

# 电话语音回声消除的研究

阎兆立<sup>1,2</sup>, 杜利民<sup>1</sup>

(1. 中国科学院声学研究所, 北京 100080; 2. 中国科学院物理研究所, 北京 100080)

**摘 要:** 为了消除电话语音系统中的反射回声, 本文通过对电话信道的研究, 提出通过估算信道冲击响应, 得到提示语音经过信道卷积后的结果, 从而抑制回声的算法. 与 LMS 自适应算法相比, 它可获得很好的回声抑制和对原始信号的保真效果. 回声抵消后, 语音识别正确率和准确率分别提高约 29% 和 60%.

**关键词:** 电话语音; 回声消除; 卷积; 梳状滤波

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 0372-2112 (2002) 11-1726-03

## Research on Telephone Echo Cancellation

YAN Zhao-li<sup>1,2</sup>, DU Li-min<sup>1</sup>

(1. Center for Speech Interaction Technology Research, Institute of Acoustics, CAS, Beijing 100080, China;

2. Institute of Physics, CAS, Beijing 100080, China)

**Abstract:** In order to cancel the reflecting echo of telephone speech interactive system, the telephone channel and line echo were studied. The channel impulse response is firstly estimated, and the convolution of channel impulse response and suggestive speech is calculated, which will be used to counteract the echo. The time delay between echo and primary signal is estimated and canceled using de-convolution or Cross-power Spectrum Phase method before the channel impulse response is estimated. This algorithm can get better echo canceling and primal speech fidelity than LMS adaptive filter. After Echo Cancellation, the speech recognition correction rate and accuracy rate are improved by about 29% and 60% respectively.

**Key words:** telephone speech; echo cancellation; correlation; comb filter

### 1 引言

电话语音系统是远程人机交互的一种实现方式, 随着语音识别技术的发展, 它有很大的应用空间, 如股票信息查询、机票查询、天气预报等, 此外我们正研究电话语音交换机也是一个很好的实例. 但是在人机交互过程中, 机器播放的提示语音到达用户耳机后, 会掺杂到用户语音中一起返回计算机. 这将严重影响系统的语音识别率, 为了识别用户语音, 必须先把提示音消除.

在研究过程中, 使用电话语音卡作语音采集, 随卡有一个回声抑制程序, 实验表明其对内线回声抑制较好, 但外线效果比较差, 而且在消除回声的同时, 对有用信号影响也比较明显. 其结果与开始做的 LMS 自适应滤波<sup>[1,2]</sup>的效果相仿, 为了得到较好的回声消除效果, 通过对电话语音的分析提出如下的回声消除算法, 即先根据回声和提示语音算出电话信道冲击响应, 然后与提示语音做卷积估计出回声信号, 从而把混合信号中的回声减去, 得到用户语音的估计.

### 2 信道估计算法

#### 2.1 信号分析

在电话网二/四线转换处使用了混合变换器, 由于阻抗不匹配等原因, 甲端四线的收话音信号会泄露到该端四线发送路径中, 作为乙端的收话音返回到乙端, 就形成电回波, 另外在用户电话端, 也会产生声回波, 它们一起形成了电话语音回声. 如图 1,  $h_1$  是话机端的回波信道响应,  $h_2$  是交换机端回波信道响应, 为了简化模型, 总回波通道响应用  $h$  统一表示, 得到

$$y = x * h + m + e \quad (1)$$

其中  $y$  是电话卡端接收的信号,  $x$  是电话卡播放的提示语音,  $m$  是用户语音,  $e$  是加性信道噪声.

用户语音  $m$

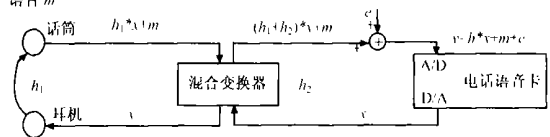


图 1 电话语音系统回声产生机理

收稿日期: 2002-02-28; 修回日期: 2002-07-10

基金项目: 国家 973 重点基础研究发展项目“图像、语音、自然语言理解和知识挖掘-汉语自然口语对话的理论和实验平台研究”资助 (No. G1998030505)

### 2.2 信道估计

在处理前,先做两个假设,即(1)通话信道是稳定或缓变的;(2)用户端在前 0.5 秒钟没有语音.事实上,不同的两次通话,可能占用的是不同的信道,其信道冲击响应是不同的,而对一次通话来讲,一旦建立联结,信道响应是基本稳定的.因此可在每次通话时先估计出冲击响应  $h$ ,并依此估计出混合语音  $y$  中提示音部分,将其去掉.同时第二个假设在实际中也是可以成立的.如果不考虑噪音,这时通过语音卡记录到的语音就是提示音经信道卷积后的结果:

$$y(n) = h(n) * x(n) = \sum_k x(k) h(n - k) \quad (2)$$

可以根据这一段录音估算出信道冲击响应.

对于卷积方程(2),其反卷积计算一般有两种最基本方法.

一是时域反卷积,也就是多项式除法.如果  $y(n)$  确是两个序列的卷积计算结果,时域反卷积可以准确的根据其中一个序列演算出另一个序列,但是在实际中,  $y(n)$  包含了各种噪音,这种情况下时域反卷积将引起严重的误差积累.在多项式除法中,某一点的误差不仅影响本次运算的系数,还将对后面的系数产生影响.实验证明,用时域反卷积求出的  $h(n)$ ,不能用来削减混合语音中的提示语音.

二是频域反卷积.在把  $x(n)$  和  $h(n)$  补足够长的零后,线卷积的计算问题  $y(n) = h(n) * x(n)$  就转换为圆卷积问题,即  $y(n) = h(n) \odot x(n)$ .设  $Y(k)$ 、 $X(k)$ 、 $H(k)$  分别是  $y(n)$ 、 $x(n)$ 、 $h(n)$  的离散傅立叶变换,则  $Y(k)$ 、 $X(k)$ 、 $H(k)$  满足

$$Y(k) = X(k) H(k) \quad (3)$$

这样频域的反卷积过程为:

$$H(k) = Y(k) / X(k) \quad (4)$$

$$h(n) = \text{IDFT}[X(k)] \quad (5)$$

但是如果  $X(k)$  中存在零点,频域的反卷积将出现  $0/0$  的奇异现象.有一些文章对这个问题进行过专门的研究<sup>[3-5]</sup>,不过在实际上由于随机噪声,频谱的各个部分都会有能量分布,大量实验数据也证明了这一点.因此作为一种合理的简化,文中采用更简单的处理方法:当  $X(k)$  中出现零点时,以一极小数值代替.

图(2)是根据录音数据计算的  $h(n)$  的幅值

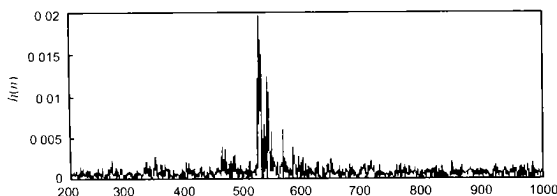


图 2 某一电话信道冲击响应的幅值

为了图示清楚,这里只是截取了其中一段,其它部分都是较低幅度的起伏,唯有  $n$  值在 500 多到 600 多的范围内有脉冲突起(这是回声的延迟引起的).

理论上分析,  $H(k)$  应该是一有限长的序列,在反卷积运算的结果中,其他部分应该是零序列.但是事实上,得到的解卷结果无论多长,不会出现零序列.这是因为做反卷积的混合

语音

$$y(n) = x(n) * h(n) + e(n) \quad (6)$$

(由于前 0.5 秒钟没有用户语音,所以和式(1)相比,没有  $m$  项),其傅立叶变换是

$$Y(k) = X(k) H(k) + E(k) \quad (7)$$

因此说信道响应计算结果只是一个估计,实际上得到的是

$$h'(n) = \text{IDFT}[E(k)/X(k)] + h(n) \quad (8)$$

很显然在  $h(n)$  为零的地方,上式中的第一项还在起作用.不过由于噪声部分所占比例不大,才出现图 2 所示的现象,即在  $h(n)$  不为零的时候,  $h'(n)$  的幅值出现一个凸起,而在  $h(n)$  为零时,  $h'(n)$  则表现为小幅值的波动.为了简化,文中仍用  $h(n)$  表示信道冲击响应的估计.

实验数据分析表明,虽然每次通话的信道冲击响应不同,但其序列长度却基本稳定,大致在 80~100 范围内.从图 2 看出,  $h(n)$  的真正有效部分从中间某处(设为  $n_0$ ,即波形中凸起的起始点)开始,这是因为交换机播出的提示语音有  $n_0$  的信道延迟.考虑到噪声对  $h(n)$  的影响,去掉延迟后可以得到更好的信道估计,因此下一步把混合语音做  $n_0$  点的移位,然后再做二次反卷积,完成信道估计.其中第一步的延迟估计也可以用互功率谱相位(Cross-power Spectrum Phase)模型<sup>[6]</sup>计算,它具有更好鲁棒性.

图 3 是移位后的解卷积结果的幅值.横轴采用对数坐标,纵轴采用线性坐标.从图中可看出,  $h(n)$  的非零部分已移至前面,这是  $h(n)$  的最终结果.

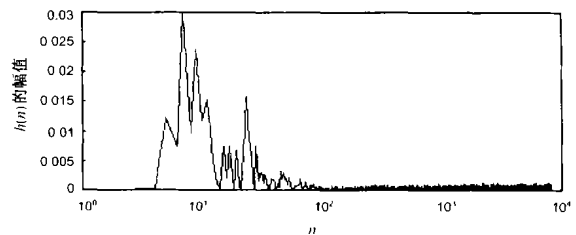


图 3 移位后的二次反卷积结果

### 2.3 回声抑制

根据估计的  $h(n)$  结果,与提示语音做卷积,得到对回声的估计,然后从混合语音中把回声减去即可.如公式(9)所示:

$$y'(n) = y(n) - h(n) * x(n) \quad (9)$$

$y'(n)$  是回声消除结果.

### 3 实验结果

为了验证算法性能,进行了如下实验:用户端设为静音(即  $m=0$ ),100 个句子作为提示语音  $x$ ,录得只有回声的 100 个文件数据  $y$ ,通过上述算法求得回声消除结果  $y'$ ,从而计算回声消除后相对于消除前的衰减量.设其为

$$i = 10 \log \left( \frac{\sum_n y'_i(n)^2}{\sum_n y_i(n)^2} \right) \quad (10)$$

图 4 是实验结果,衰减基本上在 16dB 左右,由于噪声的存在,不会衰减很大.

为了验证回声消除算法对语音识别的影响,对无回声、有回声和回声消除后的电话语音分别进行了识别实验.模型是

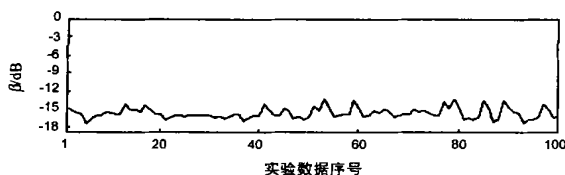


图4 电话回声抵消量

用声学所语音交互中心的电话语音数据库训练而成的,共用了200个人的语音,每人650句子;全网络识别;测试数据为集外二个人的语音(650句子/人)。表1是对比实验结果(2人结果的均值):

表1 三种情况语音识别结果对比

	无回声电话语音	有回声电话语音	回声消除电话语音
正确率	70.13 %	50.85 %	65.23 %
准确率	69.25 %	40.53 %	64.77 %

实验中,先在无提示语音情况下录得无回声电话语音,然后播放提示音录得有声电话语音,最后用本文提的算法消除回声。回声消除后,语音识别率大大提高,正确率和准确率分别提高29%和60%,仅比理想情况低4.5个百分点。

#### 4 噪声抑制

在有些情况下,电话语音可能会引入交流干扰。它是50Hz交流声及其谐波引起的,这种谐波一般是由于话筒输入插孔没很好接地导致的。这时仅仅去除50Hz交流声的干扰是不够的,这些谐波以窄脉冲的形式覆盖整个语音频谱,特别是3次至9次谐波。为此可以用陷波滤波器将谐波成分滤掉<sup>[7]</sup>。通过适当选择延迟线长度,可产生一梳状滤波器<sup>[8]</sup>,其凹陷的位置取决于交流声的谐波。传递函数为

$$A(z) = \frac{1 - z^{-T}}{1 - bz^{-T}} \quad (11)$$

电话语音中要滤的是50Hz的奇次谐波,即50Hz、150Hz、……,而公式(11)是从0Hz开始滤起,如果设梳齿间距为100Hz,陷波点在0Hz、100Hz、200Hz……等处,上式不能满足滤波要求。为此可对 $z^{-T}$ 做 的相移,即

$$A(z) = \frac{1 - z^{-T}e^{j\theta}}{1 - bz^{-T}e^{j\theta}} = \frac{1 + z^{-T}}{1 + bz^{-T}} \quad (12)$$

对于8k的采样率,令 $T=80$ 。由于传递函数中的 $b$ 接近1且小于1,所以在非陷波区域,其增益略大于1,而零点极点也几乎重合。

最后,因为电话信道的频率带宽是300Hz到3.4kHz,所以再对要处理的语音做一带通滤波。

#### 5 结论

从以上分析,这个算法能有效抑制反射回声,又能很好保留用户语音,为后续的语音识别打下较好基础。不过由于信道估计是在通话建立开始计算的,因此能够满足语音交换机或其他一般查询系统。而对于长时间对话系统,应该在用户语音间隙对信道重新计算,并和以前的信道估计进行加权,得到新

的信道估计。另外梳状滤波器虽能滤除市电干扰,却引入了一种回声,即物理空间的回声现象,因此要综合考虑。

致谢 中国科学院声学研究所语音交互信息技术研究中心的罗宇同学在本文的研究的方案实现和计算机实验方面提供了很多帮助,特此感谢。

#### 参考文献:

- [1] 吴兆熊,黄振兴,黄顺吉. 数字信号处理(下册)[M]. 北京:国防工业出版社,1985. 52 - 77.
- [2] BERNARD WIDROW, JOHN R. CLOVER, et al. Adaptive noise cancelling: Principles and Applications [C]. Proceedings of The IEEE, 1975, 63. 1692 - 1716.
- [3] 刘明亮,蔡涌泉,饶敏,等. 用卷积运算实现反卷积[J]. 电子学报, 2000, 28(5): 111 - 112.
- [4] 杨学志,严普强,晏磊. 最小平方卷积反演算法及其在桥梁动挠度测量中的应用[J]. 电子学报, 2000, 28(10): 121 - 123.
- [5] 罗平安,缪常. 反卷积定理-有零点的频谱的反卷积解法[J]. 核电子学与探测技术, 1999, 19(6): 454 - 459.
- [6] Maurizio Omologo. Acoustic event localization using a crosspower-spectrum phase based technique [A]. Proceedings of ICASSP 94 [C]. Adelaide: 1994, 2. 273 - 276.
- [7] 姚天任. 数字语音处理[M]. 武汉:武汉华中理工大学出版社, 1992. 394 - 396.
- [8] Sophocles J Orfanidis. Introduction To Signal Processing [M]. 北京:清华大学出版社, 1998. 355 - 415.

#### 作者简介:



阎兆立 男,1974年出生于河南,中科院物理所和声学所联合培养博士生,主要从事信号采集和语音增强算法方面的研究工作。



杜利民 男,1957生,四川人,博士,中国科学院声学研究所主任研究员;语言、语音和交互信息技术部主任;青海省人民政府科技顾问;中国移动互联网应用协议 QLV 专家组成员;中国电子学会理事;中国声学学会会员;中国人工智能学会自然语言理解专业委员会副主任;中国通信学会通信理论与信号处理专业委员会副主任;中国人工智能学会自然语言理解专业委员会委员;国际电子电气工程师协会(IEEE)高级会员;国际语音通信协会(ISCA)会员;《电子学报》编委。多年从事语音信号与信息处理技术的研究,在汉语话者无关的连续语音鲁棒识别和连续语音关键词检测、语音交互助理、自然口语语音对话、语音翻译、强噪声环境下语音增强和语音提取、低速率语音压缩等方面有深入研究。