Interspeech 2018
2-6 September 2018, Hyderabad, India

Estimation of Hypernasality Scores from Cleft Lip and Palate Speech

C. M. Vikram[1], *Ayush Tripathi*[2], *Sishir Kalita*[1], *and S. R. Mahadeva Prasanna*[1,3]

[1]Indian Institute of Technology Guwahati, Guwahati, India
[2]Visvesvaraya National Institute of Technology, Nagpur, India
[3]Indian Institute of Technology Dharwad, Dharwad, India

{cmvikram, sishir, prasanna}@iitg.ernet.in, ayushtripathi1811@gmail.com

Abstract

Hypernasality refers to the perception of excessive nasal resonances in vowels and voiced consonants. Existing speech processing based approaches concentrate only on the classification of speech into normal or hypernasal, which do not give the degree of hypernasality in terms of continuous values like nasometer. Motivated by the functionality of nasometer, in this work, a method is proposed for the evaluation of hypernasality. Speech signals representing two extremely opposite cases of nasality are used to develop the acoustic models, where oral sentences (rich in vowels, stops, and fricatives) of normal speakers and nasal sentences (rich in nasals and nasalized vowels) of moderate-severe hypernasal speakers represent the groups with minimum and maximum attainable degrees of nasality, respectively. The acoustic features derived from glottal activity regions are used to model the maximum and minimum nasality classes using Gaussian mixture model and deep neural network approaches. The posterior probabilities obtained for nasal sentence class are referred to as hypernasality scores. The scores show a significant correlation ($p < 0.01$) with respect to perceptual ratings of hypernasality, provided by expert speech-language pathologists. Further, hypernasality scores are used for the detection of hypernasality, and the results are compared with the nasometer based approach.

Index Terms: Posterior probability, Hypernasality, Nasal sentences and oral sentences, Nasometer.

1. Introduction

Speech of individuals with cleft lip and palate (CLP) is characterized by the presence of hypernasality and articulation errors. In speakers with CLP, the presence of velopharyngeal insufficiency (VPI) or oro-nasal fistula leads to increase in the loss of acoustic energy through the nasal cavity, which results in hypernasal speech. Hypernasality refers to the perception of excessive nasal resonances on vowels and voiced consonants [1, 2]. Hypernasality is considered as an important parameter during evaluation of the outcome of surgery and speech therapy of the individuals with CLP. Clinically, hypernasality is evaluated using perceptual methods by trained speech-language pathologists (SLPs). The perceptual evaluation is considered as a golden standard, however, the results are subjective in nature. Alternatively, different instrumental methods have been proposed for the assessment of VPI and hypernasality, which are reviewed in detail in Ref. [3].

Among the different instrumental based hypernasality assessment techniques, nasometer is used widely in the clinical and research applications [3]. Nasometer uses two microphones to collect the acoustic signals from nasal and oral cavities, where a baffle plate is used to separate between upper lip and nose. Nasometer gives an objective measure of nasality, which is referred to as nasalance score. The nasalance score has a range of 0-100, which showed good correlation with the perceptual judgment of hypernasality. Also, the nasometer can be used for the biofeedback training [3]. However, nasometer cannot be operated on stored data and also, requires subject's cooperation and technically trained persons to handle it.

Apart from perceptual and instrumental methods, speech processing based techniques are proposed for the hypernasality analysis, which do not require complex hardware and gives objective evaluation results [3, 4]. The presence of extra-nasal formants around 250 Hz and 1000 Hz in vowel spectrum, increase in the first formant bandwidth, reduction in second formant strength, and increase in the spectral flatness are considered as the important acoustic cues of hypernasality [5, 6, 7]. Mel-frequency cepstral coefficients (MFCCs), glottal source related features (jitter and shimmer), wavelet transform based features, Gaussian mixture model (GMM) and support vector based classifiers have been explored for the hypernasality detection from word and sentence level data [8, 9, 10]. Also, automatic classification of speech into normal, mild, moderate, and severe levels of hypernasality is proposed in Ref.[8, 9], where GMMs are explicitly trained for these classes. However, these methods have limitation to use for the clinical applications. Because, the results are not in the form of continuous scores like nasometer. The degree of hypernasality cannot be determined using existing acoustic methods, which is essential to verify the effect of speech therapy and surgery.

The present work mainly motivated to propose a speech processing based approach to estimate the hypernasality scores. The proposed work uses acoustic features and machine learning approaches to estimate the hypernasality scores. The paper is organized as follows. Section 2 describes the database for the analysis of hypernasality. Section 3 explains the proposed approach for the hypernasality score estimation. The results of proposed system and comparison with nasometer are presented in Section 4. Finally, Section 5 gives conclusion and mentions the scope for future work.

2. Database

In this work, 31 children with repaired cleft lip and palate/cleft palate (RCLP) of age range 6-12 years are considered. 48 typically developed children with age and gender-matched are considered as the controlled normal (CN) group. Both CN and CLP subjects have Kannada as their mother tongue (Kannada is a Dravidian language, spoken in the southern part of India). The children with RCLP are diagnosed at Unit for Structural and Cranio-facial Anomalies (USOFA), All India Institute of Speech and Hearing (AIISH), Mysore, India [11] using nasometer and perceptual assessment techniques. All children considered do not have a history of hearing loss and language disor-

10.21437/Interspeech.2018-1631

ders.

The speech stimuli consist of 6 oral and 6 nasal sentences prepared in the Kannada language. Here, oral sentences refer to sentences rich in vowels, stops, and fricatives, whereas nasal sentences contain the nasal consonants and nasalized vowels. Examples for the oral sentence: [*kaage kaalu kappu*] and nasal sentence: [*manu aneyannu nodida*]. All these stimuli are prepared by expert SLPs of AIISH and regularly used for the clinical assessment of hypernasality. These sentences are recorded in a sound treated room using sound level meter held at a distance of 15 cm from the subject. The recordings are obtained at 48000 Hz sampling frequency and digitized at 16 bit/sample. The samples are recorded at multiple sessions. In total, 416 tokens of oral and 200 tokens of nasal sentences are recorded from CN group and 145 tokens of oral and 300 tokens of nasal sentences are recorded from CLP subjects. Also, for each subject nasometric evaluation is carried out using the oral and nasal sentences. Nasometer II 6400 is used for this purpose, and mean nasalance score of each sentence is noted.

Compared to nasal sentences, oral sentences are more correlated with the degree of perceptual hypernsality [2]. The perceptual evaluation of oral sentences is carried out by three expert SLPs, using Henningsson 4 point rating scale defined in Ref. [2]. The 4 point hypernasality rating scale consists of values in the range of 0-3, where 0-normal, 1-mild, 2-moderate, and 3-severe. Correlation coefficients derived for the ratings of each pair of SLPs, i.e., 1^{st} and 2^{nd}, 2^{nd} and 3^{rd}, 1^{st} and 2^{nd} SLPs are 0.74, 073, and 079, respectively. Also, 30% of samples are given to the same SLP to measure the intra rater reliability. The correlation coefficients 0.95, 0.91, and 0.95 obtained for 1^{st}, 2^{nd}, and 3^{rd} SLPs, respectively as a measure of intra rater reliability. For each sentence, the median value of perceptual scores given by 3 SLPs is computed. Finally, out of 145 oral sentences of CLP, 13, 76, 51, and 5, are categorized into normal, mild, moderate, and severe hypernasality groups. Further, speakers are grouped into normal, mild, moderate, and severe groups based on the maximum number of sentences belonging to that particular class.

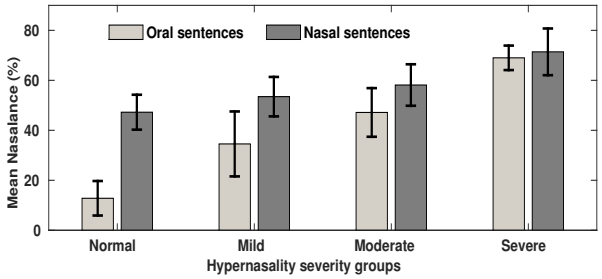

Figure 1: *Bar graph of nasalance scores for oral and nasal sentences of normal, mild, moderate, and severe hypernasal speakers.*

3. Proposed Method

The proposed work for the estimation of hypernasality score is mainly motivated by the nature of nasalance scores of nasometer for oral and nasal sentences. The nasalance score (N) can be computed using the equation given by

$$N = \frac{NE}{OE + NE} \times 100 \quad (1)$$

where NE and OE are the energies of the acoustic signals acquired from nasal and oral microphones, respectively. Mean nasalance scores derived for oral and nasal sentences are indicated in the bar graph shown in Fig. 1. From bar graph the following observations can be drawn:

1. The nasalance scores are high for nasal sentences when compared to oral sentences.
2. The discrimination of nasalance scores between the group of nasal and oral sentences is greater for normal speakers and reduces from mild to moderate and moderate to severe hypernasal speakers. Because in moderate-to-severe hypernasal speakers, most of the oral consonants (/b/, /p/) are replaced by nasals (/m/, n/) and vowels get severely nasalized [2].
3. The oral sentences of the normal group exhibit minimum nasality whereas nasal sentences of the severe group show maximum nasality.

Motivated by the behaviour of nasalance values for target oral and nasal sentences as a function of the degree of hypernasality, the current work proposes an approach for the estimation of hypernasality scores. In this work, a two-class classifier is developed for the oral class: using sentences with minimum nasality (oral sentences of normal speakers) and nasal class: using sentences with maximum attainable nasality (nasal sentences of the moderate-severe hypernasal group). During testing, features derived from the response of speaker for the target oral sentence is given for the classifier. The posterior probabilities derived for the nasal class are considered as hypernasality scores.

The proposed algorithm for hypernasality score estimation consists of different steps, i.e., glottal activity region detection, feature extraction, training of classifier for oral and nasal classes, and posterior probability computation. These steps are explained in the next subsections.

3.1. Glottal activity detection

The current work incorporates the glottal activity detection (GAD) algorithm as a preprocessing stage. GAD algorithm proposed in [12] is used, where glottal activity regions are detected by the combination of the features, namely, the strength of excitation (SoE), normalized autocorrelation peak strength (NAPS), and higher order statistics (HoS). SoE, NAPS, and HoS represent the energy, periodicity, and asymmetry of the glottal source signal, respectively. These three evidences are combined through averaging and using a heuristic threshold equal to 0.6; the speech frames are classified into voiced and unvoiced.

3.2. Feature Extraction

The speech signal is segmented into frames of length equal to 20 ms with a shift of 5 ms. Then 13-dimensional MFCCs, with delta and delta-delta variants are computed only for the frames with the presence of glottal activity. Thus, the dimensionality of each feature vector is equal to 39.

3.3. Classifiers

In this work, Gaussian mixture model (GMM) and deep neural networks (DNN) are trained for the oral sentences of normal speakers and nasal sentences of moderate-severe hypernasal speakers.

3.3.1. Gaussian Mixture Model

Let $\lambda_O = \{\mu_i, \Sigma_i, \omega_i\}_O$, $i = 1, 2, ...M$ represent the GMM model for the class of oral sentences (λ_O) with M number of

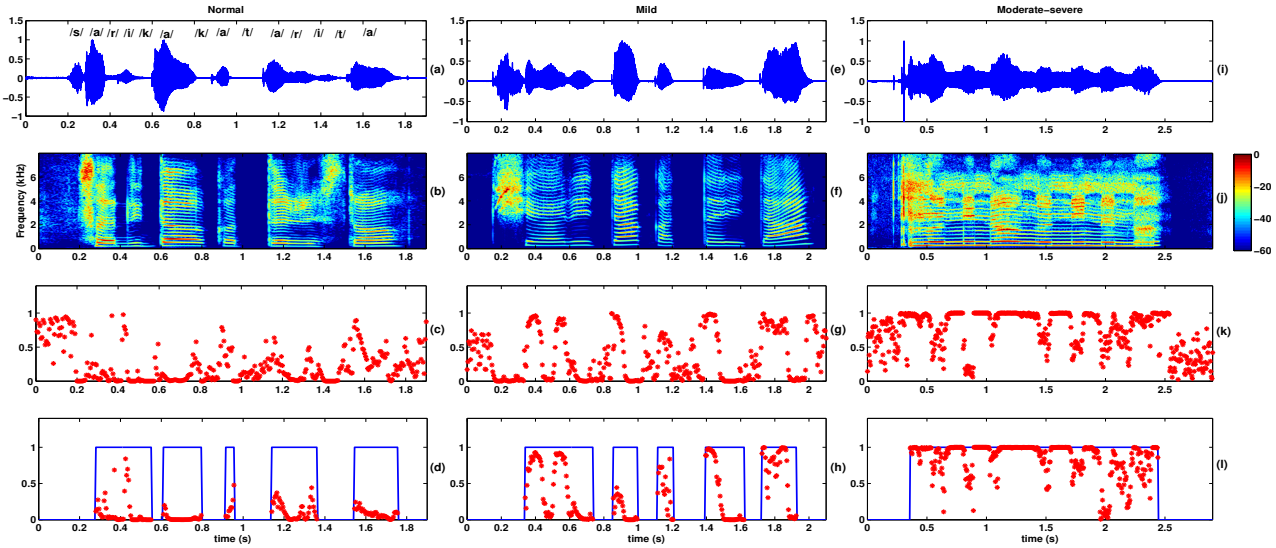

Figure 2: *Illustration of the significance of GAD in the computation of hypernasality scores for the sentence "sarita kattari taa". (a)-(d), (e)-(h), and (i)-(l) represent the speech signal, spectrogram, contour of posterior probabilities scores for hypernasal class without glottal activity and with glottal activity detection, respectively for normal, mild and moderate-severe hypernasal speech. In (d), (h), and (l) dotted lines indicate the posterior probability values and solid lines indicate the detected glottal activity regions. Application of GAD reduces the spurious scores resulting from unvoiced regions.*

mixtures. The parameters $\mu_i, \Sigma_i, \omega_i$ represent mean, covariance matrix, and weight of i^{th} Gaussian, respectively. The mixture parameters are estimated using expectation maximization (EM) algorithm. Similarly, $\lambda_N = \{\mu_i, \Sigma_i, \omega_i\}_N$, $i = 0, 1, 2, ...M$ represent the GMM for the class of nasal sentences (λ_N) belonging to the group of moderate-severe hypernasality.

The tokens corresponding to target oral sentences are used for the testing. The posterior probability of class λ_N given a feature vector x_i, $i = 1, 2...T$, is computed using

$$p(\lambda_N|x_i) = \frac{p(x_i|\lambda_N)p(\lambda_N)}{p(x_i|\lambda_O)p(\lambda_O) + p(x_i|\lambda_N)p(\lambda_N)} \quad (2)$$

where class priori probabilities $p(\lambda_N)$ and $p(\lambda_O)$ chosen equal to 0.5 by assuming two classes are equi-probable, $p(x_i|\lambda_O)$ and $p(x_i|\lambda_N)$ are likelihood values estimated from GMMs λ_O and λ_N, respectively.

The likelihood of nasal class $p(x_i|\lambda_N)$ and oral class $p(x_i|\lambda_O)$ in equation 2 are proportional to the nasal and oral sound characteristics present in the speech signal, respectively. The terms NE and OE in equation 1 are proportional to the nasal and oral sound energies, respectively. Since the nasalance value in equation 1 is proportional to the amount of nasal energy present in the speech signal. Similarly, the posterior probability value $p(\lambda_N|x_i)$ in equation 2 is proportional to the presence of nasal sound characteristics in the speech signal. Therefore, due to the existence of the similarity between the computation of nasalance score and the posterior probability of nasal class (equations 1 and 2), the posterior probability scores are expected to give a measure of nasality.

The models λ_O and λ_N are trained for the GMMs with 64 mixtures, diagonal covariance matrix using EM algorithm. Posterior probabilities $p(\lambda_N|x_i)$ computed for the target oral sentence "*sarita kattari taa*" for different levels of hypernasality are illustrated in Fig. 2. Fig. 2(a)-(d) show the waveform of normal speaker for the given target sentence, spectrogram, contours of $p(\lambda_N|x_i)$ without the inclusion of GAD pre-processing stage, and with the inclusion of GAD, respectively. Here, without GAD pre-processing stage refers to training and testing of models are carried out without the inclusion of GAD preprocessing stage. The contour of $p(\lambda_N|x_i)$ for normal speech (Fig. 2(c)) shows that the exclusion of GAD pre-processing stage results in the spurious values at unvoiced regions like fricatives, stop gaps and silence regions (example: around 0.8 seconds corresponding to stop gap of /k/). This will result in an increased hypernasality score for the normal speaker and increases the false alarm rate. Similarly, Fig. 2(e)-(h) and (i)-(l) show the speech waveform, spectrogram, contours of $p(\lambda_N|x_i)$ without and with inclusion of GAD for mild and moderate-severe hypernasal groups, respectively. The $p(\lambda_N|x_i)$ values are increased for the mild group than the normal, severe group than that of mild. This shows that the $p(\lambda_N|x_i)$ values vary in proportional to the severity of perceived hypernasality. Glottal activity region based processing significantly reduces the spurious hypernasality scores resulted from the unvoiced/silence regions and hence, reduces false alarm rate.

3.3.2. Deep Neural Network

Deep neural network (DNN) based hypernasality detection system is developed using Keras [13] toolkit. DNN structure has 3 hidden layers, with ReLU as the activation function for the hidden layers, and softmax as the activation function for the output layer. The hidden layer has 100 neurons. The 39-dimensional MFCCs are the input features for DNN. The output layer of 2 dimensions corresponding to oral and nasal classes. DNN is trained with random initialization of weights and biases. The mini-batch stochastic gradient descent (SGD) is used to optimize the cross-entropy loss function between target labels and network outputs.

Using softmax function, DNN posterior probability of the

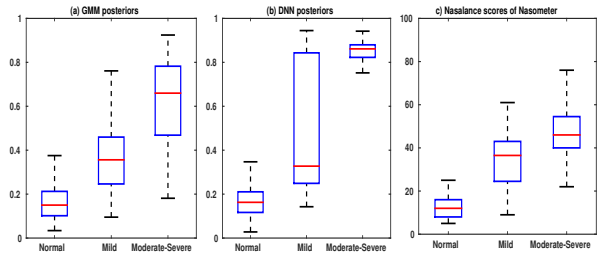

Figure 3: *Boxplots of mean hypernasality scores vs perceived severity using (a) GMM posteriors, (b) DNN posteriors, and (c) mean nasalance scores of nasometer*

test feature x_i for the nasal class λ_N is given by

$$p(\lambda_N|x_i) = \frac{e^{\beta_N y_i}}{e^{\beta_O y_i} + e^{\beta_N y_i}} \quad (3)$$

where β_N and β_O, represent the weights for nasal and oral classes, respectively and y_i is the output of hidden layer. Similar to GMM posterior equation, DNN posterior for the nasal class is also proportional to nasal sound characteristics present in the speech signal.

4. Experimental Results and Discussion

The 323 tokens of oral sentences from the CN group and 248 tokens of nasal sentences from the group with moderate-severe hypernasality are used to train the binary GMM and DNN classifiers. The test database consists of 106 tokens of oral sentences of normal (93 from CN group and 13 from CLP group, rated as normal), 76 tokens of mild, and 56 tokens of moderate-severe hypernasal groups. Due to the availability of smaller number of samples for the severe group, the moderate and severe groups are merged to form the moderate-severe group. In testing phase, the frame-wise computed posterior probability scores are averaged, which is referred as hypernasality score.

Table 1: *Statistical Test Results*

Groups	GMM posteriors	DNN posteriors	Nasalance scores
	χ^2=129.66	χ^2=158.99	χ^2=157.05
Normal vs. Mild	$p < 0.01$	$p < 0.01$	$p < 0.01$
Mild vs. Moderate	$p < 0.01$	$p < 0.01$	$p < 0.01$

4.1. Statistical analysis

The GMM and DNN based hypernasality scores for the normal, mild, and moderate-severe groups are shown in Fig. 3(a) and (b) respectively. The mean nasalance scores computed for the oral sentences of test of different hypernasality groups are shown in Fig. 3(c). The boxplots indicate that as the perceived severity increases, the posterior probability values for the nasal class also increases similar to that of nasalance values. Because, as the nasality increases the occurrence of nasal consonants and nasalized vowels in the speech increases. Kruskal-Wallis test is conducted to analyze the significance of discrimination across different groups. Table. 1 shows that the hypernasality scores significantly differentiates the different groups of hypernasality at the level of significance ($p < 0.01$). For all three methods, post hoc comparison using Tukey test revealed that between a pair of groups, the normal vs. mild, mild vs. moderate-severe the scores are discriminable at a level of significance ($p < 0.01$).

Table 2: *Performance of GMM, DNN based systems and nasometer*

Method	Correlation	Classification Performance		
		Sensitivity (%)	Specificity (%)	Accuracy (%)
Nasometer	0.78	85.99	86.91	91.06
without GAD + GMM	0.68	79.02	82.13	80.19
without GAD + DNN	0.77	91.72	93.54	91.63
With GAD + GMM	0.71	83.87	83.41	83.65
With GAD + DNN	0.82	93.10	93.54	93.34

4.2. Correlation with perceived severity

Spearman's correlation coefficients are computed for the proposed hypernasality scores with respect to perceived severity and that for nasalance values are presented in Table 2. Results indicate that the usage of glottal activity preprocessing gives better correlation with respect to perceived severity. This is observed for both GMM and DNN approaches. However, when compared to GMM, DNN based posteriors give the better correlation with the perceived severity of hypernasality. Also, the proposed system using DNN shows a better correlation with the perceived severity than that of nasometer.

4.3. Classification of normal and hypernasality groups

The significance of proposed hypernasality scores for classifying normal and hypernasal speech is analyzed. Here, the threshold on the hypernasality scores is estimated using leave-one-speaker-out criteria. In each trial, the samples of a particular speaker is left and an optimum threshold is chosen by minimizing false alarm rates of classification. The left speaker's samples are used for the testing, and classified into normal (less than the threshold) or hypernasal (equal to or greater than the threshold). The sensitivity (classification rate of hypernasal), specificity (classification rate of normal) and overall accuracy for GMM and DNN methods with and without the application of GAD pre-processing stage, and nasometer are presented in Table 2. The results indicate that DNN shows better classification accuracy when compared to GMM and nasometer.

5. Conclusion and Future Work

In this work, posterior probability based approach is proposed for the estimation of the hypernasality scores from CLP speech. MFCCs derived from glottal activity regions are used to train binary GMM and DNN classifiers for the two extremely opposite cases of nasality. DNN showed the better correlation with the perceived severity and discrimination between normal and hypernasal groups when compared to GMM based approach. Also, DNN based system outperforms the most widely used clinical instrument, i.e., nasometer. Therefore, the proposed system can be used as a tool in the clinical environments for the assessment of hypernasality.

6. Acknowledgement

Authors would like to thank Dr. M. Pushpavathi, Dr. Ajish K. Abhraham, and Mr. K. S. Girish, AIISH, Mysore for helping to record speech samples and nasometric evaluation. This work is in part supported by the project grants, for the projects entitled "NASOSPEECH: Development of Diagnostic system for Severity Assessment of the Disordered Speech" funded by the Department of Biotechnology (DBT), Govt. of India and "ARTICULATE +: A system for automated assessment and rehabilitation of persons with articulation disorders" funded by the Ministry of Human Resource Development (MHRD), Govt. of India.

7. References

[1] A. W. Kummer, *Cleft palate & craniofacial anomalies: Effects on speech and resonance*. Nelson Education, 2013.

[2] G. Henningsson, D. P. Kuehn, D. Sell, T. Sweeney, J. E. Trost-Cardamone, and T. L. Whitehill, "Universal parameters for reporting speech outcomes in individuals with cleft palate," *The Cleft Palate-Craniofacial Journal*, vol. 45, no. 1, pp. 1–17, 2008.

[3] K. Bettens, F. L. Wuyts, and K. M. Van Lierde, "Instrumental assessment of velopharyngeal function and resonance: A review," *Journal of communication disorders*, vol. 52, pp. 170–183, 2014.

[4] A. Maier, F. Hönig, T. Bocklet, E. Nöth, F. Stelzle, E. Nkenke, and M. Schuster, "Automatic detection of articulation disorders in children with cleft lip and palate," *The Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. 2589–2602, 2009.

[5] P. Vijayalakshmi, M. R. Reddy, and D. O'Shaughnessy, "Acoustic analysis and detection of hypernasality using a group delay function," *IEEE Transactions on biomedical engineering*, vol. 54, no. 4, pp. 621–629, 2007.

[6] B. J. Philips and R. D. Kent, "Acoustic-phonetic descriptions of speech production in speakers with cleft palate and other velopharyngeal disorders," *Speech and language: Advances in basic research and practice*, vol. 11, pp. 113–167, 1984.

[7] R. Kataoka, D. W. Warren, D. J. Zajac, R. Mayo, and R. W. Lutz, "The relationship between spectral characteristics and perceived hypernasality in children," *The Journal of the Acoustical Society of America*, vol. 109, no. 5, pp. 2181–2189, 2001.

[8] L. He, J. Zhang, Q. Liu, H. Yin, and M. Lech, "Automatic evaluation of hypernasality and consonant misarticulation in cleft palate speech," *IEEE Signal Processing Letters*, vol. 21, no. 10, pp. 1298–1301, 2014.

[9] L. He, J. Zhang, Q. Liu, H. Yin, M. Lech, and Y. Huang, "Automatic evaluation of hypernasality based on a cleft palate speech database," *Journal of medical systems*, vol. 39, no. 5, p. 61, 2015.

[10] M. Golabbakhsh, F. Abnavi, M. Kadkhodaei Elyaderani, F. Derakhshandeh, F. Khanlar, P. Rong, and D. P. Kuehn, "Automatic identification of hypernasality in normal and cleft lip and palate patients with acoustic analysis of speech," *The Journal of the Acoustical Society of America*, vol. 141, no. 2, pp. 929–935, 2017.

[11] "All India Institute of Speech and Hearing, Mysore, India." [Online]. Available: http://www.aiishmysore.in

[12] N. Adiga and S. Prasanna, "Detection of glottal activity using different attributes of source information," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 2107–2111, 2015.

[13] F. Chollet *et al.*, "Keras: Deep learning library for theano and tensorflow.(2015)," *There is no corresponding record for this reference*, 2015.