

Automatic Detection of Nasal Leak in Cleft Palate Speech Based on an Improved Group Delay Method

Xinyi Liu^{1, a}, Weiwei Tian^{1, b}, Heng Yin^{2, c} and Ling He^{1, d, *}

¹ College of Electrical Engineering and Information Technology, Sichuan University, Sichuan 610065, China;

² West China Hospital of Stomatology Sichuan University, Sichuan 610041, China.

^a895351929@qq.com, ^b2285675739@qq.com, ^cyinheng@scu.edu.cn, ^{d, *} ling.he@scu.edu.cn.

Abstract. Cleft palate is a kind of congenital defect, and nasal leak is one of common symptoms in cleft palate speech. The automatic detection of nasal leak in cleft palate speech can provide a certain auxiliary basis for clinical diagnosis. Aiming at the characteristics of nasal leak in cleft palate speech, an improved method based on group delay is proposed. Since group delay parameter can greatly characterize nonlinear signals. Gammatone filter combined with the characteristics of human auditory is used to demodulate speech signal, and group delay parameters are calculated for each frequency band. It is used as features to identify nasal leak in cleft palate speech. The proposed algorithm is tested on 351 speech samples provided by West China Hospital of Stomatology Sichuan University. Compared with the classical group delay method, the accuracy of this improved method is 91.74% for the detection of nasal leak in cleft palate speech.

Keywords: cleft palate speech, nasal leak, Gammatone filter, group delay.

1. Introduction

One of the most obvious dysfunctions of cleft palate is language impairment. The whimper of patients is directly connected through the nose and mouth. And it leads to the inability of patients to control the volume and direction of airflow during the pronunciation process. Due to the velopharyngeal insufficiency, the sacral flap cannot be completely closed during the pronunciation process. The airflow passes through nasal cavity and oral cavity at the same time, which produces abnormal symptoms such as abnormal resonance, nasal leak and hypernasality. Among them, nasal leak is a typical symptom of patients with velopharyngeal insufficiency. The airflow can sometimes be heard as it passes through the nasal cavity. Nasal leak mainly occurs when patients speak stop sound, match sound and rub sound [1].

At present, the detection methods of nasal leak are mainly subjective assessment and objective assessment. The subjective assessment method is mainly evaluated by speech language pathologist's observation of the actual choking situation and the judgment of patient's pronunciation [2]. The objective evaluation methods mainly rely on some experiments or instruments to check the whimper area of patients [3]. Signal processing technology is applied in various voice detection or diagnosis with its rapid development, and it also has objective and non-invasive characteristics. The application of speech signal processing technology to cleft palate speech research is almost for the detection of hypernasality [4]. The automatic detection for nasal leak in cleft palate speech is still blank in speech signal processing applications. In this paper, an improved group delay algorithm is proposed to realize the automatic detection of nasal leak in cleft palate speech.

In terms of auditory perception, Alsteris et al. [5] confirmed the important role of the short-term phase spectrum, at the same time group delay is the differentiation of the phase spectrum in frequency. Therefore, group delay can be used as a feature to represent the speech signal. At present, for the detection of voice activity, speech and non-speech segments are mostly judged by short-time energy, short-time zero-crossing rate, high-order statistical analysis, etc. [6]. However, for the detection of speech endpoints under high-noise conditions, the detection accuracy of the above features is not satisfying. Group delay has better noise robustness, therefore, the detection accuracy is high under a noisy environment. Due to the auditory characteristics of nasal leak in cleft palate speech, the patient's nasal airflow sound replaces normal vocalization. Therefore, group delay is used as a parameter to

distinguish the presence or absence of nasal leak speech. In order to obtain better results, nasal leak in cleft palate speech signal is preprocessed before the calculation of group delay parameter. Gammatone filter conforming to the human auditory model is used to demodulate the speech signal and realize the automatic detection of nasal leak speech.

2. Automatic Detection Algorithm of Nasal Leak in Cleft Palate Speech

In this paper, an improved group delay method is proposed to automatically detect nasal leak in cleft palate speech, and Gammatone filter is used to divide the speech signal into various frequency bands which are adapted to the auditory of human ear. Then group delay parameters are calculated in each frequency band. The parameters are put into the classifier, and finally the automatic detection of nasal leak in cleft palate speech is achieved. Figure 1 is a flow chart of the automatic detection algorithm for nasal leak in cleft palate speech.

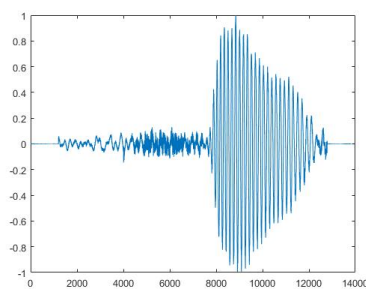


Fig. 1 Automatic detection algorithm flow for cleft palate speech nasal leak.

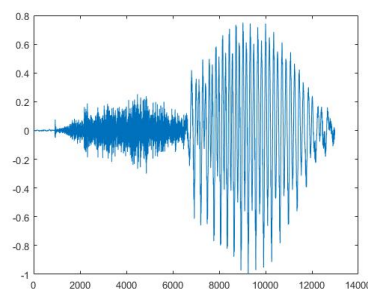
3. Acoustic Feature Extraction of Nasal Leak in Cleft Palate Speech

The extraction of acoustic parameter is an important step in the nasal leak detection algorithm in cleft palate speech. This paper extracts parameters which can distinguish normal speech and nasal leak speech, as the input parameters of classifier. KNN classifies the features by pattern recognition, then detection result is obtained.

Through the study of nasal leak in cleft palate speech, it is found that patients with cleft palate and normal people have different phonetic features when they speak consonants. Nasal leak is usually caused by velopharyngeal insufficiency, and nasal turbinate is considered to be a serious nasal leak [7]. As a result, velopharyngeal insufficiency and increased nasal resistance are the conditions for the generation of nasal sound. In the acoustic characteristic, patients with cleft palate may have a consonant abnormality during the pronunciation process. It sounds like the consonant falls off or weakens. Figure 2 is time-domain waveforms of nasal leak in cleft palate speech and control group speech. In Figure 2, an example of syllables /qiu/ is used to demonstrate the phenomenon of consonant falling off, and another example of syllables /qu/ is used to demonstrate the phenomenon of consonant weakening.



a. Consonant weakening in nasal leak speech of syllables /qu/



b. Control group speech of syllables /qu/



c. Consonant falling off in nasal leak speech of syllables /qiu/ d. Control group speech of syllables /qiu/

Fig. 2 Time-domain waveforms of nasal leak in cleft palate speech and control group speech

3.1 Gammatone Filter

The human auditory system consists of auditory organs and sensory cells in the cerebral cortex. The sound wave propagation process causes different degrees of basement membrane's vibrations. For different frequencies of sound waves, the sensitivity of basement membrane is different. If the frequency gets higher, the maximum vibration is closer to the bottom of basement membrane, and vice versa. This is the frequency selective property of basement membrane [8]. Since the amplitude of basement membrane's vibration affects the sensitivity of human ear to speech signal different bands, this paper divides speech signal into different frequency segments to distinguish the information contained in each frequency band of nasal leak speech and control group speech.

Gammatone filter (GTF) is a nonlinear filter based on the ERB scale that mimics human hearing. The impulse response is:

$$h(t) = At^{n-1} \cos(2\pi ft + \varphi) e^{-2\pi Bt} U(t) \quad (1)$$

Where n is the filter's order and B is the filter's decay speed. When B is a positive number, the larger B is, the faster the attenuation speed is, and the longer the impulse response length is. f is the center frequency. When $f=0$, GTF is called a Base Band Gammatone Filter. φ is the phase.

The impulse response of GTF is:

$$g(t) = e^{-\alpha^2 t^2} \cos(w_c t) w_c = 2\pi f_c \quad (2)$$

Where f_c is the center frequency of the filter and α determines the bandwidth of the filter.

At present, the nonlinear scale transformation based on human hearing is often used in speech technology research because it accords with the advantages of human auditory characteristics. This paper uses the frequency of ERB scale transformation modified by Moore and Glassberg [9]. The frequency approximation formula of the filter group of ERB frequency transform is as follows:

$$F_{ERB} = 21.4 \log_{10} \left(1 + \frac{4.37}{1000} \right) \quad (3)$$

The approximation formula for ERB frequency conversion bandwidth and center frequency:

$$F_{ERB} = 0.108f + 24.7 \quad (4)$$

3.2 Group Delay

Group delay is the frequency function of signal passing through the response system. It is defined as negative slope of response system's tangent on the phase-frequency characteristic diagram. The definition is as follows:

$$\tau(w) = -\frac{df(w)}{dw} \quad (5)$$

$\tau(w)$ is named group delay, as the delay time of signal passing through the response system. If the signal suddenly acts on the input of the system, its effect usually does not immediately respond at the output, but only after the delay time τ . In the continuous case, if $f(w) = -w \cdot \tau$, the delay of the response system is τ [10].

Since this paper applies group delay parameter in speech signals, the delay theory to the case of nonlinear phase is extended. It is supposed that this work evaluates the phase's effect of a time LTI systems on a narrowband input signal. The Fourier transform of the narrowband input $x(t)$ is zero or tends to be outside the small frequency range centered at $w = w_0$ infinitesimally small [11]. But the phase characteristics in the frequency band can be approximated to be linear, therefore, the phase characteristics of the system can be approximated by a linear relationship in this frequency band, that is $f(w) \approx -f - w \cdot \alpha$, from which:

$$Y(jw) = X(jw)|H(jw)|e^{-jf-jw\alpha} \quad (6)$$

Where $|H(jw)|$ is the amplitude portion, multiplied by the constant complex factor e^{-jf} and the phase shift term $e^{-jw\alpha}$ corresponding to the delay α seconds.

Thus, group delay can be directly used in discrete time systems, defined as:

$$\tau(n) = \frac{1}{2}(j(n+1) - j(n-1)) \quad (7)$$

When the signal is transmitted without distortion, τ is constant, thus the group delay is constant. Instead, it is a function of the frequency w .

4. Automatic Detection of Nasal Leak in Cleft Palate Speech

4.1 Cleft Palate Speech Data

In this paper, the data of nasal leak in cleft palate speech was collected by professional speech language pathologists of "Cleft Palate Speech Treatment Center" in West China Hospital of Stomatology, Sichuan University. The unified recording equipment was used and a database was established. The center is currently the largest cleft lip and palate treatment center in China. The professional speech language pathologists detailed label the patient data, which provides a "gold standard" for the detection of nasal leak in cleft palate. The experimental data in this paper is a total of 351 nasal leak speech samples, and 351 control group speech samples.

4.2 Automatic 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 nasal leak uses the 10 times 10-folds cross validation method in this work.

5. Experimental Results and Discussion

In this paper, the proposed improved group delay method is tested on the speech samples. In the improved group delay method, the speech signal is divided into 128 frequency bands by Gammatone filter, and group delay is calculated as a characteristic for each frequency band. As a comparison, the classical group delay method is used to extract features from speech signal without filtering. KNN classifier is used to identify nasal leak speeches, the results are listed in Table 1.

Table 1. The detection accuracy of nasal leak in cleft palate speech

The method	The accuracy of different type speeches and the average accuracy		
	The control group	Nasal leak speech	Average accuracy
Improved Group delay method	86.04%	97.44%	91.74%
Classical group delay method	83.22%	91.63%	87.43%

The automatic detection accuracy of nasal leak in cleft palate is shown in Table 1. The average detection accuracy using the classical group delay method is 87.43%. The average detection accuracy using the improved group delay method proposed in this paper is 91.74%. In particular, the detection accuracy of nasal leak is 97.44%. For the detection of nasal leak speech, the accuracy of the improved method is significantly higher than the classical method. It is proved that group delay parameter can better characterize nonlinear signals after combining the Gammatone filter, and realize the detection of nasal leak speech.

6. Summary

This paper proposes a method for nasal leak detection in cleft palate speech by calculating the group delay parameter in combination with Gammatone filter. For nasal leak speech, the sound of airflow in the nasal cavity replaces the normal vocalization of patient and group delay parameter can characterize different acoustic signals. In this paper, group delay can distinguish the characteristic between the sound of air and the sound of speaking. Gammatone filter conforms to the human auditory characteristics and can demodulate the speech signal to divide the corresponding frequency band. Calculating group delay parameter for each frequency band signal can better characterize the signal and distinguish the speech whether has nasal leak. And the average detection accuracy of the improved method is 91.74%.

Acknowledgements

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

References

- [1]. Lei Li. Cleft palate speech therapy [M]. Revision 1 edition. Wuhan: Hubei Science and Technology Press, 2004, p. 36-47.
- [2]. Chen Ren-ji. The status quo and thinking of Chinese cleft palate speech therapy [J]. Journal of International Stomatology, 2012, 39(01): 1-5+9.
- [3]. Wang Guomin. Diagnosis and treatment of pharyngeal closure dysfunction [J]. Journal of Oral and Maxillofacial Surgery, 2003 (04): p. 339-345.
- [4]. Liu Yin, Wang Xiyue, Hang Yipeng, et al. Hypemasality detection in cleft palate speech based on natural computation [J]. International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, October 19, 2016, p. 523-528.
- [5]. Alsteris, Leigh D., Paliwal, et al. Short-time phase spectrum in speech processing: A review and some experimental results [J]. Digital Signal Processing: A Review Journal, v 17, n 3, p 578-616.
- [6]. Alsteris, Leigh D., Paliwal, et al. Evaluation of the modified group delay feature for isolated word recognition [J]. Proceedings - 8th International Symposium on Signal Processing and its Applications, ISSPA 2005, v 2, p 715-718.

- [7]. Zhu Hongping, Sun Yonggang, Wang Guanghe, et al. Study on the acoustic characteristics of consonants in children with cleft palate[J]. *Journal of Modern Stomatology*,1998(03):24-26.
- [8]. Bovik A C, Maragos P, Quatieri T F. AM-FM energy detection and separation in noise using multiband energy operators[J]. *Signal Processing IEEE Transactions on*, 1993, 41(12):3245-3265.
- [9]. Kinoshita K, Delcroix M, Nakatani T, et al. Suppression of Late Reverberation Effect on Speech Signal Using Long-Term Multiple-step Linear Prediction[J]. *IEEE Transactions on Audio Speech & Language Processing*, 2009, 17(4):534-545.
- [10]. Shang Yongqiang, Yin Fuyuan, Jiang Shuangshuang, Wang Jinfang. Speech Activity Detection Based on Phase Modulation Features[J]. *Journal of Jilin University: Information Science Edition*, 2016,34(01):29-33.
- [11]. Shi Chunlei, Shu Jinlong, Zhu Zhenfu, Che Guofeng. Analysis of Delay Characteristics of Filter Groups and Its Application in Signal Processing[J]. *Systems Engineering and Electronics*, 2003(03): 260-262+266.
- [12]. Zhang Shichao, Li Xuelong, Zong Ming, “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.