Automatic Detection Algorithm of Pharyngeal Fricatives in Cleft Palate Speech Based on LPIF and Feature Selection

Jia FU¹, Xian MO¹, Shi-rui HUANG¹, Yu-xuan MENG¹, Heng YIN² and Ling HE¹,*

¹College of Electrical Engineering and Information Technology, Sichuan University, Chengdu, Sichuan, China
²Hospital of Stomatology, Sichuan University, Chengdu, Sichuan, China
*Corresponding author

Keywords: Pharyngeal fricative, Linear prediction inverse filter, Glottal waveform, Vocal tract model, Cross-correlation, Feature selection.

Abstract. Cleft palate is one of common congenital malformations that has huge impacts on the physical and psychological health of patients. Pharyngeal fricative in cleft palate speech is a kind of common compensatory articulations, which is produced by retracting tongue position to posterior pharyngeal wall and narrowing velopharyngeal opening. In this paper, based on voice mechanism of pharyngeal fricatives in cleft palate speech, linear prediction inverse filter was used to extract glottal waveform and estimate vocal tract model coefficients. Then four features were extracted, including pitch period, glottal flow derivative waveform, vocal tract area and vocal tract gain. The cross-correlation function was used to calculate the correlation between features. Glottal flow derivative waveform was removed since it had strong correlation with the others. KNN classifier was applied to realize the automatic pharyngeal fricatives detection in cleft palate speech, which reaches the detection accuracy of 98.40%.

Introduction

Cleft Palate is one of common congenital malformations, which causes the connection between oral cavity and nasal cavity, velopharyngeal insufficiency, or anterior teeth loss. Pharyngeal fricative in cleft palate speech is a kind of common compensatory articulations, which is produced by insufficient pressure in oral cavity. Cleft palate patients pronounce pharyngeal fricatives by retracting tongue position to posterior pharyngeal wall and narrowing velopharyngeal opening. Compared with normal people, patients with pharyngeal fricatives have smaller pharyngeal orifices, lower energy and longer consonants in speech signals.

At present, there are fewer targeted researches on pharyngeal fricatives in cleft palate speech. Meloy [1] first categorizes pharyngeal fricatives as compensatory dysarthria. Klatt [2] and Strevens [3] study pharyngeal fricative /h/ in Abrabic. The conclusion shows that compared with normal fricatives, such as /s/ and /sh/, pharyngeal fricatives have two distinct peaks at low frequency, and the energy at low frequency is generally higher. Weinberg [4] indicates that pharyngeal fricative /s/ has two peaks at low frequency (600~800Hz and 1000~1200Hz). Xiao Yan [5] proposes that patients with pharyngeal fricatives prolong consonants when voicing /sha/. Xiao Yan also introduces one-third octave analysis to extract acoustic features of pharyngeal fricatives. The results show that the spectra of pharyngeal fricatives have higher energy values below 4kHz, while have lower values above 4kHz than normal fricatives.

In clinical, the diagnosis of pharyngeal fricatives in cleft palate speech is mainly divided into two parts. One is subjective assessment performed by speech language pathologists. They calculate the nasal score through listening to speech and observing spectrogram [6]. However, the subject assessment depends on the experience of speech language pathologists. Another method is to utilize medical equipment, such as invasive devices and radioactive equipment [7], which is expensive and uncomfortable. Therefore, an objective, painless and inexpensive diagnosis is urgently required. In this paper, Linear Predictive Inverse Filter (LPIF) is used to extract glottal waveform and estimate...
vocal tract model coefficients, based on the acoustic model of pharyngeal fricatives in cleft palate speech. Then Pitch Period (PP), Glottal Flow Derivation Waveform (GFDW), Vocal Tract Area (VTA) and Vocal Tract Gain (VTG) are extracted as features. The GFDW is removed since it is judged as a redundant feature by cross-correlation function. And the remaining features are put into KNN classifier to realize automatic pharyngeal fricatives detection in cleft palate speech.

Materials and Methods

Materials

The speech dataset used in this paper is collected by the department of Cleft Lip and Palate, Hospital of Stomatology, Sichuan University, which has the largest number of cleft palate patients in China. The speech dataset is recorded according to the measurement table that is fully considered the characteristics of cleft palate speech in Mandarin. In this work, consonants such as /s/, /sh/, /c/, /ch/, /x/, /q/ are selected, which can reflect the difference of normal fricatives and pharyngeal fricatives in cleft palate speech. A total of 14 subjects (ten with pharyngeal fricatives and four with normal fricatives) and 306 speech samples are included.

Methods

In digital model of speech signal, speech signal is viewed as the output of a linear system that contains excitation model, vocal tract model and lip radiation model [8]. In this study, excitation model and vocal tract model are applied to realize automatic pharyngeal fricatives detection in cleft palate speech. The flowchart of automatic pharyngeal fricatives detection in cleft palate speech is shown in Fig.1. It can be divided into four steps: (1) Preprocessing. This part contains pre-emphasis and framing processing. (2) Features Extraction. The extraction of glottal waveform and the estimation of vocal tract model coefficients are carried out by using Linear Prediction Inverse Filters (LPIF). Then four acoustic features: Pitch Period (PP), Glottal Flow Derivation Waveform (GFDW), Vocal Tract Area (VTA) and Vocal Tract Gain (VTG) are extracted. (3) Features Selection. In order to select effective features, cross-correlation function is utilized to calculate the correlation between features. (4) Classification. KNN classifier is used to achieve automatic pharyngeal fricatives detection in cleft palate speech.

Preprocessing. Pharyngeal fricatives in cleft palate speech occurs in consonant parts that is high frequency component in spectrum. In order to obtain more accurate vocal tract model coefficients, pre-emphasis is used to enhance consonants.

Pre-emphasis is achieved by high-pass filter, which is calculated as follows:

\[ F(z) = 1 - \vartheta z^{-1} \]  

Here, \( \vartheta \) is the pre-emphasis parameter, usually \( 0.9 < \vartheta < 1.0 \). We use \( \vartheta = 15 / 16 \) [9].

In this study, the sample rate of speech signal is selected as 44100Hz. Frame length is set as 25ms, and frame shift is 10ms.

Features Extraction. In this paper, glottal waveform can be extracted by using LPIF. Linear predictor is used in LPIF to obtain the linear prediction value of sample points. The vocal tract model coefficients are determined when the Linear Prediction Error (LPE) is minimum. The specific implementation process is as follows:

1. Assume that the speech signal is \( x_n, n = 1,2,\ldots,N \), \( N \) is the length of speech signal. P-order linear predictor is utilized to achieve the value prediction of sample points. The predicted
value of current sample \( \hat{x}_n \) is obtained by \( p \) past values, the formulation is shown as follows:

\[
\hat{x}_n = \sum_{i=1}^{p} a_i x_{n-i}, \quad i = 1, 2, \ldots, p
\]  

(2)

Here, \( a_i \) is the \( i \)th order linear prediction coefficient, \( p \) is the order of linear predictor. Then LPE \([10]\) \( \delta[n] \) is as follows:

\[
\delta[n] = |x_n - \hat{x}_n| = \left| x_n - \sum_{i=1}^{p} a_i x_{n-i} \right|
\]  

(3)

(2) In order to obtain the minimum value of \( \delta[n] \), the p-value is adjusted. When the \( \delta[n] \) is minimum, its corresponding p-value is the order of ALP. The response of ALP-based vocal tract filter is:

\[
H(z) = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}}
\]  

(4)

When the p-value is determined, the vocal tract model coefficients is \( A = [a_1, a_2, \ldots, a_p] \).

(3) Speech signal \( x[n] \) is inverse filtered using vocal tract filter \( H(z) \), and glottal waveform \( x_0[n] \) is obtained.

The period of the glottal waveform \( x_0[n] \) is the PP \([11]\), and the first order differential of glottal waveform is GFDW \([12]\).

The vocal tract model coefficients are used to calculate the value of VTA and VTG. The specific calculation is as follows:

(1) Calculate the reflection coefficient \( k_j \), which satisfies:

\[
a_j^i = a_j^{i-1} - k_i a_{i-j}^{i-1}, \quad j = 1, 2, \ldots, i - 1
\]  

(5)

\[
a_{i-j}^i = a_{i-j}^{i-1} - k_i a_i^{i-1}, \quad j = 1, 2, \ldots, i - 1
\]  

(6)

\[
k_i = a_i^i
\]  

(7)

(2) Calculate the parameter vocal tract area \( AO_i \) \([13]\) using reflection coefficient \( k_i \):

\[
AO_i / AO_{i-1} = (1 - k_i) / (1 + k_i)
\]  

(8)

Parameter GO can be obtained by following formulation \([14]\):

\[
GO = \sum_{n=0}^{N-1} (\delta[n])^2 = \sum_{n=0}^{N-1} (x_0[n])^2
\]  

(9)

**Feature Selection.** Redundant feature has strong correlation with the other. In order to reduce the influence of redundant features, the cross-correlation function is utilized. The cross-correlation function is used to calculate the correlation between features, one of the features would be removed if the correlation between the two features is too close. Feature selection has another benefit, which can reduce the dimension of feature vectors and shorten calculation time.

**Classification.** Due to k-Nearest Neighbor (KNN) Classifier has the advantages of computational complexity, it is applied to achieve the classification of normal fricatives and pharyngeal fricatives in cleft palate speech in this paper. The algorithm of KNN classifier can be divided into three steps. First, the Euclidean distance of each data is calculated. Second, the k training samples which are closest to the new data are found. Third, the category of the new data is given, which is decided by the most category of k training samples.
Experiment Results and Analysis

In this paper, the proposed algorithm is tested by 306 speech data provided by Hospital of Stomatology, Sichuan University. The speech data contains 297 pharyngeal fricatives in cleft palate and 99 normal fricatives. The results show that when the order p of LPIF is 17, the LPE is the smallest. Using the cross-correlation function, it is found that there are strong correlation between PP and GFDW. Therefore, this paper removes GFDW, and select the others, PP, VTA and VTG. Then the three features are fed into KNN classifier, and the accuracy of algorithm is assessed using 10-fold cross-validation. The results show that the accuracy of pharyngeal fricatives detection in cleft palate speech is 98.40%.

Comparison of Different Order p

In this paper, when the LPE is minimum, the corresponding order p of linear predictor is the order of LPIF. The LPIF with different p-value is used to filter the speech signal inversely, and glottal waveforms are extracted and vocal tract model coefficients are estimated. And p-value ranges from 12 to 19. The accuracy of automatic pharyngeal fricatives detection in cleft palate speech is shown in Table 1:

| p   | Accuracy  | p   | Accuracy  |
|-----|-----------|-----|-----------|
| 12  | 98.01%    | 16  | 98.33%    |
| 13  | 98.10%    | 17  | 98.40%    |
| 14  | 98.14%    | 18  | 98.27%    |
| 15  | 98.30%    | 19  | 98.14%    |

The results show that when the p-value is set as 17, the corresponding accuracy of automatic pharyngeal fricatives detection in cleft palate speech is the highest. When the higher order of LPIF, the more sample points are needed. If the frame length is too long, it would exceed the period of glottal closure. Then the extraction of glottal waveform would be inaccurate, and the accuracy of pharyngeal fricatives in cleft palate speech would be decreased. Therefore, the order of LPIF is set as 17 in this paper.

Comparison of Feature Selection

In this study, feature selection is used to reduce the influence of redundant features. In order to verify the importance of this step, a comparative experiment is performed. The result is shown in Table 2:

| Feature selection | Accuracy |
|-------------------|----------|
| With              | 98.14%   |
| Without           | 88.73%   |

As can be seen from table 2, feature selection can improve the accuracy of automatic pharyngeal fricatives detection in cleft palate speech significantly. When the redundant features exist, the biased data of redundant features would change the value of features, resulting in the misclassification. Then the accuracy of automatic pharyngeal fricatives detection in cleft palate speech would be decreased.

Summary

In this paper, we use LPIF to extract the glottal waveform and estimate vocal tract coefficients, based on voice mechanism of pharyngeal fricatives in cleft palate speech. This method improves the signal-to-noise ratio of signal and enhances the accuracy of glottal waveform extraction. Then four features are extracted, including PP, GWDW, VTA, VTG. The cross-correlation function is used to
calculate the correlation between features. The effective features are selected, and the redundant is removed. KNN classifier is applied to realize the classification of normal fricatives and pharyngeal fricatives in cleft palate speech. The result shows that the accuracy of automatic pharyngeal fricatives detection in cleft palate speech using the algorithm proposed in this paper can reach 98.40%, which can provide aided diagnosis for speech language pathologists and doctors.

Acknowledgement

This research was financially supported by the Natural Science Foundation of China (No. 61503264).

References

[1] Trost J E. Articulatory additions to the classical description of the speech of persons with cleft palate. J. Cleft Palate J, 18(1981) 193-203.
[2] Klatt D H and Stevens K N. Pharyngeal consonants. J. Quarterly progress report, Research Laboratory of Electronics, M.I.T., 93(1969) 207-216.
[3] Strevens P. Spectra of Fricative Noise in Human Speech. J. Language & Speech, 1960, 3(1)32-49.
[4] Weinberg B, Horii Y. Acoustic features of pharyngeal /s/ fricatives produced by speakers with cleft palate. J. Cleft Palate Journal, 12(1975) 12-6.
[5] Y. Xiao, Acoustic Analysis of Compensatory Articulation in Cleft Palate Speech, Beijing Jiaotong University, 2016.
[6] Thais Alves Guerra, Viviane Cristina de Castro Marino, Nasalance at presence and absence of pharyngeal fricative, J. REVISTA CEFAC, 18(2016) 449-458.
[7] Proctor M I, Shadle C H, Iskarous K. Pharyngeal articulation in the production of voiced and voiceless fricatives. J. Journal of the Acoustical Society of America, 127(2010) 1507-1518.
[8] Y.T. Zhang, L. He, Automatic detection algorithm of hypernasality in cleft palate speech based on acoustic model, J. Computer Engineering and Design, 36(2015) 1592-1597.
[9] J.N. Zhu, Y.X. Yang, Improved algorithm of feature recognition based on the gammatone filter, J. Technical Acoustics, 34(2015) 275-278.
[10] K.H. Zhang, Research and implementation on the methods of extracting glottal wave based on inverse filtering, Jinan University, 2015.
[11] W. Zhao, S. Zhang, X.K. Zhang. Improved algorithm for pitch period detection, J. Journal of data acquisition and processing, 29(2014) 304-308.
[12] J. Sun, Source-target speaker voice conversion based on the excitation source and prosody features, University of science and technology of China, 2006.
[13] D.F. Chen, Z.L. Research on vocal tract simulation for the auxiliary treatment of dysarthria, 32(2010), 146-149.
[14] C.H. Long, On vocal tract characteristic of Chinese whispered speech and its applications in perceptual study, Soochow University, 2008.