Effects of spectral content on Horii Oral-Nasal Coupling scores in children

Lenny A. Varghese
Center for Computational Neuroscience and Neural Technology, Boston University, 677 Beacon Street, Boston, Massachusetts 02215

Joseph O. Mendoza
Department of Speech, Language, and Hearing Sciences, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

Maia N. Braden
Department of Surgery, Division of Otolaryngology, University of Wisconsin, Voice and Swallow Clinics, 1675 Highland Avenue, Mailcode C225, Madison, Wisconsin 53792

Cara E. Stepp[a)]
Department of Speech, Language, and Hearing Sciences, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

(Received 31 December 2013; revised 24 July 2014; accepted 30 July 2014)

A miniature accelerometer and microphone can be used to obtain Horii Oral-Nasal Coupling (HONC) scores to objectively measure nasalization of speech. While this instrumentation compares favorably in terms of size and cost relative to other objective measures of nasality, the metric has not been well characterized in children. Furthermore, the measure is known to be affected by vowel loading, as speech loaded with "high" vowels is consistently scored as more nasal than speech loaded with "low" vowels. Filtering the signals used in computation of the HONC score to better isolate the correlates of nasalization has been shown to reduce vowel-related effects on the metric, but the efficacy of filtering has thus far only been explored in adults. Here, HONC scores for running speech and the vowel portions of consonant-vowel-consonant tokens were calculated for the speech of 26 children, aged 4–9 yrs. Scores were computed using the broadband accelerometer and speech signals, as well as using filtered, low-frequency versions of these signals. HONC scores obtained using both broadband and filtered signals resulted in well-separated scores for nasal and non-nasal speech. HONC scores computed using filtered signals were found to exhibit less within-participant variability. © *2014 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4892791]

PACS number(s): 43.70.Aj, 43.70.Jt [BHS] Pages: 1295–1306

I. INTRODUCTION

In English speech, partial coupling of the nasal and oral cavities via the velopharyngeal port during coarticulation of nasal consonants and vowels leads to these vowels being perceived as nasalized (Ali *et al.*, 1971). Speech that is perceived as excessively nasalized—i.e., hypernasal—can arise when a speaker suffers from articulatory mislearning, anatomical irregularities, or a neurological disorder that results in abnormal amounts of resonant energy in the nasal cavity while speaking, especially during vowel production (Kummer, 2011a,b). Some populations are prone to hypernasality of speech, particularly individuals with hearing impairment (due to the changes or outright lack of in feedback regarding their speech productions; Stevens *et al.*, 1976; Kummer, 2011b) and individuals with a history of cleft palate (Kummer, 2011a). Speaking in a hypernasal manner may draw negative attention to the speaker and can have detrimental social consequences even during childhood (Blood and Hyman, 1977).

A consistent and objective way of measuring the level of nasality in speech is important not only for the diagnosis of velopharyngeal disorders but also for monitoring the progress of speech rehabilitation. Currently, most speech-language pathologists (SLPs) utilize auditory perceptual assessments most frequently, and the perceptual assessments guide the use of instrumentation-based methods such as nasometry, videofluoroscopy, nasendoscopy, pressure/flow measures, and magnetic resonance imaging (MRI; Stelck *et al.*, 2011; Kummer *et al.*, 2012). However, auditory perceptual judgments of disordered speech may show poor inter- and intra-rater reliability (Kent, 1996), and perception of hypernasality is no exception (Bradford *et al.*, 1964). In particular, ratings of hypernasality in speech have been found to depend on factors such as the rater's level of experience with clinical judgments of speech (Lewis *et al.*, 2003), as well as the speaker's vocal intensity (Counihan and Cullinan, 1972) and overall intelligibility (McWilliams, 1954). When speech therapy for resonance disorders is indicated (i.e., the resonance disorder is not the result of a severe

[a)]Author to whom correspondence should be addressed. Electronic mail: cstepp@bu.edu

structural deficiency), treatment in school and rehabilitation settings generally occurs one to three times a week with an SLP and also requires daily home practice of techniques trained in therapy. Individuals engaging in home practice, particularly individuals with hearing impairment, may not be able to reliably self-monitor their techniques in the absence of guidance from the SLP. Together, the limitations of perceptual judgments and the potential benefits of training aids that do not require supervision suggest that low-cost instrumentation serving both purposes would be beneficial to individuals with hypernasal speech. Instrumentation based primarily on acoustic recordings might be particularly well suited for diagnosing or rehabilitating hypernasal speech due to the non-invasive nature of such recordings, but current implementations of such instrumentation have shortcomings in terms of variability in their metrics of nasalization and suitability for prolonged use.

One relatively inexpensive method that has been proposed to objectively measure hypernasality is the utilization of a small accelerometer (i.e., a contact microphone) placed on the nose to record vibrations of the nasal tissue during speech. The premise for such instrumentation was first proposed in the 1970s (Stevens *et al.*, 1975) as a method of providing feedback to deaf children or their instructors regarding inappropriate nasalization of speech (Stevens *et al.*, 1976; Nickerson *et al.*, 1976). However, purely accelerometric measures of nasalization are subject to variation due to factors including vocal output (Stevens *et al.*, 1976; Garber and Moller, 1979) or accelerometer placement (Lippmann, 1981). Scaling the signals obtained from the accelerometer by another signal indicative of overall vocal output, such as that obtained from an accelerometer placed on the throat (Horii, 1980; Horii and Lang, 1981; Redenbaugh and Reich, 1985) or from a microphone at a fixed distance from the mouth (Horii, 1980; Horii and Lang, 1981; Horii, 1983) can reduce this problem. Measurements made across individuals or over time can be compared by dividing the ratio of amplitudes between accelerometer and microphone signals by the amplitude ratio of these two signals during production of a maximally nasal sound (typically an /m/). Values obtained using this scaling and calibration procedure are known as Horii Oral-Nasal Coupling (HONC) scores (Horii, 1980).

In practice, use of an oral microphone for computing the HONC score has been more popular than the use of a laryngeal microphone, perhaps influenced by the desire of researchers and SLPs to record the original speech signal for later use with a minimum amount of specialty equipment (i.e., recording devices that can record three or more inputs simultaneously). HONC scores obtained in this fashion are then conceptually similar to nasalance values obtained from Nasometry equipment in that both divide the level of a signal representative of nasal output by the level of a signal representative of combined oral-nasal output. However, the instrumentation required to compute HONC scores provides several advantages over the instrumentation used to compute nasalance scores in terms of cost (a few hundred dollars for accelerometer equipment versus several thousand dollars for Nasometry equipment) and overall obtrusiveness (a miniature accelerometer on the nose and a standard headset microphone, compared to dual microphones mounted to a large baffle or mask). The cost and size advantages of accelerometer-based instrumentation may make it a more attractive option than Nasometry for the objective measurement of nasality in children; indeed, compliance data in children indicate a preference for accelerometer-based instrumentation relative to Nasometry instrumentation (Braden *et al.*, 2013). The calibration procedure used in the computation of the HONC scores (scaling the accelerometer-to-speech ratio associated with a speech sample by the accelerometer-to-microphone ratio associated with an /m/ sound may also be considered an advantage over nasalance scores, which do not utilize any speaker-specific calibrations.

In spite of the practical advantages that HONC scores possess over other instrumented measures of nasality, there appears to be only one study to date that has examined HONC scores in children (Mra *et al.*, 1998). That study, conducted in 4–6 yr olds with typical speech, reported HONC scores showing about a 13 dB difference between speech tokens that were loaded with nasal consonants and speech tokens that were loaded with non-nasal consonants. The magnitude of the difference between HONC scores for speech that is highly nasalized compared to non-nasalized speech is similar to the difference that has been reported for the same comparison in typical adult speech (Sussman, 1995; Thorp *et al.*, 2013). The difference in HONC scores between nasalized and non-nasalized speech is also similar regardless of whether the scores were obtained from running speech (Sussman, 1995; Mra *et al.*, 1998) or whether scores were obtained only using vowel portions of words (Thorp *et al.*, 2013).

An objective acoustic measure of nasality would ideally be sensitive to the level of coupling between the nasal cavity and the rest of the vocal tract without being overly sensitive to the actions of other articulators, such as the position of the tongue. However, nasalance (Lewis *et al.*, 2000; Awan *et al.*, 2011; Thorp *et al.*, 2013) and HONC scores (Sussman, 1995; Mra *et al.*, 1998; Thorp *et al.*, 2012, 2013) each exhibit differences related to vowel content of speech. Specifically, nasalance and HONC scores tend to be higher for speech loaded with "high" vowels (vowels that are produced with a high tongue position, e.g., /i/, /u/) compared to those for speech loaded with "low" vowels (vowels that are produced with a lower tongue position, e.g., /ɑ/,/æ/). It should be noted that the direct relationship between tongue height and instrumented scores of nasality matches the relationship that has been reported between tongue height and perceptual judgments of nasality (Andrews and Rutherford, 1972). Greater perceived and measured nasality on high vowels is perhaps counterintuitive given that the size of the velopharyngeal opening that couples oral and nasal cavities is inversely related to vowel height (Moll, 1962), but can be attributed to an increase in oral cavity impedance for high vowels (Lubker and Moll, 1965; Kummer, 2011a) and the cross-velar transmission of low-frequency energy that is present in high vowels (Stevens *et al.*, 1976; Sussman, 1995; Kummer, 2011a). Such vowel-related effects in HONC scores were found to be diminished when HONC scores were computed using microphone signals that were first filtered to largely eliminate formant structure while still using a broadband accelerometer signal (Thorp *et al.*, 2012). A follow-up study (Thorp *et al.*, 2013) indicated that group-level differences in scores obtained for nasalized and non-nasalized vowels in the speech of adults were maximized when a 400–1000 Hz filter is applied to the nasal acceleration signal and a 25–425 Hz filter is applied to the combined oral-nasal signal.

Given that the speech of very young children is more spectrotemporally variable than the speech of adults (Lee *et al.*, 1999), it is unclear if HONC scores derived from the low-frequency portions of the accelerometer and acoustic signals will be more sensitive to the nasal/non-nasal distinction of vowels and/or exhibit less vowel-related variability than when broadband signals are used. As such, we obtained nasal acceleration and speech signals from a group of 26 children, aged 4–9 yrs, during production of various consonant-vowel-consonant (CVC) tokens and during production of running speech with controlled vowel and consonant loading. HONC scores were computed using broadband signals (as in most previous literature), as well as using low-frequency portions of these signals (as in Thorp *et al.*, 2013). We hypothesized that both methods would exhibit high discriminability between contextually nasalized vowels and non-nasalized vowels, as well as in running speech loaded with nasal and non-nasal consonants. Additionally, we hypothesized that effects of vowel placement would be less pronounced or eliminated when the 400–1000 Hz portion of the accelerometric and 25–425 Hz acoustic signals were isolated and used in computation of the HONC scores compared to when broadband portions of each signal were used.

II. METHODS AND MATERIALS

A. Participants

Nasal acceleration and speech signals were recorded from 26 native-English speaking children (aged 4–9 yrs, 15 female) with no reported history of speech, language, or hearing disorders. In compliance with the Boston University Charles River Campus Institutional Review Board, informed consent was obtained from a parent of each participant. Additionally, participants aged 7–9 yrs provided verbal assent. Participants were compensated for their time. A listing of the age and gender of the participants is shown in Table I.

TABLE I. Age and gender of participants.

Participant ID	Gender	Age
1	F	4
2	F	4
3	F	4
4	M	4
5	M	4
6	F	4
7	M	5
8	M	5
9	F	5
10	M	5
11	F	5
12	F	5
13	F	5
14	M	6
15	M	6
16	F	6
17	F	6
18	M	6
19	M	7
20	M	7
21	F	8
22	F	8
23	F	9
24	F	9
25	M	9
26	F	9

B. Recording equipment

One of two sets of instrumentation was used to obtain the nasal and oral signals. The first was a modified Sennheiser PC131 headset (Sennheiser GmbH, Wedemark, Germany) coupled with a BU-21771 accelerometer (Knowles Electronics, Itasca, IL). Signals from the accelerometer and headset microphone were digitized via a desktop computer's integrated sound card (RealTek ALC662). The second setup utilized a WH20 XLR microphone (Shure Incorporated, Niles, IL) for recording speech signals and a BU-21771 accelerometer wired to utilize an XLR connection to measure nasal acceleration. Both signals were digitized via an external sound card with XLR inputs and built-in microphone preamplifiers. The sound cards utilized for this purpose were either an ART USB Dual Pre (Applied Research and Technology, Buffalo, NY) or a PreSonus Fire10 sound interface (PreSonus Audio Electronics, Baton Rouge, LA). While there were slight differences in the overall frequency responses of the microphones and sound cards, we noted no obvious differences in the quality of the signals that would significantly affect computation of the filtered HONC scores.

Signals obtained from the accelerometer and the microphone were digitized at a sampling rate of 44.1 kHz using either Praat (http://www.fon.hum.uva.nl/praat) or Audacity (http://audacity.sourceforge.net:). The signals were stored as two-channel, 16-bit WAV files for analysis.

C. Testing procedure

As in previous studies, the accelerometer was placed on the side of the upper lateral cartilage near the anterior part of each child's nasal bone using medical-grade double-sided tape (Lippmann, 1981; Thorp *et al.*, 2013). The microphone was placed at approximately a 45° angle, 6 to 10 cm away from the participant's mouth.

Participants were seated in a sound-treated booth along with the experimenter and, for the youngest participants, a parent. Each participant was asked to produce the CVC tokens or sentence tokens (i.e., sentences with controlled vowel and consonant loading) listed in Table II while

TABLE II. Speech tokens utilized in the experiment; with the exception of the non-nasalized sentences (taken from Lewis *et al.*, 2000), tokens were chosen by the authors. For the CVC tokens, the vowel portion of the word in the two right columns was extracted for analysis. For the sentence tokens, silent gaps were removed using an automated procedure (see text).

Token type	Vowel place	Non-nasalized	Nasalized
CVC	Low/Front	/bæb/	/mæm/
		/dæd/	/mæm/
	Low/Back	/bɑb/	/mɑm/
		/dɑd/	/nɑn/
	High/Front	/bib/	/mim/
		/did/	/nin/
	High/Back	/bub/	/mum/
		/dud/	/nun/
Sentence	Low/Front	"Bess has dad's red cap"	"Ben scans the man's next plan"
	Low/Back	"Father got all four cards"	"Mark's blond mom yawns more"
	High/Front	"Bill sees the sleepy kid"	"Tim seems mean to Nick"
	High/Back	"Glue the old blue shoes"	"Zoom to the new home soon"

wearing the headset microphone and accelerometer. Participants were asked to produce each CVC token three times and each of the sentences once. Participants were also asked to produce a series of sustained /m/ sounds from which a calibration factor for the HONC scores was obtained (described below). Participants who were able to read were prompted to read the tokens from a printed instruction sheet, while participants who were unable to read were instructed to repeat the words after the experimenter. Participants were asked to perform these tasks in two identical experimental blocks (hereafter referred to as Block 1 and Block 2), with the headset removed and replaced between blocks.

D. Data analysis

The recorded signals were processed using the same general processing scheme described in Thorp *et al.* (2013). Visual analysis of the recorded signals and initial selection of regions of interest in the raw signals were done using either Audacity or Praat. All remaining filtering and score computations were performed in Python with the SciPy package (Oliphant, 2007).

1. Token preprocessing and extraction

For each of the produced CVC tokens, WAV samples corresponding to the middle of the vowel were visually identified and selected using the mouse cursor. These portions of the signals were visually inspected for artifacts in either channel (e.g., background speech from the parent or the experimenter, or rapid, transient signal changes obviously not related to speech processes), and the selected region of the signal was shifted or shortened such that the selected region did not contain any such artifacts. The cropped portions of the signal were re-saved to disk as 2-channel WAV files for further processing.

Sentence tokens were preprocessed by removing the silent gaps at the beginning and end of each recording in the audio editing programs. Silent gaps within each token were eliminated prior to computation of the HONC score by computing the envelope of the microphone signal and retaining only samples of the token where the amplitude of the microphone signal envelope exceeded 10% of its maximum amplitude. The speech envelope was computed by zero-phase low-pass filtering the full-wave rectified microphone signal at 20 Hz with a digital, second order Butterworth filter.

For each recording session per participant, signals corresponding to a single sustained /m/ sound were isolated in the same manner that the vowels from the CVC words were isolated.

After the process of manual selection and extraction, processed tokens that were less than 50 ms in duration after these procedures were excluded from further analyses.

2. Filtering and computation of HONC scores

All preprocessed speech tokens were filtered one of two ways: Using a 400–1,000 Hz filter on the accelerometer signal and a 25–420 Hz filter on the microphone signal (as in Thorp *et al.*, 2013; henceforth referred to as the "low-frequency" signals), or using an 80–9000 Hz filter on the accelerometer and high-pass filtering the acoustic signal at 25 Hz (roughly equivalent to computation of the broadband HONC scores used in other studies; henceforth referred to as the "broadband" signals).

For each speech token x, the HONC score H_x was computed as

$$H_x = 20 \log 10\left(k_m \frac{A_x}{M_x}\right)$$

in which A_x and M_x are the root-mean-square (rms) values of the filtered accelerometer and microphone signals, respectively, for token x. The normalizing factor k_m was computed for each experimental block as the reciprocal of the ratio between the filtered accelerometer and filtered microphone signals' rms values for a single sustained /m/ token produced within that block (Horii, 1980).

3. Statistical analyses

All statistical analyses were conducted in R (R Core Team, 2013) with add-on data manipulation (Wickham,

2009) and visualization (Wickham, 2011) packages, as well as a package to facilitate analysis of variance on repeated measures data (Lawrence, 2013).

Within-speaker reliability of the HONC scores obtained using each filtering method was determined using linear regression on all data collected in the two different blocks; i.e., a line was fit to the HONC scores obtained from speech samples collected in Block 2 plotted as a function of HONC score in Block 1 belonging to the same token and same speaker. If a score could not be computed for a given token in one or both blocks, the data points from both blocks were excluded from this analysis. For all subsequent analyses, data from Block 1 and Block 2 were combined for each participant.

Differences in HONC scores for different nasalization contexts were quantified both in dB as well as using Cohen's d, a measure of effect size, with combined standard deviation across nasalized/non-nasalized contexts (Cohen, 1992; Turner and Bernard, 2006). Scores were computed such that positive d values were indicative of greater HONC scores for nasalized tokens compared to non-nasalized tokens. Paired, two-tailed t-tests were used to perform a direct comparison between effect sizes for HONC scores computed using each filter type and separately for each token type. To additionally quantify overall nasalization context discriminability across participants, receiver operator characteristic (ROC) curves and the associated area under the curve (AUC) were computed (Bradley, 1997). ROC curves were obtained separately for each combination of signal type (broadband and low-frequency) and token type (CVC and sentence). The true positive rate (sensitivity) for a given threshold was obtained by dividing the number of nasal tokens correctly classified as "nasal" by the total number of nasal tokens. The false positive rate was obtained by dividing the number of tokens incorrectly classified as nasal by the total number of non-nasal tokens. The thresholds used for these computations ranged from -45 to $+10$ dB in steps of 0.05 dB. For each true positive vs false positive curve, AUC was computed using the trapezoidal method.

Repeated measures analysis of variance (ANOVA) on the HONC scores (averaged across all repetitions of a given vowel type across blocks for a given participant) was performed separately for each HONC score type (broadband and low-frequency) and for each type of utterance (CVC or sentences). Consonantal context (nasal or non-nasal) and vowel place/loading (high/front, high/back, low/front, and low/back) served as within-participant factors. Mauchly's test for sphericity was applied to vowel and vowel-context effects, and Greenhouse-Geisser corrections to the degrees of freedom were applied where the assumption of sphericity was found to be violated (Abdi, 2010). Generalized eta squared values (η^2_G, Olejnik and Algina, 2003; Bakeman, 2005) were computed to quantify effect sizes for each factor in the ANOVA. *Post hoc*, pairwise comparisons for the vowel place factor were conducted using t-tests with Bonferroni corrections applied. In all cases where a p value had to be adjusted (i.e., due to a violation of sphericity or for pairwise comparisons), the adjusted p value (p_{adj}) is reported.

III. RESULTS

A. Evaluation of within-subject reliability across experimental blocks

HONC scores obtained in Block 2 are plotted as a function of those obtained in Block 1 for both broadband (left panel) and low-frequency (right panel) signals in Fig. 1. The best fit line through the data points is depicted in each graph. Here, optimal test-retest reliability would be indicated by a best-fit line with a slope of 1 and an intercept of 0. While most data points for both score types appeared to lie near this diagonal, regression lines indicated that overall, HONC scores in Block 1 were not always consistent with HONC scores in Block 2, regardless of the filter settings. For both filter settings, r^2 could be characterized as moderate at best ($r^2 = 0.543$, $p < 0.001$ for the broadband data, and $r^2 = 0.463$, $p < 0.001$ for the low-frequency data).

B. Vowel-independent discrimination of contextual nasalization using HONC scores

Independent of vowel place, the mean difference across participants between broadband HONC scores for nasalized and non-nasalized CVC tokens was found to be 11.7 dB for CVC tokens utilized and 11.5 dB for the sentence tokens scores, corresponding to mean, per-speaker Cohen's d values of 2.39 and 4.68, respectively (Fig. 2, upper left). For the filtered data, the mean difference across participants in HONC scores was 12.8 dB for CVC tokens and 10.1 dB for the sentence tokens, corresponding to mean, per-speaker Cohen's d values of 5.68 and 6.14 (Fig. 2, upper right). For both CVC and sentence tokens, HONC scores computed using the low-frequency portions of the microphone and accelerometer signals had significantly higher d values than HONC scores computed using broadband signals for CVC tokens [$t(25) = 5.02$, $p \ll 0.001$], but no significant difference in means was found for the sentence tokens [$t(19) = 1.95$, $p = 0.066$].

ROC plots (Fig. 2, bottom) confirm that HONC scores, regardless of filter type, exhibit high discriminability between contextually nasalized utterances. ROC plots of the broadband data (Fig. 2, lower left) indicate that AUC was 0.889 for the CVC tokens and 0.943 for the sentence tokens. AUCs derived from the ROC plots for the filtered HONC scores (Fig. 2, lower right) were 0.934 for CVC tokens, and 0.917 for the sentence tokens.

C. Evaluation of HONC score variability due to vowel loading

1. CVC tokens

HONC scores computed for the vowel portion of the CVC tokens, broken down by vowel place, are depicted in Fig. 3 for individual participants (upper panels) and as per-vowel, per-context means in dB (lower panels). Visually, the spread of the data in each plot indicates that while inter-subject variability appears lower for the broadband HONC scores, the within-subject variability is lower for the filtered HONC scores (upper panels, compare the distribution of black and gray points across participants).

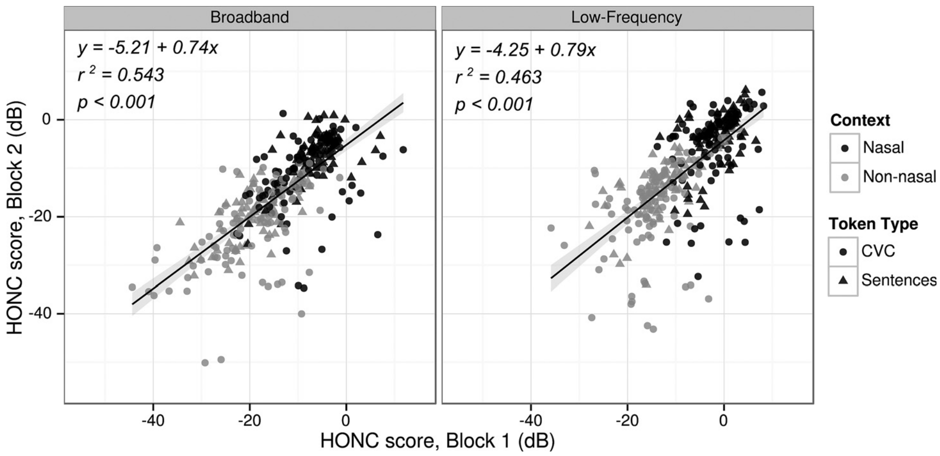

FIG. 1. HONC scores over all participants, plotted as scores obtained in Block 2 as a function of scores obtained in Block 1. HONC scores are shown when computed using broadband nasal acceleration and acoustic signals (left), and when computed using filtered nasal acceleration and acoustic signals (right). Shaded regions in each panel indicate 95% confidence intervals around the best fit line. Perfect test–retest reliability would be indicated by all points lying along the diagonal (i.e., slope of 1 and intercept of 0).

On average, the difference in scores between CVC vowels produced in nasal contexts and CVC vowels produced in non-nasal contexts was similar for both broadband HONC scores and filtered HONC scores (Fig. 3, bottom left and bottom right; difference between black and gray points is similar in each plot). Examination of the data points across vowels indicates that while the difference overall between nasal and non-nasal contexts is similar, the absolute scores within a particular context appear more similar to one another for the filtered HONC plot (lower right panel) compared to the broadband HONC plot (lower left panel).

ANOVA results are summarized in Table III. The results confirm that both broadband and filtered HONC scores were significantly higher for vowels produced in nasal contexts relative to the respective scores computed for vowels produced in non-nasal contexts (Table III, $p_{adj} < 0.001$ for factor "context" for both filter types). Significant effects of vowel place were observed for both filter types (Table III; $p_{adj} < 0.001$ for factor "vowel place" for both filter types). However, a comparison of η^2_G effect sizes indicates that the amount of variation in the data due to vowel place was larger for broadband data compared to the filtered data (Table III; compare η^2_G values of 0.373 versus 0.058). The broadband data also exhibited a significant interaction between vowel loading and consonant context; the interaction term was not found to be significant when an ANOVA was conducted on the low-frequency HONC scores (Table III; $p_{adj} = 0.007$ for

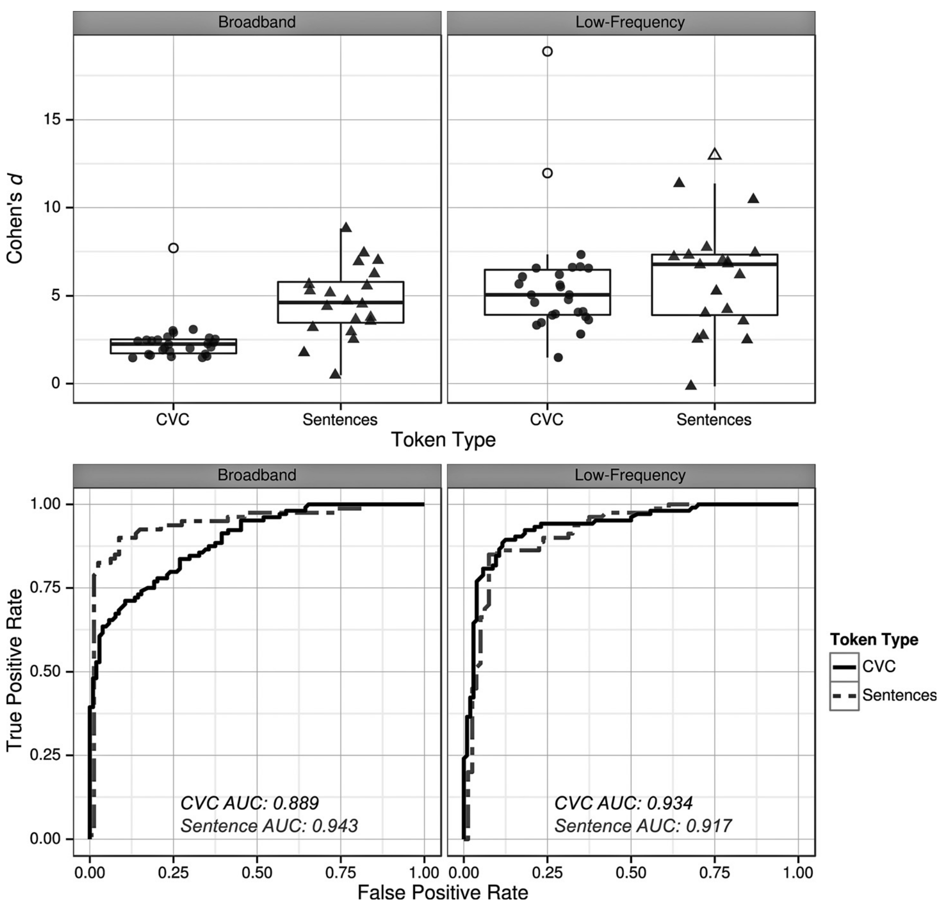

FIG. 2. Upper panels: Cohen's d for individual speakers for each computation method and for each token type. Upper and lower bounds of the boxes indicate first and third quartiles; the line through the box indicates the median of the data. Outliers are indicated by unfilled markers. Lower panels: ROC curves for each computation method for CVC tokens (solid black lines) and sentence tokens (dashed gray lines).

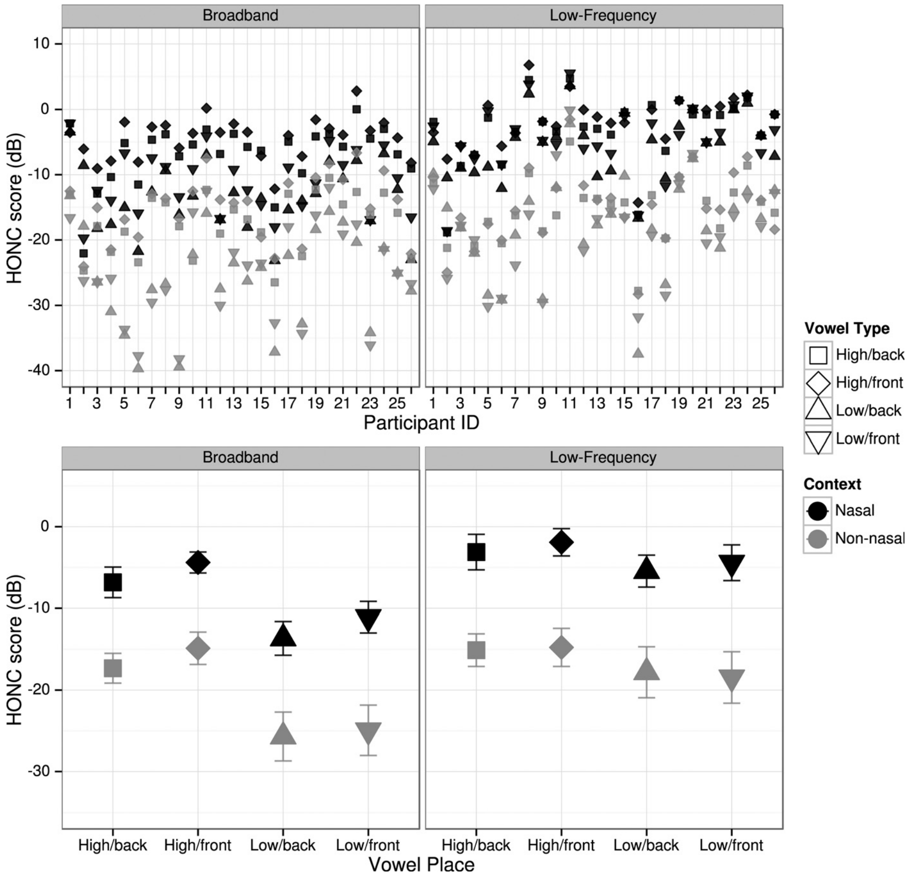

FIG. 3. Mean HONC scores for CVC tokens over all participants (upper panels) and combined across participants (lower panels, $n = 26$). Data are further broken down by context (gray markers for non-nasal tokens, black markers for nasal tokens) and vowel (square, diamond, upwards triangle, and downwards triangle for high/back, high/front, low/back, and low/front vowels, respectively). Upper left: Broadband HONC scores for all participants. Upper right: Filtered HONC scores for all participants. Lower left: Mean broadband HONC scores as a function of vowel placement. Lower right: Mean low-frequency HONC scores as a function of vowel placement. Error bars indicate 95% confidence intervals.

broadband HONC scores, $p_{adj} = 0.133$ for low-frequency HONC scores).

Post hoc pairwise *t*-tests across levels in the vowel factor indicate that there was more vowel-related variability in the broadband data than in the filtered data. In the broadband data, a significant difference was observed for all comparisons between vowel place ($p_{adj} = 0.007$ for the comparison between low/back scores and low/front scores; $p_{adj} < 0.001$ otherwise). For the filtered data, pairwise comparisons were not significant when comparisons were made between high vowels ($p_{adj} = 0.082$) and when comparisons were made between low vowels ($p_{adj} \approx 1$), but were significant when comparisons were made between data for a high vowel and data for a low vowel ($p_{adj} < 0.001$ for these remaining cases).

TABLE III. Repeated measures ANOVA results for vowel tokens. Separate ANOVAs were conducted for each broadband and filtered HONC scores. Column heading abbreviations: $DF_{N,adj}$, numerator degrees of freedom; $DF_{D,adj}$, denominator degrees of freedom; F, F-ratio; p_{adj}, p value; η^2_G, generalized eta-squared (a measure of effect size on the data due to that factor; see Olejnik and Algina, 2003; Bakeman, 2005) *denotes significant at $p = 0.05$.

Score type	Effect	$DF_{N,adj}$	$DF_{D,adj}$	F	p_{adj}	η^2_G
Broadband	Nasal context	1.00	25.00	351.57	<0.001*	0.545
	Vowel place	1.73	43.33	79.75	<0.001*	0.373
	Interaction	1.84	46.00	5.78	0.007*	0.017
Low-frequency	Nasal context	1.00	25.00	284.80	<0.001*	0.551
	Vowel place	1.93	48.27	17.97	<0.001*	0.058
	Interaction	2.21	55.14	2.06	0.133	0.004

2. Sentence tokens

HONC scores for sentence tokens broken down by vowel loading are depicted in Fig. 4, with corresponding ANOVA results summarized in Table IV. As in the CVC token data, broadband HONC scores appeared to show less across-subject variability but greater within-subject variability than filtered scores (Fig. 4, upper panels; compare the spread of black and gray points across participants and within participants in the left panel to the right panel). Visually, effects of vowel loading were observed but appeared to be less severe for the filtered data compared to the broadband data (compare within-participant distributions of black markers and gray markers in the upper panels of Fig. 4 and the position of the black and gray markers across vowel in the bottom panels of Fig. 4). ANOVA results again confirm that HONC scores are significantly higher for sentences loaded with nasal consonants compared to sentences loaded with non-nasal consonants, regardless of how signals were filtered (Table IV, $p_{adj} < 0.001$ for the nasal context factor in each ANOVA dataset). ANOVA also revealed the effect of vowel loading to be significant for both methods of score computation (Table IV, $p_{adj} < 0.001$ for the broadband data and $p_{adj} = 0.038$ for the filtered data). Examination of effect sizes indicated that the effect of vowel type for the filtered data was smaller relative to the broadband data (Table IV; compare η^2_G values of 0.106 for the broadband data versus 0.010 for the filtered data). *Post hoc* testing showed that pairwise comparisons involving high and low vowels were significant for the broadband data ($p_{adj} < 0.001$ in these four

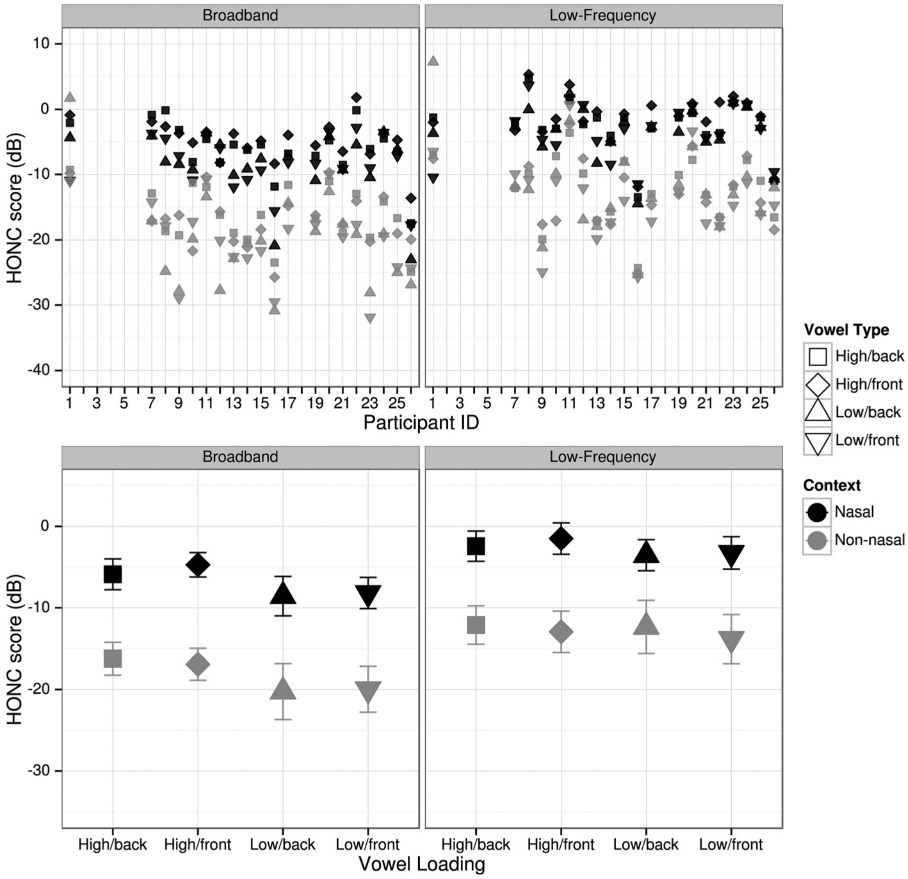

FIG. 4. Mean HONC scores for sentence tokens over all participants (upper panels) and combined across participants (lower panels, $n = 20$). Reliable sentence data were unable to be collected from six participants. As in Fig. 2, data are further broken down by nasal/non-nasal context and vowel loading. Upper left: Broadband HONC scores for all participants. Upper right: Filtered HONC scores for all participants. Lower left: Mean broadband HONC scores as a function of vowel placement. Lower right: Mean low-frequency HONC scores as a function of vowel placement. Error bars indicate 95% confidence intervals.

comparisons; $p_{adj} \approx 1$ otherwise). For the filtered data, significant differences were obtained when comparing scores for low/front vowel loading to both scores for high/back vowel loading ($p_{adj} = 0.045$) and scores for high/front vowel loading ($p_{adj} = 0.042$). All remaining pairwise comparisons were not significant ($p_{adj} \approx 1$).

IV. DISCUSSION

We computed HONC scores in a group of typically developing children for vowels and for running speech using accelerometer and microphone signals, varying the spectral content that was included in the calculations. Based on examination of effect sizes, ROC analysis, and ANOVA, we found that HONC scores could accurately discriminate between vowels produced in nasal and non-nasal contexts as well as between sentences loaded with nasal consonants and sentences loaded with non-nasal consonants, regardless of the filter settings used when HONC scores are computed. Furthermore, evaluation of effect sizes (both Cohen's d and η^2_G) indicated that variation due to vowel place within participants was diminished when filtering was used to band limit the spectral contents of the accelerometer signal from 400–1000 Hz and the microphone signal from 25–420 Hz prior to computation of the HONC score compared to when broadband signals were used in computation of the HONC score.

TABLE IV. ANOVA results for sentence tokens; separate ANOVAs were conducted for broadband and filtered HONC scores. Abbreviations as in Table III.

Score type	Effect	$DF_{N,adj}$	$DF_{D,adj}$	F	p_{adj}	η^2_G
Broadband	Nasal context	1.00	19.00	176.29	<0.001*	0.587
	Vowel place	3.00	57.00	20.59	<0.001*	0.106
	Interaction	1.97	37.51	1.09	0.346	0.005
Low-frequency	Nasal context	1.00	19.00	114.99	<0.001*	0.505
	Vowel place	3.00	57.00	3.01	0.038*	0.011
	Interaction	1.99	37.83	2.97	0.063	0.010

A. HONC scores can distinguish nasalized and non-nasalized utterances in children, but exhibit within- and across-speaker variability

The present study confirms the ability of HONC scores to distinguish between nasal and non-nasal speech found in earlier studies. The overall difference in HONC scores observed here for nasal and non-nasal speech for each filter setting (in the range of 10–13 dB) is comparable to the differences reported in previous literature examining the speech of very young children (13 dB, Mra *et al.*, 1998). We also note that this magnitude is similar to what has been reported for HONC scores in adults (also around 13 dB, as in Sussman, 1995 and Thorp *et al.*, 2013). The similarity in scores across adults and children could arise from the participant-specific scaling employed in the HONC score; scaling the accelerometer-to-microphone ratio for a given token by the same ratio for a maximally nasal sound (here, an /m/), may serve to reduce variation of the HONC score as a function of vocal tract length.

Some variation is to be expected in any sort of speech measurement data due to naturally occurring differences across repetitions of the same words or sentences as well as due to slight differences in other factors, such as microphone or accelerometer positioning. While little can be done about the former source of variability with the exception of instructing speakers to be more consistent, HONC scores should be minimally sensitive to the latter due to the per-recording calibration procedure that is employed. Yet despite the good overall discriminability between nasal and non-nasalized speech for both filtered and broadband HONC scores, we noted considerable intra- and inter-participant variability, as evidenced by the non-zero intercept and non-unity slope of the regression lines in Fig. 1, and the overall spread of individual participant data points in the remaining figures. Some of the observed variation in the current study can be attributed to the sometimes non-cooperative nature of the young participants. Anecdotally, we observed that participants, even when extremely compliant, had difficulty sitting still or reproducing speech tokens consistently even across repetitions of the same speech token, even when encouraged by the experimenter or the participant's parent. The presence of additional children-specific variability in HONC scores is supported by comparisons of effect sizes in the current study to previous studies examining HONC scores in adults. Specifically, while the difference in dB between HONC scores for nasalized and non-nasalized speech are similar for children (current study; Mra *et al.*, 1998) and adults (Sussman, 1995; Thorp *et al.*, 2013), the range of d values obtained here for the broadband HONC scores are smaller than those that have been reported for adults (see Thorp *et al.*, 2012, Fig. 3, "NATAR" condition). The same holds true for HONC scores obtained using low-frequency signals (see Thorp *et al.*, 2013, Fig. 3, "NNA" scores). The differences in the range of effect sizes across studies can therefore be traced to differences in the combined standard deviation used as the denominator in computation of d.

The inter-participant variability observed here indicates that it is difficult to determine a single "set" threshold for HONC scores that will always classify scores above it as nasal and scores below it as non-nasal, regardless of the type of signals used in its computation. This is seen most clearly in the top panels of Figs. 3 and 4; picking a threshold level that will cleanly split all of the gray and black markers in each plot is impossible. While ROC curves of the type depicted in Fig. 2 can aid in the task of determining an "optimal" split by choosing a threshold associated with the point on the curve closest to the upper left corner of the chart, the variability observed here coupled with the expected variability of children's speech in general suggests that the difference between HONC scores between speech that is supposed to be nasalized and speech that is supposed to be non-nasalized may be a more robust indicator of proper nasalization than absolute scores alone.

We noted that variability in the HONC scores in general appeared less pronounced for sentence stimuli than for vowel stimuli, regardless of filter settings. The most likely explanation is that the contribution of consonants in the signals used in computing the sentence HONC scores reduced the variability in scores arising from the vowels. The contribution of consonants toward the HONC score is also likely to be responsible for the non-significant difference in d values between low-frequency HONC scores and broadband HONC scores, as the filtering used in this study may not have appropriately diminished the spectral contributions of consonants to the score computed for running speech.

B. Appropriate filtering of the acoustic and accelerometer signals can reduce vowel-related effects on HONC scores

Numerous studies utilizing the HONC score have reported that HONC scores for high vowels tend to be larger than HONC scores for low vowels (e.g., Larson and Hamlet, 1987; Sussman, 1995; Mra *et al.*, 1998; Thorp *et al.*, 2013), a finding replicated in the current study. It has also been reported that low-frequency resonances present for high vowels (i.e., resonances associated with the low first formant frequency in high vowels) are transmitted through the soft palate and other structures surrounding the nasal cavity, resulting in greater vibrations in nasal tissue during the production of high vowels (Stevens *et al.*, 1976; Sussman, 1995; Kummer, 2011a). The effects of oral resonances on nasal tissue vibrations offers an explanation for seeing higher HONC scores in general for high vowels regardless of whether the consonantal context is nasal or non-nasal, as opposed to higher HONC scores for nasalized high vowels but not for non-nasalized high vowels.

The reduction of vowel place effects when HONC scores are computed using low-frequency portions of the accelerometer and microphone signals has been reported for adult speech (Thorp *et al.*, 2013), and was confirmed to be the case for children's speech in the current study. Within-participant variability specifically due to vowel place/loading in HONC scores was observed to be lower when filtering the signals used in its computation. The reduction in variability that filtering provides in these data is evident when comparing Cohen's d values between filtering conditions (Fig. 2, upper panels), and when examining absolute scores (Figs. 3 and 4) and effect sizes (η^2_G) computed from ANOVA on the absolute scores (Tables III and IV). Since the differences between nasal and non-nasal scores in dB are roughly consistent for both broadband and filtered HONC scores, the increase in effect size seen for HONC scores computed with low-frequency signals can be attributed to the lower combined standard deviation of low-frequency HONC scores relative to the combined standard deviation of HONC scores computed with broadband signals. Therefore, the d scores, which for the broadband HONC scores are already indicative of a "large" effect within participants (i.e., high discriminability between the two classes; Cohen, 1992) are increased even further. Additionally, the effect size η^2_G can also be classified into "small," "medium," and large effect sizes as per Cohen's guidelines (Bakeman, 2005). A comparison of effect sizes for the vowel place factor in the ANOVA on the CVC scores shows that using these guidelines, filtering reduces the effect of vowel from a large effect size ($\eta^2_G = 0.373$) to a small/medium effect size ($\eta^2_G = 0.058$). For sentence scores, filtering reduced a small/medium effect size for the vowel place ($\eta^2_G = 0.106$) to be essentially negligible ($\eta^2_G = 0.011$). This has implications for the clinical translation of HONC measures: Sentence stimuli provide more functional material for assessment, and would not require the same extent of manual offline processing (in particular, the identification and isolation of individual word tokens and/or vowels. Although vowel effects are minimal in general for sentence stimuli, our results demonstrate that they have the potential to be effectively removed with the appropriate filtering. As such, HONC scores computed using low-frequency signals could be implemented at a sentence level without specific regard to vowel contexts of speech stimuli.

The findings of reduced vowel-related variability when using low-frequency portions of the accelerometer and microphone signals in computation of the HONC score are perhaps unsurprising from a mechanistic standpoint. Previous literature on the acoustic correlates of nasalization has indicated a variety of changes to the spectra of vowels due to nasalization (e.g., see House and Stevens, 1956; Schwartz, 1968; Hawkins and Stevens, 1985; Pruthi, 2007). Some of the most obvious changes during nasalization of vowels occur in the low-frequency portion of the spectra. In particular, nasalization tends to broaden and flatten the spectral peak associated with the first formant, reducing formant energy differences between vowels below about 1 kHz (House and Stevens, 1956; Schwartz, 1968; Hawkins and Stevens, 1985). While the spectral content of the nasal acceleration signal has not traditionally been well characterized, the lower cutoff of the accelerometer filter used here (400 Hz) is above the canonical first formant frequencies associated with high vowels in the combined oral-nasal signal. Similarly, the upper cutoff used for the microphone signal (420 Hz) would largely eliminate first formant structure in low vowels. Further work is necessary to characterize the spectral content of the nasal acceleration signals and how it relates to the spectral content of combined oral-nasal output. This information would allow for the effects of filtering on the nasal acceleration signal and the effects of filtering on the microphone signal on HONC scores to be partitioned out.

C. Directions for future research

We note that while this work contributes to the literature demonstrating that HONC scores can be utilized to distinguish nasalized from non-nasalized utterances in typical speakers, it does not directly address the question of whether HONC scores are sensitive to the degree of hypernasality in disordered speakers. Furthermore, while filtering HONC scores does indeed increase the effect size between nasalized and non-nasalized utterances, it is unclear how filtering the microphone and accelerometer signals affects the relationship between the degree of hypernasality and the HONC score. It is also unclear whether the filter cutoffs used here, that were empirically determined using the speech of adults in Thorp *et al.* (2013), would be suitable for disorderd speakers. Future studies should involve the computation of HONC scores with various spectral portions of the signal isolated for children who have been diagnosed with hypernasality to better understand the correlation between HONC scores, the spectral content of the speech and accelerometer signals, the level of hypernasality in the speech, and how each of these factors relate to the structure and operation of the velum and velopharyngeal opening.

The choice of denominator in the HONC score may also be suitable for additional investigation as a way to obtain more consistent scores, particularly in children. HONC scores can be obtained by scaling the nasal acceleration signal with either the signal obtained from a laryngeal contact microphone or the signal obtained from a microphone measuring combined oral-nasal output (Horii, 1980; Horii and Lang, 1981). One advantage of the laryngeal accelerometer is that it is less influenced by the formant structure arising from vocal tract configuration, and thus may reduce vowel-related variability in the HONC score even in the absence of filtering. Second, the nasal acceleration signal may scale more linearly with an increasing laryngeal signal compared to an oral microphone signal (Horii and Lang, 1981), as the relationship between the amplitude of nasal acceleration signals and the amplitude of the acoustic waveform detected using a microphone has been reported to be nonlinear (Stevens *et al.*, 1976; Garber and Moller, 1979). Assuming children are cooperative with having another sensor placed on their body, simultaneous acquisition of the laryngeal signal should not be difficult or expensive to achieve given the proliferation of relatively inexpensive, commercially available multi-channel audio recording devices in recent years.

Finally, the relationship between articulatory kinematics and nasalization is complex and remains poorly understood. While the HONC score is a conceptually simple correlate of nasalization, it is unknown how well HONC scores correlate with velopharyngeal port area and how changes in the scores are affected by specific articulatory movements. For example, vowel-related effects on nasalization, specifically the increased HONC scores observed for high vowels relative to low vowels, may be directly related to oral cavity constriction and velar height (Lubker and Moll, 1965; Kummer, 2011a), but are inversely related to velopharyngeal port area during the production of vowels (Moll, 1962). Furthermore, nasalization may result in complex articulatory changes due to speakers modifying their articulations to counteract the self-perception of nasality (Carignan *et al.*, 2011). Toward this end, it may be a worthwhile endeavor to record the nasal acceleration signal while imaging the velopharyngeal opening using endoscopy (e.g., Ramamurthy *et al.*, 1997) or other photosensor-based methods (e.g., Dalston, 1982, 1989), as well as while performing articulatory motion tracking (e.g., Carignan *et al.*, 2011), to better characterize how nasal acceleration relates to the physiology of speech production.

D. Conclusions

HONC scores, due to the simple and inexpensive instrumentation necessary for their calculation, are promising tools for SLPs and for those undergoing speech rehabilitation for hypernasality. Filtering the signals used in the computation of HONC scores can reduce vowel-related variability present in HONC scores, perhaps by better isolating spectral features associated with nasal resonance. With a suitable user interface wrapped around the data acquisition and score interpretation, the scores have the potential to be utilized as a home or school training aid. While this idea dates back several decades (Nickerson *et al.*, 1976), the cost and ease of implementation has dropped significantly due to advances in digital recording techniques and computational capabilities of modern computers. We note that aside from a personal computer, the required equipment (accelerometer, microphone, and a compatible sound card if XLR-type connections are required) can be obtained for a few hundred dollars. Yet despite the potential utility and ease with which HONC scores can be obtained, additional research is necessary to understand the acoustic and kinematic factors contributing to variations in HONC scores.

ACKNOWLEDGMENTS

This work was supported in part by the Deborah Munroe Noonan Memorial Research Fund and Grant No. DC012651 from the National Institute of on Deafness and Other Communication Disorders (NIDCD) to C.E.S., as well as the Diane M. Bless Endowed Chair, Division of Otolaryngology-Head and Neck Surgery at University of Wisconsin, Madison. The authors would like to thank Emma Billard, Carolyn Michener, and Elias Thorp for their assistance with data collection.

Abdi, H. (**2010**). "Greenhouse-Geisser correction," in *Encyclopedia of Research Design*, edited by N. Salkind (Sage, Thousand Oaks, CA), pp. 545–549.

Ali, L., Gallagher, T., Goldstein, J., and Daniloff, R. (**1971**). "Perception of coarticulated nasality," J. Acoust. Soc. Am. **49**, Suppl. 2, 538–540.

Andrews, J. R., and Rutherford, D. (**1972**). "Contribution of nasally emitted sound to the perception of hypernasality of vowels," Cleft Palate J. **9**, 147–156.

Audacity (http: //audacity.sourceforge.net) (Last viewed July 25, 2014).

Awan, S. N., Omlor, K., and Watts, C. R. (**2011**). "Effects of computer system and vowel loading on measures of nasalance," J. Speech Lang. Hear. Res. **54**, 1284–1294.

Bakeman, R. (**2005**). "Recommended effect size statistics for repeated measures designs," Behav. Res. Methods **37**, 379–384.

Blood, G. W., and Hyman, M. (**1977**). "Children's perception of nasal resonance," J. Speech Hear. Disord. **42**, 446–448.

Braden, M. N., Varghese, L. A., and Stepp, C. E. (**2013**). "Application of normalized nasal acceleration to children with and without cleft palate," presented at the annual American Speech and Hearing Association (ASHA) Convention, November 14–16, Chicago, IL.

Bradford, L. J., Brooks, A. R., and Shelton, R. L. (**1964**). "Clinical judgment of hypernasality in cleft palate children," Cleft Palate J. **1**, 329–335.

Bradley, A. P. (**1997**). "The use of the area under the ROC curve in the evaluation of machine learning algorithms," Pattern Recogn. **30**, 1145–1159.

Carignan, C., Shosted, R., Shih, C., and Rong, P. (**2011**). "Compensatory articulation in American English nasalized vowels," J. Phonetics **39**, 668–682.

Cohen, J. (**1992**). "A power primer," Psychol. Bull. **112**, 155–159.

Counihan, D. T., and Cullinan, W. L. (**1972**). "Some relationships between vocal intensity and rated nasality," Cleft Palate J. **9**, 101–108.

Dalston, R. M. (**1982**). "Photodetector assessment of velopharyngeal activity," Cleft Palate J. **19**, 1–8.

Dalston, R. M. (**1989**). "Using simultaneous photodetection and nasometry to monitor velopharyngeal behavior during speech," J. Speech Hear. Res. **32**, 195–202.

Garber, S. R., and Moller, K. T. (**1979**). "The effects of feedback filtering on nasalization in normal and hypernasal speakers," J. Speech Hear. Res. **22**, 321–333.

Hawkins, S., and Stevens, K. N. (**1985**). "Acoustic and perceptual correlates of the non-nasal-nasal distinction for vowels," J. Acoust. Soc. Am. **77**, 1560–1575.

Horii, Y. (**1980**). "An accelerometric approach to nasality measurement: A preliminary report," Cleft Palate J. **17**, 254–261.

Horii, Y. (**1983**). "An accelerometric measure as a physical correlate of perceived hypernasality in speech," J. Speech Hear. Res. **26**, 476–480.

Horii, Y., and Lang, J. E. (**1981**). "Distributional analyses of an index of nasal coupling (HONC) in simulated hypernasal speech," Cleft Palate J. **18**, 279–285.

House, A. S., and Stevens, K. N. (**1956**). "Analog studies of the nasalization of vowels," J. Speech Hear. Disord. **21**, 218–232.

Kent, R. D. (**1996**). "Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders," Am. J. Speech Lang. Pathol. **5**, 7–23.

Kummer, A. W. (**2011a**). "Disorders of resonance and airflow secondary to cleft palate and/or velopharyngeal dysfunction," Semin. Speech Lang. **32**, 141–149.

Kummer, A. W. (**2011b**). "Types and causes of velopharyngeal dysfunction," Semin. Speech Lang. **32**, 150–158.

Kummer, A. W., Clark, S. L., Redle, E. E., Thomsen, L. L., and Billmire, D. A. (**2012**). "Current practice in assessing and reporting speech outcomes of cleft palate and velopharyngeal surgery: A survey of cleft palate/craniofacial professionals," Cleft Palate Craniofac. J. **49**, 146–152.

Larson, P., and Hamlet, S. (**1987**). "Coarticulation effects on the nasalization of vowels using nasal/voice amplitude ratio instrumentation," Cleft Palate J. **24**, 286–290.

Lawrence, M. A. (**2013**). "Package 'ez,'" Available online: http://cran.rproject.org/web/packages/ez/ez.pdf (Last viewed July 25, 2014), 43 pp.

Lee, S., Potamianos, A., and Narayanan, S. (**1999**). "Acoustics of children's speech: developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**, 1455–1468.

Lewis, K. E., Watterson, T., and Quint, T. (**2000**). "The effect of vowels on nasalance scores," Cleft Palate J. **37**, 584–589.

Lewis, K. E., Watterson, T. L., and Houghton, S. M. (**2003**). "The influence of listener experience and academic training on ratings of nasality," J. Commun. Disord. **36**, 49–58.

Lippmann, R. P. (**1981**). "Detecting nasalization using a low-cost miniature accelerometer," J. Speech Hear. Res. **24**, 314–317.

Lubker, J. F., and Moll, K. L. (**1965**). "Simultaneous oral-nasal air flow measurements and cinefluorographic observations during speech production," Cleft Palate J. **2**, 257–272.

McWilliams, B. J. (**1954**). "Some factors in the intelligibility of cleft-palate speech," J Speech Hear. Disord. **19**, 524–528.

Moll, K. L. (**1962**). "Velopharyngeal closure on vowels," J. Speech Hear. Res. **5**, 30–37.

Mra, Z., Sussman, J., and Fenwick, J. (**1998**). "HONC measures in 4- to 6-year-old children," Cleft Palate Craniofac. J. **35**, 408–414.

Nickerson, R. S., Kalikow, D. N., and Stevens, K. N. (**1976**). "Computer-aided speech training for the deaf," J. Speech Hear. Disord. **41**, 120–132.

Olejnik, S., and Algina, J. (**2003**). "Generalized eta and omega squared statistics: Measures of effect size for some common research designs," Psychol. Methods **8**, 434–447.

Oliphant, T. (**2007**). "Python for scientific computing" Comput. Sci. Eng. **9**, 10–20.

Praat (http://www.fon.hum.uva.nl/praat) (Last viewed July 25, 2014).

Pruthi, T. (**2007**). "Analysis, vocal tract modeling and automatic detection of vowel nasalization," Ph.D. thesis, University of Maryland, College Park.

Ramamurthy, L., Wyatt, R. A., Whitby, D., Martin, D., and Davenport, P. (**1997**). "The evaluation of velopharyngeal function using flexible nasendoscopy," J. Laryngol. Otol. **111**, 739–745.

R Core Team (**2013**). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria), 3349 pp.

Redenbaugh, M. A., and Reich, A. R. (**1985**). "Correspondence between an accelerometric nasal/voice amplitude ratio and listeners' direct magnitude estimations of hypernasality," J. Speech Hear. Res. **28**, 273–281.

Schwartz, M. F. (**1968**). "The acoustics of normal and nasal vowel production," Cleft Palate J. **5**, 125–140.

Stelck, E. H., Boliek, C. A., Hagler, P. H., and Rieger, J. M. (**2011**). "Current practices for evaluation of resonance disorders in North America," Semin. Speech Lang. **32**, 58–68.

Stevens, K. N., Kalikow, D. N., and Willemain, T. R. (**1975**). "A miniature accelerometer for detecting glottal waveforms and nasalization," J. Speech Hear. Res. **18**, 594–599.

Stevens, K. N., Nickerson, R. S., Boothroyd, A., and Rollins, A. M. (**1976**). "Assessment of nasalization in the speech of deaf children," J. Speech Hear. Res. **19**, 393–416.

Sussman, J. E. (**1995**). "HONC measures in men and women: validity and variability," Cleft Palate Craniofac. J. **32**, 37–48.

Thorp, E. B., Virnik, B. T., and Stepp, C. E. (**2012**). "Normalization strategies for nasal acceleration to assess velopharyngeal function," Conf. Proc. IEEE Eng. Med. Biol. Soc. **2012**, 6459–6462.
Thorp, E. B., Virnik, B. T., and Stepp, C. E. (**2013**). "Comparison of nasal acceleration and nasalance across vowels," J. Speech Lang. Hear. Res. **56**, 1476–1484.
Turner, H. M. I., and Bernard, R. M. (**2006**). "Calculating and synthesizing effect sizes," Contemp. Issues Commun. Sci. Disord. **33**, 42–55.
Wickham, H. (**2009**). *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York), 213 pp.
Wickham, H. (**2011**). "The split-apply-combine strategy for data analysis," J. Stat. Software **40**, 1–29.