

非语音声音在人机交互技术中的应用

The Applications of Non-speech Audio in Human-computer Interaction Technology

方志刚 吴晓波 徐新民 金文光

(浙江大学 CAD/CG 国家重点实验室 杭州 310027) (杭州大学电子工程系 杭州 310028)

摘要 Presenting information in non-speech audio could take advantages of the superiority of human auditory sensory modality. This paper discusses the user listening model, the applications of non-speech in human-computer interaction technology, and some audio display technology.

关键词 Non-speech audio, Human-computer interaction, Visualization, Scientific computing

1 引言

人机交互技术已经历了字符用户界面(CUI),图形用户界面(GUI)和多媒体用户界面阶段,目前正朝着三维,多通道用户界面以及虚拟现实(VR)技术的方向发展。从人的特点考虑,人机交互技术发展的几个阶段可分别归结为:(1)符号阶段,用户面对的只有单一文本符号,虽然有视觉参与,但视觉是非本质的,本质的东西只有符号和概念;(2)视觉阶段,借助计算机图形学技术使人机交互大量利用色彩,形状等视觉信息;(3)听觉阶段,早期计算机系统只有单调的蜂鸣声,虽然多媒体技术将音频及视频形式带进计算机,但仍缺少听觉交互手段,即人处于被动收听状态,声音缺少位置、方向等变化。上述每个阶段的早期阶段所用的交互手段在后续的阶段中仍然存在。人们在日常交往和信息传递过程中大量运

用视觉信息、听觉信息、语言和概念符号,人学会运用视觉听觉可能不分先后,但语言及概念符号却一定是后天获得的。显然人机交互的历史基本上与人认识世界的过程相反,这可归结为以下原因:(1)计算机作为计算工具具有高度抽象性;(2)技术限制;(3)更重要的是声音通讯的抽象性和短暂性。当前,在人机交互中结合进视觉的和听觉的以及更多的通道是必然趋势,特别是将听觉通道作为补充的或替换的信息通道已显示出重要性和优越性^[1]

研究表明,听觉通道有许多优越性,如:(1)听觉信号检测快于视觉信号检测的速度;(2)人对声音随时间的变化极其敏感;(3)声音信号所具有的全向特性可作为引导视觉对目标进行细调分析的粗调机制,即所谓“听觉是视觉的眼睛”;(4)听觉信息与视觉信息同时提供可使人获得更强烈的存在感和真实感;等等。

- [2] Z. Yang et al., Study on the Method and System Implementation of An Isolated Word Recognition System Used in Noise Environment, Chinese Journal of Acoustics(English Edition), 15(2) 1996
- [3] J. A. Nolasco Flores, S. J. Yong, Continuous Speech Recognition in Noise Using Spectral Subtraction and HMM Adaptation, ICASSP, 1994
- [4] F. Liu et al., Environment Normalization for Robust Speech Recognition Using Direct Cepstral Comparison, Same to [3]
- [5] D. Naik, Pole-Filtered Cepstral Mean Subtraction, ICASSP, 1995
- [6] H. Hermansky et al., RASTA -PLP Speech Analysis Technique, ICASSP, 1992
- [7] M. Rahimi, B. Juang, Signal Bias Removal for Robust Telephone Based Speech Recognition in Adverse Environments, Same to [3]
- [8] S. Dvorak, T. Hornum, High Performance Speech Recognition in Noise by Continuously Updated Reference Templates, Eurospeech, 1992
- [9] K. Takagi et al., Rapid Environment Adaptation for Robust Speech Recognition, ICASSP, 1995
- [10] A. P. Varga, R. K. Moore, Hidden Markov Model Decomposition of Speech and Noise, ICASSP, 1990

国外对听觉通道应用于人机交互技术的研究工作至少开始于八十年代,研究者在科学计算可视化、人机接口及虚拟现实技术等方面开展了许多有意义的工作。本文总结国外有关研究并结合我们初步的研究工作,对如下主题进行探讨:(1)非语音声音信号的声学特征和用户听觉模型;(2)人机交互技术中的听觉通道扩展;(3)听觉显示原理,技术和系统实现。

2 非语音声音的声学特征和用户听觉模型

声音有语音和非语音之分。语音是人类漫长的进化历史的产物,其发生和识别机制非常复杂,目前人们还不十分了解。就我们的研究主题而言,语音是语言的物化,就其本质而言属于概念符号的范畴,即语音所携带的信息本质是抽象的而非形象的。本文所探讨的非语音声音则集中于形象的听觉信息表达方面,下文所述声音均指非语音声音。

2.1 声音的声学特征

人能够根据声音信号的许多声学特征来区别不同的声信号,从而获得所携带的信息。用声音表达信息的有效性取决于人对声音特征的感受性,感受性越强则检测信号越容易。人们常用如下的声学特征表达信息:幅度;频率;谐波成分;频率调制;增强(衰减)速率;空间位置,人的双耳定位能力可确定声源的位置(方向和距离)。

2.2 听觉的音乐模型^[2]

声音的声学特征(即物理特征)并不等同于人的听觉特征,物理特征有显著差异的两个声音信号不一定会引起显著的听觉差异,而且听觉特征量与物理特征量之间一般没有线性关系,而是符合韦伯定律或对数律。几个世纪以来,作曲家们尽力解决我们现在所面临的同样一些问题,即如何把一些单一的、独立的旋律组合成一个可听的整体,使得其中的各个部分以及整体都同样可辨别并且有效。音乐家们为我们回答了如下问题:首先,听觉系统对声音的感知能力如何?其次,如何使同时呈现的多个旋律可被听众感知?我们现今的一些常识来源于音乐及音乐理论,如:(1)低音会产生滞重感,通常变换缓慢(否则会造成混浊和不清晰),而高音则产生轻盈感,快速变化不会影响可听度;(2)由长笛奏出的缺少泛音的单元(正弦波)显得清冷清晰,而由双簧管奏出的泛音丰富的声音(锯齿波)则显得热烈;等等。第二个

问题在复调音乐中已获得很好的解决。复调音乐由和声与旋律组成,其中同时出现两个或多个同等重要的旋律。两种主要类型的复调音乐是:(1)幕声复调音乐,其中各个声部之间精确地或不精确地幕声(如二重轮唱);(2)非幕声复调音乐,其中各个声部与主旋律完全无关。

2.3 用户听觉模型¹

声音信号的构成必须定义一组声学特征(如音色,幅度,位置),究竟哪一类声学特征能使信号检测能力最强呢?这必须以一般的用户听觉模型和特殊的用户听觉特征为依据。听觉的音乐模型反应了人的一般听觉特征,而特殊用户听觉特征需要自适应听觉模型的支持。用户听觉模型可在音乐的听觉模型的基础上建立,这样做的好处是:(1)可以利用音乐理论的成果;(2)使用户不经训练或少训练即可适应目标系统。建立用户听觉模型的方法是:

(1)为用户放送位于用户听力空间 ULS(User Listening Space)的声音信号,不同的信号由相异的声学特征组合得到;

(2)用户对听到的声音信号作出反应:①强存在,②弱存在,③平均存在;

(3)对收集的用户数据进行统计处理,导出声学特征值;

(4)根据这些声学特征值和音乐的听觉模型综合得到目标用户的用户听觉模型。

用户听觉模型以知识库的形式存储于计算机系统中,作为以后组合产生声音信号的依据。

3 非语音声音在人机交互技术中的应用

声音可用于表示离散的和连续的信号,在一些简单的应用场合人们已广泛用声音表示离散信号,而在连续信号表示方面仍处于探索阶段,下文侧重后者介绍几种典型应用。

3.1 科学计算可视化^[4-6]

八十年代中后期科学计算可视化成为计算机科学最活跃的研究领域之一,科学计算可视化给科学研究和工程设计带来巨大的生产率。近十年中,科学计算可视化技术获得很大发展,人们已越来越不能满足于单纯的“可视化”,而希望利用声音来增强甚至代替科学计算可视化,这是因为听觉通道有许多视觉通道所不及的优点,如人对声音的不随意注意,可避免重要信息的遗漏。科学计算可视化和可听化

的根本目标是追求人对数据特征本能的和自发的感知过程。几种典型的应用研究包括:

(1)Yeung 利用声音对矿石样品分类,这些矿石分别来自加利福尼亚的 4 个矿区,分类特征数据涉及 7 维化学特征数据,分别映射到不同的声学特征。实验表明,被试不经训练就能达到 90% 的分类精度,而经训练则可达 98% 以上。

(2)Lunney 将红外频谱峰值映射为相应声音频率的幅度,以便让被试对 10 多种有机化合物作模式匹配,实验表明被试识别准确率接近 100%。该实验很有趣的是,由红外频谱映射而成的声音信号中存在“旋律”与“和声”,非常类似可演奏的音乐。

(3)Bly 尝试用声音来表达多维数据,试验将 6 维数据样本的 6 个变量映射为各种声音特征,另将其中 2 个变量映射为 X-Y 平面图形显示。实验表明,被试单独根据视觉信息标识样本正确率为 62%,单独根据听觉信息标识样本时正确率为 64.5%,组合两者则为 69%。

3.2 虚拟现实技术

虚拟现实技术是对人机交互新范型的探索,它追求的目标是使计算机生成的虚拟环境“看上去逼真,动作逼真,听起来逼真,感觉逼真”,这要求系统允许用户使用多种感官协作工作。虚拟现实技术的典型应用有:卫星电视会议、共享的电子工作空间、远程或危险环境下机器人监控或遥控、娱乐环境、驾驶舱模拟器、化合物虚拟模型等等。

在虚拟环境中不仅仅强调声音的使用,更强调真实感声音的作用^[7]。在这里,用户可与虚拟声音“交互”,声音有方向感和距离感,用户在虚拟环境中的不同位置会引起声音变化。Lowell 大学开发的可视化工作站^[8]向用户呈现一个矩形窗口和一个同心的矩形声频显示窗口。可视的和可听的显示空间可扩展到物理显示窗口之外。系统允许用户作可视的和可听的快速缩放,当用户在一个显示区域内快速移近时,该区域的图形显示看上去变得较大,而且相应的声音听上去也更真切。

3.3 交互式人机对话

在交互式人机对话设计中,人们不仅关心人从计算机获得信息的可能性,而且关心获得信息的效率和自然性。因此,在交互技术中除利用可视的图形文字外,还有必要用声音作为附加的或可替换的反馈通道,这在多媒体系统中已得到发挥和证明。声音

在人机接口中除传统地用于告警,还用于增强人对信号的感知,特别在为弱视者设计人机接口时,需考虑用声音可作为替换的通道^[9]。

4 听觉显示技术和系统

4.1 声音信号设计与布局

非语音信号可用于表示离散事件(如报警)和表示数据,两种情况下都要借助声音的不同维度的值的差异来区分不同信号。令声音信号 $S=(A, f, F, m, P)$, 其中, A 为幅度, f 为频率, F 为音品, m 为调制度, P 为空间位置(二维或三维)。理论上,一组不同声学特征可表示一个唯一的信号,但实际应用中存在几个问题:(1)信号可辨性;(2)用户对信号的解释;(3)信号优化设计。信号可辨性要求声信号的物理特征要有足够的差别以引起人的感觉差异;用户解释会受到日常经验等因素影响,因此在设计信号时必须以用户听觉模型为依据;信号优化设计是在保证信号可辨性和用户正确解释的前提下,使信号在用户听力空间(ULS)的优化分布。自适应听力系统具有对 ULS 作自适应处理的能力,使多个甚至可能是并行的事件信号在 ULS 达到最优分布。

4.2 三维虚拟声音技术⁷

声信号的幅度、频率等维度通常可用简便的手段直接控制,而空间位置和方向却不易控制,事实上要复杂得多。在科学计算可听化和虚拟现实技术中必需采用三维真实感声音生成技术。一般而言,三维声音具有如下特征:(1)三维定位特性;(2)三维实时跟踪特性;(3)临境感;(4)交互特性。三维声音生成技术的主要理论依据是双耳感知模型,人的听觉的空间定位能力取决于声音到达双耳的时间差,声级差和双耳滤波效应。根据声音感知模型用计算机模拟而再现声音。

4.3 声音信号发生器

为了产生由不同声学特征构成的声音信号,需借助实时声音信号发生器。声音信号发生器的实现要综合考虑性能成本等因素。

4.3.1 Convolvotron^[10]。是美国 Crystal River 工学院的 Scott Foster 设计的实时数字信号处理器,由两块卡组成,用于 IBM-PC 或 PC-AT 机。Convolvotron 使用外耳与方向有关的滤波特性合成空间的虚拟三维声音,可接近实时地将 1 至 4 个声源置于三维空间。它还提供磁性头部跟踪系统,因而可实现交互式声音显示,即声显示与用户头部运动保持一致的变化,这种声音的头部耦合方式有利于加

强声显示的真实感。Convolutron 在美国政府、大学及工业研究实验室等部门得到应用,包括 NASA、VPL、华盛顿大学人机接口实验室等,但价格相当昂贵。

4.3.2 Gravis Ultrasound. 是一种市售的三维声音卡,价格可以接受且与声霸卡兼容。Gravis Ultrasound 允许用户将普通声音文件转换为三维空间中任何虚拟位置的声音文件,其软件开发工具箱还提供声音记录、三维声音转换、测试、输出等功能,用户对它进行适当编程可实现简单的三维声音定位,建立虚拟声源。

4.3.3 MIDI. MIDI 设备现在已在声霸卡这类产品中非常廉价地提供,但 MIDI 主要目的是满足流行音乐市场的需要,并不十分适合科学研究工作,而且要充分挖掘 MIDI 的声音资源(包括生成虚拟三维声音)需要相当复杂的软件算法并需要系统独占。

结束语 非语音声音在人机交互技术中有着广泛的应用前景,相比语音更具直观性和直接性,特别在科学计算可听化方面有其突出的优点,与科学计算可视化相结合而形成科学计算视听化则一定具有单个通道所不具备的优势。我们目前正进行在医学图像处理技术中引入非语音声音表示数据的研究工作。

参考文献

- [1] B. H. Deatherage, Auditory and Other Sensory Forms of Information Presentation, Human Engineering Guide to Equipment Design, Washington DC U. S. Govern-

ment Printing Office, 1972

- [2] M. M. Blatner, et al., Communicating and Learning through Non-Speech Audio, Multimedia Interface Design in Education, Springer Verlag, New York, NY, NATO, Series F., 1992
- [3] R. D. Riecken, Auditory Adaptation, In: Proc. of IEEE Intl. Conf. on Systems, Man and Cybernetics, 1991
- [4] E. S. Yeung, Pattern Recognition by Audio Representation of Multivariate Analytical Data, Analytical Chemistry, 52(7), 1980
- [5] D. Lunney and R. C. Morrison, High Technology Laboratory Aids for Visually Handicapped Chemistry Students, J. of Chemical Education, 58(3), 1981
- [6] S. A. Bly, Presenting Information in Sound, In: Proc. of the CHI '82 Conf. on Human Factors in Computer Systems, 1982
- [7] E. M. Wenzel, et al., A System for Three-dimensional Acoustic Visualization in a Virtual Environment Workstation, In: Proc. of the IEEE Visualization'90 Conf., 1990
- [8] G. Grinstein, et al., EXVIS, An Exploratory Visualization Environment, Graphics Interface '89, London, 1989
- [9] A. D. N. Edwards, Soundtrack, An Auditory Interface for Blind Users, Human Computer Interaction, 4, 1989
- [10] S. H. Foster and E. M. Wenzel, Virtual Acoustic Environments, The Convolutron, SIGGRAPH '91, 18th ACM Conf. on Computer Graphics and Interactive Techniques, 1991

(上接第 69 页)

利用浏览工具^[4],比如 Mosaic CGI 和 Netscape 的 Plug-In-API 建立特定的浏览器,但是这样做限制了客户程序的多平台性。HTTP 代理方式^[5],当 Web 客户需要检索文档时,它不是直接与有关的 URL 连接,而是连接到某个 HTTP 代理或者称为协调程序(Mediator),由它从服务器取回数据返回给 Web 客户,这样做的好处是提供更多的丰富的服务,同时可以借助于防火墙对系统的数据进行保护,可以实现数据库系统应用程序与 WWW 的集成。在数据库系统与 WWW 的集成中,用户身份的确认、信息的保密、请求的合法性与操作的安全性的限制也是需要重新研究的问题。一些新的 WWW 服务器提供了将代码直接连接到服务器的 API,可以将通用的 WWW 服务器改造成代理服务器。面向 Java 的类库和快速开发工具还比较少,因此 Java 也还远非成

熟。然而它的对象特色和 WWW 上的广泛的适用性,在 WWW 上建立开放的计算环境方面,利用 Java 的方法是最有希望的途径。

参考文献

- [1] Stephen E. Dossick, WWW Access to Legacy Client/Server Applications, Fifth Intl. World Wide Web Conf., May 1996
- [2] Berners-Lee T., The World-Wide Web, Comm. ACM, Aug. 1994
- [3] December John and Ginsburg M., HTML&CGI Unleashed, Sam. net Publishing(1995)
- [4] Steven P. Reiss, Connecting Tools, Using Message Passing In The Field Environment, IEEE Software, July 1990
- [5] Kappe F., Maintaining Link Consistency in Distributed Hyperwebs, Proc. of INET, 1995
- [6] P. D. Stotts, Dynamic Adaptation of Hyper Text Structure, Proc. of Hypertext Conf. (ACM), Dec. 1991