ICSV14
Cairns • Australia
9-12 July, 2007

BINAURAL CEPSTRUM COEFFICIENT AND ITS APPLICATION TO GROUND TARGET RECOGNITION

Guan Luyang, Bao Ming, Li Xiaodong and Tian Jing

Institute of Acoustics, Chinese Academy of Sciences, Beijing 100080, China
guanluyang@mail.ioa.ac.cn

Abstract

Stereausis is a biologically motivated model proposed by Shamma which encodes both binaural and spectral information in a unified framework to simulate the processing of human binaural auditory system. In this paper, a new type of cepstrum coefficient is proposed based on this model. Two-channel acoustic signals are first processed by the stereausis binaural model to synthesize the spectral information and reduce the interference of noise signal. The binaural cepstrum coefficient is then extracted based on the diagonal vector of the stereausis model's output pattern, and is applied as feature to the multi-class acoustic target recognition. Learning Vector Quantization (LVQ) algorithm is implemented as the classifier and is tested by samples of vehicle acoustic signals. Experimental results show that binaural cepstrum coefficient improves both the performance and generalization of the classifier, especially at low SNR.

1. INTRODUCTION

Acoustic ground target recognition is of great interest to many applications due to the fact that the targets always have distinctive acoustic signatures. However, the targets' acoustic signals are often mixed with interferences, such as wind noise. Therefore, it is important for the recognition system to extract robust features

Various techniques have been applied to feature extraction. Sampan used the short time strength of the acoustic signal to classify vehicles into four classes [1]. However, it is hard for the time domain features to distinguish different vehicles with nearly the same size and engine power. For more precise classification, features extracted from the frequency domain must be considered. A 50-dimentional FFT based feature was used in [2], and the maximal percentage of correct classification is about 70%. Cheo *et al*. distinguished two types of vehicles and get a 98% correct classification by combining the short time Fourier transform (STFT) and wavelets as feature [3]. But for more types of vehicles, their method can not guarantee the same performance. The peripheral auditory model was used to extract features in [4], and the best performance is 92.61% without decision fusion. Data processed by these feature extraction algorithms mentioned above are all collected from mono-channel acoustic signal. Since the target information is limited and the interferences are difficult to be eliminated for mono-channel acoustic signal, it is hard to guarantee the robustness of the feature extracted from this signal.

Considering the advantages of array signal processing techniques and the limits on the costs and sizes of equipments used in practical scenarios, a two-microphone array with about 35 cm inter-microphone distance is designed to collect acoustic signals. For this small aperture microphone array, the binaural model is a good choice to measure the similarities and differences between the two input signals. Based on these, the feature extraction algorithm is designed to get robust feature to improve the performance of the recognition system [5].

2. STEREAUSIS BINAURAL MODEL

The stereausis model is a binaural hearing model proposed by Shamma [6]. As shown in Fig 1, the two input signals are first processed by the cochlea model, and then the outputs of the two cochlea models are cross-correlated inside the binaural model to obtain the output pattern *Y* of the stereausis model.

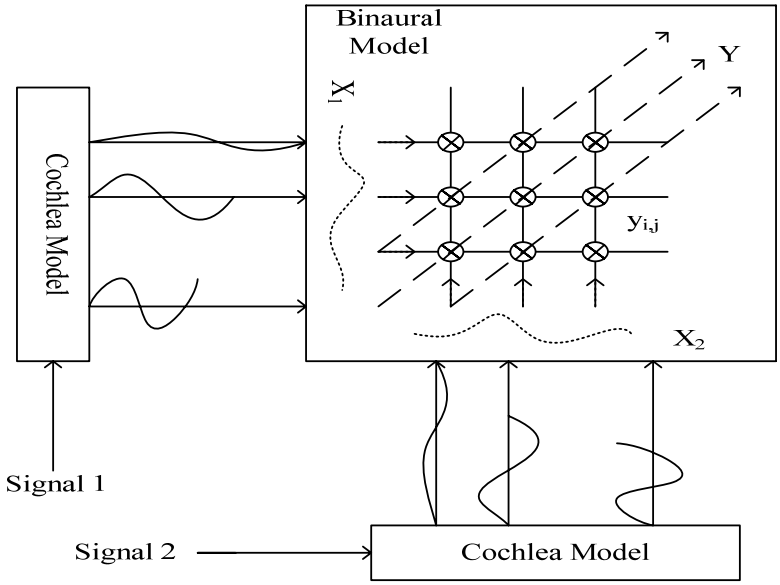

Figure 1. Structure of stereausis model

2.1 Cochlea Model

The core of the cochlea model is a constant Q filter bank served as the cochlear filters. The biological counterpart of this filter bank is the spatially distributed basilar membrane along the cochlea. The basilar membrane at different location of the cochlea appears to be a bandpass filter sensitive to particular frequency stimuli.

In the cochlea model, an acoustic signal *x* is filtered by the cochlear filters:

$$y_1(n,m) = x(n) \otimes h(n,m) \tag{1}$$

where *h(n,m)* is the impulse response of the *m*th cochlear filter and y_1 are the spatiotemporal patterns of basilar membrane vibrations. Then it is followed by a nonlinear transformation that simulates the transduction from basilar membrane vibration into intracellular hair cell potential [7]. This transformation is described as:

$$y_2(n,m) = g(y_1(n,m)) \tag{2}$$

where *g(x)* is a sigmoid function.

2.2 Binaural Model

As shown in Fig.1, the two output patterns of the cochlea model are fed in stereausis binaural network simultaneously and the output pattern *Y* of stereausis model is an M×M matrix, where M is the number of filters in the cochlear filter bank. Each (*i,j*)th node of the binaural network performs the following operation:

$$y_{i,j} = \frac{1}{N}\sum_{n=1}^{N}\left(X_{1,i}(n)\bullet X_{2,j}(n)\right)^2 \qquad (3)$$

where $X_i(n)$ and $X_j(n)$ are the ipsilateral and contralateral inputs at time *n*, *N* is the length of data frame.

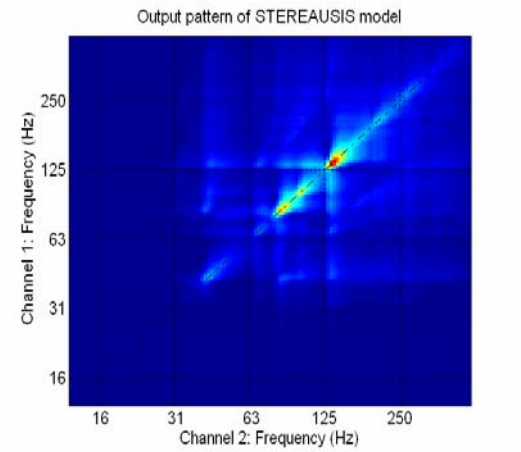

Figure 2. Output pattern of stereausis model

Fig 2 shows the output pattern of stereausis model obtained from 256 ms vehicle signal at 1 kHz sampling rate.

3. BINAURAL CEPSTRUM COEFFICIENT

The stereausis pattern (Fig 2) describes many characteristics of the binaural signals, such as spectrum, interaural level differences (ILD), and interaural time difference (ITD). A dominant peak appears along the main diagonal if the two inputs are identical to each other. it is equivalent to the audio spectrum described in [7]. Fig 3 shows that compared with FFT, the main diagonal reduces the interference of noise and maintains main spectral characteristics of the target by synthesizing the target information contained in the two inputs. As an output of the binaural model, the main diagonal mimics the result of spectral analysis of human hearing system. Thus it can be used as feature in acoustical target recognition.

Because of the ITD and ILD between two input signals, the dominant peak often deviates from the main diagonal. Then the modified main diagonal could be obtained by averaging the several diagonals besides the main one as shown in Fig 1. If the aperture of two sensors is small, the departure will be negligible.

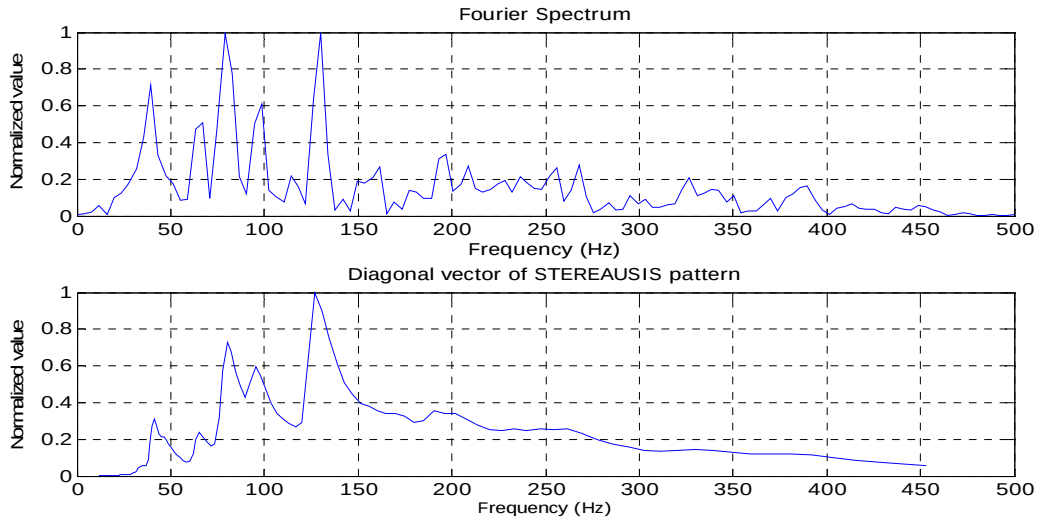

Figure 3. Contrast between Fourier spectrum and the main diagonal of stereausis pattern

MFCC (Mel Frequency Cepstrum Coefficient) is one of the most popular features used by researchers in the speech recognition field. It provides a great improvement over real cepstrum because it introduces the mel filter bank motivated by perceptual characteristics of human hearing. However, the stereausis binaural model simulates the structure of human audio system more precisely and it makes use of binaural signals to enhance the robustness of the output pattern, thus the mel filter bank could be replaced by the stereausis binaural model.

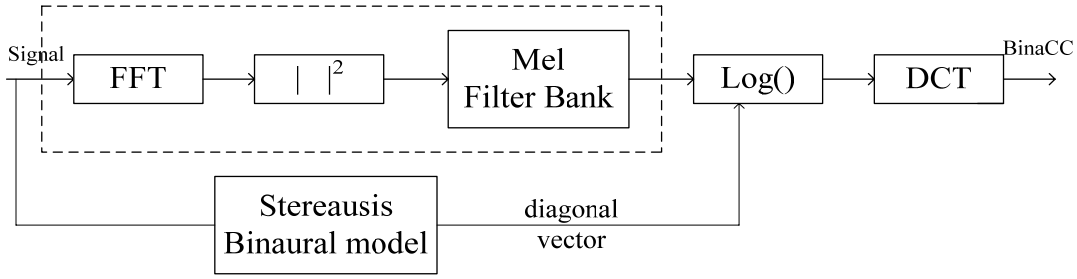

Figure 4. Computation of binaural cepstrum coefficient

As shown in Fig 4, acoustic signals collected by two microphones are fed into the stereausis binaural model, and then binaural cepstrum coefficient is computed from the log-magnitude diagonal of stereausis pattern using discrete cosine transform (DCT).

4. EXPERIMENTAL RESULTS

The main goal of the ground target recognition system in this paper is to classify four types of vehicles correctly based on the target acoustic signals.

The vehicle acoustic signal is approximately confined to the range of 20 to 400Hz, and the nonstationarity is typical for this signal. Approximately, vehicle signal can be recognized as stationary in 250ms or less time. Thus in our recognition system, acoustic signal is collected by two microphones at a sample rate of 1 kHz.

The signal is segmented into 256-point frames. A 32-dimensional binaural cepstrum coefficient is extracted from each frame of data with the method described above. In addition, a 24-dimensional MFCC is also obtained with the VoiceBox toolbox, including 12-dimentional MFCC and 12-dimensional △MFCC. And a 32-dimensional wavelet is calculated based on the db6 wavelet.

Learning VQ2.1 [8] is adopted as the classifier in this recognition system. Its initial codebook consists of sample selected randomly from the sample set [9], and the size of the codebook is defined by users.

4.1 Classification Performances of Different Features

In the experiment, 3-folder cross validation is applied to estimate the performances of the three types of features [10]. The mean value and standard deviation of correct rate listed in Table 1 are obtained from the results of 15 independent runs.

Due to the low dimensionality of MFCC feature, stable performance is obtained with a smaller codebook. Increasing in the size of the codebook improves the performance of binaural cepstrum coefficient and wavelet features due to the complexity of high dimensional feature space. However, a stable performance independent on the size of the codebook is obtained finally and it indicates the difference of intrinsic separability between the three types of features. It is obvious that binaural cepstrum coefficient achieves better performance.

Table 1. Performances of 3 types of features

Size of codebook	Binaural CC (%)	MFCC (%)	Wavelet (%)
100	92.01 (0.195)	92.68 (0.644)	86.24 (1.085)
300	94.75 (0.272)	92.66 (0.415)	89.57 (0.576)
400	95.27 (0.409)	92.51 (0.638)	90.22 (0.505)

4.2 Classification Performances at Different SNR

In this experiment, we add white noise to the original signal and extracted the three types of features mentioned above at different SNR. The size of the codebook is set to be 300, and the average performances of the three types of features are obtained through the same method described in section 4.1.

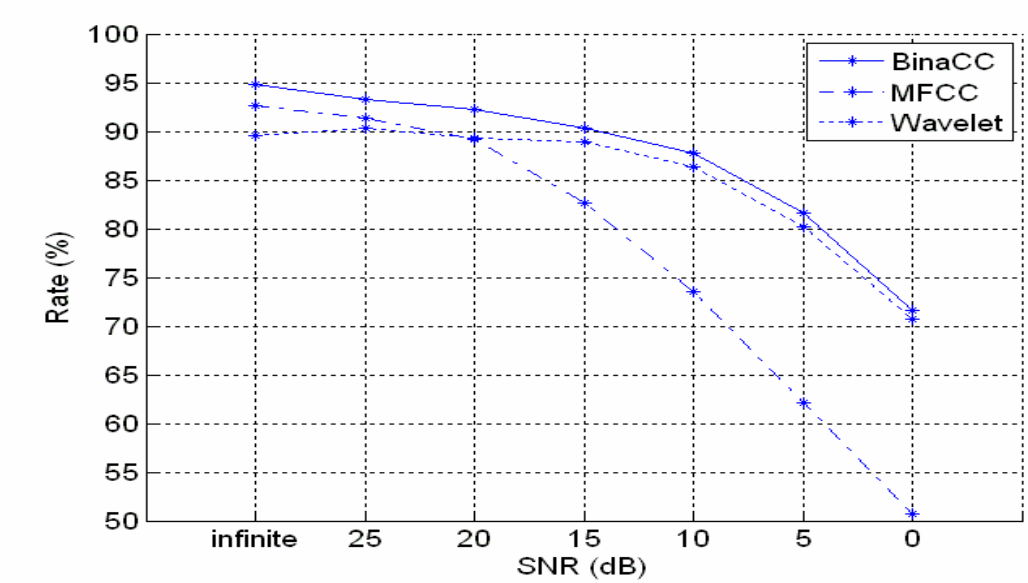

Figure 5. Performances of 3 types of features at different SNR

As shown in Fig 5, at low SNR, binaural cepstrum coefficient is more robust than MFCC and the performance of the recognition system descends more slowly. For SNR above 10 dB, it outperforms the wavelet feature obviously. Binaural cepstrum coefficient improves the

robustness and performance of the recognition system evidently.

5. CONCLUSIONS

Experimental results show that binaural cepstrum coefficient achieves a better balance between robustness and performance of the recognition system. At high SNR, it guarantees better performance than MFCC feature; and it is more robust than wavelet feature at low SNR.

Evidently, this feature extraction method based on binaural model is easier to reduce the interference of noise signal, and it guarantees the feature's robustness. We could try to fuse more sensors' data and design new method for more robust feature. On the other hand, binaural model also measures the difference between the two input signals, and the binaural pattern contains both ITD and ILD cues. So target recognition and orientation could be achieved synchronously by the system based on the binaural model.

REFERENCES

[1] S. Sampan *Neural fuzzy techniques in vehicle acoustic signal classification*. Blacksburg VA: Virginia Polytechnic Institute and State University; 1997

[2] M.F. Duarte, YHH, "Vehicle classification in distributed sensor network" *Journal of Parallel and Distributed Computing* 64, 826-838(2004)

[3] H. Choe, G. Gerhart and T. Meitzler, "Wavelet based ground vehicle recognition using acoustic signals" *Proceedings of SPIE 1996*, pp.434-445

[4] L. Li, "Ground vehicle acoustic signal processing based on biological hearing models", *Master paper*, University of Maryland, 1999

[5] M.P. DeSimino, "Phoneme recognition with a model of binaural hearing", *IEEE Transactions Speech and Audio Processing*, 157-166(1996)

[6] S.A. Shamma, "Stereausis: binaural processing without neural delays", *Journal of the Acoustical Society of America* **86**, 989-1006(1989)

[7] X. Yang, K. Wang, and S.A. Shamma, "Auditory representations of acoustic signals", *IEEE Transactions Information Theory*, **38**, 824-839(1992)

[8] T. Kohonen, "Improved versions of learning vector quantization", *IEEE Transactions Neural Network*, **1**, 545-550(1992)

[9] R.M. Gray, "Vector quantization", *IEEE ASSP Mag*, **1**, 4-29(1984)

[10] D. Li, K.D. Wong, Y.H. Hu, and A.M. Sayeed, "Detection, classification, and tracking of targets", *IEEE Signal Processing Mag*, **19**, 17-29(2002)