文章编号:1001 - 893X(2012)08 - 1286 - 05

能量比检测与连续学习速率结合的改进双话处理算法*

黄 瑛,唐 昆,崔慧娟

(清华大学 电子工程系 清华信息科学与技术国家实验室,北京 100084)

摘 要:利用连续可变学习速率处理回声抵消双话情况时,随着近端语音能量的提高,学习速率的估计偏差增大,导致残留回声增加。提出了一种利用短时能量比显示检测与连续学习速率相结合的改进双话处理算法。该算法利用近端与远端语音的短时能量比,对学习速率估计中的泄露因子参数进行自适应修正,从而调整连续学习速率。实验证明,该算法使得回声抵消双话情况下,自适应滤波器发散程度下降,语音质量得到提升。在近远端能量比 - 6 ~ 6 dB 范围内,回声返回损失增加度(ERLE)提高 0 ~ 11 dB,平均意见得分(MOS)提高 0.02 ~ 0.45 分。

关键词:回声抵消;双话检测;泄露因子;能量比检测

中图分类号:TN912.3 **文献标志码**:A doi:10.3969/j.issn.1001 - 893x.2012.08.013

An Improved Method to Handle Double Talk Using Energy Ratio Detection Combined with Continuous Learning Rate

HUANG Ying, *TANG Kun*, *CUI Hui-juan*

(Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)

Abstract: When handling double talk in echo cancellation with continuous variable learning rate, estimation of learning rate is biased as energy of near end signal improves, thus causing more residual echo. This paper proposes an improved method to handle double talk using explicit detection of energy ratio combined with variable continuous learning rate. Short time energy ratio of near end to far end signal is used to adaptively modify leakage parameter in learning rate estimation, thus adjusting learning rate. Experimental results show that this algorithm can decrease divergence of adaptive filter and improve sound quality in double - talk situation of echo cancelation. With energy ratio of near to far end signal between - 6 dB and 6dB, ERLE(Echo Return Loss Enhancement) is improved by 0 ~ 11 dB and MOS(Mean Option Score) is improved by 0.02 ~ 0.45 point.

Key words: echo cancellation; double talk detection; leakage parameter; energy ratio detection

1 引 言

声回声抵消中一个重要的问题是回声路径较长,且易受近端语音的干扰。回声抵消多以自适应滤波器来模拟真实回声路径,通过残差信号来自适应调整滤波器的系数[1]。当近端存在语音时,将引起滤波器的发散。双话处理一直是回声抵消中的一个重要问题。Geigel 算法基于能量进行检测,该算法简单,但是随着路径的改变或者未知情况下,性能变差;H.Ye[2] 提出了正交方法,该方法利用自适应滤波器在收敛状况下残差信号与远端输入信号正交的特点来检测滤波器收敛状况而不是检测双话,此算法运算量大,且固定门限的判决容易造成双话与路径变化之间的误判;Jacob Benesty[3-4] 等利用参考信号与远端信号或残差信号的互相关系数。算法缺

* 收稿日期:2012 - 03 - 12;修回日期:2012 - 04 - 25

点是同样需要固定的门限,且需要进行矩阵运算,在回声路径较长的情况下更不利于实时实现。针对大多数回声抵消双话检测算法需要显式检测门限,误判或者漏判会导致收敛速度下降或者双话过程中滤波器发散的问题,本文基于多延时块频域自适应回声抵消器研究了一种可变学习速率算法,该算法不显示检测回声,而是采用一种可变的连续学习速率的调整方法。将速率建模成残留回声与输出信号能量的比重。由于输出信号残留回声能量的估计值比较困难,将残留回声建模成为一种难以估计的变化缓慢的泄露因子和一种容易估计快速的回声副本的能量。原有算法采用线性回归系数估计泄露因子,然而,随着近端语音能量增加,估计偏差增大,导致双话性能下降。本文引入了一种基于近端语音和远端语音能量比修正泄露因子的方法,即能量比显式检测与连续学习速率相结合。实验证明,本文算法改进了原算法的缺陷,双话跟踪效果更好,发散减小。

2　回声抵消的原理与多延时块频域(MDF)算法

回声主要分为两种:一是在电话网络中,由于用户端与交换局之间二、四线转换时阻抗不匹配产生的线路回声,也称电回声;二是伴随着免提电话,由于麦克风与扬声器之间的耦合产生的声回声。图 1 为声回声产生的原理框图。

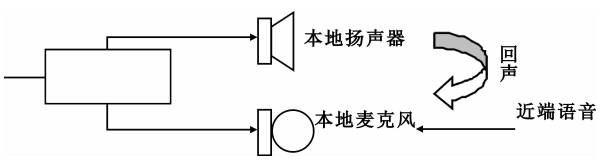

图 1　声回声产生原理图

Fig.1 Diagram of acoustic echo production

为了避免回声的存在引起语音质量的下降,如何处理回声十分重要。通常通过自适应滤波器来模拟实际的回声路径,然后从回声信号中减去模拟的回声从而达到消除回声的目的。自适应滤波算法分为时域算法和频域算法两种[1]。在处理延时较长的声回声信号时,对滤波器阶数要求较高,为节省运算保证实时实现,通常采用频域自适应滤波算法。Soo J. – S.[5]提出了一种多延时块的频域自适应滤波算法,通过将滤波器分成多个延时块,该算法比整块频域滤波的 FLMS 算法收敛速度更快,延时更小。计算线性卷积和相关通常采用重叠保留法和重叠相加法。以重叠保留 50% 为例,设 N 是权值总数,M 是延迟块数,每块大小 N/M,每次输入一个延时块即 N/M 个数据样点,并保留前一块的输入值,做 FFT 运算转换到频域,$N'=2N/M$ 是 FFT 算法的大小:

$$\boldsymbol{X}(M,j)=\mathrm{diag}\{\mathrm{FFT}[x_0(j-1),\cdots,x_{N'/2-1}(j-1),x_0(j),\cdots,x_{N'/2-1}(j)]^{\mathrm{T}}\} \tag{1}$$

式中,j 是块标号。每帧只需要进行一次 FFT 运算,通过块标号的移位保留前面 $M-1$ 个延时块的 FFT[5]:

$$\boldsymbol{X}(m,j)=\boldsymbol{X}(m+1,j-1),m=1,2,\cdots,M-1 \tag{2}$$

$\boldsymbol{W}(m,j)$为第 m 个延时块的权值向量,即用来模拟回声路径滤波器的系数向量,频域乘积进行滤波得到估计的回声副本[5]为

$$\hat{\boldsymbol{y}}(j)=\left\{\mathrm{FFT}^{-1}\left[\sum_{m=1}^{M}\boldsymbol{X}(m,j)\boldsymbol{W}(m,j)\right]\right\}\text{的后 } N'/2 \text{ 个点} \tag{3}$$

将真实语音中去除回声副本所得残留回声的误差向量变换到频域,用来更新频域滤波器的系数[5]:

$$\boldsymbol{E}(j)=\mathrm{FFT}\{\underbrace{0,0,\cdots 0}_{\frac{N'}{2}},\underbrace{[\boldsymbol{d}(j)-\hat{\boldsymbol{y}}(j)]^{\mathrm{T}}}_{\frac{N'}{2}}\}^{\mathrm{T}} \tag{4}$$

$\boldsymbol{d}(j)$是远端信号经过实际回声路径后产生的回声,即期望向量,频域滤波器系数更新是基于最小均方误差准则,使得 $E(|\boldsymbol{E}(j)|^2)$最小[5]:

$$\oslash(m,j)=\{\mathrm{FFT}^{-1}[\boldsymbol{X}^*(m,j)\boldsymbol{E}(j)]\text{前半部分}\} \tag{5}$$

$$\Phi(m,j)=\mathrm{FFT}\left[\oslash(m,j),\underbrace{0,0,\cdots 0,}_{\frac{N'}{2}}\right]^{\mathrm{T}} \tag{6}$$

$$W(m,j+1)=W(m,j)+\boldsymbol{\mu}(j)\Phi(m,j) \tag{7}$$

3　MDF 的最优可变学习速率的形式

为避免近端语音或者噪声引起的滤波器发散,Jean – Marc Valin[6]提出了一种动态调整学习速率的算法,通过最小化系数的误调程度,求解得到最优学习速率表示为残留回声能量占输出信号能量的比重:

$$\mu_{\mathrm{opt}}\approx\frac{\sigma_r^2}{\sigma_e^2} \tag{8}$$

将该速率用于 MDF 算法,得到第 l 延时块的频点 k 处学习速率[6]为

$$\mu_{\mathrm{opt}}(k,l)\approx\frac{\sigma_r^2(k,l)}{\sigma_e^2(k,l)} \tag{9}$$

残留回声的能量估计表示成了一个缓慢变化但是难以估计的量 $\hat{\eta}(l)$和一个快速变化但是容易估计的 $\hat{\sigma}_y^2(k,l)$:

$$\hat{\sigma}_r^2(k,l)=\hat{\eta}(l)\hat{\sigma}_y^2(k,l) \tag{10}$$

其中,$\hat{\eta}(l)$是泄露因子,近似回声返回增强损失 ELRE 的倒数。为了使得学习速率对于双话情形能

够给出快速的响应,通常采用瞬时估计

$$\hat{\sigma}_y^2(k,l)=|\hat{Y}(k,l)|^2,\hat{\sigma}_e^2(k,l)=|\mathrm{E}(k,l)|^2$$

由于残留回声与回声副本相关度较高,而近端语音与回声副本不相关,因此以回声副本和输出信号的功率谱的线性回归来估计泄露因子[6]:

$$P_Y(k,l)=(1-\gamma)P_Y(k,l-1)+\gamma(|\hat{Y}(k,l)|^2-|\hat{Y}(k,l-1)|^2) \tag{11}$$

$$P_E(k,l)=(1-\gamma)P_E(k,l-1)+\gamma(|\mathrm{E}(k,l)|^2-|\mathrm{E}(k,l-1)|^2) \tag{12}$$

泄露因子可以表示成 $P_Y(k,l)$、$P_E(k,l)$的线性回归系数[6]:

$$\hat{\eta}(l)=\frac{\sum_k R_{EY}(k,l)}{\sum_k R_{YY}(k,l)} \tag{13}$$

相关值可以通过下列递归运算得到[6]:

$$R_{EY}(k,l)=(1-\beta(l))R_{EY}(k,l)+\beta(l)P_Y(k)P_E \tag{14}$$

$$R_{YY}(k,l)=(1-\beta(l))R_{YY}(k,l)+\beta(l)(P_Y(k))^2 \tag{15}$$

其中:

$$\beta(l)=\beta_0\min\left\{\frac{\hat{\sigma}_y^2(l)}{\hat{\sigma}_e^2(l)},1\right\} \tag{16}$$

β_0 是泄露系数的学习速率,$\hat{\sigma}_y^2(l)$、$\hat{\sigma}_e^2(l)$分别是回声副本与输出信号的方差。$\beta(l)$可以防止估计值在没有回声的时候被调整。

4 改进的泄露因子估计算法

上文将残留回声建模成一个缓慢变化但是难以估计的量 $\hat{\eta}(l)$和一个快速变化但是容易估计 $\hat{\sigma}_y^2(k,l)$。假设的前提是近端语音与回声副本之间是独立的,而残留回声与回声副本能量之间却是高度相关的。因此在双话情况下,泄露因子 $\hat{\eta}(l)$较小,学习速率较低,从而有效避免双话下的滤波器发散。然而在实际语音测试中我们发现,随着近端语音的增加,上述方法并不能精准地估计泄露因子。本文将一段语音经过回声路径以后,在回声中1 900~2 900帧(160 样点/帧)的范围,加入近端语音。近端语音与远端能量比分别为6 dB、0 dB、-3 dB、-6 dB。从图 2 中可以看出,随着近端语音能量的增加,泄露因子越来越大,导致学习速率也随之提高,失调增大。如图 3 所示,在近端语音比较高的时候,意味着残留回声随之增大,回声抵消性能下降。

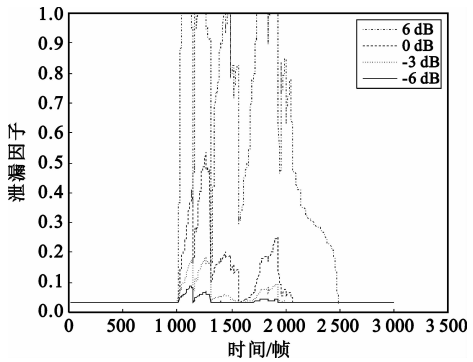

图 2 在不同近远端能量比下原算法的泄露因子

Fig.2 Leakage of original algorithm with different near to far end energy ratio

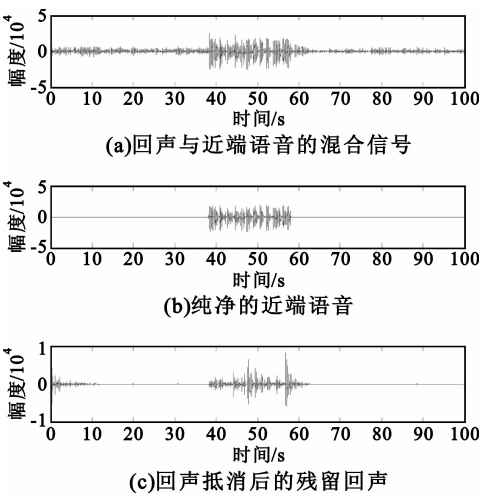

图 3 近远端能量比 6 dB 情况下,回声抵消后残留回声输出图

Fig.3 Residual echo with near to far end energy ratio at 6 dB

因此,在本文中,我们采用近端与远端信号短时能量比修正泄露因子,从而调整连续学习速率。首先估计近端语音和远端语音的本帧能量:

$$S_{xx}(n)=\sum_i x^2(n-i) \tag{17}$$

$$S_{dd}(n)=\sum_i d^2(n-i) \tag{18}$$

通过一阶平滑估计短时平均能量如下式,其中平滑因子 $\lambda\in(0,1)$:

$$\bar{S}_{xx}(n)=(1-\lambda)\bar{S}_{xx}(n-1)+\lambda S_{xx}(n) \tag{19}$$

$$\bar{S}_{dd}(n)=(1-\lambda)\bar{S}_{dd}(n-1)+\lambda S_{dd}(n) \tag{20}$$

利用两个短时平均能量修正泄露因子如下,当近端语音能量小于远端语音能量,采用原来的线性回归,当近端语音能量大于远端语音能量,利用两者的能量比以及修正因子 α 相结合,进行修正:

$$\hat{\eta}(l)=\begin{cases}\dfrac{\sum_k R_{E\hat{Y}}(k,l)}{\sum_k R_{\hat{Y}\hat{Y}}(k,l)}, & \bar{S}_{xx}(n)\geqslant\bar{S}_{dd}(n)\\ \alpha\dfrac{\bar{S}_{xx}(n)}{\bar{S}_{dd}(n)}\dfrac{\sum_k R_{E\hat{Y}}(k,l)}{\sum_k R_{\hat{Y}\hat{Y}}(k,l)}, & \bar{S}_{xx}(n)<\bar{S}_{dd}(n)\end{cases}\tag{21}$$

式中,α 为修正因子,$\alpha\in(0,1)$。该改进实际上是一种原可变学习速率与短时能量比显式检测的结合。当近端语音短时能量高于远端信号时,判决为近端语音存在,利用能量比与自适应因子衰减原学习速率,从而减小滤波器的发散。

5　测试结果

测试条件:本文基于 800 阶 160 延时块的 MDF 算法,采用本文修正的泄露因子估计法。回声路径采取 G.168 中的 mode1,冲激响应和频响特性如图 4 和图 5 所示。通过引入延时使得延时约为70 ms。采用8 kHz采样、16 bit量化的标准语音库语音材料。在回声中加入近端语音,近端语音/远端语音能量比分别为6 dB、0 dB、－3 dB、－6 dB。

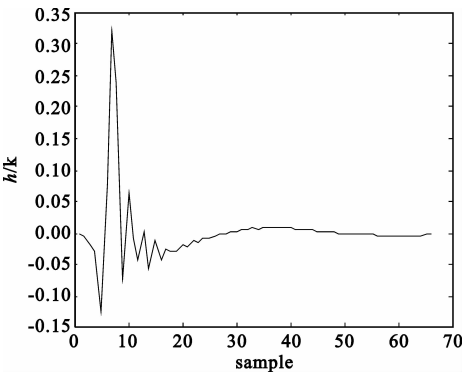

图 4　回声路径的冲激响应

Fig.4 Impulse response of echo path

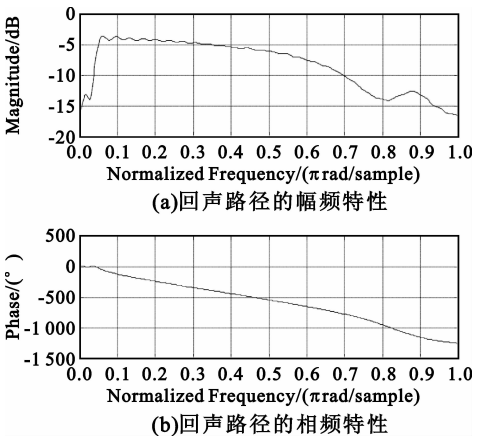

图 5　回声路径的频响特性

Fig.5 Frequency response of echo path

双话性能可以通过多种指标的测试结果来反映[7-8],本文主要从以下三方面对算法性能进行测试。

(1)泄露因子与残留回声波形

图 6 是在不同近远端能量比下改进算法的泄露因子,图 7 是 6 dB 近远端能量比下改进前后残留回声比较。

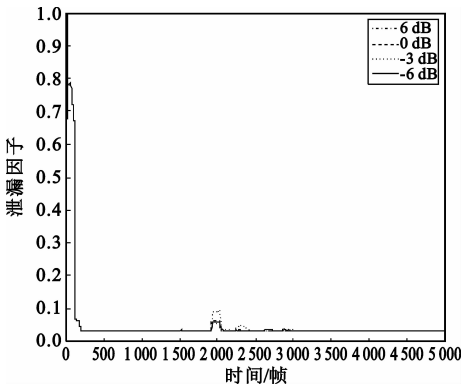

图 6　在不同近远端能量比下改进算法的泄露因子

Fig.6 Leakage of proposed algorithm with different near to far end energy ratio

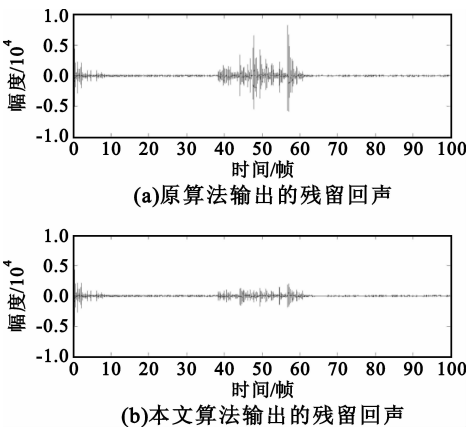

图 7　6 dB 近远端能量比下,改进前后残留回声比较

Fig.7 Comparison of residual echo with near to far end energy ratio at 6dB

由图 6 和图 7 可知,采用改进的算法,泄露因子并不随着近端语音能量的增加而增加,只有少许偏差。在近端语音能量较高(如近/远端能量比6 dB)的双话情况下,残留回声的幅度相比原算法大大衰减。

(2)回声返回损失增加度(ERLE)

ERLE 表征经过回声抵消处理以后残留回声相对于原来回声衰减的分贝值。该数值越大,表明回声衰减程度越大。该参数的表达式如下:

$$ERLE=10\lg\frac{\mathrm{E}[d(n)^2]}{\mathrm{E}[\mathrm{e}(n)^2]}$$

实际测试中,将上式中语音能量的期望值采用短时绝对能量代替,以计算每0.01 ms时间长度的语

音段的能量为例,对于8 kHz采样的语音来说,即连续 80 个样点的能量比,具体计算公式如下:

$$ERLE = 10\lg\frac{\sum_{i=0}^{79} y(n-i)^2}{\sum_{i=0}^{79} e(n-i)^2}$$

表 1 是不同近远端能量比下 ERLE 的比较。

表 1 不同近远端能量比下 ERLE 比较

Table 1 Comparison of ERLE with different near to far end energy ratio

近远端能量比/dB	*ERLE*/dB		
	原算法	本文算法	提高
6	16.223 0	27.249 1	11.026 1
0	28.343 7	34.210 9	5.867 2
-3	33.690 5	35.392 4	1.701 9
-6	36.750 4	37.123 2	0.372 8

从表 1 可以看出,采用本文算法残留回声返回损失增强度有较大提高,其中近端语音与远端语音能量比6 dB和3 dB情况下,提高11 dB和5.8 dB。随着近端语音能量越高,改进效果越明显。

(3)平均意见得分

对双话情况下近端语音的主观听觉质量进行了测试,用以区分不同双话检测算法下近端语音的失真度量。采用 ITU 标准 P.862 软件测试平均意见得分(Mean Opinion Score,MOS),该软件通常用于语音编解码系统或者降噪系统的语音质量性能评估,在本文的实验中能够反映残留回声的能量大小。

表 2 是不同近远端能量比下 MOS 分比较。

表 2 不同近远端能量比下 MOS 分比较

Table 2 Comparison of MOS with different near to far end energy ratio

近远端能量比/dB	MOS/dB		
	原算法	本文算法	提高
6	3.463	3.912	0.449
0	3.468	3.703	0.235
-3	3.481	3.550	0.069
-6	3.401	3.422	0.021

从表 2 可以看出,采用本文算法双话情况下近端语音 MOS 分有较大提高,其中近端语音与远端语音能量比6 dB和3 dB情况下,提高 0.449 和 0.235。随着近端语音能量越高,改进效果越明显。

6 结束语

本文基于多延时块频域自适应回声抵消算法,采用改进的短时能量比显示检测与连续可变学习速率结合的方法来处理双话,解决了原可变速率中由于残留回声估计的泄露因子随着近端语音能量的增加偏差增大,引起残留回声增加的问题。实验证明,该算法能较大程度上修正原算法的问题,提高双话情况下的 ERLE 和近端语音的 MOS 分。在近远端能量比 -6 ~ 6 dB范围内,两者分别提高 0 ~ 11 dB和 0.02 ~ 0.45 分。因此,在声回声抵消过程中,固定门限双话与连续可调学习速率两者结合起来,可以使得双话性能更好。

参考文献:

[1] Haykin S.自适应滤波器原理 [M]. 4 版.郑宝玉,译.北京:电子工业出版社, 2006.
Haykin S. Adaptive Filter Theory[M]. 4th ed. Translated by ZHENG Bao - yu. Beijing: Publishing House of Electronics Industry, 2006.(in Chinese)

[2] Ye H, Wu B X. A new double - talk detection algorithm based on the orthogonality theorem [J]. IEEE Transactions on Communications, 1991, 39(11):1542 - 1545.

[3] Benesty J, Morgan D R, Cho J H. A new class of doubletalk detectors based on cross - correlation[J]. IEEE Transactions on Speech and Audio Processing, 2000, 8(2):168 - 172.

[4] Iqbal M A, Stokes J W, Grant S L. Normalized Double - Talk Detection Based on Microphone and AEC Error Cross - Correlation[C]//Proceedings of 2007 IEEE International Conference on Multimedia and Expo. Beijing: IEEE, 2007: 360 - 363.

[5] Soo J S, Pang K. Multidelay block frequency domain adaptive filter [J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1990, 38(2):373 - 376.

[6] Valin Jean - Marc. On Adjusting the Learning Rate in Frequency Domain Echo Cancellation With Double - Talk [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(3):1030 - 1034.

[7] Cho J H, Morgan D R, Benesty J. An objective technique for evaluating doubletalk detectors in acoustic echo cancellers [J]. IEEE Transactions on Speech and Audio Processing, 1999, 7(2): 718 - 724.

[8] Ahgren P, Jakobsson A. A study of double - talk detection performance in the presence of acoustic echo path changes [C]//Proceedings of 2005 Internatinal Conference on Acoustics, Speech and Signal Processing. Vienna, Austria: IEEE, 2005:141 - 144.

作者简介:

黄 瑛(1986—),女,湖南沅江人,2009 年获学士学位,现为硕士研究生,主要研究方向为语音信号处理;

HUANG Ying was born in Yuanjiang, Hunan Province, in 1986. She received the B.S. degree in 2009. She is now a graduate student. Her research direction is speech signal processing.

Email: ying - huang09@mails.tsinghua.edu.cn

唐 昆(1945—),男,江苏宜兴人,教授,主要研究方向为数字通信、语音编码等领域;

TANG Kun was born in Yixing, Jiangsu Province, in 1945. He is now a professor. His research interests include communication and speech coding.

崔慧娟(1945—),女,辽宁沈阳人,教授,主要研究方向为信源编码和多媒体通信系统等。

CUI Hui-juan was born in Shenyang, Liaoning Province, in 1945. She is now a professor. Her research interests include signal source coding and multimedia communication system.