Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Acoustic Speech Analysis for Hypernasality Detection in Children

G. Castellanos, G. Daza, L. Sánchez, O. Castrillón, J. Suárez
Grupo de Control y Procesamiento Digital de Señales
Universidad Nacional de Colombia Sede Manizales
cgcastellanos, gdazas, lgsanchezg, odcastrillong, jfsuaresc **@unal.edu.co**

Abstract— **Here, an analysis of different acoustic features and their influence in automatic identification of hypernasality is shown. Effective feature selection method includes preprocessing of the initial feature space based on statistical independence analysis. Simultaneously, the synthesis of a specialized diagnostic feature is proposed based on analyzing the acoustic emission of the hyper nasal speech. As a result, It is obtained the acoustic features can differentiate with enough precision the pathology. However, the proposed feature does not require training samples and less computational power, as well.**

I. Introduction

In the treatment of children with fixed Clip Lip and Palate (CLP), problems might appear related to vocal emission and resonance, such as: Hipponasality and hypernasality. Nevertheless, according to the report presented in [1], it is more frequent to find hypernasality cases (90 %) than hippo nasality (10 %). The interest shown in the hypernasality detection is due to its occurrence, points out problems of anatomical and neurological sort, and also related to peripheral nervous system [2]. The presence of hypernasality, understood as the leak of nasal air and compensatory joints, leads to low intelligibility of speech which causes declining of communication capability of the subject with its environment who may develop changes in their interpersonal attitude and behavior.

In velopharyngeal learning, the distortion of the acoustic production leads up to nasalized voice. Besides, since the air loss or nasal leak is massive, articulator mechanisms are compromised. The patient can not speak clearly and intelligibly, so then they replace their velum palatine sphincter by glottal articulation that let clearer articulation: /*p*/, /*t*/, /*k*/, /*b*/, /*d*/, /*g*/ come from beatings of the glottis, while sounds like /*ch*/, /*s*/, /*t*/, /*j*/ are substituted by hoarse blowings [3]. Although hard palate has been repaired surgically, yet it might not provide velopharyngeal competence for a normal speech production. Even if the palate is potentially capable after surgery, previous speech habits could have headed to developing errors of compensatory articulation or physiologic compensation which aim is to approx to intelligibility; enhancing then the number of pathological patterns in speech. As a result, compensatory articulations persist generally, even after undergoing post technical or post surgical managing that had forecasted a plenty shutting. Thus, they have to be fixed before increasing the performance of the velopharyngeal sphincter throughout language therapy.

In the last years, it has grown the interest for acoustic speech analysis (ASA) as an alternative method for diagnosis and treatment in identifying functional disorders in children's voice [4]–[6]. This sort of analysis exposes great advantages over traditional methods due to its non-invasive nature and potential to obtain a quantitative measure on clinic state of larynx and vocal tract.

Acoustic features or objective parameters are frequently used to represent the pathologic voice on held vowels [7]–[9]. However, such vectors are limited in their robustness because of their estimation complexity in real conditions with perturbations of non-stationary structure [10]. Although several analysis of effectiveness have been made around the different kinds of proposed features for objective evaluation of speech disorders [7], [11], they can not be taken a standard set of parameters for hypernasality identification because each disorder affects differently diverse aspects of speech emission.

In the present work is analyzed the statistical effectiveness of the different acoustic features in the automatic identification of hypernasality. Acoustic features reflect part of information contained in perceptual analysis; in part, due to their estimation is derived directly or indirectly from the vocal cords behavior. Consequently, it is convenient to apply multivariate analysis techniques in determining the effectiveness of voice features. The effectiveness is studied by using multivariate analysis techniques such as MANOVA in a heuristic focus of feature selection to train classifiers. Moreover, it is analyzed the use of Wavelet Transform to estimate the quality of vocal emission.

II. Methods

II-A. Nasalization and nasal emission

Nasalization is defined as the link between nasal cavity and the rest of the vocal tract; while nasal emission refers to abnormal air loss through nasal route. This abnormal leakage reduces intra-oral pressure causing distortion in consonants. When air loss turns into an audible re-blowing; the nasal emission is more obstructive and speech is seriously affected. Nasality commonly named hypernasality refers to low speech quality, which results from inappropriate adding of the resonance system to vocal tract. Conversely to nasal emission, nasality does not involve large flows of nasal air, so that there is no significant change in intra-oral air pressure. For this pathology identification two methods have been proposed. The first one, based on the signal modeling (specialized diagnosis) and the second one is founded on pattern recognition techniques.

1-4244-0033-3/06/$20.00 ©2006 IEEE.

II-B. Specialized diagnosis feature

Considering that the normal voice is made of resonances at different frequencies Formants F_k, it is proposed the following acoustic model [2]:

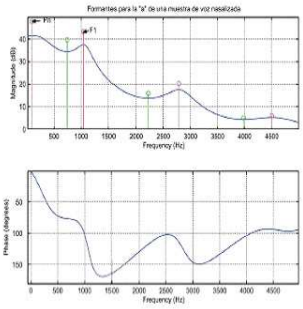
$$S_n(\omega) = \sum_{k=1}^{K} F_k(\omega) \tag{1}$$

In contrast to normal voice, nasalized voice is the appearance of the anti-formants $\hat{F}'$ and nasal formants $\hat{F}$:

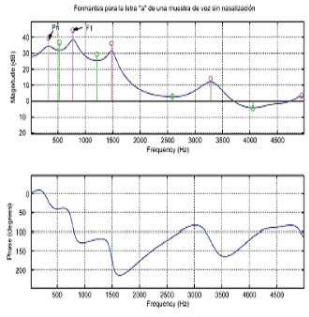
$$S_h(\omega) = \sum_{k=1}^{K} F_k(\omega) - \sum_{l=1}^{L} \hat{F}'_l(\omega) + \sum_{m=1}^{M} \hat{F}_m(\omega) \tag{2}$$

It has been suggested that the intensity in the reduction of the first formant is a primary indicator of nasality. In [2], the superior formants are filtered in such a way that filtered normal voices will have one component, while nasalized voices will correspond to a signal with several components, which can be estimated using Teager's instant power operator.

In a general way, patients with CLP manifest nasality with a deficiency of the velopharyngeal port, coupling their nasal cavity to vocalic sounds that gives place to appearance of an additional resonance in the amplitude-frequency feature of the vocal tract, and so, a noticeable decrease in the formants F_1 and F_2 [12] as shown in Fig. 1(a) and 1(b).



(a) Pitch Energy (Pathological case).



(b) Pitch Energy (Normal case).

Fig. 1. Pitch Energy for distinguishing glottal beating.

II-C. Acoustic features

They can be split into two categories according to the acoustic properties to be measured. Based in additive noise, among them: HNR (*Harmonic to Noise Ratio*) that corresponds to the mean value of the vocal emission noise component; GNE (*Glottal Noise Excitation*) defined as the noise estimation and it is based in the assumption that resulting glottal pulses from collisions of vocal folds head to a synchronous excitation of the different band frequencies; and NEP (*Normalized Error prediction*) that can be expressed as the relationship between geometric and arithmetic means of spectral model.

Other acoustic features are associated to frequency modulation noise, among them Pitch or fundamental period of the signal and Jitter defined as the average variation percentage between two consecutives values of Pitch. In addition, there are considered features associated to parametric models of speech generation. Among them: Cepstral coefficients derived from linear prediction analysis (LPC *Linear Prediction Coefficients*), Cepstral coefficients over pounded frequencies scale (MFCC *Mel-Frequency Cepstrum Coefficients*), and RASTA coefficients (*Relative Spectral Transform*) [13].

III. Experimental background

III-A. Database

The sample is constituted by 90 children. Classes are balanced (45 patients with normal voice and 45 with hypernasality), evaluated by specialist. Each recording is conformed by five words of Spanish language: */coco/*, */gato/*, */jugo/*, */mano/* and */papá/*. Signals are acquired under low noise conditions using a dynamic, unidirectional microphone (cardioide). Signal range is between $(-1, 1)$.

III-B. Initial feature space

Complete set of considered features are: Pitch (F_0), Jitter, percent Jitter (J_P), Tone perturbation coefficient, NEP, GNE, HNR, Energy, Zero crossings, Cepstrum , LPC y MFCC. Given the feature, which is corresponded by $\mathbf{x}$, a vector formed by the measurement observations $\{x_i : i = 1, \ldots, n\}$ for each of the k classes, it is carried out the estimation of the moments as descriptors of randomness structure of empirical distributions. The analysis moments for each ξ feature are the followings:

1. *Position parameters.* Initial moment of first order (mean value $\tilde{m}_{1\xi}$) is estimated in order to be removed from observations, this is $\mathbf{x} - \tilde{m}_{1\xi}$, it is considered that such value has no relevant information, but it could generate an inappropriate slant to understand the results. In practice, there are other estimations for mean value like the median.
2. *Scale parameters.* The followings are considered: variance, quadratic mean value, variation coefficient $\tilde{m}_{1\xi}/\tilde{\sigma}_\xi$, peak to peak value and absolute median deviation $\operatorname{med}|x_i - \tilde{m}_{1\xi}|$. Additionally, there are considered centralized moments of order $l = 3, 4, 5, 6$
3. *Shape parameters.* Correspond to asymmetry and excess coefficients (obliquity and kurtosis respectively). Besides this preceding moments, it can be considered the cumulative of order $m = 1, \ldots, 4$.

128 variables per word are analyzed. hence, initial feature space increments up to 640 speech features.

III-C. Data preprocessing

Its main target is to reduce or even eliminate the influence of measurement errors. Among them: systematic errors during acquisition, occasional failures in the measurement instruments and so on. It is also used for controlling the homogeneity principle of the different statistical properties of phenomena in analysis. Data preprocessing consists on analyzing odd logs for each feature and vivificating their normality.

III-D. Effective feature selection

The proposed methodology for reducing the initial spaces is based on strong relevance analysis, which studies the correlation among features. Reduction methodology considers combining different heuristic techniques (DFCC) [13] to generate features sets and multivariate analysis for further evaluation of such sets. Particularly, Multivariate analysis of variance (MANOVA) is used as the intrinsic cost function; therefore, finding the feature subset that displays better discriminative power among classes.

III-E. Classification

The employed classifier is Bayesian. Five classifiers of this kind are used; each of them for analyzing the error between classes (hyper-nasal and normal or control) for each one of the words previously mentioned. Besides, cross-validation was conducted to see the variation in the classifier's parameters and its generalization capability.

IV. Results and Discussion

IV-A. Acoustic features analysis

Acoustic feature effectiveness can be measured according to classification performance. Next, the results are displayed for each one of the stages.

The abnormal value detection is conducted for each feature ξ. It allows making clear the quality of the measurements. For acoustic features results are shown on Table I. Although average reduction for word-set is between 25 to 30 %, the classifier does not converge due to the high dimensionality of the train set. Covariance matrixes tend to be singular. In

TABLE I

Word of analysis	Reduced space dimensionality [%]	Classifier's performance
Coco	70.3	No convergence
Gato	71.9	No convergence
Jugo	61.7	No convergence
Mano	71.1	No convergence
Papá	68	No convergence

what concerns to normality verification, Gaussian structure of data can be achieved by using hypothesis test. In this case, the Kolmogorov-Smirnov test. If the verification tests of the distribution result in rejecting normality, means of transforming the observation and so, accepting the hypothesis are needed. For such aim, *Box-Cox* transform is carried out. The results for dimension reduction and classification performance can be seen in Table II.

TABLE II

Word of analysis	Reduced space dimensionality [%]	Classifier's performance
Coco	61.7	No convergence
Gato	64.8	No convergence
Jugo	53.1	No convergence
Mano	64.8	No convergence
Papá	60.9	No convergence

The dimension reduction achieved by these two methods can be compared in terms of percentage. Yet such reductions are not enough to reach the classifier's convergence because high dimensionality is still a problem to calculate the Covariance matrixes. Moreover, another problem is the scale disparity that still exists among acoustic features $\left(10^{-3}, 10^{6}, 10^{15}\right)$.

The paramount purpose of effective feature selection is to find those features that allow a better discrimination among classes. In this case it is accomplished by using a heuristic technique for growing called "filter". MANOVA becomes then, the cost function. The effective feature selection is applied twice (Table III): a) Without data preprocessing, b) After data preprocessing. Results show that even though the reduction percentage could be similar, the classifier's performance is substantially different; being better after data preprocessing.

TABLE III

	Space without preprocessing	Performance without preprocessing	Preprocessed space	Preprocessed space performance
Coco	3.9	83.3	3.1	99.9
Gato	9.3	88.9	3.1	99.9
Jugo	7	91.1	3.1	99.9
Mano	6.3	93.3	3.1	99.9
Papá	3.9	91.1	3.1	99.9

The analysis on Table IV, which displays selected features for each word, bring as a result that effective features are related to the following acoustic parameters: GNE that is present for each word and is defined as a noise estimation, HNR is the average of the vocal emission noise component, NEP as the relationship between geometrical and arithmetic mean values of spectral model. All these features are associated to the measurements of additive noise. This can be linked to the additional components for hypernasality, which were previously described in the mathematical model shown in (2). Another type of features is the ones connected to parametric models of speech generation (LPC and Cepstral coefficients); particularly ARMA models that are stimulated by noise, as well. In the mean case, linear prediction coefficients (LPC) with the most influence are the first ones. They have higher amplitudes, therefore less sensitive to measurement variability.

The results of the concordance study between evaluators and the automatic recognition system are posted on Table V. It can be seen the system performance is highly significant.

TABLE IV

Word of analysis	Selected Features
Coco	Average of GNE, median of GNE, standard deviation of Cepstral coefficients, third centralized moment of Cepstral coefficients.
Gato	Covariance matrix's mean value of GNE, standard deviation of HNR, Cepstral coefficients kurtosis, sixth Cepstral coefficient.
Jugo	Covariance matrix's mean value of GNE, standard deviation of HNR, Cepstral coefficients kurtosis, sixth Cepstral coefficient.
Mano	Standard deviation and variance of GNE, LPC 3, LPC 5.
Papá	Maximum GNE, standard deviation of GNE, Cepstral coefficients variance, range of LPC.

Test were made using *Jackknife* for a likelihood level of 95 %.

TABLE V

Concordance between specialists (Phone-audiologists) and automatic recognition system [%]	
	Automatic recognition system
Specialist 1	98
Specialist 2	97.7
Specialist 3	86.2
Specialist 4	88.7

IV-B. Specialized diagnostic feature

For this case the test results are shown on Table VI. Concordance study carried out under equal conditions exposed an effectiveness of 69 %.

TABLE VI

	No Nasality [%]	Nasality [%]		
		Low	Moderate	Severe
Control patients (normal)	3.1	30.1	59.8	7
Pathological patients	2.1	65.1	28.1	4.7

V. Conclusions

Acoustic features let discriminate with enough precision the pathology. One of the main problems for clustering voice features in Hypernasality analysis is their sensibility to acoustic properties. This mutual dependency might be one of the reasons for their contradictory interpretation along literature. Consequently, phonemes of velar articulation patterns ought to be identified (lip posture detection) and vocal emission (acoustic analysis) with enough discriminating power.

In the case of hypernasality detection, it is better to design specialized diagnostic features that reflect the nature of the irregularities in the functional state of speech. For such purpose, a specialized diagnostic feature is proposed. This feature offers a similar detection level compared to acoustic features, but being less computationally complex. Moreover, there is no need of training and also allows to differentiate the compromise degrees of resonance of the pathology.

Detecting the hypernasality degrees is important because in diagnosis as well as in treatment and surgery; the specialists need to have available a taxonomic reference system that can be useful for relating the different compromise degrees of resonance to velopharyngeal patterns; generating then, different alternatives for treatment while excluding others.

Acknowledgments

The present word is enclosed by the project called "*Acústica de labio y paladar hendido en la zona centro del país*"(Acoustics of clip lip and palate in center zone of the country). It is funded by the Caldas University and the National University of Colombia. Likewise, thanks to the Dr. Colombia Quintero, the phone-audiologists Beatriz Mejía and Ana Maria Escandón, and the epidemiologist Arnoby Chacón whom collaborated in the expertise of the database used for this work.

References

[1] G. Castellanos, F. Prieto, and C. Quintero, "Análisis acústico de voz y de posturas labiales en pacientes de 5 a 15 años con labio y/o paladar hendido corregido en la zona centro del país," Colciencias and Universidad Nacional and Universidad de Caldas, Tech. Rep., 2004.

[2] D. A. Cairns, J. H. L. Hansen, and J. E. Riski, "A noninvasive technique for detecting hipernasal speech using a nonlinear operator," *IEEE Transaction on Biomedical Engineering*, 1996.

[3] A. Habbaby, *Enfoque integral del niño con fisura labiopalatina*. Buenos Aires: Médica Panamericana, 2002.

[4] G.Ñiedzielska, "Acoustic analysis in the diagnosis of voice disorders in children," *International Journal in Pediatric Otorhinolaryngology*, 2000.

[5] J. González, T. Cervera, and J. L. Miralles, "Análisis acústico de la voz: Fiabilidad de un conjunto de parámetros multidimensionales," *Acta Otorrinolaringol*, 2002.

[6] G.Ñiedzielska, E. Glijer, and A.Ñiedzielski, "Acoustic analysis of voice in children with noduli vocales," *International Journal in Pediatric Otorhinolaryngology*, 2001.

[7] P. Yu, M. Ouaknine, J. Revis, and A. Giovanni, "Objective voice analysis for dysphonic patients: A multiparametric protocol including acoustic and aerodynamic measurements," *Journal of Voice*, 2001.

[8] S. Hadjitodorov and P.Mitev, "A computer system for acoustic analysis of pathological voices and laryngeal diseases screening," *Medical Engineering and Physics*, 2002.

[9] R. D. Kent, H. Vorperian, and J. Kent, "Voice dysfunction in dysarthria: application of the multidimensional voice program," *Journal of communication Disorders*, 2003.

[10] M. Gupta and A. Gilbert, 2001.

[11] M. Frohlich and D. Michaelis, "Acoustic voice analysis by mean of the hoarseness diagram," *Journal of speech, language and hearing research*, vol. 3, no. 43, 2000.

[12] R. J. Baken, *Clinical Measurement of Speech and Voice*. Thomson Delmar Learning, 1996.

[13] G. Castellanosa, O. Castrillón, and E. Guijarro, "Multivariate analysis techniques for effective feature selection in voice pathologies," in *CASEIB*, 2004.