http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 11 (1): 154-159, 2012
ISSN 1812-5638 / DOI: 10.3923/itj.2012.154.159

Speaker Recognition Based on Mathematical Morphology

Zhou Ping, Du Zhi-Ran and Wang Run-Duo
Guilin University of Electronic Technology, Guilin 541004, Guangxi, China

Abstract: In the field of speaker recognition, the noise of the speech interfered with the correct rate of speaker recognition largely. In order to filter out the noise in the noisy speech and further to improve the correct rate of speaker recognition, a weighted mathematical morphological filter was proposed in this study which can denoise the one-dimensional noisy speech signal. The morphological pre-filtering was conducted before training the speech to improve the matching abilities between the speech to be recognized and the training template. Experiment results showed that, compared with conventional spectral subtraction method, the method presented in this study suppressed the noise effectively and the signal to noise ratio of the output speech was significantly improved and the error rate of the speaker recognition system was further reduced.

Key words: Mathematical morphology, weighted morphology filter, speech feature parameters, recognition rate, denoise, signal to noise ratio

INTRODUCTION

Speaker recognition is to extract the personality features of people from part of the speech and to analyze and recognize these features for the purpose of recognizing and confirming the speakers (Wu, 2009). With the development of electronic computer technology, researches of the speaker recognition methods attack more and more attention.

In actual environment, the speech is always interfered with outside noise which is from the surroundings, the transmission medium or the electrical equipment and so on (Guo *et al*., 2010; Yang *et al*., 2011). In applications of speaker recognition, the background noise makes the voice quality decrease a lot, resulting in lower correct rate of speaker identification. Thus, eliminating noise to improve speech signal to noise ratio is an important issue in speech recognition research (Abdullah *et al*., 2008; Geravanchizadeh and Rezaii, 2009).

Mathematical morphological filter is based on strict mathematical theory and has been successfully applied to image analysis and processing (Dinesh, 2007; Nougrara *et al*., 2006), computer vision and other fields successfully (Strauss and Loquin, 2009; Dinesh, 2008). At present, mathematical morphology has been gradually applied to one-dimensional signal processing field. It is nonlinear transformation of signal processing, having the ability to simplify the signal and to eliminate the small component while keeping the basic shape and characteristics of the signal. This study introduced the mathematical morphology to the field of speaker identification. A new weighted morphological filter was designed to denoise the noisy speech and improve the correct rate of speaker recognition. Besides, a speech database was established to do experiments on the comparison between the traditional method and the morphological method.

MEL FREQUENCY CEPSTRUM COEFFICIENT SPEECH FEATURE PARAMETER

It is important to extract appropriate feature parameters in order to improve the recognition rate of the speaker recognition system (Frikha and Ben Hamida, 2007). LPCC (Linear Predictive Cepstral Coefficient) is based on the pronunciation characteristics of people without concerning the auditory characteristics of human ears. In fact, the human auditory system is a special nonlinear system and its sensitivity to signal response of different frequencies is different which is basically a logarithmic relationship. So, the MFCC (Mel Frequency Cepstrum Coefficient) is more in accord with the characteristics of human ears (Zhao, 2003; Kachouri *et al*., 2007).

The conversion relation of the Mel frequency and linear frequency (Aiming, 2006) is shown in the following Eq. 1:

$$f_{mel} = 2595 \log_{10} (1+f/700) \quad (1)$$

The calculation process of MFCC is shown in Fig. 1.

The special step of extracting MFCC is shown as follows:

Corresponding Author: Zhou Ping, Guilin University of Electronic Technology, Guilin 541004, Guangxi, China

Step 1: Pre-filter the original signal with mathematical morphology method

Step 2: Pre-emphasize, frame and window the signal to get the signal in time domain of each speech frame

Step 3: Transform the signal in time domain with DFT (Discrete Fourier Transformation)/FFT (Fast Fourier Transform) to get linear frequency spectrum

Step 4: Transform the linear frequency spectrum with Mel frequency filter bank to get Mel frequency spectrum, then conduct logarithmic energy operation to get logarithm frequency spectrum

Step 5: Transform logarithm frequency spectrum to cepstrum domain with DCT (Discrete Cosine Transformation), then MFCC based on morphological filter is obtained (Rodriguez *et al.*, 2005)

The filter bank in Fig. 1 is obtained by adding triangle band-pass filter to Mel coordinates in Mel frequency domain. It is shown in Fig. 2.

MFCC is a static feature parameter which just concerning the information in each frame but ignoring the dynamic information between different frames. In order to describe the dynamic features, this study combines MFCC, its first-order difference coefficients and second-order difference coefficients to be a vector and defines the vector as the feature parameter of a frame.

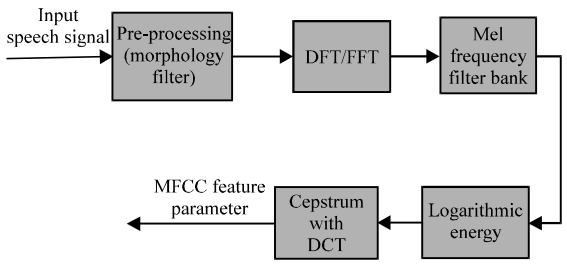

Fig. 1: The process of MFCC

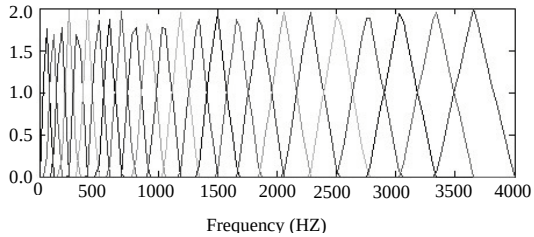

Fig. 2: Mel-scaled filter bank

THE APPLICATION OF MATHEMATICAL MORPHOLOGY IN SPEECH SIGNAL PROCESSING

The speech signal processing is a broad intercrossing subject which is one of the core technologies in the pattern recognition and the electronic computer. Eliminating the noise signal has been considered to be important in the speech signal processing (Marr, 2001; Durak and Arikan, 2003). Because the speech signal is complicated nonlinear and non-stationary, some traditional methods based on linear and stationary signal processing have a lot of limitations.

Morphology filter based on mathematical morphology is a new non-linear filter. This filter has a good performance of filtering. It can effectively suppress the noise in speech signal processing and maintain the characteristics of speech signal edge details. Therefore, this study introduces the non-linear signal processing method of mathematical morphology into the speech signal processing, compensating for the deficiency of the traditional linear methods.

As the digital speech signal is one-dimensional discrete and for the convenience of analyzing problems, the multi-valued morphological transformation which is in one-dimensional discrete condition of mathematical morphology is proposed. It also has the similar properties with the binary morphological transformation (Gong and Shi, 1997).

Let f (n) and g (n) be discrete function which are, respectively defined in F = {0, 1,... N-1} and G = {0, 1,... M-1}, N»M where, f (n) is input function and g (n) is structuring element. The erosion and dilation operations of f (n) on g (n) are defined as follows (Li and Chongzhao, 2008):

$$(f\Theta g)(n) = \min\{f(n+m)-g(m)\} \quad (m\in G) \tag{2}$$

$$(f\oplus g)(n) = \max\{f(n-m)+g(m)\} \quad (m\in G) \tag{3}$$

where, Θ and $\oplus$ are, defined as the morphological erosion and morphological dilation operations, respectively.

Morphological opening and closing operations are composed of the dilation operation cascading erosion operation in different order. They are compound extremum operations of the functions. From Eq. 2 and 3, the opening and closing operations of function f(x) on g(x) are defined as follows:

$$(f\circ g)(x) = [(f\Theta g')\oplus g](x) \tag{4}$$

$$(f\bullet g)(x) = [(f\oplus g')\Theta g](x) \tag{5}$$

in which ◦ and • separately stand for opening and closing operations.

Generally, the opening and closing operations can constitute various morphology filters. They are basic filters of themselves and can smooth the speech signal in different ways. That is the dilation operation increases the signal's valley value, extending the peak domain. While the erosion operation decreases the signal's peak value, extending the valley domain.

SPEECH SIGNAL PROCESSING BASED ON WEIGHTED MATHEMATICAL MORPHOLOGICAL FILTER

The basic principle of mathematical morphological filter is taking advantage of pre-selected structuring elements to match with the waited speech signal and then filtering the noise which is smaller than structuring elements through mathematical morphological transformation, finally achieving the purpose of extracting the signal characteristics and filtering the noise.

The erosion and dilation operations, morphological opening and closing operations of mathematical morphology compose basic morphological filters. In actual application, in order to get better morphological filters which can filter out both positive noise and negative noise in the speech signal, a variety of operations are combined to constitute integrated cascaded filters. The classical opening-closing filter and closing-opening filter (Guo *et al*., 2002; Chen and Li, 2005) are defined as follows:

$$F_{OC}(f(x)) = ((f\circ g)\bullet g)(x) \quad (6)$$

$$F_{CO}(f(x)) = ((f\bullet g)\circ g)(x) \quad (7)$$

The expansibility of opening operation and the anti-expansibility of closing operation make the statistical probability of the opening and the closing operations shift to some extend. So it's difficult to obtain the best filtering effect only using the filters separately. This study analyzed the result of the two filters synthetically and made optimization adjustment of the structuring elements in the filtering process, significantly improving the effect of the filtering.

In order to suppress the noise in the speech signal effectively, this article improves the traditional morphological filter and present a weighted morphological filter, in which w (i) and w (j) are defined as weighted coefficients. According to different situations in the actual speech signal, the F_{OC} (f(x)) and F_{CO} (f(x)) use different weighted coefficients, respectively which w (i) corresponding to F_{OC} (f (x)) and w (i) corresponding to F_{CO} (f (x)). The value range is $0<w(i)<1$, $0<w(j)<1$ and $w(i)+w(j) = 1$. The specific weighted morphological filter is defined as follows:

$$y(x) = w(i)\,F_{OC}(f(x)) + w(j)\,F_{CO}(f(x)) \quad (8)$$

The specific cascaded diagram of weighted opening-closing and closing-opening filters is shown in the Fig. 3.

Because weighted combinational operations which adopting opening-closing operation and closing-opening operation filters are more in line with the actual speech signal processing, especially in the speaker recognition system. We can choose different weighted coefficients of the morphological filter due to the actual speech. On the one hand, it can reduce the noise in the environment with the method of morphological filter effectively. On the other hand, it can also highlight the speech differences between different speakers, thus enhancing the ability to distinguish different speakers, further reducing the error rate of speaker identification.

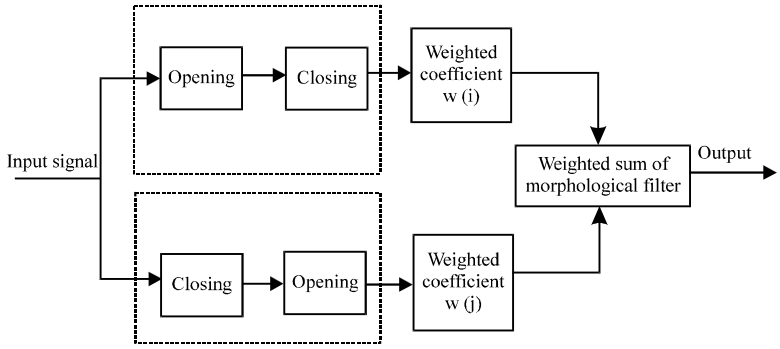

Fig. 3: Diagram of weighted morphological filter

EXPERIMENT ANALYSES

In the experiment, the first step was to reduce the noise and the second step was to extract the dynamic feature parameters from the customized speech. The content of the recording was number 1 and the length of time was 1 sec. Besides, the sampling frequency was 8 kHz and the sampling accuracy was 16 bit. The experiment preprocessed the speech signal with first-order filter $H(z) = 1-0.935z^{-1}$ and then calculated the 12-dimensional MFCC parameters and its first-order difference cepstrum coefficients and second-order difference cepstrum coefficients. Finally, compose them to be 36-dimensional dynamic speech feature parameters. This study adopted traditional spectral subtraction method (Ogata and Shimamura, 2001) and the morphological filter method to extract dynamic feature parameters (Xu and Sun, 2006) of the speech, respectively. The resulting dynamic speech feature parameters are shown in Fig. 4.

Figure 4 indicates that the surface of dynamic feature parameters which extracted with the method of morphological filter is smoother than that with the method of traditional spectral subtraction, especially has significant smoothing and denoising effects on the raised peaks. So, the morphological filter method has a better effect on eliminating the noise in the speech than the traditional method.

In addition, this study introduced the morphological filter method of the speech to the speaker identification rate experiments. There were 50 undergraduates in the experiment, all boys, marked with S1-S50. Each of them read 10 sentences in a magazine at random and the recording contents were different. The length of each sentence was 4 sec. Besides, the students were all from different cities in the province of Guangxi and they were all sophomores, aged from 19 to 22. In this way, it can improve the similarities of the speech between the imitators and the original speakers. This experiment adopted MFCC feature parameters with 12-order vector. The sampling frequency was 8 kHz and the sampling accuracy was 16 bit, besides the length of the frame was 32 msec and the frame shift was 16 msec. Each frame was handled with hamming windows.

Before the experiment, people marked with S1-S25 were defined as a set M which was a closed set. The flowing people marked with S26-S50 were defined as a set N which was an opening set. In the experiment, 10 people who would be imitated were marked with P1-P10, in order to be distinguished from other imitators. The 25 individuals in the open set N claimed themselves to be the people marked with p1-p10 to do the experiment, so did the 25 people in the closed set M.

Suppose P is one of the speakers we want to identify. We defined a formula X|Y, in which Y represents the original speaker and X represents the result of judgment. Y has two values that is s and n, in which s means the original speaker is P and n means the original speaker is not P. X also has two value that is s and n, in which s means the result of judgment is P and n means the result of judgment is not P. To sum up, there are 4 kinds of conditions that is s|s, s| n, n| s and n|n. The corresponding probability is, respectively recorded as P (s|s), P (s|n), P (n|s) and P (n|n), in which P (s|s) stands for true acceptation rate (TA) and P (s|n) stands for false acceptation rate (FA), as well as P (n|s) stands for false rejection rate (FR) and P (n|n) stands for true rejection rate (TR). Currently the most wildly used indicators of speaker recognition are FR and FA. The experiments applied FR and FA to judge the performance of speaker

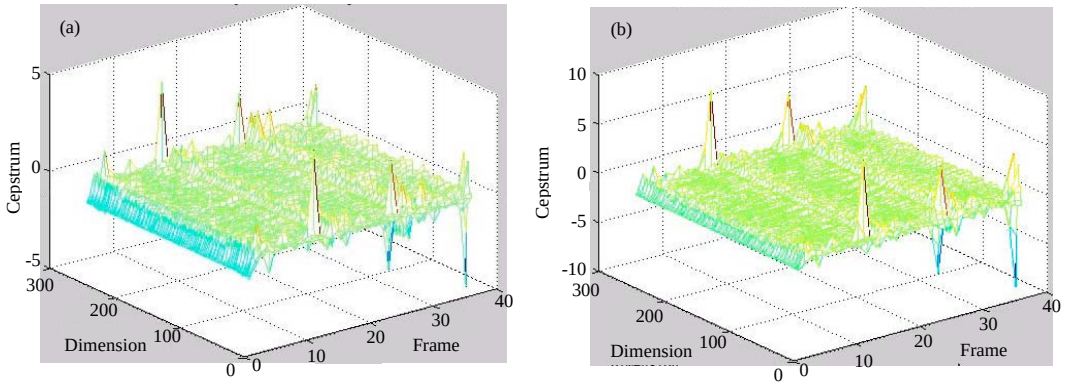

Fig. 4 (a-b): Dynamic feature parameters of the speech adopting the methods of traditional spectral subtraction and morphological filter; (a) Traditional dynamic feature parameter and (b) Morphological dynamic feature parameter

Table 1: Comparison of the experiment results (define P1-P10 as the people imitated)

Method	Traditional VQ method		Morphology method	
test index	FR (%)	FA (%)	FR (%)	FA (%)
P1				
Closed set	6.6	7.7	0.1	0.0
Open set	-	9.4	-	3.5
P2				
Closed set	7.9	11.7	4.5	5.5
Open set	-	11.3	-	5.6
P3				
Closed set	12.0	12.1	4.8	6.2
Open set	-	18.0	-	9.8
P4				
Closed set	14.2	15.4	8.3	6.4
Open set	-	18.7	-	10.0
P5				
Closed set	14.3	15.8	8.4	6.7
Open set	-	20.2	-	11.5
P6				
Closed set	15.6	16.2	9.5	7.1
Open set	-	23.0	-	12.3
P7				
Closed set	17.4	16.7	11.1	7.8
Open set	-	23.5	-	12.4
P8				
Closed set	18.1	19.4	11.5	10.0
Open set	-	24.3	-	14.0
P9				
Closed set	18.8	21.0	12.5	11.8
Open set	-	25.3	-	15.6
P10				
Closed set	25.0	29.0	15.4	15.7
Open set	-	31.8	-	20.9

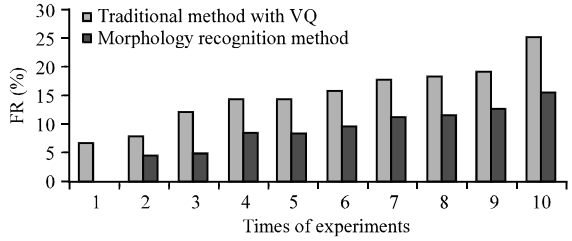

Fig. 5: Comparison figure of false rejection rate FR on the closed set

recognition. The lower the FR and FA are, the lower error rate and higher recognition rate of the system will be. There is no false rejection rate FR and no true acceptation rate TA in the case of the open set. So the false rejection rate and the true acceptation rate are only meaningful in the case of closed set. In the study, the symbol "-" means no such situation existing.

In order to observe the date curve clearly, the experimental date FR and FA are kept in the order of small to large. Table 1 compares the traditional speaker recognition method based on VQ (Vector Quantization) (Tan, 2010; Ilyas *et al.*, 2010) with that based on mathematical morphology.

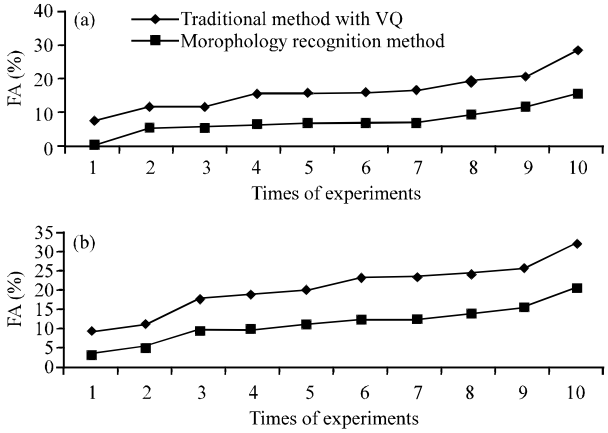

Fig. 6 (a-b): Comparison figures of false acceptation rate FA on the closed set and that on the open set; (a) False acceptation rate on the closed set and (b) False acceptation on the open set

Figure 5 shows that the FR of speaker recognition based on mathematical morphology is lower than that based on traditional methods. So, the speaker recognition based on mathematical morphology has a lower false rejection rate.

The corresponding rejection acceptation rate is shown in Fig. 6.

Figure 6 shows that the FA of speaker recognition based on mathematical morphology is lower than that based on traditional methods no matter on the closed set or on the open set. So, the speaker recognition based on mathematical morphology has a lower false acceptation rate.

From the Fig. 5 and 6, we can see that the speaker recognition based on mathematical morphology has a lower false acceptation rate and false rejection rate than that based on the traditional VQ. So, the speaker recognition based on mathematical morphology is an effective system, with lower error date and higher recognition rate.

CONCLUSION

This study focused on the mathematical morphology which was applied to one-dimensional speech signal processing and denoised the speech for the experiment of speaker recognition. According to the features of the speech signal, this study designed a morphological filtering algorithm for speech signal processing and made use of the weighted morphological filter to denoise the speech signal. The experiments showed that the weighted mathematical morphological filter can eliminate the noise

in noisy speech effectively and had a notable effect on the SNR improvement of the mixed noise compared with the traditional spectral subtraction. Because the mathematical morphological filter has better denoising characteristic, the article introduced it into speaker recognition system for the purpose of reducing the speaker recognition error rate and improving the popularity of the system.

REFERENCES

Abdullah, S., S.N. Sahadan, M.Z. Nuawi and Z.M. Nopiah, 2008. Fatigue road signal denoising process using the 4th order of daubechies wavelet transforms. J. Applied Sci., 8: 2496-2509.

Aiming, D., 2006. Process of extracting the MFCC in speaker recognition. Electron. Eng., 32: 51-53.

Chen, P. and Q.M. Li, 2005. Design and analysis of mathematical morphology-based digital filters. Proc. CSEE, 25: 60-65.

Dinesh, S., 2007. Fuzzy classification of simulated droughts and floods of water bodies. J. Applied Sci., 7: 2610-2616.

Dinesh, S., 2008. Computation of surface roughness of mountains extracted from digital elevation models. J. Applied Sci., 8: 262-270.

Durak, L. and O. Arikan, 2003. Short-time fourier transform: Two fundamental properties and an optimal implementation. IEEE Trans. Signal Process, 51: 1231-1242.

Frikha, M. and A. Ben Hamida, 2007. Noise robust isolated word recognition using speech feature enhancement techniques. J. Applied Sci., 7: 3935-3942.

Geravanchizadeh, M. and T.Y. Rezaii, 2009. Transform domain based multi-channel noise cancellation based on adaptive decorrelation and least mean mixed-norm algorithm. J. Applied Sci., 9: 651-661.

Gong, W. and Q.Y. Shi, 1997. Mathematical Morphology in Digital Space-Theory and Application. Science Publisher, Beijing, China.

Guo, J.F., G.X. Sheng and S.X. Zheng, 2002. Application of mathematical morphology in digital filter. Chin. J. Mech. Eng., 38: 144-147.

Guo, S.N., H.J. Cui and K. Tang, 2010. Speech enhancement based on short-time spectral amplitude estimates in low SNR. J. Tsinghua Univ. Sci. Technol., 50: 149-152.

Ilyas, M.Z., S.A. Samad, A. Hussain and K.A. Ishak, 2010. Improving speaker verification in noisy environments using adaptive filtering and hybrid classification technique. Inform. Technol. J., 9: 107-115.

Kachouri, A., T. Hdiji, Z. Sakka and M. Samet, 2007. Contribution to the vocal print recognition in Arabic language. J. Applied Sci., 7: 2560-2567.

Li, D. and H. Chongzhao, 2008. Research on edge detection of gray-scale image based on mathematical morphology algorithm and rough sets. Proceedings of the International Conference on Computer and Electrical Engineering, Dec. 20-22, Phuket, Thailand, pp: 852-855.

Marr, D., 2001. A computational investigation into the human representation and processing of visual information. Signal Process., 81: 2171-2255.

Nougrara, Z., A. Benyettou and N.I. Bachari, 2006. Methodology for road network extraction in satellite images. J. Applied Sci., 6: 2185-2192.

Ogata, S. and T. Shimamura, 2001. Reinforced spectral subtraction method to enhance speech signal. Proceedings of the 10th International Conference on Electrical and Electronic Technology, Aug. 19-22, IEEE, Singapore, pp: 242-245.

Rodriguez, R., T.E. Alarcon and O. Pacheco, 2005. A strategy for atherosclerosis image segmentation by using robust markers. J. Applied Sci., 5: 1316-1327.

Strauss, O. and K. Loquin, 2009. Linear filtering and mathematical morphology on an image: Abridge. Proceedings of the 16th IEEE International Conference on Image Processing, Nov. 7-10, Cairo, Egypt, pp: 3965-3968.

Tan, R.L., 2010. Speaker recognition based on VQ. Inform. Technol., 8: 103-104.

Wu, C.H., 2009. Speaker Recognition Models and Methods. Tsinghua University Publisher, Beijing, China.

Xu, Y. H. and J.N. Sun, 2006. Anti-noise research of the various feature coefficient on the speech recognition system. J. Jingling Inst. Technol., 22: 35-37.

Yang, G.Q., H.L. Xu and M.Z. Tang, 2011. Method for speech enhancement based on short-time fractional fourier transform. Measur. Control Technol., 30: 42-44.

Zhao, L., 2003. Speech Signal Processing. Mechanical Industry Publisher, Beijing, China.