The analysis and detection of hypernasality based on a formant extraction algorithm

Jiahui Qian¹, Fanglin Fu¹, Xinyi Liu¹, Ling He¹, Heng Yin² and Han Zhang¹*

¹School of Electrical and Information Sichuan University, Chengdu, Sichuan, 610065, China.
²West China Hospital of Stomatology, Sichuan University, Chengdu, Sichuan, 610041, China. Email: ling.he@scu.edu.cn

Abstract. In the clinical practice, the effective assessment of cleft palate speech disorders is important. For hypernasal speech, the resonance between nasal cavity and oral cavity causes an additional nasal formant. Thus, the formant frequency is a crucial cue for the judgment of hypernasality in cleft palate speech. Due to the existence of nasal formant, the peak merger occurs to the spectrum of nasal speech more often. However, the peak merger could not be solved by classical linear prediction coefficient root extraction method. In this paper, a method is proposed to detect the additional nasal formant in low-frequency region and obtain the formant frequency. The experiment results show that the proposed method could locate the nasal formant preferably. Moreover, the formants are regarded as the extraction features to proceed the detection of hypernasality. 436 phonemes, which are collected from Hospital of Stomatology, are used to carry out the experiment. The detection accuracy of hypernasality in cleft palate speech is 95.2%.

1. Introduction

Cleft palate is a birth defect with a gap in the roof of the mouth, which leaves a hole between nasal cavity and oral cavity. Velopharyngeal inadequacy is that the soft palate cannot separate the oral cavity and nasal cavity sufficiently, which can cause abnormal resonances due to the additional airflow through the oral cavity to nasal cavity during speech production. Thus, velopharyngeal inadequacy can result in hypernasal speech.

The cleft palate can be corrected with a series of surgeries. For many cleft palate patients, their treatments might last several years. If the opening of velopharyngeal port is not fully repaired after one-stage operation, the cleft palate patient will still have the resonance speech disorder. A further repair surgery is needed under this situation. Hypernasality is a symptom of resonance speech disorder. In the clinical environment, the grades of nasal speech is proportional to the opening size of velopharyngeal.

The hypernasality assessment in cleft palate speech can be classified into two categories: 1) invasive and 2) non-invasive techniques. In the clinical environment, the invasive techniques use invasive instruments which detect the movement of velopharyngeal. Non-invasive techniques can further be classified into two categories, a) clinical assessment and b) digital signal processing based techniques. The clinical assessment widely uses the nasometer, which is an expensive device. Another clinical assessment is the subjective judgment by experienced and well-trained speech-language pathologists. However, the result of subjective judgment
can easily vary among the speech-language pathologists. Digital signal processing based techniques are fast and objective, thus making the detection process comfortable, simpler and more convenient.

Formant is an important cue to hypernasality detection. Three acoustic cues can be summarized from the literatures [1, 3]: 1) the existence of a new formant (around 250Hz for /a/, around 1000Hz for /u/ and /i/); 2) broadening of formant bandwidths; 3) presence of antiformants. Based on the above cues, in this paper, estimated formant is used to detect the hypernasality.

Formant extraction has been widely used for decades in the field of speech signal processing. The formant extraction consists of spectral peak picking and root extraction [4-7]. The two methods often suffer the problem of peak merger (two formants, which are closed to each other, are regarded as one formant). Dellar et.al and Kim et.al [8, 9] propose a method to resolve the peak merger using prediction-error polynomial obtained from LPC extraction. That method involves too much computation. Moreover, the Snell et.al put forward a method which uses the number of poles in a sector in z-plane to calculate the formant [4]. However, that method does not discuss the condition for peak merger. It may resolve the roots which peak merger does not occur to. Thus, the method in [4] may calculate the wrong formant from the LPC root. Peak merger is a serious problem of formant extraction. Thus most of current researches on the study of hypernasality use other acoustic parameters to detect hypernasality instead of formant. Dubey et.al [2] indicate that the changes in the vocal tract characteristics, and the extra nasal formant is a cue of hypernasality detection. They use the impulse-like ZTW (Zero Time Windowing) function, which is a high temporal resolution method, to resolve peak spectrum of hypernasal vowel sounds. Vjiayalakshmi et.al finds an appearance of a new resonant in the low frequency region. They use group delay function to resolve two closely spaced formants and define the Group Delay function-based Acoustic Measure (GDAM) to detect hypernasality [1]. Vjiayalakshmi et.al [3] put forward a selective pole-defocusing based technique and use maximum of cross-correlation between the input and the resynthesized speech to detect hypernasality.

In this work, a method is proposed to calculate the formant using the LPC root extraction and the root classification algorithm. In the root classification algorithm, the condition for peak merger is discussed and the method mentioned in article [4] is used to classify the roots into three cases. Then, the estimated formants are used as acoustic features to detect hypernasality in cleft palate speech.

2. The Proposed Formant Extraction Method
In this work, a method is proposed to estimate the first, second and third formants in cleft palate speech. The flow chart is illustrated in figure 1. The LPC roots are extracted firstly. Then these LPC roots are classified into three cases. The final step is to calculate the formants.
2.1. LPC Roots Extraction
The vocal tract system can be modelled as an all-pole system [4, 5]:

\[ H_v(z) = \frac{G_v}{\sum_{k=0}^{\ell} a_k z^{-k}} \]  

(1)

The method presented in article [5] is used to obtain linear prediction (LP) coefficients and the prediction-error filter \( A(z) \):

\[ A(z) = \sum_{k=0}^{N_p} a_k z^{-k} \]  

(2)
Where \( G_v \) is the gain factor, \( N_{LP} \) is the prediction order, and \( a_k \) is the coefficients of Linear Prediction (LP).

The roots of \( A(z) \) correspond to the poles of the vocal tract system and relate to formants and formant bandwidths. The formant frequency \( F \) and formant bandwidth \( B \) can be calculated from the roots of \( A(z) \) by equations (3) and (4).

\[
F = \frac{f_s}{2\pi} \phi_0
\]
\[
B = -\frac{f_s}{\pi} \ln(r_0)
\]

In equations (3) and (4), \( r_0 \) is the magnitude of the pole, \( \phi_0 \) is the phase of the pole. And \( f_s \) is the sampling frequency, \( F \) is the formant frequency, and \( B \) is the 3-dB bandwidth.

This method could not solve the peak merger. As mentioned before, an additional nasal formant is an important cue in hypernasality speech. Therefore, resolving the peak merger problem and obtaining the accurate formants is crucial to detect the hypernasality.

2.2 Three Cases of LPC Roots

In order to solve the peak merger problem and separate the nasal formant, roots obtained from linear prediction coefficients are classified into three cases judged by their phase and magnitude. The root is defined in equation (5).

\[
r_v = r_0 |e^{j\theta}|
\]

2.2.1. Case I. In case I, the roots do not form the formants. The range of formant bandwidths is given by Dunn [9].

Criterion i: \( B > 700 \text{Hz} \)

The root obey to criterion i is the root which do not form the formant. If a LPC root forms a formant, its formant bandwidth should smaller than 700Hz.

2.2.2. Case II. In case II, the roots form the formants, and the peak merger does not exist. The roots obey to criterion ii and do not obey to criterion I.

Criterion ii: \( |\theta_{\text{nearest}} - \theta_{r}| \geq 2\cos^{-1}(\frac{2r}{1+r^2}) \) \( \theta_{\text{nearest}} \) is the phase of the nearest root to \( r_v \). The \( r \) is decided by equation (4), when \( B \) is 700Hz.

To illustrate the condition for peak merger and the criterion ii, \( |H_v(e^{j\omega})| \) can be considered as a two pole system as illustrated in equation (6).

\[
|H_v(e^{j\omega})| \approx \frac{G_v}{|1-r_v e^{j\omega}|(1-r_v e^{-j\omega})}
\]

To obtain the largest phase difference, the two-pole system \( |H_v(e^{j\omega})| \) can share the same magnitude \( r \). Equation (6) can be substitute by equation (7).

\[
|H_v(e^{j\omega})| \approx \frac{G_v}{\sqrt{(1+r^2-2r\cos(w-\theta_j))(1+r^2-2r\cos(w-\theta_j))}}
\]

If the peak merge occurs, Equation (6) should have a single maximum value. The derivative of the squared value of equation (4) can be simplified into equation (8).
It is evident that only when the two poles satisfy equation (9), that the peak merger happens to the \([\theta_1, \theta_2]\)

\[
|\theta_1 - \theta_2| < 2 \cos^{-1}\left(\frac{2r}{1 + r^2}\right)
\]

Because the root do not obey to criterion I. The bandwidth B of the roots in case II is smaller than 700Hz. Using equation (4) and (9) can calculate the criterion II easily.

2.2.3. Case III. In case III, the roots form the formants, and the peak merge may occur, if the roots do not obey to criterion1. To solve the problem of peak merge, a method is proposed to calculate the exact number of formants from LPC analysis in article [4]. Moreover, that work presents an algorithm to calculate the location of all formants. The algorithm which can obtain the number of the roots in a sector region in z-plane is discussed here. Return to the roots of LPC root extraction in z-plane. The change of \(\theta\) is regarded as \(r\) moves along a curve \(\tau\). Moreover, the times of a closed curve circles the origin noted as \(n(\tau)\). As the article [10] states, the number of roots between the \(\theta_n\) and \(\theta_{\text{nearest}}\) is equal to \(n(\tau)\). \(\tau\) is a closed curve, which is consisted of \(\theta_n\), \(\theta_{\text{nearest}}\) and an arc as shown as figure.2. In the figure.2, there is a sector which contain two roots. The figure.3 shows \(A(\tau)\) in equation (2) where \(\tau\) is a sector bounded by the \(\theta_1\), \(\theta_2\) and the arc as shown as figure.2. From the figure.3 the \(A(\tau)\) circle the origin two times.

![Figure 2. The root in the sector in z-plane](image-url)
In order to avoid all polynomial evaluations on the arc, the R=1 is replaced by a large radius. As [4] notes, R=2 serve very well.

Separate the closed curve into two straight lines and an arc. Thus, n of the two straight lines can be calculated as follows:

(i) Substitute the two straight lines into a sequence \( r_n \) of 100 points with same distance. Then, take the sequence \( r_n \) into \( A(z) \) as \( z \).

(ii) Transform the discrete curve \( A(z) \) into the polar quadrant and calculate the position in polar quadrant;

(iii) Count the times of the discrete curve \( A(z) \) pass through the X negative axis \( N_- \) and positive axis \( N_+ \);

(iv) Calculate the number of formants:

\[
n(\theta) = N_+ + N_- + \left[ \frac{N_{\text{LP}} \theta}{2\pi} \right] \tag{10}
\]

where \( N_{\text{LP}} \) is the number of LPC analysis.

Criterion iii: \( n(P(\tau)) = 2 \)

If the roots in case III obey to the criterion iii, the peak merger actually occurs to the root. The algorithm which can obtain the number of a sector in z-plane is extended to determine the location of the two formants. The dichotomy in this sector is used here. Narrowing the area of the sector \([\theta_1, \theta_2]\) into two sub-sectors \([\theta_1^*, \theta_2^*]\) which are small enough and just obtain one root in one sub-sector. The way to divide the sub-sector is dichotomy. And the condition to stop the dichotomy is listed in equation (11).

\[
|\theta_1^* - \theta_2^*| < \varepsilon \tag{11}
\]

When the sub-sector \([\theta_1, \theta_2]\) meets the equation (11), the dichotomy is finished. Moreover, the \( r = |r| e^{i\theta} \) is the final root, where the \( \theta \) equal to \((\theta_1 + \theta_2) / 2 \). In this work, the \( \varepsilon \) equal to 0.001.

3. The Proposed Hypernasality Detection Method

The phonemes /a/, /i/ and /u/ are used to carry out hypernasal judgment in the papers [1] to detection the hypernasality. As shown in paper [1], the nasal formant of /a/ is around 250Hz, which is closed to F1. However, the nasal formant of the phonemes /i/ and /u/ is around 1000Hz. Merger peak may not occur to /i/ and /u/ as often as /a/ do. Thus, in this paper, the phonemes /a/ are used to carry out the experiment. In this paper, the 426 phonemes /a/ collected from Hospital of Stomatology are used to carry out the experiment. 216 speech utterances collected from normal speakers and 210 speech utterances collected from cleft palate patients are used in this work. Moreover, peak merger always occurs to the spectrum of
nasal speech in low-frequency region. The nasal formant is close to the first formant F1. Thus, the peak merger in the second formant F2 and the third formant F3 is not discussed here.

In this work, a k-nearest neighbour’s algorithm (k-NN) is used for classification. The classification of hypernasality applies the 10 times 10-folds cross validation method.

4. Experiments and Results
Table 1 shows the number of speech recordings with peak merger, for normal speech and cleft palate speech. It is observed that, 129 of 210 speech data from cleft palate speakers generate peak merger. For normal speakers, the 76 of 216 speech data form peak merger.

| Total number of speech recordings | 216 | 210 |
|----------------------------------|-----|-----|
| The number of speech recordings  | 216 | 210 |
| Percentage of the speech data   | 76  | 129 |
| with peak merger                | 35.1% | 61.4% |

Table 1. The number of speech data with peak merger

Figure 4 shows the frequencies of F1, F2 and F3. The frequencies of formants are calculated from the classical LPC root extraction method and the method presented in this paper. Take the ∆F=F1-F1’, where F1 is obtained by LPC root extraction and F1’ is obtained by our method to resolve the peak merger. The speech data of the cleft palate speakers have an additional formant which is around 250Hz. From figure 3, ∆F of the nasal speech data is obviously larger than the normal one. That is because that F1 which is substituted into formant in low-frequency region. It verifies the theory mentioned in the article [1] that there is a nasal formant in low-frequency region in the spectrum of nasal speech.

**Figure 4.** The comparison of the LPC root extraction method and our method

Formant frequency is regarded as an important feature to detect normal/hypernasal speech. The result of the detection of hypernasality is shown in Table 2. The accuracy is calculated by $\frac{n'}{n} \times 100\%$. The $n'$ is the number of the speech data which are classified into the right type and the $n$ is the total number of this type.
The accuracy of hypernasality detection using the formant calculated by LPC root method is 57.4%. The average detection accuracy of this classification using the formant with our method is 88.3%.

| Method                                | Normal speech | Hypernasal speech | Average accuracy |
|----------------------------------------|---------------|-------------------|-----------------|
| Method proposed in this paper          | 81.5%         | 95.2%             | 88.3%           |
| LPC root extraction method             | 53.2%         | 58.1%             | 57.4%           |

5. Discussions and Conclusions
This paper proposes a method to detect hypernasality in cleft palate. Formant is used as the acoustics feature. This method is based on the roots calculated from LPC root extraction method. These roots are classified into three cases, which are determined by the condition of peak merger, to test if peak merger occurs. In order to resolve the roots with peak merger, an algorithm which can determine the number of a sector in z-plane is proposed. Then, the formants can be calculated from the roots. This method can resolve the roots with peak merger to obtain an accurate formant frequency. As shown in Table 2, the detection accuracy of normal speech is slightly lower than that of hypernasal speech. As shown in table 1, peak merger occurs to 35.1% normal speech. If the peak merger happens, the formant for normal speech may be resolved into two formants, which are closed to the nasal formant. This normal speech is classified as nasal speech. Thus, the classification accuracy of normal speech is lower than that of hypernasal speech.

6. Conference
[1] P. Vijayalakshmi, M. Ramasubba Reddy and Douglas O'Shaughnessy, “Acoustic analysis and detection of hypernasality using a group delay function” IEEE Trans. Biomedical Engineering, Vol. 53, No. 4, April 2007.
[2] Akhilesh Humar Dubey, S.R. mahadeve Prasanna and S. Dandapat, “” Communication (NCC), 10.1109/NCC.2016.7561188.
[3] Caplar R and Kulisic P 1973 Proc. Int. Conf. on Nuclear Physics (Munich) vol 1 (Amsterdam: North-Holland/American Elsevier) p 517
[4] P. Vijayalakshmi, T. Nagarajan and Jayanthan Ra.V, “Select pole modification-based technique for the analysis and detection of hypernasality” TENCON 2009-2009 IEEE Region 10 Conference, pp. 1-5, 2009
[5] 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.
[6] 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.
[7] L. Welling and H. Ney, “Formant estimation for speech recognition,” IEEE Transactions on Speech and Audio Processing, vol. 6, no. 1, pp. 36–48, 1998.
[8] Chanwoo, “A robust formant extraction algorithm combing spectral peak picking and root polishing” EURASIP Journal on Applied Signal Processing, vol.2006, pp1-16, 2006
[9] J.R Dellar Jr., J.G Proakis, and J.H.L Hanson, “Discrete-Time processing of speech Signal,” Macmillan, New York, NY, USA, Article in a conference proceedings:
[10] H. K. Dunn, “Methods of measuring vowel formant bandwidths,” Journal of the Acoustical Society of America, vol. 33, no. 12, pp. 1737–1746, 1961.
[11] P. Henrici, “Applied and computational complex analysis”, Applied and ComputationalComplex Analysis. Volume 1. New York: Wiley, 1974