A novel double-talk detector based on pattern classification

CAO Jian-lin[1], FANG Jian[2], ZOU Cai-rong[2]

(1. Jiangsu College of Iinformation Technica, Wuxi 214061, China; 2. Engineering Research Center of Information Processing and Application, Southeast University, Nanjing 210096, China)

Abstract: This paper presents a novel double-talk detector (DTD) based on a nearest neighbor line (NNL) classifier. The underline idea is to use the feature information sufficiently and to design the DTD with pattern classification method. This paper analyzes 2 main kinds of conventional DTD, Geigel and DTD based on correlation, from the perspective of pattern classification, and then gives a new design method of DTD. A novel nonparametric classifier called NNL classifier is introduced to detect double-talk. NNL classifier has low computation cost and good performance. With NNL classifier, we fuse several conventional DTD and avoid the problem of making a fixed threshold, which exists in most of the conventional DTD. So the NNL-DTD is robust in adverse conditions. Experiments show that the proposed approach is more effective than conventional methods.

Key words: speech communication; double-talk detector; nearest neighbor line classifier

一种基于模式识别的新型双端检测器

曹建林[1], 方　健[2], 邹采荣[2]

(1.江苏信息职业技术学院，无锡 214061; 2 东南大学信息处理与应用工程研究中心，南京 210096)

摘要：本文提出了一种基于最近邻线分类器的新的双端检测器(DTD)。主要的思想是充分地利用特征信息以及用模式识别方法来设计 DTD。本文从模式分类的角度分析了二种主要的传统 DTD(Geigel 和相关 DTD)并给出了新的设计方法。一种称为 NNL 分类器的新的非参数分类器被用来检测双端通话。NNL 分类器具有低运算量和优良的性能。用 NNL 分类器，我们熔合了几种传统的 DTD 并且避免了存在于大多数传统 DTD 中的固定阈值带来的问题。因此 NNL-DTD 在各种条件下是鲁棒的。实验结果也显示出了这个方法比传统方法更有效。

关键词：语音通信；双端检测器；最近邻线分类器

中图分类号：TN911.7　　　文献标识码：A　　　文章编号：1000-3630(2006)-06-0644-03

1 INTRODUCTION

Echo canceling is an essential portion of speech communication systems. Adaptive echo cancellers (AEC) of finite impulse response filter type have been dominant among the echo cancellers. A double-talk detector (DTD) plays a very important role in a practical AEC. According to the decision

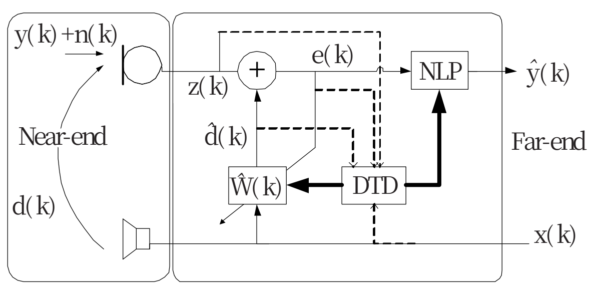

Fig.1 A AEC system with DTD

of the DTD, the adaptive filter updates its coefficients during single-talk periods and freezes adap-

Received Jun. 10, 2006; Revised Aug. 30, 2006

tation during double-talk periods to avoid unwanted divergence. Fig.1 shows an AEC system with double-talk detection.

The paper is organized as follows: In section 2, we review the Geigel algorithm and correlation algorithm from the perspective of pattern classification. The proposed algorithm (NNL-DTD) is presented in section 3. Section 4 is devoted to the experiments. Discussion and conclusion are given in section 5.

2 CONVENTIONAL DTD

In this section, we try to explain the design process of DTD from the perspective of pattern classification. The double-talk detector is viewed as a binary classifier. The design process of a classifier contains follow phases:

2.1 Geigel

(1) Collecting a group of training data in the form(ω, C):

$$\omega=(\boldsymbol{\omega}_1)=(r_{zd}(k)),\quad C=\{0,1\} \tag{1}$$

$$r_{zd}(k)=\frac{|z(k)|}{\max\{|x(k)|,|x(k-1)|,\cdots|x(k-L+1)|\}} \tag{2}$$

(2) Design a classifier $g_{geigei}(\omega)$ from the training data:

$$g_{geigei}(\omega)=H(\omega_1-1) \tag{3}$$

$$H(x)=\begin{cases}1 & (x\geq 0)\\ 0 & (x<0)\end{cases} \tag{4}$$

c is chosen according to the training data set $\{(\omega_i{}^j, C_i)\}_{i=0,1;\ j=1,\cdots,N_i}$.

2.2 Correlation DTD

Seon Joon park et al proposed a DTD based on two cross-correlations[2]:

(1) The cross-correlation coefficient between the microphone input and the estimated echo $\rho_{z\hat{d}}(k)$.

(2) The cross-correlation coefficient between the microphone input and the residual error of echo canceller $\rho_{ze}(k)$.

The cross-correlation coefficients $\rho_{z\hat{d}}(k)$ and $\rho_{ze}(k)$ are defined as follows:

$$\rho_{z\hat{d}}(k)=\frac{P_{z\hat{d}}(k)}{\sqrt{P_z(k)\cdot P_{\hat{d}}(k)}} \tag{5}$$

$$\rho_{ze}(k)=\frac{P_{ze}(k)}{\sqrt{P_z(k)\cdot P_e(k)}} \tag{6}$$

where $P_{\hat{d}}(k)$ is the power of the estimated acoustic echo signal, $P_z(k)$ is the power of the microphone input signal, $P_{z\hat{d}}(k)$ is the cross-power between the microphone input and the estimated acoustic echo signals, $P_e(k)$ is the power of the error signal, and $P_{ze}(k)$ is the cross-power between the microphone input and the error signals.

Fig.2 shows the decision flow of this DTD.

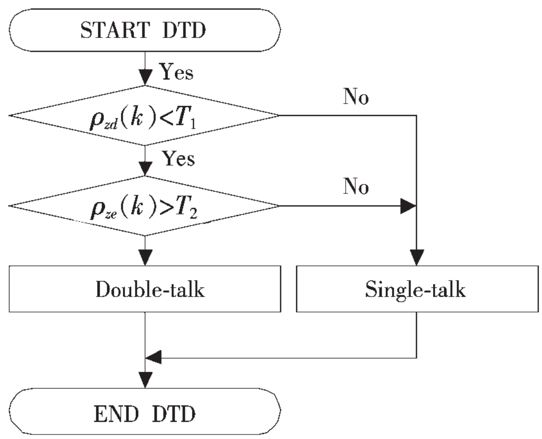

Fig.2 The design flow of Park's DTD

3 NNL-DTD

We use 4 features to training classifier. The 4 features are $r_{zd}(k)$, $\rho_{z\hat{d}}(k)$, $\rho_{ze}(k)$ and $\rho_{zx}(k)$.

$$\rho_{ze}(k)=\frac{P_{zx}(k)}{\sqrt{P_z(k)\cdot P_x(k)}} \tag{7}$$

$$\omega=(\omega_1,\omega_2,\omega_3,\omega_4)$$
$$=(r_{zd}(k),\rho_{z\hat{d}}(k),\rho_{ze}(k),\rho_{zx}(k)) \tag{8}$$

Given a sample point set $\{\omega_i{}^j\}_{i=0,1;\ j=1,\cdots,N_i}$ belonging to 2 classes, where $\omega_i{}^j$ represents the jth point of the ith class, and N_i is the number of the points of the ith class. Let ω be the query point. For the ith class. For any two sample points ω_i^m and ω_i^n, a feature line (FL) $\overline{\omega_i^m\omega_i^n}$ passing through them is generalized and the FL distance between ω and $\overline{\omega_i^m\omega_i^n}$ is given as:

$$d(\omega,\overline{\omega_i^m\omega_i^n})=\|x-p^j_{mn}\| \tag{9}$$

The nearest neighbor line (NNL) is the NL with the lowest neighbor line distance over all 2 classes. Suppose that the NNL is $\overline{\omega_{c^*}^{N(1)}\omega_{c^*}^{N(2)}}$, then we have $\overline{\omega_{c^*}^{N(1)}\omega_{c^*}^{N(2)}}=\arg\min_{i} d(x, \overline{\omega_i^{N(1)}\omega_i^{N(2)}})$. The index number c^* is used as the classification result of the unknown query point.

4 EXPERIMENTS

To test the performance of the proposed method, we extracted training data from 8 segments of speech in different environments. 4 points in double-talk period and 4 points in single-talk period were picked from each segment of speech. NLMS algorithm is used for adaptive filter. 10 other segments of speech were used for test. Fig.4 show one of the results.

The experiments are implemented with the MATLAB. The microphone input mixes local speech and the echo of far speech. The far speech exists from the 2-3.8 second, so the local speech indicates the double-talk period. The black bars in lower 3 graphs indicate the happening of the double-talk. The Geigel DTD is inclined to judge some double-talk periods (low voice level) as single-talk periods. The Correlation DTD is inclined to judge some single-talk periods as double-talk periods. Comparing the conventional DTD, The proposed algorithm obtained better performance.

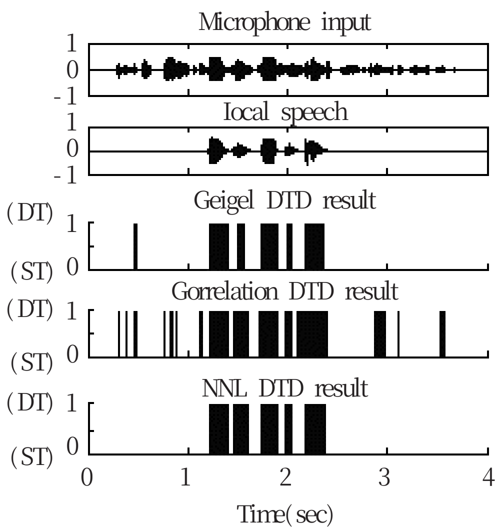

Fig.3 Comparison of 3 kinds of DTD

5 CONCLUSIONS

In this paper, an efficient DTD, called NNL-DTD, is presented for AEC. Different from conventional DTDs, it is designed by a new method which make it fuse more feature to help judge double-talk periods and avoid the problem of fixed threshold. The use of NNL classifier makes the DTD easy to implement and to have a low computational cost. In the future, if the new feature of double-talk is found, it can be easily add in the proposed DTD. The development of pattern classification also could offer a high efficient and low cost method to implement the classifiers. Not only the proposed DTD, but also the proposed design method, contributes to the design of DTD.

References

[1] Zheng W, Zhao L, Zou C. Face recognition using tow two novel nearest neighbor classifiers[J]. ICASSP 2004, 2004, 5: 725-728.

[2] Park S J, Cho C G, Lee C, Youn D H. Integrated echo and noise canceler for hands-free application[J]. IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II, 2004, 49(3): 188-195.

[3] Duda R O, Hart P E, Stork D G. Pattern Classification[M]. China Machine Press, 2004, 161.

[4] Duttweiler D L. A twelve-channel digital echo canceller [J]. IEEE Trans, Commun. 1978, 26(5): 647-653.

[5] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290: 2323-2326.

[6] Saul L K, Roweis S T. An Introduction to Locally Linear Embedding. http://www.cs.toronto.edu/lle/publications.html.

[7] Sondhi M M. An adaptive echo canceller[J]. Bell Syst. Tech. J, 1967, 46(5): 497-511.

[8] Butterweck H. Iterative analysis of the steady-state weigh fluctuations in LMS-type adaptive Filters[J]. IEEE Trans. on Signal Processing, 1999, 47(9): 2558-2561.

[9] Widrow B, Stearns S. Adaptive Signal Processing[M]. Prentice Hall, New Jersey, 1985.

[10] Haykin S. Adaptive Filter Theory, Fourth Edition[M]. 2002.

[11] Ho K C. A minimum misadjustment adaptive FIR filter[J]. IEEE Trans., on Signal Processing, 1996, 44 (32): 577-585.

[12] Saul B. Gelfand, Wei Yongbei. The stability of variable step-size LMS algorithms[J]. IEEE Trans. on Signal Processing, 1999, 47(12): 3277-3288.