

# 基于双耳模型的倒谱系数在声目标识别中的应用

管鲁阳<sup>†</sup> 鲍明 李晓东

(中国科学院声学研究所 北京 100190)

**摘要** 在声目标识别中,基于单路信号的特征提取算法难以保证特征在低信噪比条件下的鲁棒性。针对该问题,本文使用由两个传声器组成的小孔径阵列采集信号,采用 Stereausis 双耳听觉模型对两路信号进行综合处理,在此基础上设计了基于双耳模型的倒谱系数,并将其作为识别特征应用于多类地面目标识别。试验表明该特征可有效提高识别系统的识别性能和抗噪声能力。

**关键词** Stereausis, 双耳模型, 倒谱, 目标识别

## Application of novel cepstrum coefficient based on binaural model to acoustic target recognition

GUAN Lu-Yang BAO Ming LI Xiao-Dong

(Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190)

**Abstract** Stereausis model encodes both binaural and spectral information in a unified framework. In this paper, a new type of cepstrum coefficient is proposed based on this model and is applied to multi-class ground target recognition. Experimental results show that this cepstrum coefficient based on binaural model improves both the performance and generalization of the classifier, especially at low SNR.

**Key words** Stereausis, Binaural model, Cepstrum, Target recognition

### 1 引言

目标的辐射噪声往往可以反映该声源目标区别于其他目标的特性,因此基于声信号的目标识别被广泛应用于众多领域。然而,目标所处环境不可避免地存在自然背景噪声或人为噪声的干扰,这些干扰降低了传声器所采集

信号的信噪比,从而严重影响声目标识别系统的识别性能。因此提高特征的抗噪声能力是声目标识别系统非常重要的工作。

在地面目标识别应用中,由于设备工作环境存在风噪声等多种噪声的影响,设计具有较强抗噪声能力的识别特征显得尤为重要,各种声信号的分析处理技术被广泛应用于特征提取。

2007-05-22 收稿; 2008-07-14 定稿

作者简介:管鲁阳(1979 - ),男,河南省南阳市人,中国科学院声学研究所博士研究生,研究方向:信号处理、机器学习。

鲍明(1973 - ),男,博士研究生。李晓东(1966 - ),男,研究员。

<sup>†</sup> 通信联系人 E-mail: guanluyang@mail.ioa.ac.cn

Sampan 组合多种时域信号短时能量的相关参数作为特征区分五种车辆类型,最高正确率约 79%<sup>[1]</sup>。由于时域信号所含信息有限,多数车辆识别系统提取信号的频域特征以获得更准确的分类。在威斯康辛大学的 SensIT 项目中,无线传感网络的前端识别系统采用基于 FFT 的频域特征区分轮式车和履带车,其识别正确率约 70%<sup>[2]</sup>。在文献[3]中,实时车辆识别系统组合了信号的短时傅里叶频谱特征和基于小波分析的特征,对两类车辆的识别正确率达到 98%,但对于多类车辆识别,难以保证同样的正确率。文献[4]则利用听觉外周模型提取信号的频域特征,取得 91% 左右的正确率。以上所述文献中,虽然特征提取算法不同,但是都基于单个传声器所采集的数据提取特征。由于单路信号所包含的声源目标信息有限且消除噪声干扰相对比较困难,因此基于单路信号的特征提取算法难以保证所提取特

征在低信噪比条件下的分类性能。

本文借鉴人耳听觉对噪声环境的自适应能力以及阵列信号处理的优势,使用由两个传声器组成的小孔径阵列用于采集目标的声信号。对于此小孔径阵列,采用模拟人双耳听觉感知功能的 Stereausis 模型综合处理两路信号<sup>[5]</sup>并在此基础上设计新的特征提取方法以提高识别系统的分类性能和抗噪声能力。

## 2 Stereausis 双耳模型

Shamma 提出的 Stereausis 双耳模型<sup>[6]</sup>通过检测双耳同步输入信号的空间差异来测度其总体的耳间差异,在输出模式中同时反映两路信号的共性和差异。Stereausis 模型如图 1 所示,双耳采集的两路声信号分别经耳蜗模型分析后,由双耳模型模拟人综合处理双耳信息的生理机能,得到包含两路信号信息的输出模式 Y。

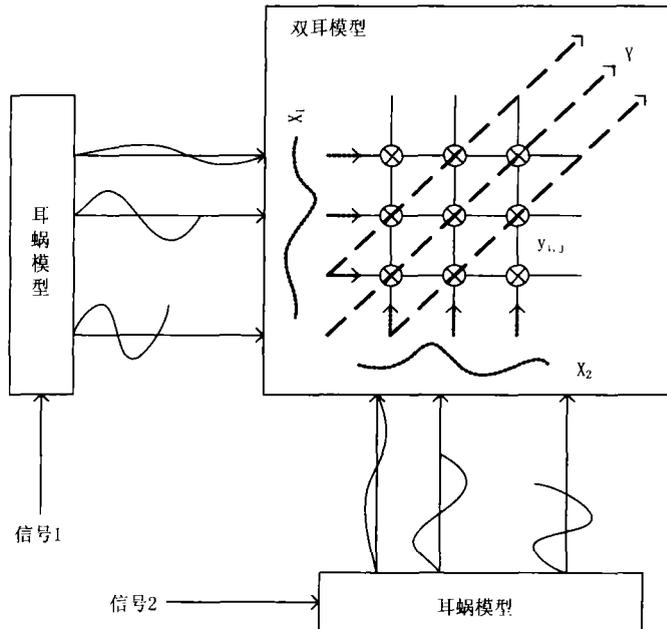


图1 Stereausis 双耳模型结构示意图<sup>[6]</sup>

### 2.1 耳蜗模型

耳蜗模型的核心是耳蜗滤波器组,它模拟了耳蜗基底膜对声信号频域信息的分析功

能<sup>[7]</sup>。本文所使用的耳蜗滤波器组由多个带通滤波器组成的恒 Q 滤波器组实现,图 2 是部分滤波器的冲击响应。

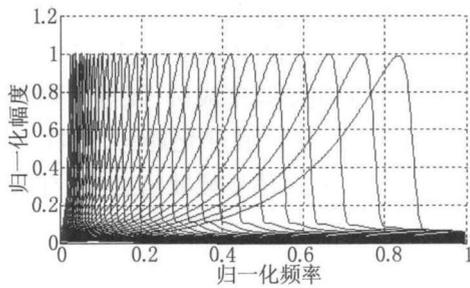


图 2 耳蜗滤波器组部分滤波器的频率响应

经过预处理的一帧声信号首先经过耳蜗滤波器组的滤波

$$y_1(n, m) = x(n) \otimes h(n, m) \quad (1)$$

其中  $h(n, m)$  是第  $m$  个滤波器的冲击响应。然后通过非线性变换(式 2)得到耳蜗模型的输出  $y_2$ , 该输出即双耳模型的输入  $X$ , 图 1 所示  $X_1$  和  $X_2$  分别表示来自两个耳蜗模型的输入矢量。该非线性变换模拟信号由基底膜震动转换为内毛细胞电势的信号模式变换过程<sup>[7]</sup>。

$$y_2(n, m) = g(y_1(n, m)) \quad (2)$$

其中  $g(x)$  是 sigmoid 函数, 定义为

$$g(x) = \frac{1}{1 + e^{-\gamma x}} - \frac{1}{2}, \quad \gamma \text{ 是非线性增益}^{[7]}。$$

### 2.2 Stereausis 双耳模型

双耳模型综合两个耳蜗模型输出模式所包含的信息, 得到模拟人双耳听觉功能的输出模式。如图 1 所示, 对两个输入  $X_1$  和  $X_2$ , 双耳模型输出模式  $Y$  的每个输出单元  $Y_{ij}$  计算如下:

$$Y_{i,j} = \frac{1}{N} \sum_{n=1}^N (X_{1,n}(i) + X_{2,n}(j))^2 \quad (3)$$

其中  $X_1(i)$  和  $X_2(j)$  分别是两个耳蜗模型响应中第  $i$  个和第  $j$  个通道的输出值,  $N$  是每帧信号的长度。

图 3 为相距 35cm 的两个传声器以 1kHz 采样率同时采集的履带车声音时域信号。从第 35s 开始, 截取 256ms 数据为一帧, 经 Stereausis 模型处理后得到的输出模式为  $128 \times 128$  的矩阵, 如图 4 所示。

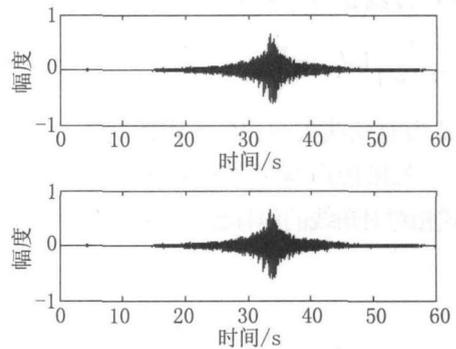


图 3 两个传声器同步采集的履带车声音信号 (上图:左通道;下图:右通道)

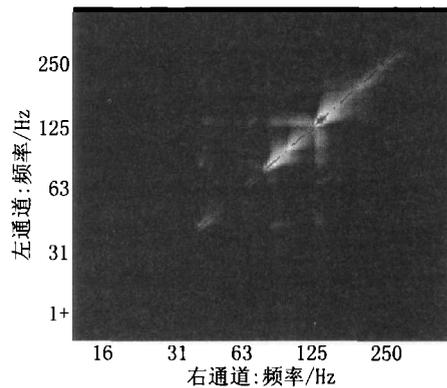


图 4 Stereausis 模型对一帧信号的计算结果

### 3 基于双耳模型的倒谱系数

Stereausis 模型输出模式(图 4)的形状、对称性, 峰值的位置及分布等包含了信号的频谱、相位差等很多信息。其对角线主要反映了信号的频域特征, 和通常的傅立叶频谱不同的是, 由于 Stereausis 模型是模拟人的双耳对声信号的分析过程, 对角线矢量综合了两路信号的信息, 并在一定程度上体现了人耳对声信号频域特征的理解和提取, 因此可以作为特征应用于目标识别。

由于双耳信号之间的差异, 导致 Stereausis 模型的输出模式中, 每个通道对应的峰值相对于对角线会产生不同的偏移, 因此, 在抽取对角线矢量时, 可对相邻的对角线矢量进行累加

取均值,以纠正峰值偏移的影响,如:

$$\bar{d}_i = \frac{1}{2C+1} \left( d_i + \sum_{j=1}^C (y_{i-j,i+j} + y_{i+j,i-j}) \right) \quad (4)$$

其中,  $d$  为对角线矢量;  $2C$  为相邻对角线的数量(图1)。如果传声器距离较近, Stereausis 模型输出模式相对对角线的偏移很小,可忽略此偏移量。

以滤波器的中心频率为坐标,该对角线矢量和傅里叶频谱的对比(图5)表明对角线矢量保留了信号的主要频率成份,而对次要频率成份进行了平滑,保留其变化趋势,这在一定程度上反映了人对声信号频域信息的分析、理解特性。

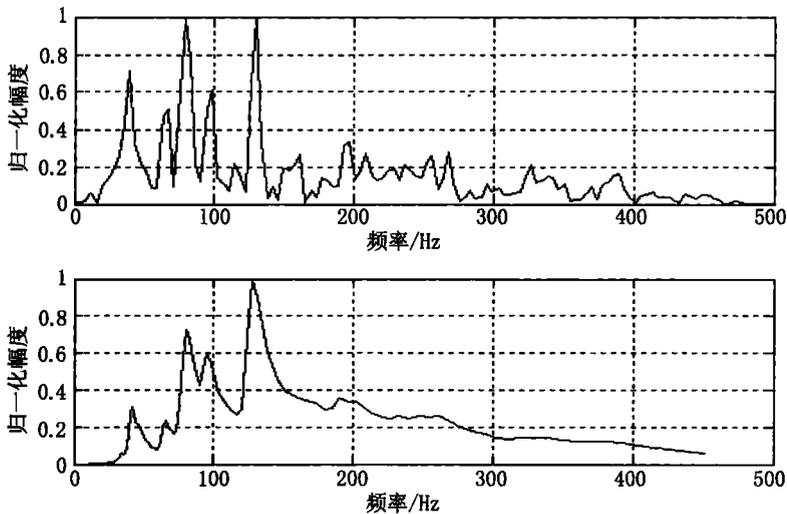


图5 Stereausis 输出模式的对角线矢量和傅立叶频谱的对比  
(上图:傅立叶频率; 下图:Stereausis 模型二维输出模式的对角线矢量)

在语音识别系统中,常用的特征是 Mel 频率倒谱系数 (Mel Frequency Cepstrum Coefficient, MFCC)。Mel 频率倒谱系数相对于实倒谱系数的重要改进是引入反映人听觉感知特性的 Mel 频率滤波器组。双耳模型和 Mel 频

率滤波器组相比,更精细地模拟了人耳对信号频域信息的分析功能,并综合了两路信号的信息,可以有效降低噪声的影响<sup>[5]</sup>。因此本文用双耳模型取代 Mel 频率滤波器组,得到基于双耳模型的倒谱系数。

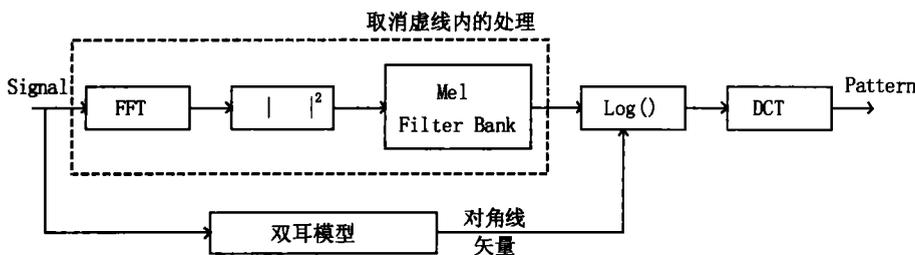


图6 双耳模型倒谱系数的计算框图

两路同步采集的声信号首先经过双耳模型处理,然后对其输出模式的对角线矢量求对数和离散余弦变换得到双耳模型倒谱系数。

图6显示了本文提出的双耳模型倒谱系数的计算过程。其中离散余弦变换(DCT)和计算MFCC时的DCT相同,定义为<sup>[8]</sup>:

$$c(i) = \sqrt{\frac{2}{L}} \sum_{m=1}^L \log(\hat{s}(m)) \cos\left(\frac{\pi i}{L}(m-0.5)\right)$$

$$i = 1, 2, \dots, C \quad (5)$$

其中  $L$  为对角线矢量的维数,  $C$  为倒谱系数的维数。

## 4 地面目标识别系统

本文所述地面目标识别系统的目标是根据声信号正确区分四种车辆类型,包括两种履带车和两种轮式车。因为车辆行驶声音的能量主要集中在 500Hz 以下的低频部分且实时系统计算能力有限,所以本系统利用距离 35 cm 的两个传声器组成的小孔径阵列以 1kHz 采样率采集车辆行驶的声信号。

### 4.1 特征提取

车辆行驶的声信号是非平稳的,但在小于 200 ~ 300ms 的瞬间内,可近似认为是平稳的<sup>[4]</sup>。所以本系统中每 256ms 数据截取为一帧,每帧数据经预处理后提取上文所述基于双耳模型的倒谱系数特征,其维数设定为 32,计算过程如图 6 所示。

另提取信号的 Mel 频率倒谱系数特征(MFCC)和小波特征作为对比。使用语音信号处理工具箱 VoiceBox 得到 24 维 MFCC 特征向量(包括 12 维 MFCC 和 12 维一阶差分 MFCC)。小波特征则采用基于 db6 小波的 32 维特征。四类车辆的特征数量见表 1。

表 1 特征数量说明

车辆类型	履带车 1	履带车 2	轮式车 1	轮式车 2
特征样本数量	2550	2569	515	614

### 4.2 分类器

本系统采用基于矢量量化的 Learning VQ2.1 算法<sup>[9]</sup>作为分类算法。当  $\lim_{N \rightarrow \infty} (K/N) \rightarrow 0$  ( $K$  为码元数量,  $N$  为样本数量), LVQ 的分类错误率将趋近于 Bayes 最小分类错误率<sup>[10,11]</sup>。因此, LVQ 可以作为一个很好的多

元分类算法应用在模式识别中。

为简单起见, LVQ 算法从样本集中随机挑选预定数量的样本作为初始码本<sup>[12]</sup>, 虽然得到的结果未必最优,但是多次独立运行得到的平均正确率可以反映所提取特征的分类性能。

## 5 试验结果

### 5.1 不同特征的分类结果比较

对所有特征样本集使用  $m$ -folder 的交叉校验法( $m=3$ )<sup>[13]</sup>, 每次从样本总集中随机挑选 1/3 的样本作为训练集,其余 2/3 的样本作为测试集。重复运行 15 次取总体正确率的均值和标准差。试验结果见表 2,括号内为标准差。

表 2 不同特征分类结果统计表

VQ 码元数量	双耳模型倒谱系数 (%)	MFCC + $\Delta$ MFCC 特征 (%)	小波特征 (%)
100	92.01 (0.20)	92.68 (0.64)	86.24 (1.81)
300	94.75 (0.27)	92.66 (0.42)	89.57 (0.58)
400	95.27 (0.41)	92.51 (0.64)	90.22 (0.51)

由于 MFCC 特征维数较低,采用较少的码元,分类器即可得到比较稳定的性能,其他两种特征的分类正确率随着码元数量的增加而上升,高维特征空间的复杂性使得分类器往往需要比较多的码元才能取得较好的分类结果。最后三种特征的分类正确率都保持在一个比较稳定的水平,不再随码元的增加得到明显提升。这个比较稳定的分类正确率显示了对于地面目标,不同特征的可区分性的差别。

### 5.2 不同信噪比条件下的分类结果比较

对信号添加白噪声,在不同信噪比条件下对信号提取上述三种特征,使用 LVQ 分类器,设定码元数量为 300,采用  $m$ -folder 交叉校验( $m=3$ ),重复运行 15 次取平均正确率。试验结果见下图。

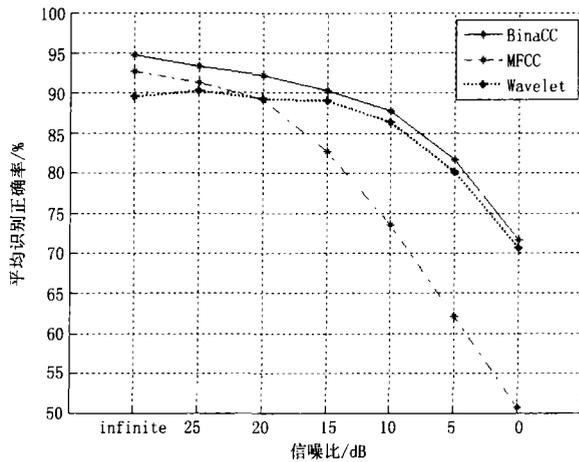


图7 不同信噪比条件下三种特征分类结果比较

图7显示在高信噪比条件下,基于双耳模型的倒谱系数的分类性能明显好于小波特征而略高于MFCC。随着信噪比的降低,双耳模型倒谱系数的分类性能下降最缓慢。当信噪比低于15dB, MFCC特征的分类性能显著下降,双耳模型倒谱系数的分类性能则略好于小波特征。

比较三种特征对噪声的鲁棒性可知,当信号的信噪比大于10dB,双耳模型倒谱系数的分类性能具有比较明显的优势,提高了识别系统在噪声环境中分类性能,其次是小波特征,而MFCC特征的抗噪声能力较差。

基于双耳模型的倒谱系数识别和抗噪声性能较好,其原因主要在两个方面:首先,特征的维数很大程度上由耳蜗滤波器组决定,此滤波器组类似信号处理中的非均匀子带分析,根据目标信号的特点在不同时频段采用不同的分辨率,既满足频率分辨率的要求也控制了子带数目。其次,双耳模型对耳蜗模型处理后的两路信号进行综合处理,得到的双耳模型输出模式中可能存在峰值位置相对于对角线的偏移。此偏移量既和两路信号之间的时间差(ITD)有关,也受到不相关噪声,如风噪声等背景噪声的影响<sup>[6]</sup>。式(4)纠正此偏移量即在一定程度上消除了类似噪声的影响,从而提高了特征的鲁棒性。

## 6 结论

针对声目标识别,本文采用 Stereausis 双

耳听觉感知模型综合处理两路声信号,在此基础上设计了基于双耳模型的倒谱系数作为识别特征,并将其应用于地面目标识别系统。与其他类型特征在不同信噪比条件下的对比试验表明基于双耳模型的倒谱系数在对噪声的鲁棒性和识别正确率之间取得较好的平衡,使声目标识别系统在不同信噪比条件下保持了较好且相对稳定的识别性能。

## 参 考 文 献

- [1] S Sampan. Neural fuzzy techniques in vehicle acoustic signal classification. Ph. D Thesis, Virginia Polytechnic Institute and State University, 1997.
- [2] M F Duarte, Y H Hu. Vehicle classification in distributed sensor networks. *Journal of Parallel and Distributed Computing*, 2004, 64(7): 826-838.
- [3] P K H Choe, G Gerhart, T Meitz. Wavelet based ground vehicle recognition using acoustic signals. *Proc SPIE*, 1996, 434-445.
- [4] L Liu. Ground vehicle acoustic signal processing based on biological hearing models. Master Thesis, University of Maryland, 1999.
- [5] M P Desimio. Phoneme recognition with a model of binaural hearing. *IEEE Trans. Speech and Audio Processing*, 1996, 4(3): 157-166.
- [6] S A Shamma. Stereausis: binaural processing without neural delays. *J Acoust Soc Am*, 1989, 86(3): 989-1006.
- [7] X Yang, K Wang, S A Shamma. Auditory representations of acoustic signals. *IEEE Trans. Information Theory*, 1992, 38(2): 824-839.
- [8] G E S Young. The HTK Book (for HTK Version 3.2.1). Cambridge University Engineering Department, 2002.
- [9] T Kohonen. Improved versions of learning vector quantization. *Proc Int Joint Conf Neural Network*, 1990, 1: 545-550.

- [10] J S Baras, A L Vigna. Convergence of a neural network classifier. *Decision and Control, Proc of the 29th IEEE Conference on*, 1990, 3: 1735 - 1740.
- [11] S Haykin. *Neural Networks: A Comprehensive Foundation (2nd Edition)*. Prentice Hall, 2001.
- [12] R M Gray. Vector quantization. *IEEE ASSP Magazine* 1984, 1(1): 4 - 29.
- [13] D Li, K D Wong, Y H Hu, A M Sayeed. Detection, classification, and tracking of targets. *IEEE Signal Processing Magazine*, 2002, 19(2): 17 - 29.

## 中国声学学会六届三次常务理事扩大会议召开

中国声学学会六届三次常务理事扩大会议于 2008 年 7 月 12 - 15 日在张家界召开,参加会议的常务理事、常务理事代表、各分会和地方学会的代表 32 人。名誉理事长、中国科学院院士张仁和研究员应邀出席了会议。

会议由副理事长杨德森教授主持,首先由副理事长宗健研究员向代表传达了 2008 年学会工作会议精神,并传达了 2008 年中国科协所属全国学会秘书长会议上的讲话、在中国科协第三届学术交流理论研讨会上的报告精神。代表们对冯长根书记的讲话进行了热烈的讨论,并结合声学学会的学术活动,如何提高学术论文质量,办好学会期刊发表了积极的建议。

理事长田静研究员向会议代表阐述设立马大猷声学奖的由来,我国物理学家、声学家马大猷先生在我国物理学、声学事业和创建中国声学学会做出了杰出贡献,为发扬光大马大猷先生的治学精神,也为促进声学事业的发展,特设立马大猷声学奖和中国声学学会青年奖。此奖项由北京声望声电技术有限公司赞助。中国声学学会成立马大猷声学奖和中国声学学会青年奖评选委员会。

参会代表各抒己见,踊跃发表意见,对设立马大猷声学奖和中国声学学会青年奖表示赞同。会议决定:马大猷声学奖每两年评选一次,申报人年龄限在 60 岁以下,名额 1 - 2 人,在当年的学术年会上颁奖;每两年评选一次中国声学学会青年奖,申报人年龄限在 35 岁以下,名额 5 - 10 人,在当年的青年学术会议上颁奖。

六届三次常务理事会议还决定,第六届理事会第三次理事会议将于全国学术年会召开的前一天同期举行。

会上田静理事长向各位代表通报了第 37 届国际噪声控制工程大会暨展览会筹备情况,会议将在上海国际会展中心召开,日期是:2008 年 10 月 26 - 30 日。

学会办公室罗燕主任向各位代表通报了中国声学学会 2008 年全国声学学术会议的筹备情况以及论文的收录情况。2008 年全国声学学术会议自 2008 年

1 月起征文,截止至 6 月 5 日,共收到论文(含综述) 330 篇,经审稿后录用 321 篇。其中,综述报告 3 篇;物理声学 15 篇;水声学 55 篇;超声学、量子声学、物理效应 55 篇;声波测井 32 篇;噪声、噪声效应及其控制 16 篇;结构与建筑声学 8 篇;语言声学、语音通讯 12 篇;生理与心理声学 14 篇;气动声学、大气声学 1 篇;声学测量、信号处理与分析的方法、仪器 80 篇;声学换能器 30 篇。本次会议的主题为《创新与和谐——新世纪的声学》。

此次会议是由上海市声学学会负责协办,在上海市声学学会的积极努力下,会议的筹备工作已基本就绪。会议日期是:2008 年 10 月 20 - 25 日,会议地点是:上海物贸大厦。

张仁和院士向各位代表通报了 2008 年 ICA 会议情况以及 ICA 的组成情况,目前我国仍是该组织的常任委员。

2009 年将在我国召开的部分国际会议:

2009 年 4 月 13 - 17 日在上海召开第二届国际浅海声学会议;2009 年 9 月 21 - 23 日在北京召开第十届西太平洋地区声学会议;2009 年 9 月 23 - 25 日在西安召开环太平洋声学会议。

参会代表对学会会员发展工作提出了很多意见和建议,一致认为要充分利用网络,做到网上发展会员,并把现有的会员名单发布到网上,给会员优惠。同时,也要把团体会员交费情况在网上公示。要及时更新学会动态,还要调动大家写一些科普文章并挂在网上,让更多的人来了解声学。到会的分会代表也作了工作介绍,参会的学术刊物负责人对所办期刊情况作了汇报。各地方学会相互介绍了各自的工作动态、办会经验以及学术交流的经验体会。地方学会的学术活动非常活跃,他们由一个地方学会拓展到三至四个地方学会联合、地方学会与分会联合,共同召开学术会议;对联合办会、学科交叉的做法是值得提倡的。

(中国声学学会办公室)