

有混响条件下的室内语音定向技术的研究

胡跃辉, 周康源, 陈昕, 施俊, 周平, 黄以华

(中国科学技术大学电子工程与信息科学系 合肥 230027)

摘要: 文章讨论了一种新的基于回波分离算法的改进算法,并将其应用于室内有较强混响干扰下的语音定向。算法在有效语音片段检出时,不仅采用了回波估计,还同时考虑了语音特征,最后利用数字波束形成法来计算检出的有效语音片段的方位角,并且,回波估计是在语音短时能量分析的基础上进行的。试验表明,所获得的精度和人耳定向精度相接近。

关键词: 语音定向; 室内混响; 回波分离; 波束合成; 语音处理

中图分类号: TB559 文献标识码: A 文章编号: 1000-3630(2004)01-0043-06

Study on speech localization technique in a reverberation room

HU Yao-hui, ZHOU Kang-yuan, CHEN Xin, SHI Jun, ZHOU Ping, HUANG Yi-hua

(Department of Electronic Engineering and Information Science,
University of Science and Technology of China, Hefei 230027, China)

Abstract: In this paper an improved algorithm based on echo separation is discussed. This algorithm not only uses echo estimation method but also takes the speech characteristics into account when the effective speech segment is separated. A beam-forming method is used to calculate the azimuth of the effective speech segment, and echo estimation is based on a short time energy analysis. The results show that the localization accuracy obtained is very close to that of the human ear.

Key words: speech localization; reverberation; echo free; beam-forming; speech processing

1 声音定向的基本原理

对于一般的噪声背景干扰下的时延估计即定向的问题目前已有许多方法^[1]可以解决,但室内语音定向不仅要考虑一般背景噪声干扰的问题,更困难的是如何处理由于室内各个反射面反射和源信号瞬态叠加而引起的在同一频谱的高度相关的混响声干扰。

研究表明^[1-2],时差和强度差是人耳用来判断声源方向的基础;同时人耳对声源方向的判别依声源频率而不同,低频时主要依靠时差,高频时主要是由强度差来判定;低频时判别声源方位所费时间较高频时短;人耳对噪声源的定向,水平偏差为 3.1° ,垂直偏差为 4.7° ,综合偏差为 5.7° 。目前采用的声

音定向(又称声音追踪)技术都基于双耳定向这一原理,最简单的声音定向系统就是利用距离固定的两个声接收传感器同步接收声源发出的信号,对接收信号进行时差和强度差计算,并进而求出声源方位。

而室内声音定向需采用统计声学来进行处理。对室内声场(扩散声场)进行统计声学处理需遵循如下原理^[3]:

(1) 声以声线方式,声速 C_0 直线传播,声线所携带的声能向各方向传递的几率相同;

(2) 各声线是互不相干的,声线在叠加时,它们的位相变化是无规的;

(3) 室内平均声能密度处处相同。

这三条原理之间都有内在联系,是互相制约的,缺少其中任何一条,扩散声场的假定就会受到全面破坏。

通常,利用计算互相关的方法可以方便的求得两路信号的时间差^[4]。设两个距离固定为 D 的传

收稿日期: 2002-07-15; 修回日期: 2002-09-08

作者简介: 胡跃辉(1970-),男,博士生,主要从事超声技术应用方面研究。

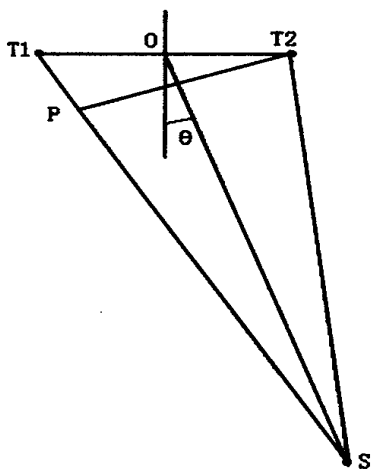


图 1 利用两个传感器确定 S 点声源方位示意图

传感器 T_1, T_2 ; 声音采样速率为 F_s ; 室温下空气声速为 C ; 如图 1 所示。

$$R(k) = \sum_{n=0}^N T_1(n) * T_2(n+k) \quad (1)$$

式(1)为本文所用的互相关公式, $T_1(n), T_2(n)$ 为同时接收到的来自声源 S 的接收信号序列, N 为要进行分析的声音信号序列片段长度, $R(k)$ 为 T_2 通道相对 T_1 通道延迟 k/F_s 秒时的互相关系数。假设互相关系数 $R(k)$ 最大时, $k = \Delta n$, 并且假定距离 $OS \gg D$ 时 $\angle T_1 T_2 P \approx \theta$, 则有, 声源 S 相对 T_1, T_2 的夹角 θ 为:

$$\theta = \sin^{-1} \left(\frac{C \Delta n}{F_s D} \right), \quad k \in \left[-\frac{F_s D}{C}, \frac{F_s D}{C} \right] \quad (2)$$

但是在室内存在混响的情况下, 互相关法遇到巨大的障碍。首先由于扩散声场中声线在叠加以形成室内空间一点瞬时声压时, 它的位相变化是没有规律的, 而互相关法要求两路信号之间有较严格的线性相位关系。其次, 由人耳判别声源的方式来看, 互相关法仅利用的接收信号的时差信息, 对强度信息, 互相关法未能在公式中表达出来。最后, 如前所述, 时差法仅适用低频信号, 而在室内条件下, 声源信号中的低频分量较高频更易受到混响的污染, 因而导致低频分量检测存在较大误差。原因是墙壁、地面等反射体对低频分量的吸收系数以及空气中传播衰减系数要比高频分量小, 导致反射信号(含多次反射)中低频分量仍然很强, 在和直达声信号中低频分量相叠加后足以导致系统误判。

由数字多波束合成理论^[5-6]可知, 无指向性的发射阵元 T_1, T_2 发射同一声音序列, 欲使确定的空间(对于本文讨论的二阵元阵则是一个平面)某一方

向 θ 能量最大, 只需发射阵元 T_1 和 T_2 发送信号的延迟 Δn 满足式(2)即可。由此, 可写出波束合成法求方位角的基本公式:

$$I(k) = \sum_{n=0}^N [T_1(n) + T_2(n+k)]^2 \quad (3)$$

求出使 $I(k)$ 最大的 $k = \Delta n$, 利用式(2)即可求得方位角 θ 。将式(3)展开整理可得:

$$I(k) = 2R(k) + \sum_{n=0}^N [T_1^2(n) + T_2^2(n+k)] \quad (4)$$

可见, 波束合成法含有互相关项, 同时又含有能量成分, 因此对声音信号高低频分量都有较好的兼顾。

和互相关法相比, 波束合成法虽然可以利用高通滤波器避开接收信号中混响成分强烈的低频分量, 但是, 由于该基本公式没有包含对室内声场特性的表达项, 也即混响参数, 所以在室内声场中, 仍然不能获得理想的结果。

对上述两种方法我们做了如下试验:

(1) 在一个长宽高乘积为 $4.1 \times 3.4 \times 3.2 \text{m}^3$ 有一定吸声效果的空旷封闭录音室内, 墙壁、天花板铺设细孔吸声材料, 地板铺设地毯。对距离接收阵前方 $0.5 \text{m} \sim 3.5 \text{m}$ 每间隔 0.5m 的各网格点用上述两种方法进行测试, 滤波器选用 $800 \text{Hz} \sim 4000 \text{Hz}$ 带通 FIR 滤波器, 波束形成法可以达到平均 $\pm 2.5^\circ$ 的误差, 最大误差 5° , 有一个点误判, 相关法则有 22% 的误判(指方向判定完全相反的情况)。说明波束形成法较相关法优越, 同时也说明该录音室的室内声学结构仍然对两种测试方法有很大影响。

(2) 同一房间, 对利用 $800 \text{Hz} \sim 4000 \text{Hz}$ 带通滤波器时相关法和波束形成法均不产生误判并且精度在 ± 3 度以内的点上, 做了一组带通滤波器试验, 3dB 低频截止点范围 $200 \text{Hz} \sim 1500 \text{Hz}$, 高频 3dB 截止点范围 $3000 \text{Hz} \sim 5000 \text{Hz}$, 步进为 50Hz , 各滤波器均采用 Kaiser-Bessel 窗函数, 长度 1023 点, 采样率 500kHz , 直流衰减均为 150dB 。结果表明 $1000 \text{Hz} \sim 4000 \text{Hz}$ 带通滤波器精度最佳。也就是说用语音信号高频分量进行方位判定效果最好, 对于被动式语音定向, 采用能量性的判别函数较采用时间差为主的判别函数效果要好。

(3) 在另一个长宽高乘积为 $6 \times 5.5 \times 4.3 \text{m}^3$ 的普通空旷封闭房间, 墙壁、天花板为摸灰水泥面, 地面铺设地毯, 窗户用窗帘遮挡, 选用 $1000 \text{Hz} \sim 4000 \text{Hz}$ 带通滤波器对距离接收阵前方 $0.5 \text{m} \sim 4 \text{m}$ 每间隔 0.5m 各网格点用两种方法进行测试, 结果相关法

80% 以上的点误判, 采用波束合成法则有 23% 的点误差 $> \pm 6$ 度, 其中 16% 为误判, 其它各点平均误差在 ± 6 度。

上述各试验选用的语音材料为中文新闻录音, 并通过扬声器循环播放, 以下实验也相同。

经过对相关法和波束形成法的分析和试验, 我们可以得出一个结论, 在室内条件下, 声音定向必须考虑室内声场的影响, 也就是说, 判别函数必须包含反映室内声场特性的数学模型。

2 室内声场脉冲响应模型

J. Huang^[7] 等人利用室内声场脉冲回波响应模型来检出回波声音片段并将这些片段剔除, 对剩余的语音片段则认为是原声信息, 并利用这些原声片段进行声源定向。较好的解决了有混响情况下的室内声音定向问题。

一声线自声源发出, 经过且只经过一次反射后的反射波幅度模型可以用如下公式来描述:

$$i_{ech}(t) = \begin{cases} 0, & 0 < t < T_e \\ \alpha_e e^{-(t-T_e)/\tau}, & \end{cases} \quad (5)$$

记 $t = 0$ 为声线发出时刻; 声线发射幅度归一化为 1; T_e 为声线自发射经反射面反射后再被传感器接收的传播时间; τ 为时间常数; α_e 为回波接收幅度。

基于式(5), J. Huang 提出了以下的回波检出方法:

a. 对接收到的声音片段经带通滤波器后进行幅度包络检波, 获得幅度信号序列 $s(t)$ 。

b. 对 $s(t)$ 任一时刻 t_0 的幅值, 假设在 t_0 之后的时刻 t , 幅值变为 $s(t_0) i_{ech}(t - t_0)$, $t_0 < t$;

而对于 t 时刻的回波幅值按式(6)来获得:

$$ech(t) = \text{Max}\{s(t_0) i_{ech}(t - t_0)\}, \quad 0 < t_0 < t \quad (6)$$

至此, 可构建出反映各时刻回波强度影响因子:

$$r(t) = \frac{s(t)}{ech(t)} \quad (7)$$

c. 利用 $r(t)$ 来进行回波信号和原声信号的分离, 对 $r(t)$ 大于一定值的点就认为是原声信号, 反之则认为是回波信号予以剔除。最后对检出后的原声信号片段, 利用过零点检测法来获得通道间信号时差并进而求出声源位置。

该方法由于引入了室内声场模型, 从而在有混响条件下, 获得了较好的实验结果。但是过零点检

测法对通道间两个信号序列是否存在相似关系无法判别, 而是认为两信号序列过相同零点(或者相同相位)的时刻差就是所求的两信号序列声程时差, 但是式(7)检出的原声信号只是代表所受回波(或称混响)干扰较小的情况, 而实际上由于存在电路噪声等随机干扰, 根据前面所述的原理: 室内声场声线在叠加时它们的位相变化是无规的, 因而产生误判的可能性很大。

同时, 基于有混响条件下的室内语音定向, 除了考虑室内声场特性以外, 还应该对语音特点加以研究并在定向判别中加以体现, 这些都是上述算法所需要改进和完善的。由此, 我们提出了如下的改进算法。

3 基于室内声场脉冲响应模型和语音识别处理技术的数字波束合成语音定向算法

显然, 直达声最先达到传感器, 而经过反射等其它途径到达传感器的时间要比直达声有所延迟。这个延迟一方面和后续的声源直达声叠加影响测位精度; 另一方面, 也说明如果两个语音片段间隔足够长(比如大于室内混响时间), 那么第二个语音片段起始部分可以认为就是直达声, 对于利用这部分语音片段测量的角度就比较精确, 因此在最后统计处理输出结果时可以给予比较大的一个权重。反之, 对于一个连续语音片段的中后部分, 则有很大可能是前面语音信号回波信号产生的, 即使利用式(7)检测表明是原声信号而进行了测量, 其测量结果也要考虑在最后统计处理输出时取较小的权重。而我们对于语音片段的划分以及某声音片段位于语音片段的位置可以利用语音信号处理中的端点检测技术。还有, 中文清音频谱较广, 高频成分能量较强, 和浊音频谱能量相比清音能量要小很多, 且频谱类似噪声谱。但是由前面讨论和实验结果来看, 利用高频分量较强的清音音节定向精度应该较浊音音节要好。而文献^[7]对所有声音信号都采用同一方式进行处理, 因为清音信号频谱广且信号能量较小, 就有可能将随机噪声信号等同于清音信号进行处理, 从而引起定向误差, 或增加统计处理结果的时间和长度(比如需增加每次输出结果所要的测定方位角的样本数目), 影响了实时性和准确度。为此, 我们引入语音信号中常用的清音和浊音判别技术和端点检测技术^[8~10, 12~14]一起构成本算法的语音识别处理部分

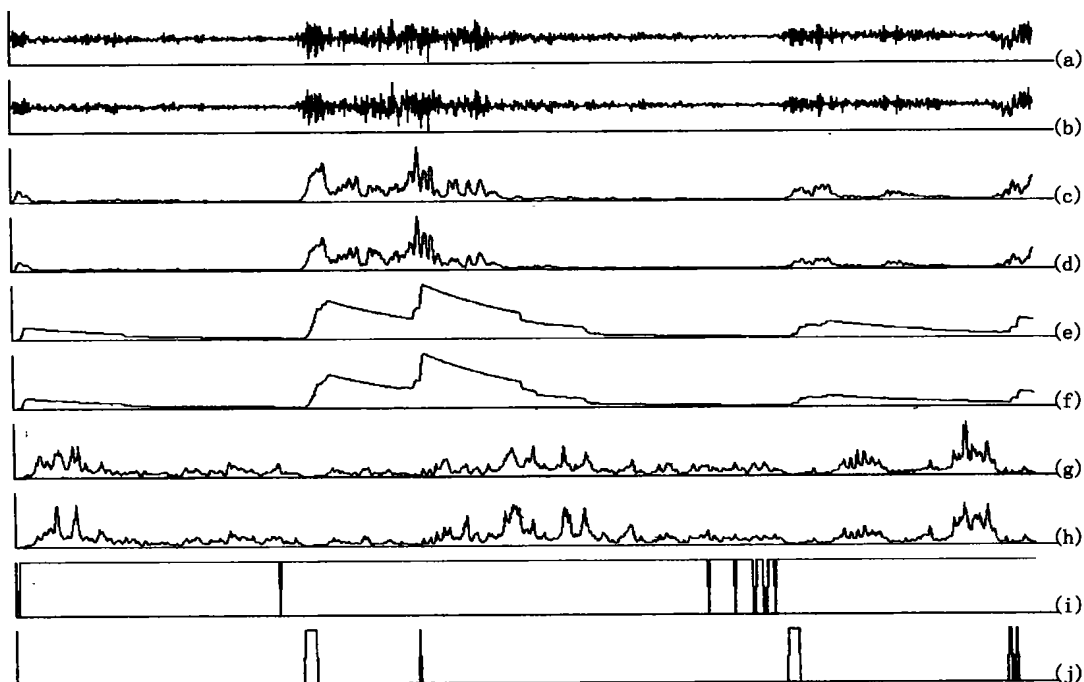


图2 一次采用语音识别端点检测和回波估计联合检出可用语音片段过程中各阶段的输出结果

再配合回波强度影响因子(公式(6))来联合检出原声语音片段。最后利用数字波束形成法求出一个样本方位角,并利用截尾和加权平均对若干样本点进行统计处理从而获得一次结果输出。

图2是一次在以语音识别端点检测和回波估计两者联合检出过程中对原始数据进行处理所获得的各阶段结果示意图谱,各图谱横轴为时间,长度均为1秒。以下对整个检出算法作一介绍。

整个语音定向系统分成两个部分,即有效语音片段检出模块和方位角计算模块。有效语音片段模块用于计算出可供方位角计算模块用的语音片段的起始和结束位置,方位角计算模块仅用有效语音片段来计算方位角,对于本文来说,即波束形成法,也可以利用其它时延估计方法。使用 AD-Link PCI9812 12 位四通道同步采集卡,最大模拟带宽 20MHz。为避免不必要的误差,自 T_1 和 T_2 接收到的同步声音信号在送 A/D 进行数字化以前,经过 20kHz 模拟低通滤波器作抗混叠处理。在使用波束法进行时延估计时,合成后的波束频谱将大大扩宽^[5-6],因此需要更高的采样频率,本文情况采样速率为每通道 600kHz。每路通道采用一个抗混叠滤波器,挑选高精度参数的元器件以保证各抗混叠滤波器特性基本一致,并在使用前用一路信号同时输入各抗混叠滤波器后经 PCI9812 采样至计算机内,

以计算出各滤波器间的固有相位差。利用各通道固有相位差即可求出以后系统运行过程中自采集卡读取各通道数据时所需的地址偏移量,以便处理器利用相对位移寻址指令在读出数据的同时也消除因采用抗混叠滤波器而产生的固有相位差对测量精度的影响。如此获得读取的各路采样数据信号序列为 $T_1[n]$ 、 $T_2[n]$ 。

对最后方位角计算所用语音片段仍采用 600kHz 数据 $T_1[n]$ 、 $T_2[n]$ 作为输入信号进行处理,只不过起始和结束位置由有效语音检出模块计算决定,这一部分将在稍后的几节再作介绍。而对于有效语音片段检出模块,如果也利用 600kHz 采样率输入数据进行计算,则计算量太大,因此本文将 $T_1[n]$ 、 $T_2[n]$ 减抽样至 40kHz。但是,如果采用正常的减抽样处理方式,需要对 600kHz 信号先做 20kHz 低通抗混叠滤波,然后按 15:1 获得所需减抽样序列。而 600kHz 的抗混叠滤波仍然要消耗大量硬件资源。由于有效语音检出模块完成的功能是给出有效语音的起始和结束位置,其位置的偏差对于检出的有效语音片段满足一定长度后所能产生的影响可忽略不计,特别是考虑到语音信号能量绝大部分集中在 5kHz 以下,如果在数字化前又经过 20kHz 抗混叠模拟滤波后,其直接 15:1 抽取产生的混叠影响可能会更小,出于这一设想,我们加装了模拟抗混叠滤

波器, 并对数字化后的 600kHz 采样信号作直接抽样 y 以及按正常减抽样处理抽样获得的信号序列分别送有效语音检出模块进行处理的对比试验, 最后得到的位置几乎没有差别, 因此对方位角计算模块就不产生影响。经过对比试验后, 我们认为直接抽样是可行的, 前提是在前端加装 20kHz 抗混叠滤波器, 并且在 151 进行减抽样后再经过 200Hz~5000Hz 数字带通滤波器获得 40kHz 减抽样滤波信号 $T_{sub1}[n]$ 、 $T_{sub2}[n]$ (图 2 中(a)、(b)), 以便只对 5kHz 以下信号分量部分作分析。此外, 本文所提到的数字带通滤波器均是利用专业软件设计的直流衰减为 -200db 的 FIR 滤波器。

对 $T_{sub1}[n]$ 、 $T_{sub2}[n]$ 按式(8) 计算短时能量得 $E_1[n]$ 、 $E_2[n]$ (见图 2 中(c)、(d))

$$E[n] = \sum_{k=n-N+1}^n W_H[n-k] T_{sub}^2[K] \quad (8)$$

式中, $W_H[n]$ 为汉明窗, 为避免较多高频分量损失, 并考虑到最后检出的有效语音片段较短(一般在几 ms), 汉明窗长度 N 仅取 4ms, 而不是语音识别(删去) 端点检测中常用的 10ms~20ms 长度:

$$W_H[n] = \begin{cases} 0.54 - 0.46\cos[2\pi n/(N-1)], & 0 \leq n \leq N-1 \\ 0, & \text{其它} \end{cases} \quad (9)$$

按式(5)、式(6) 方法对 $E_1[n]$ 、 $E_2[n]$ 作回波能量估计得 $Echo_1[n]$ 、 $Echo_2[n]$ (图 2 中(e)、(f))。式(5) 中 τ_e , 当接收阵列距离任一反射界面 $\geq 0.5m$ 时可设定为 0.3ms, 其它参数可以根据经验和试验结果来选取。

然后, 由式(7) 求出回波能量影响因子 $r_1[n]$ 、 $r_2[n]$ (图 2(g) 和图 2(h) 为回波能量影响因子倒数谱 $1/r_1[n]$ 、 $1/r_2[n]$)。

同时, 利用 $E_1[n]$ 、 $E_2[n]$ 作语音端点检测。在信噪比良好的实验室环境中, 本文选用的是能量和过零率法。在整个系统初始化时, 保持 10s 的静寂时间, 同时对这段时间进行 40kHz 采样和 200Hz~5000Hz 滤波, 再利用式(8) 统计出平均噪声能量 EMN , 利用下式(10) 求出最小有声门限和最大有声门限:

$$\begin{cases} ETL = 4EMN \\ ETU = 5ETL \end{cases} \quad (10)$$

并且在 200Hz~5000Hz 滤波后统计平均噪声过零率 ZC 并计算其方差 σ_{ZC} 计算出过零率门限 ZT :

$$ZT = ZC + 2\sigma_{ZC} \quad (11)$$

各点的过零率按(12):

$$Z[n] = \sum_{k=n-N+1}^n | \operatorname{sgn}(T[k]) - \operatorname{sgn}(T[k-1]) | \quad (12)$$

$$\operatorname{sgn}(T[n]) = \begin{cases} 1, & T[n] \geq 0 \\ -1, & T[n] < 0 \end{cases} \quad (13)$$

式中 N' 长度取 15ms。

对两个传感器分别利用 ETU 、 ETL 求出有声语音片段起始 N'_{1i} 和结束端点 N'_{2i} (下标 i 用来标记不同的端点对), 使端点对内各点能量均 $\geq ETL$, 且片段中最少包含有一点能量 $\geq ETU$ 。然后对 N'_{1i} 向前, N'_{2i} 向后搜索, 直至该点的过零率低于 ZT , 此刻可认为是最终两个端点 N_{1i} 、 N_{2i} 。接着将两个传感器所求出的有声片段进行合并, 求出有声片段的交集, 且用二值函数形式表示(图 2 中 i), 作为语音识别端点检测部分的结果。

本例对 $r_1[n]$ 、 $r_2[n]$ 中小于 0.3 的点认为是有效的语音片段, 同样对 $r_1[n]$ 、 $r_2[n]$ 检出的有效片段求交集, 作为回波估计部分的结果。最后将语音识别端点检测部分和回波估计部分的结果合并求交集获得最终的检出结果(图 2 中 j)。

对照图 2 可以看出, 有效语音片段在有混响条件下占整个声音信号序列的比重很小, 有效语音片段往往出现在一段语音的开始位置, 这和本节开始时预计的情况相吻合。此外, 对于出现在一段语音较后位置的检出段, 如本节开始所指出的, 即使回波估计认为是有效的, 仍然应该在该段信号求出角度后乘以一个较小的权重后再加入最后结果输出计算中, 权重由试验来确定。同时由图 2 也可以看出, 整个声音序列在不考虑室内声场特性以及语音特征的情况下单纯使用信号处理方法来进行定向, 误差是相当大的且会产生很多错判的结果。

对检出的减抽样 40kHz 有效语音片段端点, 通过修正因 200Hz~5000Hz 滤波器以及检出算法过程各窗函数引起的线性相移后, 求出在 600kHz 下对应的位置。由于采用波束合成法对要计算的信号长度有一定要求, 本例我们只对检出信号长度大于 2ms 的进行计算。对符合要求的 600kHz 信号作 1023 点 800Hz~5000Hz 带通滤波。图 3 给出了一次上述处理的结果, 图 3 中的 a 和 c 为自 T_1 和 T_2 中提取的 600kHz 语音片段, 图 3 中的 b、d 则为做了 800Hz~5000Hz 带通滤波的结果。可以看出, 滤除低频信号后, 信号相似性得到进一步改善(图 3 中(b)、(d))。

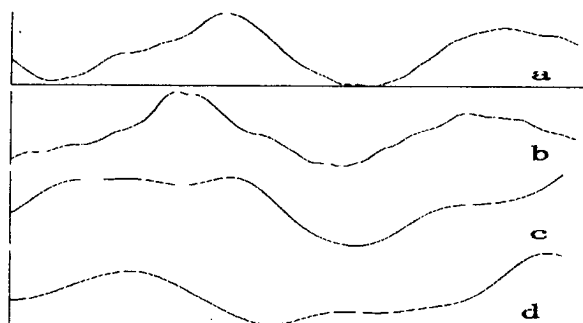


图3 检出的有效语音片段及其 800Hz-5000Hz 滤波后的情况

在长宽高乘积为 $6 \times 5.5 \times 4.3 \text{m}^3$ 的普通空旷封闭房间, 墙壁、天花板为摸灰水泥面, 地面铺设地毯, 窗户用窗帘遮挡。上述方法进行测试, 测试角度为 $\pm 70^\circ$, 距离 1m~4m, 采用 3 传感器阵元组成接收阵列 (算法同前面介绍方法相类似), 平均获得的水平精度为 $\pm 3.5^\circ$, 最大误差 7° (出现在房间对角线方向上), 大角度情况下误差变大, 解决的办法之一, 是遇到大角度值时就转动整个接收阵使之接近中心位置。角度、距离、绝对误差($^\circ$) 测量数据见表 1。没有误判现象。

表 1 不同角度、距离下的绝对误差

角度 $^\circ$	绝对误差/ $^\circ$			
	1m	2m	3m	4m
0	3.3	1.3	1.4	4.1
10	3.8	1.6	3.8	1.1
20	1.2	1.3	1.3	1.5
30	1.1	3.0	3.1	1.7
40	4.4	2.4	3.5	5.6
50	2.9	2.8	6.9	5.8
60	4.1	5.3	5.4	5.3
70	4.3	5.1	4.2	5.7

4 结 论

有混响室内语音定向技术是智能机器人的人耳仿真, 以及电视会议系统摄像定向, 乃至法庭说话人识别及定向等方面的关键技术。在室内有混响条件

下, 语音定向不仅要考虑室内声场的特点, 还要利用语音的特征对混响进行检出, 特别是在说话人识别方面更要利用说话人的语音特征进行说话人语音分离并定向, 单纯采用普通的信号处理方法及抗干扰方法是不行的。

参 考 文 献

- [1] John D Pitt, Martin Baxley. Directional sensitivity of the human ear [J]. Journal of Biological Education, 1985, 4(19): 275-276.
- [2] 崔庚寅. 人耳对声源定位的时差研究[J]. 河北师范大学学报(自然科学版), 1994, 4(18): 84-86.
- [3] 杜功焕, 朱哲民, 龚秀芬. 声学基础(下)[M]. 上海: 上海科学技术出版社, 1981. 167-200.
- [4] Bendat JS, Piersol A G 著, 凌福根 译. 相关分析和谱分析的工程应用[M]. 北京: 国防工业出版社, 1983. 254-273.
- [5] 李启虎. 声呐信号处理引论[M]. 海洋出版社, 1985. 65-117.
- [6] 冯若, 刘忠齐, 姚锦钟, 等. 超声诊断设备原理与设计[M]. 北京: 中国医药科技出版社, 1993. 251-279.
- [7] HUANG J, Ohnishi N, and Sugie N. Sound Localization in Reverberant Environment based on the Model of the Precedence Effect [J]. IEEE Trans. Instrum. Meas., 1997, 4(46): 842-846.
- [8] 王仁华, 陈永彬. 语音信号处理[M]. 合肥: 中国科学技术大学出版社, 1990. 202-235.
- [9] Lawrence Rabiner. 语音识别基本原理[M]. 北京: 清华大学出版社, 1999. 104-167.
- [10] 胡光锐. 语音处理与识别[M]. 上海: 上海科学技术出版社, 1994. 123-156
- [11] 王成友, 朱庶, 马建华, 梁甸农. 一种新型的语音分析编辑合成系统[J]. 声学技术, 1997, 16(1): 14-17.
- [12] 张杰, 黄志同. 无端点检测汉语识别算法的实现及改进 + 动态时间规整和隐马尔可夫统一模型的应用[J]. 声学技术, 1998, 17(4): 181-185.
- [13] 钱城. 微机和数字信号处理器模拟室内脉冲响应的音质仿真[J]. 声学技术, 1995, 14(4): 172-191.

译作《音乐厅和歌剧院》获全国优秀科技图书三等奖

同济大学出版社出版的《音乐厅和歌剧院》, 是国际著名的建筑声学专家白瑞奈克(Leo L. Beranek) 教授论著的中译版, 由我国著名声学专家王季卿教授领衔翻译, 中国工程院资深院士马大猷教授审定。书中总结了作者 30 年来科学研究成果, 以主要篇幅描述全世界 22 个国家和地区的 76 座音乐厅和歌剧院的音质设计经验, 并由此总结出一系列评价标准。除一般附图外, 还有各厅、院的大量优美的内部构造照片和结构图。该书被公认为关于建筑声学实际经验的巨著。原书作者为了该译本的出版, 特意为全书作了详细的校对, 并把原书出版后的最新资料和数据补充寄给译者, 使得译著的局部章节与原著相比有较大的变动, 中国读者有幸比英文书的读者早两年读到该书英文修订再版本的内容。这种情况在译著中是不多见的。

日前揭晓的第十一届全国优秀科技图书奖中《音乐厅和歌剧院》获全国优秀科技图书三等奖。