Automatic identification of hypernasality in normal and cleft lip and palate patients with acoustic analysis of speech

Article *in* The Journal of the Acoustical Society of America · February 2017

DOI: 10.1121/1.4976056

Automatic identification of hypernasality in normal and cleft lip and palate patients with acoustic analysis of speech

AQ1 Marzieh Golabbakhsh[a)]
Medical Image and Signal Processing Research Center, Isfahan University of Medical Sciences, Hezarjarib Street, 81745-319, Isfahan, Iran

Fatemeh Abnavi
Craniofacial Anomalies and Cleft Palate Research Center, Isfahan University of Medical Sciences, Hezarjarib Street, 81745-319, Isfahan, Iran

Mina Kadkhodaei Elyaderani
Medical Image and Signal Processing Research Center, Isfahan University of Medical Sciences, Hezarjarib Street, 81745-319, Isfahan, Iran

Fatemeh Derakhshandeh and Fatemeh Khanlar
Craniofacial Anomalies and Cleft Palate Research Center, Isfahan University of Medical Sciences, Hezarjarib Street, 81745-319, Isfahan, Iran

Panying Rong
Department of Communication Sciences and Disorders, MGH Institute of Health Professions, 36 First Avenue, Boston, Massachusetts 02129, USA

David P. Kuehn
Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, 901 South Sixth Street, Champaign, Illinois 61820, USA

(Received 21 May 2016; revised 21 January 2017; accepted 26 January 2017; published online xx xx xxxx)

Hypernasality is seen in cleft lip and palate patients who had undergone repair surgery as a consequence of velopharyngeal insufficiency. Hypernasality has been studied by evaluation of perturbation, noise measures, and cepstral analysis of speech. In this study, feature extraction and analysis were performed during running speech using six different sentences. Jitter, shimmer, Mel frequency cepstral coefficients, bionic wavelet transform entropy, and bionic wavelet transform energy were calculated. Support vector machines were employed for classification of data to normal or hypernasal. Finally, results of the automatic classification were compared with true labels to find accuracy, sensitivity, and specificity. Accuracy was higher when Mel frequency cepstral coefficients were combined with bionic wavelet transform energy feature. In the best case, accuracy of 85% with sensitivity of 82% and specificity of 85% was obtained. Results prove that acoustic analysis is a reliable method to find hypernasality in cleft lip and palate patients.
 [http://dx.doi.org/10.1121/1.4976056]

[JFL] Pages: 1–7

I. INTRODUCTION

Diagnostic methods which use acoustic analysis of speech are non-invasive in nature and have been used to assess different pathologies of the larynx and vocal tract.[1] Hypernasality has been studied mostly by evaluation of perturbation, noise measures, and cepstral analysis. These studies have led to acoustic features like energy operators, Mel frequency cepstral coefficients (MFCC), and noise features such as harmonics to noise ratio (HNR).[2–10] Nonlinear dynamic features like correlation dimension, largest Lyapunov exponent, Hurst exponent, and Lempel-Ziv complexity have also been considered as a complement to classical techniques that are mainly focused on the analysis of acoustic and perturbation content in voice.[11,12]

Glass and Zue[2] worked on modeling various acoustic features to detect nasalization in English. They built a database composed of 200 words. However, the accuracy to classify oral and nasalized speech was 74%.[2] Cairns *et al.*[3] started to pay attention to detection of hypernasality in patients. The authors introduced correlation of a Teager energy operator as a feature that is able to detect hypernasality. Later they applied hidden Markov models (HMM) to separate simulated hypernasal speech from speech with normal resonance. After comparing the results of the proposed system to subjective medical results, accuracy levels of 100% for hypernasal speech (simulated) and 98.8% for speech with normal resonance were obtained.[3]

A feature which is commonly used in speech analysis is the set of MFCC. Dibazar *et al.*[4] used automatic classification of pathologic speech. They used a sustained phonation of the English vowel /a/. Results showed an accuracy of

[a)]Electronic mail: marzgol@gmail.com

98.5%. Pruthi and Espy-Wilson[5,6] implemented a spectral analysis of English speech to detect hypernasality. Since 2005, Vijayalakshmi and Reddy[7] have been working on the signal spectrum for hypernasality detection. In 2009 they offered a method based on the modification of the poles of linear prediction spectrum for automatic detection of hypernasal voices. Their database contained 25 normal and 25 unrepaired cleft lip and palate (CLP) patients who pronounced sustained vowel /a/. They reported 100% accuracy employing this method.[8]

Kataoka *et al.* made quantitative evaluation of hypernasality in cleft palate patients by cepstrum analysis.[9] They also examined the relationship between spectral characteristics and the perceptual evaluation of hypernasality in speakers with cleft palate.[10]

Later in 2011 Murillo *et al.*[11] used noise features such as HNR, cepstral harmonics to noise ratio (CHNR), normalized noise energy(NNE), and glottal to noise excitation ratio(GNE) to find signs of hypernasality in CLP children. Their database included Spanish vowels. Using a linear Bayes classifier, they reached success rates between 80% and 90%. Orozco-Arroyave *et al.*[12] used non-linear dynamic features and compared the performance of acoustic features with non-linear dynamic ones. To reduce computational load in classification stages, they first did automatic feature selection and eliminated irrelevant and/or redundant information. Two techniques were employed for this purpose, principal component analysis (PCA) and also a heuristic technique named sequential floating forward selection (SFFS). They could reach an accuracy of more than 85%.[12]

In a study in Iran, the prevalence of velopharyngeal insufficiency in cleft lip and palate patients who had undergone primary cleft palate repair at the hospitals affiliated to Isfahan University of Medical Sciences between 2005 and 2009 were reviewed. Results showed that the prevalence of velopharyngeal insufficiency was very high compared to other studies.[13] Hypernasality is one of the main disorders caused by velopharyngeal insufficiency. The prevalence of hypernasality in Persian speaking people who suffered from cleft lip and palate is reported to be 70.9%.[14]

Choosing the right medical treatment for velopharyngeal insufficiency is based on speech assessment by speech therapists and measuring the level of hypernasality. In fact, detection of the level of hypernasality is one of the important factors for medical decision making. Because if high levels of hypernasality makes the speech quality very low, the patient may need to do surgery. Assessment of hypernasality is also very important during patient's follow-up.

Resonance is the process by which the basic product of phonation is enhanced. Hypernasality is a resonance disorder caused by velopharyngeal insufficiency. Speech therapists use two different methods for evaluation of resonance; Perceptual and instrumental. In the perceptual method, the evaluator listens to the patients' speech and then using his/her audio skills and clinical experience judges about it. Perceptual judgments of hypernasality are original sources for therapists to evaluate resonance because they have a high level of content validity. However these evaluations have problems that reduce their reliability because the judgments are subjective, the scoring systems are not consistent, and also there are independent speech characteristics which have an effect in the increase or decrease of hypernasality perception.[15]

Because hypernasality changes with articulation, amplitude, and frequency of speech, the judgment of the evaluator is somewhat ambiguous and it is better to complement perceptual findings with instrumental and objective means. Instrumental evaluations to judge hypernasality include direct and indirect means and are important because they may lead to quantitative analysis of hypernasality. Direct means of assessing the velopharyngeal mechanism include videofluoroscopy and nasoendoscopy. Nasoendoscopy is a common but invasive method. Videofluoroscopy is a non-invasive method but it has risk of putting the patient under radiation.[15–17]

As Horii[18] has mentioned, a good analysis of velopharyngeal sphincter should be non-invasive and should not have an effect on phonation and articulation and should not alter sensory feedback for speech. Therefore, there is a strong need to find objective and non-invasive means like acoustic analysis and aerodynamic analysis to measure hypernasality. In acoustic analysis, no device is applied to the patient and the speech would not be altered by the method of analysis. The audio signal is recorded by a microphone placed near the mouth of the subject and therefore, it is the actual signal received by the listener which is analyzed. Most of the previous studies were experimented on vowels and/or words, however, it is very important to do acoustic analysis during running speech. Therefore, in this study, six different sentences were employed while each of them had emphasis on one of the consonants.

II. METHODS

A. Dataset

The data were recorded from 15 normal subjects and 15 patients with cleft lip and palate who were under evaluation at the Cleft Lip and Palate Center, Rehabilitation Department, Isfahan University of Medical Sciences. Normal subjects were recruited at an elementary school. Data were recorded using a microphone(C417 omnidirectional condenser Lavalier microphone, AKG Acoustics, Austria) connected to a portable digital sound recorder (Edirol R-44, Japan).The microphone was placed at 20 cm distance from the mouth.

The recorder amplified and digitized the sound signal at a sampling rate of 44.1 kHz. This database contains, for each speaker, a recording of six sentences which contain stop consonants and fricative consonants (see the Appendix). These sentences are used by speech therapists to assess quality of speech from a perceptual point of view.

B. Feature extraction

Time and frequency features, which are defined based on the characteristics of the human's hearing system were used in this study; MFCC, jitter, shimmer, bionic wavelet transform energy, and bionic wavelet transform entropy.

These features were used because the plan was to do experiments on running speech. Jitter and shimmer are routinely used by speech pathologists for speech analysis. MFCC coefficients have been widely used in speech recognition and speech technology. The reason that wavelet transform was used is because it is a time-frequency analysis and as we want to examine running speech it is important that the selected feature could enhance analysis over time while focusing on frequency information too. Bionic wavelet transform was applied because it was designed based on human auditory model. Combining the features gives the chance to have higher accuracy in detecting hypernasality.

Due to the non-stationarity of the speech signal, the signal was divided to smaller frames. We chose 40 ms frames so that each frame would include at least two formant periods of a signal so that we could calculate jitter and shimmer based on the behavior of the signal. The frames had an overlap of 20 ms to smooth feature calculation in a frame-to-frame basis. Hamming window was applied to focus on the information in the center of the frames. The mentioned features are calculated for each time frame, forming five feature vectors per speech recording. Mean values are then calculated for each vector.

1. Jitter

Jitter is defined as the parameter of fundamental frequency variation from period to period. Relative jitter is the average absolute difference between successive periods divided by the average period.[19] It is calculated as

$$\frac{\dfrac{1}{N-1}\displaystyle\sum_{i-1}^{N-1}|T_i - T_{i+1}|}{\dfrac{1}{N}\displaystyle\sum_{i-1}^{N} T_i}, \qquad (1)$$

where T_i are the extracted F_0 period length. N is the number of extracted F_0 periods where F_0 is the fundamental frequency.

In this study we used relative jitter for measuring jitter.

2. Shimmer

Relative shimmer is defined as the average absolute difference between the amplitudes of successive periods divided by the average amplitude,[19] calculated as

$$\frac{\dfrac{1}{N-1}\displaystyle\sum_{i-1}^{N-1}|A_i - A_{i+1}|}{\dfrac{1}{N}\displaystyle\sum_{i-1}^{N} A_i}, \qquad (2)$$

where A_i are the extracted peak-to-peak amplitude of ith fundamental frequency period and N is the number of extracted fundamental frequency periods.

3. Mel frequency cepstral coefficients

MFCC are nowadays widely used in automatic speech recognition. The main idea in MFCC is based on Mel scale which is derived from the human ear characteristics in receiving and understanding of speech. The hearing system in humans is such that the perceived pitch is different from the actual speech frequency.

Mel scale relates perceived frequency, or pitch, of a pure tone to its actual measured frequency. Humans are much better at discerning small changes in pitch at low frequencies than they are at high frequencies. Incorporating this scale makes our features match more closely with what humans hear.[20]

The formula for converting from frequency to Mel scale is

$$M(f) = 1125 \ln\left(1 + \frac{f}{700}\right). \qquad (3)$$

For calculation of MFCC the sound signal was segmented into intervals of 40 ms with an overlap of 50%. For each interval, a spectrogram of the signal was computed by taking the fast Fourier transform (FFT) after applying a hamming window. In the next step the Mel-spaced filterbanks are computed. This is a set of 26 triangular filters that were applied to the periodogram power spectral estimate from step 2. Later the logarithm of each of the 26 energies from step 3 were calculated which leaves us with 26 log filterbank energies. After taking the discrete cosine transform (DCT) of the 26 log filterbank energies, 26 cepstral coefficients are obtained. For automatic speech recognition, only the lower 12–13 of the 26 coefficients are kept.[20] The resulting features (12 numbers for each frame) are called Mel frequency cepstral coefficients.

4. Bionic wavelet transform

To apply wavelet transform (WT)on a signal $f(t)$, which simply means the inner product between the shifted and dilated versions of mother wavelet and the signal itself, we have

$$\mathrm{WT}_f(a,\tau) = \frac{1}{\sqrt{a}}\int f(t)\tilde{\varphi}^*\left(\frac{t-\tau}{a}\right)\exp\left[-jw_0\left(\frac{t-\tau}{a}\right)\right]dt, \qquad (4)$$

where a is scale and τ is the time shift. WT_f stands for wavelet transform of the signal f. φ is the mother wavelet or the basis function and w_0 is the initial center frequency of the mother wavelet.

Yao and Zhang proposed the bionic wavelet transform (BWT) as a novel time frequency method which is designed based on human auditory model.[21] The BWT is different from the standard WT in that the resolution in the time-frequency domain achieved by the BWT can be adaptively changed not only by the signal frequency changes but also by the signal's immediate amplitude and its first order differential.[21] In essence, the idea of the BWT is that the envelope of the mother wavelet varies in accordance with the characteristics of the target signal using the time varying T function.

To simulate the dynamic control utility of the hair cells of the auditory system, the T function regulates the mother wavelet $\varphi(t)$ in the following way

$$\varphi(t) = \frac{1}{T\sqrt{a}}\tilde{\varphi}\left(\frac{t}{T}\right)\exp(jw_0 t). \tag{5}$$

The first T alters the scaling factor a and the second T adjusts the envelope of the mother wavelet $\varphi(t)$ without adjusting its center frequency w_0. Applying the WT definition with our new adaptive mother wavelet, we have the bionic wavelet transform:

$$\mathrm{BWT}_f(a,\tau) = \frac{1}{T\sqrt{a}}\int f(t)\tilde{\varphi}^*\left(\frac{t-\tau}{Ta}\right)\exp\left[-jw_0\left(\frac{t-\tau}{a}\right)\right]dt, \tag{6}$$

where a is scale and τ is the time shift. BWT_f stands for bionic wavelet transform of the signal f. φ is the mother wavelet or the basis function and w_0 is the initial center frequency of the mother wavelet. T is the varying time function invented by Yao and Zhang. For detailed information about how this function is made and the underlying mechanism, please see Refs. 22 and 23.

BWT of a signal is calculated by passing it through a series of filters called filterbanks. In the software implementation, 22 filters have been considered due to the number of electrodes in cochlear implants. The central frequencies of the filters are defined in Tables I and II.

In this study two bionic features were used; bionic wavelet entropy and bionic wavelet energy.

To obtain BWT energy, each speech frame is decomposed into 22 bands with the use of 22 filterbanks (Table II). BWT coefficients are then calculated for each band and energy of each frame is obtained using root mean square calculations.

BWT entropy is given by

$$e(i) = -\sum_j s^2(j)\log_{10}s^2(j), \tag{7}$$

where $e(i)$ is entropy of ith band and $s(j)$ is the jth bin of the histogram of the BWT coefficients of this band.

C. Support vector machines

Support vector machines (SVM) are supervised learning algorithms that analyze data and recognize patterns used for classification and regression analysis. They are simple and efficient methods in machine learning and have many applications in classification.[24,25]

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.[24–26]

SVM has shown success in similar works that address the problem of the automatic detection of pathological speech signals.[27,28]

TABLE I. Description of data.

Patient No.	Age	Gender	Degree of hypernasality	Palate's type
1	4	M	Mild	Bilateral cleft lip + left alveolar cleft
2	5	F	Mild	Submucous cleft
3	10	F	Severe	Soft cleft palate
4	25	F	Moderate	Cleft palate
5	15	M	Moderate	Bilateral cleft lip and palate
6	20	F	Severe	Cleft palate
7	6	F	Moderate	Cleft palate
8	4	M	Moderate	Bilateral cleft lip and palate
9	21	F	Moderate	Bilateral cleft lip and palate
10	6	M	Severe	Bilateral cleft lip and palate
11	4	M	Mild	Cleft palate
12	17	F	Severe	Unilateral cleft lip and palate
13	28	M	Moderate	Unilateral cleft lip and palate
14	4	M	Severe	Unilateral cleft lip and palate
15	15	F	Moderate	Unilateral cleft lip and palate
16	6	M	None	None
17	28	F	None	None
18	24	F	None	None
19	23	F	None	None
20	20	F	None	None
21	18	F	None	None
22	12	M	None	None
23	6	M	None	None
24	13	F	None	None
25	7	F	None	None
26	6	M	None	None
27	9	M	None	None
28	5	F	None	None
29	6	M	None	None
30	6	M	None	None

TABLE II. Filterbanks and their corresponding center frequencies used to calculate BWT of speech signal.

Filter	Center frequency
1	5291
2	4551
3	3916
4	3369
5	2899
6	2494
7	2145
8	1846
9	1588
10	1366
11	1175
12	1011
13	870
14	748
15	644
16	554
17	476
18	410
19	352
20	303
21	261
22	221

TABLE III. Set of features that was used for classification of speech to normal or hypernasal.

MFCC
(MFCC,BWT energy)
Jitter
Shimmer
(MFCC,BWT entropy)

D. Experiments

Our preliminary studies showed that if we use some features together and perform classification in a higher dimensional space, better results could be obtained. Therefore, MFCC, jitter, and shimmer were used separately in a one dimensional space. MFCC and BWT energy, and also MFCC and BWT entropy were used in a two dimensional space. Table III shows the set of features which were used for training and classification. The classification process was implemented with SVM. This classifier is trained using a radial basis Gaussian kernel with bandwidth σ and it was tested using a repeated random sub-sampling validation strategy.

In each round, 15 data, seven from patients and eight from normal people, were randomly selected to train and the remaining data was used for test. This was done five times and the values of accuracy, sensitivity, and specificity of the classifier were obtained by taking mean over the values obtained in each round.

III. RESULTS AND DISCUSSION

Results are presented in terms of the overall accuracy of the system. The subjects are labeled hypernasal (abnormal) if they are chosen from the list of patients and they are considered normal if they are from the healthy group. Classification is performed using SVM method and finally results of the automatic classification are compared with true labels to find accuracy, specificity and sensitivity. Specificity and sensitivity are provided to indicate the probability of a normal subject to be correctly identified (specificity) and the probability of an abnormal signal to be correctly labeled (sensitivity).

Tables IV–IX show the values of accuracy, sensitivity, and specificity obtained for each sentence. These values are obtained by averaging over five repeated random sampling tests.

TABLE IV. Statistical measures of the performance of the classification for sentence “baba sib bede.”

“baba sib bede”	Accuracy	Sensitivity	Specificity	No. of false positive ($N = 35$)
MFCC	72	67	71	10
(MFCC,BWT energy)	85	82	85	5
Jitter	48	50	45	19
Shimmer	54	55	48	18
(MFCC,BWT entropy)	62	67	62	13

TABLE V. Statistical measures of the performance of the classification for sentence “tupe puya pare ʃod.”

“tupe puya pare ʃod”	Accuracy	Sensitivity	Specificity	No of false positive ($N = 35$)
MFCC	59	60	51	16
(MFCC,BWT energy)	83	85	80	7
Jitter	52	50	48	18
Shimmer	49	52	42	20
(MFCC,BWT entropy)	49	47	57	18

TABLE VI. Statistical measures of the performance of the classification for sentence “yek keyke kuc̆ak.”

“yek keyke kuc̆ak”	Accuracy	Sensitivity	Specificity	No. of false positive ($N = 35$)
MFCC	73	77	71	10
(MFCC,BWT energy)	86	85	77	8
Jitter	50	50	42	20
Shimmer	50	52	51	17
(MFCC,BWT entropy)	61	60	57	15

TABLE VII. Statistical measures of the performance of the classification for sentence “sepehr sos dare.”

“sepehr sos dare”	Accuracy	Sensitivity	Specificity	No. of false positive ($N = 35$)
MFCC	67	67	62	13
(MFCC,BWT energy)	77	82	68	11
Jitter	53	55	51	17
Shimmer	44	52	51	17
(MFCC,BWT entropy)	73	75	71	10

TABLE VIII. Statistical measures of the performance of the classification for sentence “reza saz zæd.”

“reza saz zæd”	Accuracy	Sensitivity	Specificity	No. of false positive ($N = 35$)
MFCC	73	80	71	10
(MFCC,BWT energy)	81	85	80	7
Jitter	51	50	48	18
Shimmer	52	50	48	18
(MFCC,BWT entropy)	55	50	48	18

TABLE IX. Statistical measures of the performance of the classification for sentence “ye gæle gorg.”

“ye gæle gorg”	Accuracy	Sensitivity	Specificity	No of false positive ($N = 35$)
MFCC	69	75	68	11
(MFCC,BWT energy)	69	90	77	8
Jitter	48	45	48	20
Shimmer	58	60	54	16
(MFCC,BWT entropy)	49	45	51	17

This is the first study which evaluates hypernasality during running speech in Persian speaking people. This study is very valuable in recognizing what sentence will better indicate hypernasality.

It is seen in Tables IV–IX that accuracy and sensitivity of the test is higher when MFCC is combined with BWT energy. MFCC coefficients are one of the most important characteristics of speech signal and they are widely used in speech analysis. As both MFCC and BWT are based on the human hearing system and they were extracted from small time frames, together they could easily follow variations in speech signal in running speech.

IV. CONCLUSION

Development of automatic methods of hypernasality detection in speech signal is very important. Acoustic methods could be used as a substitute or complementary tool to identify needed therapy, how we proceed with the disease treatment, and follow-up of results of previous medical actions. Different acoustic parameters have been used for hypernasality detection in previous studies: HNR, NNE, GNE ratio, amplitude and frequency perturbation parameters, MFCC.

Most researchers believe that there are non-linear processes during speech production which could not be retrieved by using each of above-mentioned methods alone.

Therefore, children with cleft palate use compensatory movements in the vocal tract as a solution to compensate dysfunction created as a result of structural problems. This strategy leads to compensatory articulation errors and non-linear characteristics in vocal tract movements. Based on non-linear effects on hypernasality, utilizing new features such as BWT characteristics and also combining features is necessary in order to diagnose hypernasality in speech signal.

According to the accuracies in Table IV–IX, better accuracy results are in sentences containing anterior consonants that are useful for characterization of hypernasality. In clinical trials, the use of sentences in identifying hypernasality is very important.

For clinical application, the use of speech sentences plays an important role in determining degree of hypernasality. Because in some cases, hypernasality at the level of words is not diagnosed. Therefore, evaluation of sentence and connected speech is required. In addition, based on the universal parameters and caps, tests which are widely used for perceptual assessment, one of the main tasks in speech evaluation process for patients with cleft palate is the sentence repetition. According to the importance of completing the perceptual judgment with other methods, the acoustic study for all levels and assignments is substantially required. This study has provided the possibility of using recognition at the level of sentences.

In the future, we would like to test patients during conversation or while reading passages. Other techniques like HMM should be examined too. HMM might be able to predict hypernasality with higher accuracy.

APPENDIX

baba sib bede
tupe puya pare ʃod
yek keyke kuc̆ak
sepehr sos dare
reza saz zæd
ye gæle gorg

[1]O. Fujimura and J. Lindqvist, "Sweep-tone measurements of vocal-tract characteristics," J. Acoust. Soc. Am. **49**(2B), 541–558 (1971).
[2]J. Glass and V. Zue, "Detection of nasalized vowels in American English," in *IEEE International Conference on Acoustics, Speech, and Signal Processing* (1985), Vol. 10, pp. 1569–1572.
[3]D. Cairns, J. H. Hansen, and J. E. Riski, "Detection of hypernasal speech using a nonlinear operator," in *Proceedings of the 16th Annual International Conference of the IEEE, Engineering in Medicine and Biology Society, Engineering Advances: New Opportunities for Biomedical Engineers* (1994), pp. 253–254.
[4]A. A. Dibazar, S. Narayanan, and T. W. Berger, "Feature analysis for automatic detection of pathological speech," in *Engineering in Medicine and Biology, 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, Proceedings of the Second Joint* (2002), Vol. 1, 182–183.
[5]T. Pruthi and C. Y. Espy-Wilson, "Acoustic parameters for automatic detection of nasal manner," Speech. Commun. **43**(3), 225–239 (2004).
[6]T. Pruthi and C. Y. Espy-Wilson, "Acoustic parameters for the automatic detection of vowel nasalization," in *Interspeech* (2007), pp. 1925–1928.
[7]P. Vijayalakshmi and M. R. Reddy, "The analysis on band-limited hypernasal speech using group delay based formant extraction technique," in *Interspeech* (2005), pp. 665–668.
[8]P. Vijayalakshmi, T. Nagarajan, and J. Rav, "Selective pole modification-based technique for the analysis and detection of hypernasality," in *TENCON 2009-2009 IEEE Region 10 Conference* (2009), pp. 1–5.
[9]R. Kataoka, K. Takahashi, Y. Yamashita, S. Imai, K. I. Michi, K. Okabe, H. Hamada, and T. Miura, "Quantitative evaluation of hypernasality in cleft palate patients," J. Acoust. Soc. Am. **84**(S1), S115–S115 (1988).
[10]R. Kataoka, R. D. W. Warren, D. J. Zajac, R. Mayo, and R. W. Lutz, "The relationship between spectral characteristics and perceived hypernasality in children," J. Acoust. Soc. Am. **109**(5), 2181–2189 (2001).
[11]S. M. Rendón, J. O. Arroyave, J. V. Bonilla, J. A. Londoño, and C. C. Domínguez, "Automatic detection of hypernasality in children," in *International Work-Conference on the Interplay Between Natural and Artificial Computation* (Springer, Berlin, 2011), pp. 167–174.
[12]J. R. Orozco-Arroyave, J. F. Vargas-Bonilla, J. D. Arias-Londoño, S. Murillo-Rendón, G. Castellanos-Domínguez, and J. F. Garcés, "Nonlinear dynamics for hypernasality detection in Spanish vowels and words," Cognit. Comp **5**(4), 448–457 (2013).
[13]H. H. Hosseinabad, F. Derakhshandeh, F. Mostaajeran, H. Abdali, H. A. Davari, A. Hassanzadeh, and A. W. Kummer, "Incidence of velopharyngeal insufficiency and oronasal fistulae after cleft palate repair: A retrospective study of children referred to Isfahan Cleft Care Team between 2005 and 2009," Int. J. Pediatr. Otorhinolaryngol. **79**(10), 1722–1726 (2015).
[14]M. Zahedi, "Determining the frequency of post operative hypernasality in children with cleft palate born between 1380–1387," Ph.D. dissertation, Isfahan University of Medical Sciences, Isfahan, Iran, 2011.
[15]J. Van Doorn and A. Purcell, "Nasalance levels in the speech of normal Australian children," Cleft Palate Craniofac. J. **35**(4), 287–292 (1998).
[16]K. V. Lierde, F. L. Wuyts, M. D. Bodt, and P. V. Cauwenberge, "Nasometric values for normal nasal resonance in the speech of young Flemish adults," Cleft Palate Craniofac. J. **38**(2), 112–118 (2001).
[17]M. Pannbacker, N. J. Lass, G. F. Middleton, E. Crutchfield, D. S. Trapp, and K. A. Scherbick, "Current clinical practices in the assessment of velopharyngeal closure," Cleft Palate J. **21**(1), 33–37 (1984).
[18]Y. Horii, "An accelerometric approach to nasality measurement: A preliminary report," Cleft Palate J. **17**(3), 254–261 (1980).
[19]J. P. Teixeira, C. Oliviera, and C. Lopes, "Vocal acoustic analysis–Jitter, shimmer and HNR parameters," Procedia Technol. **9**, 1112–1122 (2013).
[20]X. Huang, A. Acero, and H. W. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development* (Prentice Hall PTR, Upper Saddle River, NJ, 2001), p. 980.

[21]J. Yao and Y. T. Zhang, "Bionic wavelet transform: A new time-frequency method based on an auditory model," IEEE Trans. Biomed. Eng. **48**(8), 856–863 (2001).
[22]J. Yao and Y. T. Zhang, "Cochlear is an inhomogeneous, active and non-linear model," in *Proceedings of the 1st Joint Meeting of BMES & IEEE/EMBS*, Atlanta, GA (1999), p.1031.
[23]J. Yao and Y. T. Zhang, "From otoacoustic emission modeling to bionic wavelet transform," in *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Chicago (2000), Vol. 1, pp. 314–316.
[24]V. Vapnik, *The Nature of Statistical Learning Theory* (Springer Science and Business Media, Berlin, Germany, 2013), pp. 267–290.
[25]C. J. Burges, "A tutorial on support vector machines for pattern recognition," Data Min. Knowl. Disc. **2**(2), 121–167 (1998).
[26]B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, Cambridge, MA, 2002), p. 625.
[27]A. Maier, T. Haderlein, U. Eysholdt, F. Rosanowski, A. Batliner, M. Schuster, and E. Nöth, "PEAKS–A system for the automatic evaluation of voice and speech disorders," Speech. Commun. **51**(5), 425–437 (2009).
[28]J. D. Arias-Loñdo, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and G. Castellanos-Dominguez, "Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients," IEEE Trans. Biomed. Eng. **58**(2), 370–379 (2011).