

The Automatic Detection of Hypernasality in Cleft Palate Speech Based on an Improved Cepstrum Method

Fang-Ling FU^{1,a}, Jia-Hui QIAN^{1,b}, Fei HE^{1,b}, Heng YIN^{2,c}, Xi-Yue WANG^{1,d},
Ling HE^{1,e,*}

¹School of Electrical Engineering and Information, Sichuan University, Sichuan, China

²West China Hospital of Stomatology Sichuan University, Sichuan, China

^aFufangling0201@gmail.com, ^bQianjiahur1997@163.com, ^c1067823139@qq.com

^dyin.heng@scu.edu.cn, ^eling.he@scu.edu.cn

Keywords: Hypernasal Speech, Formant Extraction, Cepstrum, Cleft Palate.

Abstract. Cleft palate is a birth defect with the resonance between the nasal cavity and oral cavity, which causes an additional nasal formant. And the accurate formant extraction method contributes to hypernasality detection. In this sense, we present a method for formant extraction. The proposed approach uses multi-resolution analysis of automatic segmented filtering combining cepstrum method to accurately extract the formants. It can increase the resolution in the low-frequency region, which is helpful for extracting the addition nasal formant around 250Hz. The proposed technique is tested on 416 vowels /a/ recorded by the West China Hospital of Stomatology, including 216 vowels for normal speech and 210 vowels for cleft palate speech. The formant frequency and formant amplitude are used as acoustic features for hypernasality detection, and the detection accuracy of hypernasality in cleft palate speech is 97.41%.

Introduction

Cleft palate can lead to a series of problems: feeding problems, speech disorders, appearance of the face, ear infections and psychological disorders. The cleft palate can affect the movement of the velopharyngeal valve, thus resulting in Velopharyngeal inadequacy (VPI). Velopharyngeal inadequacy is the incomplete closure of the velopharyngeal sphincter between the oropharynx and the nasopharynx [1]. The cleft palate patient with velopharyngeal inadequacy is characterized by hypernasal resonant speech [2].

Repair of a cleft palate often requires multiple surgeries. Sequence of the cleft palate treatment is a long procedure, and the initial surgery creates a functional palate [3]. If the functional palate works undesirably, the abnormal resonance still exists during speech production. In this situation, a further surgery is required. In the clinical environment, professional doctors state that the degree of nasal is proportional to the opening size of velopharyngeal. So accurate assessment of hypernasality grade is very important for the judgment whether a further surgery is needed.

In the clinical environment, the invasive techniques use invasive instruments like nasendoscopy, videofluoroscopy to assess hypernasality [4]. For the non-invasive technique, a voice and resonance evaluation completed by the Speech Language Pathologist (SLP) can provide informative data for hypernasal assessment [5]. This method is non-invasive and subjective, but the result assessment can easily be influenced by the subjective perception of SLP. Digital signal processing based techniques provided a convenient and accurate way to hypernasal assessment, which can be applied only with a high-quality microphone and a computer.

Researchers try to detect hypernasality by analyzing the speech of children with cleft palate, synthesized speech, and nasalized vowels of normal speech. And the main methods of hypernasality detection are divided into three categories using: 1) acoustic parameters; 2) vocal tract model; 3) nasal formant. Castellanos et al. use a set of parameters including pitch, jitter, tone perturbation coefficient, energy, zero crossings, LPC and MFCC to detect hypernasality [6]. Ehsan et al. put forward an ARMA model with typical number of poles and appropriate number of zeros to model

the hypernasal speech signal. And the geometric distance between the cepstral sequences of AR model and ARMA model is used to detect hypernasality [7]. Vijayalakshmi et al. find an additional nasal formant in the low-frequency region (around 250Hz). They use group delay function to resolve two closely spaced formants and define the group delay function-based acoustic measure (GDAM) to detect hypernasality [8].

Due to the appearance of an additional nasal formant, the formant is used as the acoustic feature for hypernasality detection in this work. However, classical methods of formant extraction are hard to separate the additional nasal formant and the first normal formant because of the weak spectral resolution. In this work, a new method for formant extraction is presented using a combination of automatic segmented filtering and Cepstrum technique. The objective of this method is to increase the spectral resolution especially in the low-frequency region, which is helpful to separate the additional nasal formant and the first normal formant. The formant frequency and formant amplitude are applied as the acoustic features for hypernasality detection.

Proposed Formant Estimation Method

The proposed formant estimation method combines the automatic segmented filtering and Cepstrum technique. The speech signals are framed and windowed firstly. Then each framed signal is decomposed into different frequency sub-bands using automatic segmented filtering. Each sub-band is further processed by Cepstrum technique. The formants can be extracted from the spectral envelope using peaks picking method.

Automatic Segmented Filtering

In normal speech, the resonance is the strengthen part of spectrum, which forms peak in power spectrum. The location of the peak in frequency is called formant. In other word, locating the resonance is helpful for finding the accurate formant frequency. Based on the relationship between the formant and resonance, a frequency-band segmented algorithm is proposed. In this algorithm, power spectrum is calculated by Linear Prediction Coefficient (LPC).

The vocal tract system can be modelled as an all-pole system clarified in [9, 10] using Eq.1

$$H_v(z) = \frac{G_v}{\sum_{k=0}^l a_k z^{-k}} \quad (1)$$

Using the coefficients of LP and Fourier transform, the power spectrum can be calculated as Eq.2:

$$P(f) = |H(f)|^2 = \frac{G^2}{1 - \sum_{k=1}^p a_k e^{(-j \frac{2\pi k \square fs}{f})^2}} \quad (2)$$

The location of resonance changes with the position of nostril, pharyngeal cavity and oral cavity. Moreover, in article [8], the speech whose sample frequency is 8000Hz have five formants. Thus, in order to propose a robust algorithm which is suitable for normal speakers and cleft palate speakers, the segment of frequency is judged by the average power of the five segments.

The segmented algorithm is introduced as follows:

Sept 1: Calculate the sum of the power spectrum:

$$P = \sum_{k=1}^{fs/2} P(f) \quad (3)$$

Step 2: In this paper five formants should be obtained. Therefore, the formant spectrum should be divided into five segments.

$$N_p = P / 5. \quad (4)$$

Step 3: Divide the whole frequency spectrum into five segments. The starting and stopping frequencies in each segment have to meet the following condition:

$$\sum_{k=f_{start}^i}^{f_{stop}^i} P(f) \leq N_p \quad (5)$$

Where the f_{start}^i is the starting frequency of i th segment and f_{stop}^i is the stopping frequency of i th segment.

Based on the segmented algorithm proposed above, the whole frequency spectrum can be divided into five segments. However, the nasal formant, which is an important cue for hypernasality detection, is very closed to the first normal formant F1. The nasal formant may be divided into the same segment with F1. As the article [8] shows, the nasal formant is around 250Hz. Thus, the first segment and the second segment are merged as one segment, from which two formants can be calculated.

Cepstrum Calculating

In this section, the calculating steps of cepstrum is specifically described. The concept and principle of cepstrum are resolved in detail in [11]. The cepstrum sequence $\hat{x}_i^m(n)$ can be gained from the method in [11]. In order to separate the glottal pulse and vocal tract impulse response, a low-pass window function is additionally used. The low-pass window function is set up on the cepstrum sequence $\hat{x}_i^m(n)$. A rectangular window is applied, which is represented by Eq.6

$$w(n) = \begin{cases} 1 & n \leq n_0 - 1 \text{ and } n \geq N - n_0 + 1 \\ 0 & n_0 - 1 < n < N - n_0 + 1 \end{cases}, n = 0, 1, \dots, N-1 \quad (6)$$

where n_0 is the length of the low-passing filtering window. And it is related to the sampling frequency and the points of Fast Fourier Transform.

And, the whole flow chart of the cepstrum technique is shown in Fig.1.

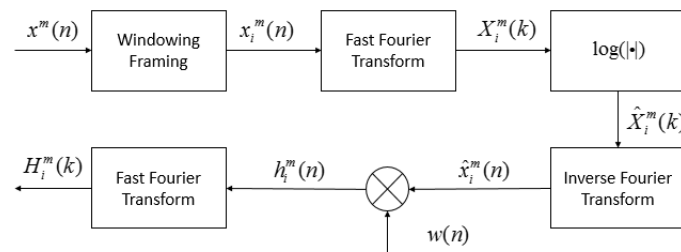


Fig.1. The flow chart of the cepstrum technique

Automatic Hypernasality Detection in Cleft Palate Speech

Cleft Palate Speech Data

In this paper, a total of 426 phonemes /a/ are used for hypernasality detection, which are recorded by the West China Hospital of Stomatology. And 216 out of 426 speech utterances are recorded from normal speakers, and 210 out of 426 speech utterances are recorded from the cleft palate speakers. For each speech utterance, the formants are extracted from the four sub-bands respectively. And the formant frequency and formant amplitude are used as acoustic features for hypernasality detection.

Automatic Hypernasality Detection

In this work, the K-Nearest Neighbors (KNN) method is employed for classification. KNN is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (distance functions) [12]. The classification of hypernasality use the 10 times 10-folds cross validation method in this work.

Experimental Results

In this paper, the automatic segmented filtering combining with cepstrum technique and the classical cepstrum technique are respectively applied to extract the formant frequency and formant amplitude. The spectral envelope of the automatic segmented filtering combining cepstrum method obtained from the first sub-band D1 is shown in Fig.2 (a). And the spectral envelope of classical cepstrum method obtained from the whole frequency band is shown in the Fig.2 (b). Comparing this two figures, it clearly shows that the proposed technique employing the multi-resolution analysis of the automatic segmented filtering has superior capability to extract the formant. Both Fig.2 (a) and Fig.2 (b) are painting with the same speech frame spoken by a cleft palate patient with hypernasality, pronouncing the phoneme /a/. In Fig.2 (a), two formants are extracted from the sub-band D1 (0-850Hz). An additional nasal formant around 220Hz is found and the first normal formant around 650Hz is found.

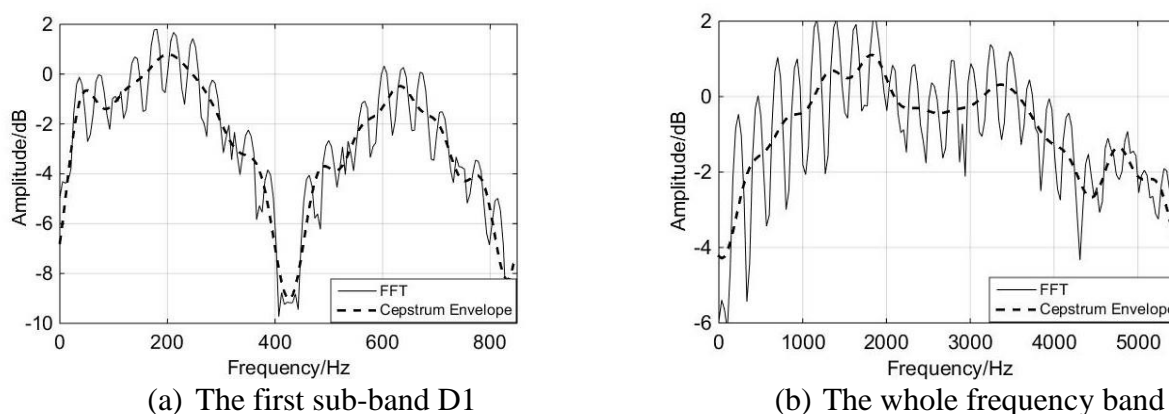


Fig.2 The spectral envelope of: (a) the first sub-band D1 by proposed method (b) the whole frequency band by classical method

The formant frequency and formant amplitude are used as the acoustic features for classification of hypernasality detection. And the accuracy of hypernasality detection is shown in Table 1. The average accuracy of hypernasality detection using classical cepstrum method is 80.80%. The average detection accuracy of this classification using the proposed method is 97.41%.

Table 1. The accuracy of hypernasality detection

The method	The accuracy of different type speech and the average accuracy		
	Normal speech	Hypernasal speech	Averagy accuracy
Automatic segmented filtering combining cepstrum method	98.15%	96.67%	97.41%
Classical cepstrum method	77.31%	84.29%	80.80%

Discussion and Conclusion

This paper has proposed a method to detect hypernasality in cleft palate using the automatic segmented filtering combining cepstrum method. The additional formant around 250Hz for phoneme /a/ is found in the nasalized vowel, which is because of the adjacent nasal resonance. However, the nasal formant is hard to detect because the classical methods of formant extraction have poor frequency resolution. So in this paper, the automatic segmented filtering method is used to improve frequency resolution, especially in the low frequency region. Firstly the whole frequency spectrum is automatically divided into four sub-bands based on the power spectrum computed by LPC. Speech is respectively filtered in four sub-bands. There are a nasal formant and the first normal formant in the first sub-band, and the left three sub-bands contain one normal

formant each. Then cepstrum method is applied in each sub-band, and the formant frequency and formant amplitude are extracted from the spectral envelope using peaks picking method. The formant frequency and formant amplitude are used as acoustic features to detect hypernasality. For the hypernasal speech /a/, a nasal formant around 220Hz and the first normal formant around 650Hz are found in the first sub-band. And the average detection accuracy of this classification using the formant with the proposed method is 97.41%.

Acknowledgement

This work is supported by the National Natural Science Foundation of China 61503264.

References

- [1] S. Ha, H. Sim, M. Zhi, D. P. Kuehn, "An acoustic study of the temporal characteristics of nasalization in children with and without cleft palate", *The Cleft palate-craniofacial journal*, vol. 41, no. 5, pp.535-543, 2004.
- [2] A. W. Kummer, J. L. Strife, W.H.Grau, N. A. Creaghead, "The effects of Le Fort I osteotomy with maxillary movement on articulation, resonance, and velopharyngeal function", *The Cleft palate journal*. vol. 26 , no.3, pp. 199-200, 1989.
- [3] Robert J. Shprintzen, Eileen Marrinan, "Velopharyngeal insufficiency: diagnosis and management", *Current Opinion in Otolaryngology & Head and Neck Surg* . 2009 Aug; 17(4): 302–307.
- [4] Jones MC, "Prenatal diagnosis of cleft lip and palate: detection rates, accuracy of ultrasonography, associated anomalies, and strategies for counseling", *The Cleft Palate-craniofacial Journal*. 2002 Mar;39(2):169-73.
- [6] Castellanos G, Daza G, Sánchez L, "Acoustic speech analysis for hypernasality detection in children", *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2006;1:5507-10.
- [5] P. Vijayalakshmi, T. Nagarajan, V. Jayanthan Ra, "Selective pole modification-based technique for the analysis and detection of hypernasality", *TENCON 2009-2009 IEEE Region 10 Conference*, 2009, pp. 1-5.
- [7] Ehsan Akafi, Mansour Vali, Negin Moradi, "Detection of hypernasal speech in children with cleft palate", *Biomedical Engineering (ICBME) of 2012*. 10.1109/2012.6519688.
- [8] Vijayalakshmi P1, Reddy MR, O'Shaughnessy D, "Acoustic analysis and detection of hypernasality using a group delay function", *IEEE Transaction on Bio-medical Engineering*. 2007 Apr;54(4):621-9.
- [9] R. C. Snell and F. Milinazzo, "Formant location from LPC analysis data," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no.2, pp. 129–134, 1993.
- [10] S. S. McCandless, "An algorithm for automatic formant extraction using linear prediction spectra," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 22, no. 2, pp. 135–141, 1974.
- [11] D. Gargouri, M. A. Kammoun, A. B. Hamida, A comparative study of formant frequencies estimation techniques, *Proceeding of the 5th WSEAS international conference on Signal processing of 2006*.
- [12] Shichao Zhang, Xuelong Li, Ming Zong, "Efficient kNN Classification With Different Numbers of Nearest Neighbors", *IEEE Transactions on Neural Networks and Learning Systems*. vol.99, no. 2, pp. 1-12, 2017.