

Assessment of Hypernasality for Children with Cleft Palate Based on Cepstrum Analysis

Ehsan Akafi, Mansour Vali¹, Negin Moradi², Kowsar Baghban³

Department of Engineering, Shahed University, Tehran, Iran, ¹Department of Electrical and Computer Engineering, K.N Toosi University of Technology, Tehran, Iran, ²Musculoskeletal Rehabilitation Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran, ³Department of Rehabilitation, Tehran University of Medical Sciences, Tehran, Iran

Submission: 15-12-2012 Accepted: 05-08-2013

ABSTRACT

Hypernasality is a frequently occurring resonance disorder in children with cleft palate. In general, an operation is necessary to reduce the hypernasality and therefore an assessment of hypernasality is imperative to quantify the effect of the surgery and design the speech therapy sessions, which are crucial after surgery. In this paper, a new quantitative method is proposed to estimate hypernasality. The proposed method used the fact that an autoregressive (AR) model for vocal tract system of a patient with hypernasal speech is not accurate; because of the zeros appear in the frequency response of the vocal tract system. Therefore in our method, hypernasality was estimated by a quantity calculated from comparing the distance between the sequences of cepstrum coefficients extracted from AR model and autoregressive moving average model. *K*-means and Bayes theorem were utilized to classify the utterances of subjects by means of proposed index. We achieved the accuracy up to 81.12% on utterances and 97.14% on subjects. Since the proposed method needs only computer processing of speech data, compared with other clinical methods it provides a simple evaluation of hypernasality.

Key words: *Cepstrum, cleft palate, hypernasality, speech processing, speech therapy*

INTRODUCTION

Hypernasality frequently occurs in children with cleft palate due to excessive nasal resonance perceived during the speech because the oral cavity is not properly separated from the nasal cavity. In these cases in addition to surgical interventions like palatoplasty, patients should receive speech therapy. Therefore, assessment of nasality is necessary to facilitate the evaluation of the operation efficacy and help the therapist to manage the speech therapy sessions. Approaches for the assessment of hypernasality classified into two categories of invasive and non-invasive techniques. The invasive techniques involve the assessment of velopharyngeal function using invasive instruments such as nasendoscopy, videofluoroscopy, etc., in the clinical environment. Non-invasive techniques include clinical assessment and digital signal processing-based techniques. These approaches describe concisely as below.

Invasive Techniques

Multi view videofluoroscopy allows experts to observe the vocal tract structures during connected speech from

several spatial planes.^[1] Flexible fiber-optic nasendoscopy allows direct observation of velopharyngeal movements during connected speech. In these methods quantitative results cannot be acquired and also they need expensive equipment.

Non-invasive Clinical Assessments

In many types of clinical equipment such as nasometer, pressure, vibration and nasal flow were used as quantities for assessing hypernasal speech. The nasometer uses two separate microphones; first one is placed in front of the mouth and the other in front of the nostrils. The microphones record both the oral and nasal sound pressures and the index of nasometer called nasalance, is defined as the ratio of nasal sound pressure to total sound pressures assessed at the nostrils and mouth. This device is widely used in the clinical field, but it is an expensive device and it is uncomfortable to use for children. Moreover, Horii and Lang introduced the Horii oral nasal coupling index that measured the nasal coupling.^[1] This index was derived from signals measured by an accelerometer attached to the outside of the nares and by a microphone placed in front

Address for correspondence:

Dr. Mansour Vali, Department of Electrical and Computer Engineering, K.N Toosi University of Technology, Tehran, Iran.
E-mail: mansour.vali@eetd.kntu.ac.ir

of the mouth. For use these kinds of devices, subjects must attend clinics that make control of the therapy process hard, moreover, uncomfortable condition of using these devices may cause the speech of children be unnatural and the results of assessments became unreliable.

Perceptual Judgment

Perceptual judgments performed with scoring the quality of speech by experienced or well-trained listeners. Since hypernasal speech occurs in conjunction with abnormalities in pitch, loudness, voice quality and these coexisting features affect the perception of nasality, judgment results vary among the listeners.^[2] To overcome the problems of invasive methods and non-invasive clinical methods and to avoid disagreement of perceptual judgments, signal processing techniques were proposed.

Signal Processing Techniques

In previous studies, researchers tried to detect hypernasality by analyzing the speech of subjects with cleft lip or palate, synthesized hypernasal speech or nasalized vowels of normal speech.^[2] In signal processing-based techniques for hypernasality detection, the assessment is usually carried out by finding the deviation of the spectrum of hypernasal speech from the normal speech.^[2-5] Researchers claimed that nasalization increase the first formant bandwidth and intensity and also introduce nasal formants and antiformants.^[3] They compared the output of a low pass filter with cut-off frequency between the first and second formant and a band pass filter that just filter the first formant that both applied to speech samples and found a distinctive difference for nasalized vowels, whereas the normal vowels do not show any remarkable difference. Another proposed method, estimated hypernasality by comparing the distance between sequences of cepstrum coefficients extracted from low order and high order linear predictive (LP) model.^[4] The fact that a LP model with a typical order for the human vocal tract system is not accurate when the vocal tract system has zeros in its frequency response was used in this work. This approach assessed by calculating the Pearson correlation coefficient between the nasalance scores and the obtained distances. Lee *et al.* proposed new index for detect hypernasality called voice low tone to high tone ratio (VLHR).^[5] VLHR was defined as the division of power of low frequency part of voice spectrum, into power of high frequency part of voice spectrum. They calculated correlation of nasalance scores with values of VLHR on speech samples of subjects to assess their proposed index. Vijayalakshmi *et al.* introduced a new quantity based on acoustic analysis on nasalized vowels that showed the existence of a new resonance in the low frequency region (around 250 Hz) of speech spectrum,^[2] this fact also was used in another work of these authors, where a LP-based pole modification technique was introduced.^[6] They used a higher order LP

spectrum to select and weaken the pole corresponding to strongest peak in the low frequency region and then resynthesized new signal. Maximum of the cross-correlation value between the original signal and resynthesized speech signal was taken as a measure for the detection of hypernasality.

Some other works were focused on extracting famous speech features and employing different classification methods. Castellanos *et al.*, Delgado-Trejos *et al.* and Maier *et al.* used set of features including pitch, jitter, tone perturbation coefficient, harmonic to noise ratio, energy, zero crossings, linear predictive coefficients (LPC), mel-frequency cepstral coefficients and wavelet transform, also they employed Bayesian classifier, Gaussian mixture model and support vector machine (SVM).^[7-9] Orozco-Aroyave *et al.* and Arias-Londoño *et al.* concentrated on features related to non-linear dynamics of speech such as correlation dimension, largest Lyapunov exponent and Hurst exponent.^[10-12]

In this paper, the fact that the vocal tract system of a patient with cleft palate has additional zeros in its frequency response was used. Since an autoregressive (AR) model for frequency response of the vocal tract system of these patients is not accurate, in our method the hypernasality was estimated by comparing the distance between the sequences of cepstrum coefficients of AR model and cepstrum coefficients of autoregressive moving average (ARMA) model.

MATERIALS AND METHODS

During the pronunciation of a vowel, the excitation signal from the vocal cord is assumed to be an ideal impulse train. Under this assumption, the transfer functions of the vocal tract system, which is from the glottis to lips and acts like a filter, considered as all-pole system. Thus, an AR model has been widely used in modeling speech signal analysis. AR coefficients a_k used in an AR model of order M are defined as

$$s(n) = \sum_{k=1}^M a_k s(n-k) + Gu(n) \quad (1)$$

Where $s(n)$ is the speech signal and $u(n)$ is the excitation signal produced by the glottis. G is the gain of the excitation signal and also the root mean square value of the residual error between the predicted and original speech signals.^[4] Figure 1 represents a simplified model of the human vocal tract. Model of the human vocal tract is assumed to be composed of several lossless acoustic tubes.^[13] Nasality depends upon how much and when the velum that separate nasal and oral cavity is open, during pronunciation. Thus, the pronunciation of a speaker with cleft lip or palate or subjects with defective velopharyngeal mechanism is hypernasal because the velum cannot separate the nasal and oral cavities appropriately or because poor timing of velopharyngeal. As mentioned in previous works,

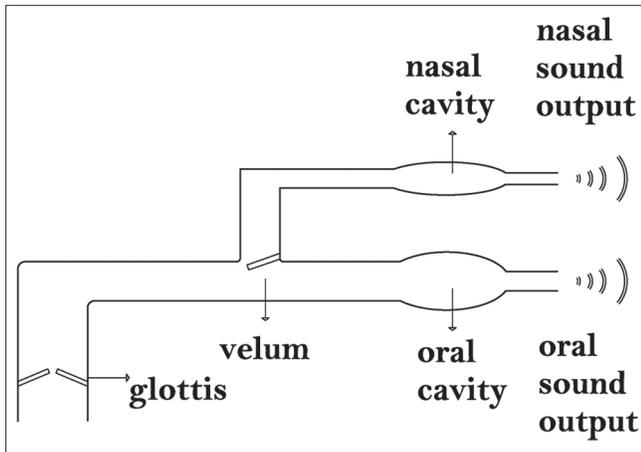


Figure 1: Simple model of human vocal tract

hypernasality is characterized by: (1) Amplitude reduction of the first formant, (2) presence of zeros in the spectrum due to the coupling of the nasal cavity and oral cavity, (3) presence of reinforced harmonics (nasal formant) resulting from the sound resonance in the nasal cavity and (4) shift of formants.^[4] The characteristic (2) was utilized to estimate hypernasality in this study. If we want to model the vocal tract system of a normal speaker, AR model of order 8-10 is typically used.^[14] Since the hypernasal speech signal cannot be modeled accurately by an AR model we can expect that an ARMA model with typical number of poles and the appropriate number of zeros give a more accurate representation of the hypernasal speech. Based on this fact, we propose an algorithm to estimate the hypernasality of speech using the AR and ARMA coefficients. Actually, there will be a significant difference between the cepstrum coefficients obtained from the AR model and from the ARMA model, in the case of the hypernasal speech.

Methods

Speech signal should pass the pre-processing stage; including data normalization and a pre-emphasis filtering. Pre-emphasis filter with pre-emphasis coefficient $\alpha = 0.98$ defined as

$$p(z) = 1 - 0.98z^{-1} \quad (2)$$

The reasons for employing a pre-emphasis filter consist of eliminating the scattering effect that is introduced when the speech signal is transmitted from the lips through the air; and also removing the spectral component of the larynx from the speech signal.^[15] After pre-processing, the signal should pass the windowing stage; The hamming window was applied to speech data for the frame length of 30 ms and the frame was shifted by 15 ms for 50% overlap. Each frame with a length of 30 ms can be assumed to be stationary.^[16] The order of 10 was selected as the order of AR model and the number of poles of ARMA model. To choose the number of zeros for ARMA model, we tested our

method with different number of zeros. The general form of the ARMA model of vocal tract is defined as

$$s(n) + a_1 s(n-1) + \dots + a_{n_a} s(n-n_a) = b_1 u(n) + \dots + b_{n_b} u(n-n_b) \quad (3)$$

Where $s(n)$ and $u(n)$ are respectively the speech signal and the excitation signal, n_a represents the number of poles and n_b is the number of zeros plus 1. The Z-transformed version of Eq. 3 is

$$S(z) = \frac{B(z)}{A(z)} U(z) \quad (4)$$

Where $A(z)$ and $B(z)$ are polynomials, defined as below

$$A(z) = 1 + a_1 z^{-1} + \dots + a_{n_a} z^{-n_a} \quad (5)$$

$$B(z) = b_1 + b_2 z^{-1} + \dots + b_{n_b} z^{-n_b} \quad (6)$$

In our method, ARMA coefficients were calculated with a widely used iterative prediction-error method.^[17] This method ensures that all poles and zeros are inside the unit circle. AR coefficients were calculated with a same method by considering n_b equal to zero. As our proposed algorithm used the difference between the spectrums obtained using AR and ARMA coefficients to estimate hypernasality, we needed a distance measure between the spectrums. Therefore we used cepstrum coefficients of AR and ARMA model to have two comparable and equal sequences in length. The zero-pole form of the vocal tract filter cited in Eq. 4, is defined as

$$\frac{S(z)}{U(z)} = \frac{Kz^r \prod_{i=1}^{n_b} (1 - d_i z^{-1})}{\prod_{i=1}^{n_a} (1 - p_i z^{-1})} \quad (7)$$

Where p_i and d_i are respectively i th pole and i th zero of vocal tract filter. All poles and zeros were considered to be inside the unit circle. Since the system is minimum phase, its cepstrum can be uniquely determined as^[18]

$$c(m) = \begin{cases} \log(k) & m = 0 \\ \sum_{i=1}^{n_a} \frac{p_i^m}{m} - \sum_{i=1}^{n_b} \frac{d_i^m}{m} & m > 0 \\ 0 & m < 0 \end{cases} \quad (8)$$

It is easy to see from Eq. 8 that the cepstrum coefficients could be consider as a decaying sequence, which is the reason that a finite number of coefficients are sufficient to approximate it and therefore we can refer to truncated cepstrum as a cepstrum vector.^[18]

Let $c_{AR}(m)$ and $c_{ARMA}(m)$ be cepstral sequences of AR and ARMA models, respectively. Then, the geometric distance between the cepstral sequences is calculated by

$$DI = \sum_{m=0}^M [c_{ARMA}(m) - c_{AR}(m)]^2 \quad (9)$$

In order to compute the distance using Eq. 9, it has been reported that a sufficient accuracy can be obtained if M parameter is at least set to 3 times the order of the AR model.^[19] In this paper, M was set to 40; also we estimated our method by using 120 cepstrum coefficients, where the variations of cepstrum coefficients become very small, then we compared the two cases to ensure the accuracy of our method. During calculation of distances using Eq. 9, we found that normalizing AR or ARMA cepstrum coefficients on total frames of a speech sample, may lead to a more separable DI value for normal speakers and cleft palate speakers, so we estimated our method with both normalized and non-normalized cepstrum coefficients and compared the results. DI was calculated for each frame of signal, final decision for an utterance made by the average of DI on all of its frames called $DI_{average}$, also final decision for a subject attained by computing the mean of $DI_{average}$ on all of his or her utterances. Figure 2 shows a flow chart of our algorithm for calculation of distances and detection of hypernasality.

Speech Samples

Oral consonants require velopharyngeal closure to accomplish the separation of the oral and nasal cavities. In contrast, nasal consonants involve velopharyngeal opening that allows the propagation of sound energy into the nasal cavity. In children with cleft palate early onset and delayed offset of velar movement occurs before and after the oral cavity occlusion causes the vowel preceding and following nasal consonants to be nasalized for certain durations.^[20] Therefore, in this study vowels (/a/) extracted from 392 utterances consisted of disyllables (/pamap/) that uttered by 22 normal subjects and 13 subjects with cleft palate, were used. A series of /p/ and /m/ before and after vowels in the test word requires velopharyngeal closing and opening movements; this context was considered useful for measuring the amount of nasalization. Because both oral phoneme (/p/) and nasal phoneme (/m/) were produced at the labial place of articulation, the influence of the change of articulation position on nasal resonance could be controlled.^[20] The age range of the subjects for this study was 4-12 years. Children with cleft palate had the palate repaired through primary surgical correction and also they exhibited moderate or severe hypernasality. In order to collect the acoustic signals a high quality microphone (Shure Beta 54, USA) was used. The microphone was attached to a headset and positioned at a fixed distance of 3 cm away from the right side of the subject's mouth. The signal to noise ratio of all recordings was more than 30 dB. The sampling rate was 44.1 kHz with 16 bits of resolution.

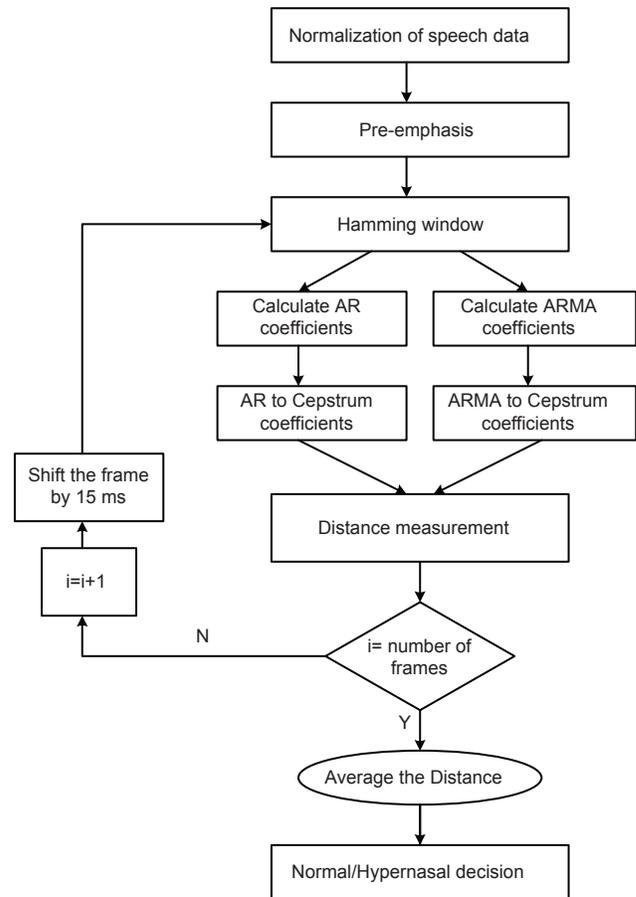


Figure 2: Flow chart of hypernasality detection method

RESULTS AND DISCUSSION

To evaluate the proposed algorithm, we applied the algorithm to each frames of utterances. We expected the $DI_{average}$ could be an appropriate index for detecting hypernasality and we could simply separate normal and hypernasal samples by setting a threshold value for $DI_{average}$.

In order to find the optimum number of zeros for ARMA model and also the best upper limit for the summation of Eq. 9 (M parameter), we performed a t -test. This test is employed to study the null hypothesis that data in the two groups of hypernasal and normal, are random samples with equal means, against the alternative that the means are not equal.^[21] The P value which is the result of this test could be used as a quantity that showed us how much our data is separable with proposed index; a large P value shows that the calculated indexes for two groups do not have a significant difference.

Hence, we calculated P values for $DI_{average}$ values of all utterances that were obtained by different number of zeros for ARMA model and different values of M parameter and chose the cases which have smaller P values to continue

our study with them. Table 1 shows the calculated P values in different cases; it is obvious that utilizing 120 normalized cepstrum coefficients had the best result among all. For better assessment of our method we plotted the distributions of $DI_{average}$ values of two groups of normal and hypernasal samples by means of boxplot for the best two cases selected from Table 1, in Figure 3.

As mentioned before we also calculated the means of $DI_{average}$ on all utterances of each subject and plotted the results for the best two cases in Figures 4 and 5. These

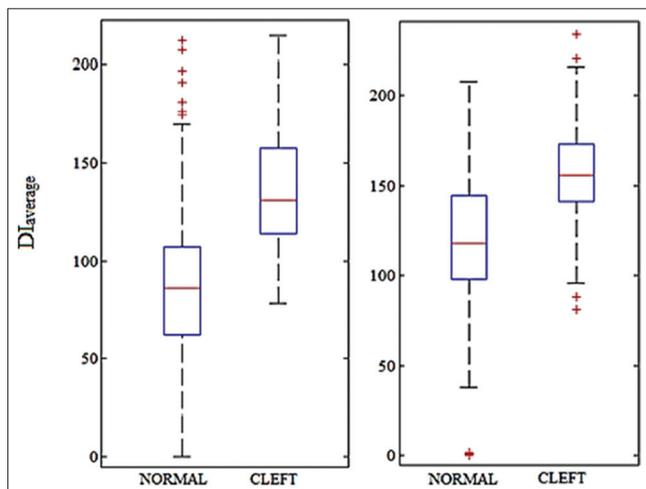


Figure 3: Boxplot of $DI_{average}$ for subjects with cleft palate and normal subjects by using 120 normalized cepstrum coefficients and autoregressive moving average model with two zeros (left), five zeros (right)

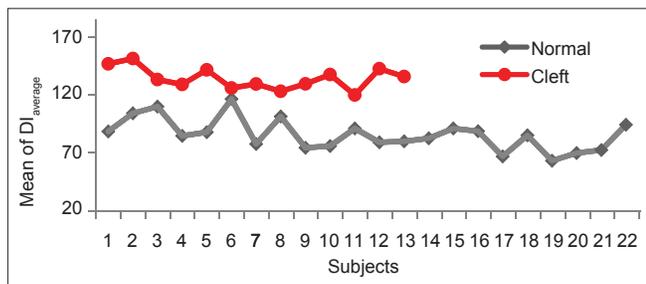


Figure 4: Mean of $DI_{average}$ for each subject with 120 normalized cepstrum coefficients by using autoregressive moving average model with two zeros

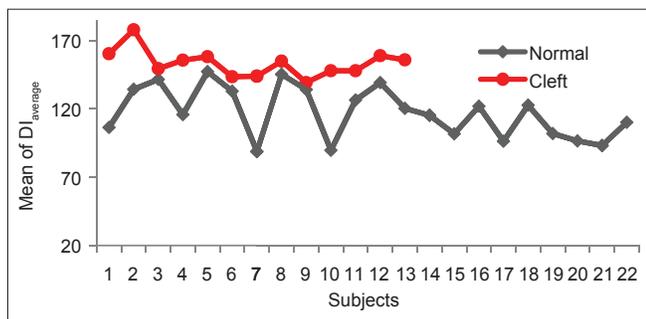


Figure 5: Mean of $DI_{average}$ for each subject with 120 normalized cepstrum coefficients by using autoregressive moving average model with five zeros

figures imply that for the proposed method, using two zeros in ARMA model for 120 Normalized cepstrum coefficients had promising results. Figure 4 shows that by setting an appropriate threshold value, we can simply separate two groups of subjects. In order to find the proper threshold value, we applied two well-known classification methods, k -means and Bayes and compared the results.

K -means partitions the data points into k clusters. This iterative partitioning minimizes the sum of the within-cluster summation of distances of points to the cluster centers, over all clusters. As in our approach each data point represent with one dimension value ($DI_{average}$ for each utterance), we consider the average of final values of cluster's centers obtained by k -means approach as the threshold value. Actually in our approach k was considered 2. Based on Bayes theorem, another approach was applied for threshold estimation; we fit a Gaussian distribution function for each group and select the intersection of the two groups' function as the threshold value.

In order to evaluate these two classifiers the leave-one-out cross-validation was employed. That means that for 35 (the number of subjects) separate times, the classifiers were trained on all data except for one subject and a prediction was made for that subject and related utterances. It means that in each step, k -means was trained on utterances of 34 subjects and the average of the centers of two obtained clusters settled as the threshold value, then the utterances of the rest subject compared with this threshold value. For the approach based on Bayes theorem the same procedure were applied. Actually, the classifiers were trained on utterances, but the results were presented for two levels of utterances and subjects. Table 2 shows the result of classification for all 392 utterances, which

Table 1: P values for different parameters of our method

M parameter	Number of zeros of ARMA model				
	1	2	3	4	5
40	1.4e-7	8.1e-2	9.5e-6	2.3e-4	4.4e-6
120	1.6e-7	8.6e-2	8.6e-6	2e-4	3.4e-6
40 (normalized)	3.1e-2	2.8e-16	8.2e-8	3.3e-9	1.8e-8
120 (normalized)	1.3e-1	4.1e-35	4.3e-22	1.6e-19	1.7e-22

ARMA – Autoregressive moving average

Table 2: Confusion matrix for utterances classification

Classifier	Results of classifier	Expert diagnosis		Threshold value
		Hypernasal	Normal	
K-means	Hypernasal	tp: 126	fp: 54	106.2491
	Normal	fn: 20	tn: 192	
	Total	146	246	
Bayes	Hypernasal	tp: 115	fp: 47	110.6190
	Normal	fn: 31	tn: 199	
	Total	146	246	

146 of them were hypernasal and 246 utterances were normal through the expert judgments. The represented results are the summation of the results obtained for each step of cross validation test. The average of threshold values of all steps also showed in this table. In the case of subjects, we compared the mean of $DI_{average}$ for utterances of the out subject (the subject who is out of the training phase, based on the leave-one-out cross validation) to the obtained threshold value and made the hypernasal or normal decision. The confusion matrix values for the case of subjects represented in Table 3.

For better comparison of the results of two classifiers, sensitivity, specificity and accuracy were calculated and represented in Table 4 for the case of utterances and also in Table 5 for the case of subjects. Note that the balanced accuracy, which is the average of sensitivity and specificity values, is not equal to the accuracy due to the unbalance of the datasets for normal and hypernasal subjects. Furthermore, note that both, the values of sensitivity and specificity, are a tradeoff that depends upon the threshold chosen. These results lead us to establish a protocol for assessment of hypernasality. This protocol starts with recording of several utterances of a nasalized vowel for each subject and continues with calculating $DI_{average}$ for all utterances with the method proposed in this paper and finally comparing the mean of $DI_{average}$ with settled threshold value and making the hypernasal or normal decision.

To assess the ability of our approach, some comparison with previous works would be useful. As mentioned in previous sections, some works tried to introduce a quantitative index for assessment of hypernasality. These works represented their results by values of correlation coefficients of their proposed index and either nasalance scores or results of perceptual judgments, which could not be compared with our results. In some other works results of classification with classifiers like SVM that trained on different acoustic feature, were announced. Since these works used different unreachable datasets of subjects with various degrees of hypernasality, we could not compare our results with them.

Therefore to have some comparable results, the algorithm proposed by Rah *et al.* was simulated and applied on our dataset, the obtained results presented in Tables 6 and 7. Within this algorithm after pre-processing and windowing, LPC coefficients of order 10 and 36 were extracted. The geometric distance between the high and low order LPC considered as an index for hypernasality. Better performance of our approach is obvious from the tables.

CONCLUSION

In this paper, we proposed a straightforward method to evaluate hypernasality. The method is much less troublesome

than the current clinical methods. Our method introduces an index proportionate to the amount of hypernasality that could be used for evaluating the effects of surgery for velopharyngeal insufficiency and cleft palate patients and helps the therapist to control the speech therapy process of these patients. We found out the best results were obtained with comparing the 120 normalized cepstrum coefficients of AR model with 10 poles to ARMA model with two zeros and the same number of poles. In this situation, we achieved classification accuracy of up to 81.12% for utterances and up to 97.14% for subjects. These results show grate improvement of classification accuracy compare with the results of a similar method based on LPC (74.49% for utterances and 77.14% for subjects). In addition, the $DI_{average}$ Index or sequences of distances, which were introduced in this paper could be used as a feature or representation vector for using with other classification methods.

Table 3: Confusion matrix for subjects classification

Classifier	Results of classifier	Expert diagnosis	
		Hypernasal	Normal
K-means	Hypernasal	tp: 13	fp: 2
	Normal	fn: 0	tn: 20
	Total	13	22
Bayes	Hypernasal	tp: 13	fp: 1
	Normal	fn: 0	tn: 21
	Total	13	22

Table 4: Result of the classification on utterances (given in %)

Classifier	Sensitivity	Specificity	Accuracy	Balanced accuracy
K-means	86.3	78.05	81.12	82.18
Bayes	78.77	80.89	79.83	80.1

Table 5: Result of the classification on subjects (given in %)

Classifier	Sensitivity	Specificity	Accuracy	Balanced accuracy
K-means	100	90.91	94.29	95.45
Bayes	100	95.54	97.14	97.72

Table 6: Result of the classification on utterances, LPC method (given in %)

Classifier	Sensitivity	Specificity	Accuracy	Balanced accuracy
K-means	69.86	77.24	74.49	73.55
Bayes	51.37	88.21	74.49	69.79

LPC – Linear predictive coefficients

Table 7: Result of the classification on subjects, LPC method (given in %)

Classifier	Sensitivity	Specificity	Accuracy	Balanced accuracy
K-means	76.29	77.27	77.14	76.78
Bayes	53.85	90.91	77.14	72.38

LPC – Linear predictive coefficients

REFERENCES

1. Horii Y, Lang JE. Distributional analyses of an index of nasal coupling (HONC) in simulated hypernasal speech. *Cleft Palate J* 1981;18:279-85.
2. Vijayalakshmi P, Reddy MR, O'Shaughnessy D. Acoustic analysis and detection of hypernasality using a group delay function. *IEEE Trans Biomed Eng* 2007;54:621-9.
3. Cairns DA, Hansen JH, Riski JE. A noninvasive technique for detecting hypernasal speech using a nonlinear operator. *IEEE Trans Biomed Eng* 1996;43:35-45.
4. Rah DK, Ko YL, Lee C, Kim DW. A noninvasive estimation of hypernasality using a linear predictive model. *Ann Biomed Eng* 2001;29:587-94.
5. Lee GS, Wang CP, Yang CC, Kuo TB. Voice low tone to high tone ratio: A potential quantitative index for vowel a: And its nasalization. *IEEE Trans Biomed Eng* 2006;53:1437-9.
6. Vijayalakshmi P, Nagarajan T, Jayanthan RV. Selective pole modification-based technique for the analysis and detection of hypernasality. *Proc. of IEEE Conf. Region 10; TENCON 2009*. p. 1-5.
7. Castellanos G, Daza G, Sanchez L, Castrillón O, Suarez J. Acoustic speech analysis for hypernasality detection in children. In: *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*; p. 5507-10.
8. Delgado-Trejos E, Sepúlveda-Sepúlveda FA, Castellanos-Domínguez G. Robustness improvement of hypernasal speech detection by acoustic analysis and the Rademacher complexity model. *Advances in biomed research*; 2009. p. 159-62.
9. Maier A, Reuß A, Hacker C, Schuster M, Nöth E. Analysis of hypernasal: Speech in children with cleft lip and palate. *Text, Speech and Dialogue*. Springer Berlin Heidelberg: Springer; 2008. p. 389-96.
10. Orozco-Arroyave JR, Murillo-Rendón S, Alvarez-Meza AM, Arias-Londono JD, Delgado-Trejos E, Vargas-Bonilla JF, et al. Automatic selection of acoustic and non-linear dynamic features in voice signals for hypernasality detection. In: *Proceedings of Interspeech*. Florence, Italy: This article presented in 12th annual conference of the international speech communication association (Inter Speech); 2011. p. 529-32. Available from: <http://www.ISCA.org>.
11. Arias-Londoño JD, Godino-Llorente JI, Sáenz-Lechón N, Osma-Ruiz V, Castellanos-Domínguez G. Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients. *IEEE Trans Biomed Eng* 2011;58:370-9.
12. Orozco-Arroyave JR, Vargas-Bonilla JF, Arias-Londoño JD, Murillo-Rendón S, Castellanos-Domínguez G, Garcés JF. Nonlinear dynamics for hypernasality detection in spanish vowels and words. *Cognit Comput* 2012;4:1-10.
13. Rabiner LR, Schafer RW. *Digital processing of speech signals*. Vol. 100. Englewood Cliffs, NJ: Prentice-Hall; 1978.
14. Deller J, Proakis J, Hansen JH. *Discrete-time processing of speech signals*. New York, USA: John Wiley and Sons; 2000. p. 266-342.
15. Gray A Jr, Markel J. A spectral-flatness measure for studying the autocorrelation method of linear prediction of speech analysis. *IEEE Trans Acoust* 1974; 22:207-17.
16. Mammone RJ, Zhang X, Ramachandran RP. Robust speaker recognition: A feature-based approach. *IEEE Signal Process Mag* 1996;13:58.
17. Ljung L. *System identification: Theory for the user*. Vol. 7632. New Jersey: Prentice Hall Inf and System Sciences Series; 1987.
18. Huang X, Acero A, Hon HW. *Spoken Language Processing*. Vol. 15. New Jersey: Prentice Hall PTR; 2001.
19. Gray A Jr, Markel J. Distance measures for speech processing. *IEEE Trans Acoust* 1976; 24:380-91.
20. Ha S, Sim H, Zhi M, Kuehn DP. An acoustic study of the temporal characteristics of nasalization in children with and without cleft palate. *Cleft Palate Craniofac J* 2004; 41:535-43.
21. Gibbons JD, Chakraborti S. *Nonparametric statistical inference*. Vol. 168. New York: Marcel Dekker, Ink; 2003. p. 247-61.

How to cite this article: Akafi E, Vali M, Moradi N, Baghban K. Assessment of Hypernasality for Children with Cleft Palate Based on Cepstrum Analysis. *J Med Sign Sens* 2012;3:195-209-16.

Source of Support: Nil, **Conflict of Interest:** None declared

BIOGRAPHIES



Ehsan Akafi was born in Isfahan, Iran, in 1987. He received the B.S in Electronic Engineering from Shiraz University and M.S degree in Biomedical Engineering from Shahed University. His research interests include Signal and Speech Processing.

E-mail: ehsan_ek12@yahoo.com



Mansour Vali received the B.S in electronic engineering from Isfahan university of technology and M.S degree in Biomedical Engineering from Sharif University of technology. He received PHD degree in Biomedical Engineering in 2006 from Amirkabir university of technology. He is now assistant Professor at K.N Toosi university of technology. His research interests include Biomedical signal processing, speech processing and artificial neural network.

E-mail: mansour.vali@eetd.kntu.ac.ir



Negin Moradi recived her B.S. and M.S degree in Speech and Language Pathology in Iran University of Medical Sciences respectively in 2003 and 2005. She has finished her Ph.D. in Speech and Language

Pathology at Tehran University of Medical Sciences in 2013. She is now a Speech and Language Pathologist and assistant professor in department of Speech Therapy in Ahvaz Jundishapur University of Medical Sciences.

E-mail: neginmoradist@gmail.com



Kowsar Baghban received the B.Sc degree in Speech and Language Pathology from Isfahan University of Medical Sciences, Isfahan, Iran in 2010 and her M.Sc degree from Tehran University of Medical Sciences in 2012. She is now a member of department of Speech Therapy in Hamedan University of Medical Sciences.

E-mail: baghban.kowsar@gmail.com