundamental frequency of each speaker only. Given that speakers consisted of both men and women, this threshold value may have allowed increased energy in the filtered signal for male speakers with particularly low fundamental frequencies. For instance, a speaker with a fundamental frequency of 100 Hz would have a second harmonic at 200 Hz, easily in the range of the filtered signal. In the future we plan to create automatic user-specific filtering processes based on average fundamental frequency values. This technique may further improve the performance of this strategy.

C. Summary

The normalization strategies pursued show high discriminability of nasalization in vowels produced by healthy speakers as evidenced by their sensitivity and specificity. In our future work we will explore the relationships between these measures and standard clinical measurements of VP function, nasalance and nasality, in both healthy speakers as well as speakers with VPD. In all, the current results are promising for the use of acceleration and acoustics as reliable measures of VP function appropriate for home-use.

Acknowledgments

The authors thank Margaux Canevari for her assistance.

References

[1] L. L. Kummer A, "Evaluation and Treatment of Resonance Disorders.," *Language, Speech, and Hearing in School,* vol. 27, pp. 271-281, 1996.

[2] W. B. Mc, "Some factors in the intelligibility of cleft-palate speech," *J Speech Hear Disord,* vol. 19, pp. 524-8, Dec 1954.

[3] M. Hodge and C. L. Gotzke, "Preliminary results of an intelligibility measure for English-speaking children with cleft palate," *Cleft Palate Craniofac J,* vol. 44, pp. 163-74, Mar 2007.

[4] Y. Kim, R. D. Kent, and G. Weismer, "An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria," *J Speech Lang Hear Res,* vol. 54, pp. 417-29, Apr 2011.

[5] L. Barr, S. L. Thibeault, H. Muntz, and L. de Serres, "Quality of life in children with velopharyngeal insufficiency," *Arch Otolaryngol Head Neck Surg,* vol. 133, pp. 224-9, Mar 2007.

[6] S. L. Boulet, C. A. Boyle, and L. A. Schieve, "Health care use and health and functional impact of developmental disabilities among US children, 1997-2005," *Arch Pediatr Adolesc Med,* vol. 163, pp. 19-26, Jan 2009.

[7] I. Hocevar-Boltezar, A. Jarc, and V. Kozelj, "Ear, nose and voice problems in children with orofacial clefts," *J Laryngol Otol,* vol. 120, pp. 276-81, Apr 2006.

[8] A. P. Vanderas, "Incidence of cleft lip, cleft palate, and cleft lip and palate among races: a review," *Cleft Palate J,* vol. 24, pp. 216-25, Jul 1987.

[9] S. Pratt, Hricisak, I., "Commerically Available Cmputer-Based Speech Feedback Systems," *Journal of Medical Speech-Language Pathology,* vol. 9, pp. 257-274, 1994.

[10] T. U. Brancamp, K. E. Lewis, and T. Watterson, "The relationship between nasalance scores and nasality ratings obtained with equal appearing interval and direct magnitude estimation scaling methods," *Cleft Palate Craniofac J,* vol. 47, pp. 631-7, Nov 2010.

[11] S. G. Fletcher and J. M. Higgins, "Performance of children with severe to profound auditory impairment in instrumentally guided reduction of nasal resonance," *J Speech Hear Disord,* vol. 45, pp. 181-94, May 1980.

[12] M. Brunner, A. Stellzig-Eisenhauer, U. Proschel, R. Verres, and G. Komposch, "The effect of nasopharyngoscopic biofeedback in patients with cleft palate and velopharyngeal dysfunction," *Cleft Palate Craniofac J,* vol. 42, pp. 649-57, Nov 2005.

[13] Y. Horii, "An accelerometric measure as a physical correlate of perceived hypernasality in speech," *J Speech Hear Res,* vol. 26, pp. 476-80, Sep 1983.

[14] G. E. Peterson, Barney, H. L., "Control Methods Used in a Study of the Vowels," *J Acoust Soc Am,* vol. 24, pp. 175-184, 1952.

[15] S. N. Awan, K. Omlor, and C. R. Watts, "Effects of computer system and vowel loading on measures of nasalance," *J Speech Lang Hear Res,* vol. 54, pp. 1284-94, Oct 2011.

[16] K. N. Stevens, *Acoustic phonetics*. Cambridge, Mass.: MIT Press, 1998.