

物联网声呐感知研究综述



陈超¹ 赵春蕾¹ 张春祥¹ 罗辉²

1 天津理工大学计算机科学与工程学院 天津 300384

2 解放军信息工程大学电子技术学院 郑州 450002

(476052121@qq.com)

摘要 近年来,随着科技的快速发展,智能移动设备已经成为人们生活的一部分。智能移动设备的普及为实现声呐感知理论的应用做好了充分的物理支持。声呐信号在传播时受到传播空间和生命活动的调制,因此携带着丰富的生命状态和空间信息。智能手机的普及、通讯技术的成熟以及声学信号的创新型应用,使得声呐感知设备可以实现低成本细粒度的感知收集和计算。利用声学信号在声呐感知技术中无需额外特制硬件的支撑,以及其特有的隐蔽性和高精度等优点,通过提取声学信号的有效特征信息,就可对周边空间信息进行定位计算。文中详细阐述了声学信号在空间定位感知使用技术领域中的已有研究,并对主要技术的基本原理进行了总结,最后分析了声学信号在移动感知应用技术中存在的问题以及未来的发展趋势。

关键词: 声呐;超声波信号;声音阵列;解调;智能手机

中图法分类号 TP393

Review of IoT Sonar Perception

CHEN Chao¹, ZHAO Chun-lei¹, ZHANG Chun-xiang¹ and LUO Hui²

1 School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300384, China

2 School of Electronic Technology, PLA Information Engineering University, Zhengzhou 450002, China

Abstract In recent years, with the rapid development of technology, smart mobile devices have become a part of people's lives. The popularity of smart mobile devices provides sufficient physical support for the realization of sonar perception theory. When the sonar signal is propagated, it is modulated by the propagation space and life activities, so it carries a wealth of life state and space information. The popularity of smartphones, the maturity of communication technology, and the innovative use of acoustic signals have enabled sonar sensing devices to achieve low-cost, fine-grained sensing collection and calculation. Utilizing acoustic signals in sonar sensing technology does not require the support of special hardware. With the unique concealment and its typical feature of high accuracy, the acoustic signals can calculate the surrounding space information. This article elaborates on the research history of acoustic signals in the field of spatial positioning and sensing technology, summarizes the basic principles of the main technologies, and finally analyzes the problems and future development trends of acoustic signals in mobile sensing application technology.

Keywords Sonar, Ultrasonic signal, Voice array, Demodulation, Smart phone

1 引言

传统声呐感知利用声波信号在水下传播的特性,来实现水下测距、测绘等目的。过去,声呐设备在水下探测领域中发挥了巨大的作用,同时也证明了声波探测的可行性。在物联网高速发展的今天,由于智能手机的普及以及其内置传感器功能的愈发强大,声呐感知在理论和实际应用上具备更强的普适性、细粒度性和隐蔽性等优点,客观上为物联网声呐感知的实现提供了更多的可能,也吸引了更多的学者去关注和研究物联网声呐感知。

物联网声呐感知是一种利用传感器发射声波信号去探测生命体征和空间信息数据的方式,是物联网的一种表现形式。目前,基于复杂场景的多样性和不确定性,声呐感知设备正向微小化、细粒度化发展。物联网声呐感知的发展经历了从传统声呐到声音阵列感知再到基于手机的音频近场感知。基于智能手机的声呐实现原理是借助已有智能手机的内置传感器,通过麦克风抓捕音频数据的方式来进行数据传输。在感知手段中,常见的有利用手机内置加速度计^[1-2]、摄像头^[3-5]、WIFI热点^[6-7]等来实现感知目的。这些方案皆具有较高的可行性,但仍有不足,如对外界环境有较高的需求。运动传感器

收稿日期:2020-03-23 返修日期:2020-07-29 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:天津市自然科学基金(18JCQNJC69900)

This work was supported by the Tianjin Natural Science Youth Fund (18JCQNJC69900).

通信作者:赵春蕾(zeltjut@gmail.com)

需要提取手机传感器参数,辅以带标记的数据进行训练,受限于手机空间状态,数据不够精准;摄像头的感知方式需要在用户附近进行拍摄,分析用户手势等行为,检测精度低;WIFI热点技术则是通过分析 WIFCSI 数据来推测用户按键,优点是具有较高的推测精度,缺点是设备周围不能存在其他干扰,对环境的要求极为苛刻。

声学信号技术的应用给感知领域带来了新的思路。由于其隐蔽性^[8]、普适性和细粒度探测等优点,被广泛应用于空间感知等技术。1906年第一台声呐设备的诞生标志着声学技术应用正式登上历史的舞台。该技术一被提出就被广泛应用于军事领域,借助声音在水中良好的传播特性,侦测敌对舰队位置并进行打击。此后,声呐设备逐渐被广泛地应用于海洋测绘、医学B超诊疗、工业产品质检等方面。其广泛的使用范围在很大程度上帮助了人们更好地了解世界,解决自身问题。但是随着时间的推移,单一的大型声呐探测设备已经无法满足这个多元化的社会,声源感知技术在复杂的应用场景下,正向具有移动性、便携性和普适性等方向发展。经研究发现,单一的音频传感器经过特殊的排列组合可以组成性能更加强大的声音阵列来感知空间目标,随着智能移动设备的爆炸式普及以及声学技术的发展,其为声源定位感知技术带来了新的解决思路。研究人员借助移动智能终端内置的扬声器、麦克风等传感器,将智能终端设计成便携式声呐类型设备。通过提取移动智能终端采集到的声学信号(这些信号受空间物体的调制),可以进行空间定位,使声源感知步入一个新的高度和方向。图1给出了物联网声呐的历史发展轨迹。

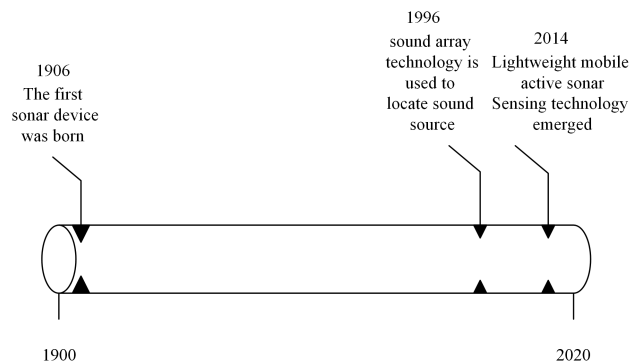


图1 物联网声呐发展过程

Fig. 1 Development process of IoT sonar

目前,物联网声呐类感知技术的研究主要分为3个方向,分别是传统声呐、声音传感器阵列和基于移动设备的音频近场感知。本文第2节依次从这3个方向介绍了物联网声呐类感知技术,首先介绍了传统声呐的知识;然后介绍了基于声音传感器阵列空间感知方向的技术;最后着重介绍了轻量级主动式手机声呐感知技术;第3节从不同方向对研究方法进行了比较;第4节总结了研究的方向以及发展前景;最后总结全文。

2 基于声学技术的相关内容

2.1 声呐的原理

声呐设备的出现正式开启了声学探测的科研道路,在生活中,人们无处不在使用声源感知的天赋能力,例如我们时刻

都在借助双耳效应来判断音源与我们的距离。目前,大多数人认为双耳效应的产生依赖于双耳的时间差和声级差。由于“人头遮蔽效应”的影响以及声波频率的不一致,使得影响双耳效应的因素不一而足。简单来说,声波在中低频率时是受时间差和声级差的共同影响,在频率高于4kHz时,主要受双耳声级差的影响。在自然界中,音源感知并非人类的专利,很多生物都拥有更高级的“声呐”感知系统,例如蝙蝠可以利用喉头发出的超声波感知猎物和空间的障碍物,飞蛾会听到40m外的蝙蝠发出的超声波,以此来躲避攻击。海豚等海洋生物则利用超声波探寻食物,甚至还能借此完成通讯。而声呐探测仪的出现,使得我们拥有了超越动物的音频感知能力,这是声学技术应用史上的一个巨大飞跃。声呐探测设备最早被广泛应用于水下探测领域,我们知道雷达是利用电磁波发现并测量目标的空间位置,但是在水中,由于趋肤效应的存在,导致电磁波在传播时会趋于导体的表层,使其在水中的衰减速率极高,无法感知目标的来源。而声波信号不同于电磁波,从本质来看,声波信号是一种机械波,其传播方式是依靠介质质点之间的相互推动来传播的。事实上,在深海声道中引爆一个炸弹,20000km外也可以收到信号,低频声波甚至还可以穿透海底几千米的地层,并得到地层中的信息。因此,利用声波制成的声呐类型设备成为了水下探测中最为重要的手段。

声呐是一种声学探测设备,不同的应用场景需要有不同的类型的声呐。按照工作原理的不同,声呐设备可以分为被动声呐和主动声呐。1906年由刘易斯·尼克森发明的第一代声呐仪就是被动声呐的一种。被动声呐指在工作中,自身不发射音频信号,单纯被动地接收目标物体所产生的声波信息,从而测定目标的具体方位和距离,就像飞蛾感应超声波来躲避蝙蝠一样,其优点是可以非常隐蔽安全地感知目标物体所在的位置而不被对方所知。主动声呐技术则恰恰相反,其依靠主动发射声波来感知目标,通过目标障碍物的回波信号与发射信号的时延来判断目标的距离和方位等要素,从而实现感知功能。声呐工作示意图如图2所示。

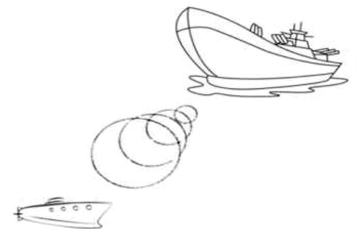


图2 声呐工作示意图

Fig. 2 Sonar working diagram

传统意义的声呐通常由基阵、电子机柜和辅助设施组成,其中基阵是由各种排列组合的水声换能器构成,电子机柜包含发射、接收、显示控制等系统,辅助设备则指电源、增音机等设备。它们的组成往往庞大且复杂,主要被应用于各类水下作业,在军事、海洋测绘、渔业和水声通信等领域发挥着重要作用。

2.2 声音传感器阵列

过去,声呐设备被广泛应用于水声领域,利用声波对水下

目标进行感知探测和定位等。近年来,中国的科研水平取得了很大的进步,大量的声呐类型设备进入到了人们的生活中。传统声呐虽然足够强劲有效,但其设备组成往往非常复杂,设备体积庞大拥挤,无论是人机交互的体验还是群体布置的场合需求,在复杂多样的情境下传统声呐已无法满足新一代的社会需求。为了满足这些需求并解决问题,声音传感器阵列这一科研思路被提出。

声音传感器阵列利用的是一组声音传感器,并以某种几何图案部署,共同收集和处理声波信号,通过判断不同声音传感器声信号之间的相位关系来判断声源目标位置。相比单独传感器的使用,声音传感器阵列的部署优点更明显。该方法为探测增加了新的维度,多维角度的解析使得阵列具有更强的空间选择性,甚至实现声源单位的自适应检测、定位以及追踪等,在军事国防、人机交互、市内车辆鸣笛定位和声音监控等领域有着卓越的科研价值。根据麦克风阵列定位原理的不同,可将其分为3类:1)可控波束形成定位法;2)高分辨率谱估计定位法;3)时延估计定位法。

2.2.1 基于可控波束形成技术

利用传感器阵列获取音频时,可控波束形成的技术是比较成熟的,其中心思想围绕的是数学的加权求和,是一种最大似然估计的过程。该方法基于声音到达各个麦克风的相位延迟,给每个麦克风阵元赋予不同的权值,当数值达到最大时,则认为形成的波束所指向的位置就是声源的位置,其原理示意图如图3所示。Buckley等^[9]开创性地提出了一种新的宽带源空间谱定位方法 BASS-ALE,他们采用宽带时空协方差矩阵和宽带源模型的特征结构,借助识别宽带源的时空观测的低特征值,证明了噪声子空间的存在。文献^[10]提出了一种基于向量长线性阵列的改进版 RCB算法,该算法对数据矩阵进行奇异值分解,从而避免了不完全秩的协方差矩阵的出现,减小了计算量和存储量,减轻了运算的负担,成功解决了各种转向矢量误差引起的性能降低。

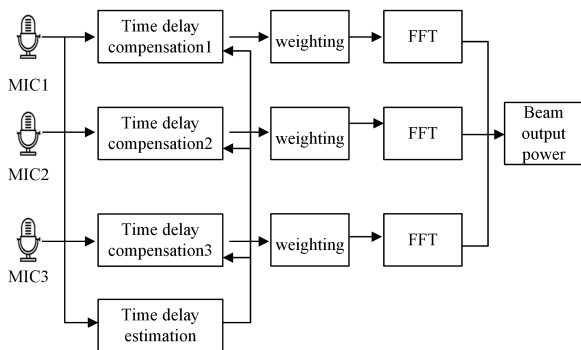


图3 波束形成示意图

Fig.3 Schematic diagram of beamforming

在音源定位的研究中,除噪是一个很重要的课题,文献^[11]设计了一种恒方向性波束形成器 LCA-CDB,该系统可以自适应地将暂态噪声降至最低,使波束在幅度和角度上均可得到控制,从而提高去除噪声的能力。文献^[12]提出了一种基于压缩波束形成的被动合成孔径方法,用于音频的定位。在低频声源定位领域中,小孔径声音阵列常常因为孔径不够大而影响空间的分辨率。该系统利用一个固定的传感器来校

正移动阵列的相位,从而实现低频信号的感知定位。

2.2.2 基于高分辨率谱估计技术

相比可控波束形成技术,高分辨率谱估计技术可以对多个声源目标进行定位跟踪,且不会受到麦克风阵列的波束宽度的限制。高分辨率谱估计技术包含 AR 模型法、最小方差估计法(MVE)和特征值分解法等,其利用的是阵列各个阵元所接受到的信号矩阵谱,通过计算相关矩阵来判断目标生源的位置。

Collobert等^[13]利用图像识别和麦克风阵列技术,通过视觉和听觉来共同实现人脸的追踪识别。该麦克风阵列由9个按哈蒙嵌套子阵列原则排列的单向麦克风组成,为了保证阵列在宽频带上的均匀性,每个子阵列都有固定的空间安排,并采用线性相位滤波器进行过滤。

Gustafsson等^[14]针对房间混响环境中的声源定位,基于多径效应以及诸多杂音干扰的情况,做出了房间内函数统计模型,并在此基础上统计分析了源局部化方法的特性。

2.2.3 基于时延估计的定位技术

基于时延的估计定位方法是应用得最多的一种方法,也是运算开销最小的方法,相比前两种算法,该技术的最大优势在于实时性强。时延定位技术首先借助麦克风阵列收集音频信号,并对其进行相关的音频处理,求得各个接受阵元收集到的时间差,从而求出时延值;然后根据计算的时间差,通过声速方程及阵列的空间排布来计算声源位置。根据这套理论,可以在二维和三维空间中进行定位计算。

2014年,Duan等^[15]研究了一种改进版的互功率谱相位加权算法,借助不同麦克风阵列的拓扑结构,该系统在强噪环境中拥有很强的定位性能。Liu等^[16]在对比传统时延的估计算法的基础上,提出了改进的GCC时延估计算法,最后利用球形插值法进行定位。该系统拥有更好的抗高噪声以及抗高混响的功能,且时延精度更加细腻。

2.3 基于手机音频信号的近场感知

基于声音传感器阵列的定位技术可以解决很多静态的空间感知问题,相比声呐设备,声音传感器阵列的设备规模已经足够小巧,但其设备阵元的排布规则导致其大部分依然不适合随身携带。从是否方便携带使用的角度来看,声音传感器阵列是一种静态感知的方法,而且众多传感器阵元的使用增加了大范围部署的难度和成本,对空间和成本需求提出了极大的挑战。在移动近场感知的领域中,声音传感器阵列稍显力不从心。幸运的是,近年来物联网的快速发展使得智能手机拥有了愈来愈多的功能,例如移动支付、影音娱乐和通信交友等。可以说,手机已经影响到人们生活的各方面,而且随着手机装备的传感器性能愈发强大,计算能力的增强给移动近场感知带来了新的可能。

基于手机音频近场感知的技术有诸多好处,包括部署成本低,无需额外的硬件设备支撑,其单纯利用手机内置的麦克风收录环境周围中的声音,在排除多径效应以及环境中存在的干扰音之外,加以数学分析就可以实现动态近场感知。此外,现在主流的手机系统基本上都是免费开源的。基于此,我们可以调用智能手机的内置传感器来部署近场感知的系统。目前,已经出现了部分基于智能手机声音信号的近场感知系

统。与声呐设备的区分方法类似,根据是否需要自身发射探测声波信号,我们将这类手机应用划分为被动感知系统和主动感知系统。图4展示了两者感知目标方式的区别。主动感知系统则借助设备的扬声器发出经过调制的音频信号,经目标物体反射后被自身麦克风捕获,如图4(a)所示;被动感知系统借助麦克风直接捕获环境音或者特定的设备音频信号,如图4(b)所示。文献[17]提出的 SoundWrite 就利用了手机内置的麦克风去捕获用户手指划过桌面的声音,辅以对时间频率特征的提取,用预先准备的模式分类定位用户手势笔画。其进阶版^[18] SoundWrite II 增加了阈值限定来进一步识别噪声中的有效原始信号,并采取了 MFCC 来增强提取稳定手势行为的能力。文献[19]中的 iPand 系统也是被动感知系统的一个代表,对于被麦克风捕捉到的信号,其利用带通滤波器对声音进行过滤,并利用短时傅里叶变化对手势行为特征进行提取。

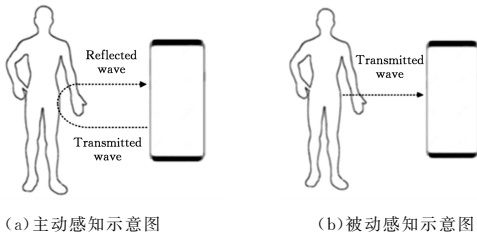


图4 声信号识别示意图

Fig. 4 Schematic diagram of acoustic signal recognition

相比主动感知系统,被动感知近场目标的声音极具专一性,而不具有普适性,因此只能识别目标物体接触特定的表面所发出的特殊声音。主动感知系统的特点是接收自身发出的特殊的超声波去感知目标的状态信息,这样的声波比较自然,因此更容易被识别。基于主动感知系统的优点及其技术相通之处的考量,接下来将介绍基于超声波信号的智能手机主动感知系统的概况,包括其基本框架以及技术原理,并对当前最新的科研成果系统做出总结。

2.3.1 系统框架

如图5所示,我们先对基于超声波的手机主动感知系统的工作场景进行还原。

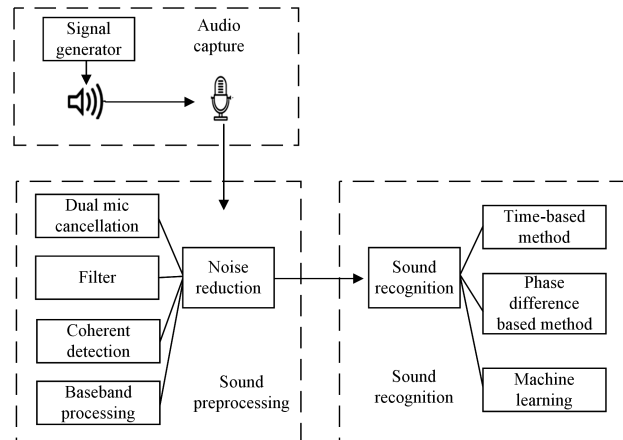


图5 声信号的识别技术流程简图

Fig. 5 Schematic diagram of acoustic signal recognition technology

首先,智能手机的扬声器发出一种人类听不见的超声波

信号,声波遇到目标物体经反射被手机的麦克风捕获,当目标移动时(如手),其会不断地改变声波的传播路径,捕获的音频信号经过除噪和傅里叶变化等手段后,就是我们需要的目标音频信号;再根据良好的数学模型和计算,可以不断推测出目标的位移等信息。这一系列流程皆不需要额外的硬件支持,智能手机本身可以很好地完成这些任务。接下来会围绕该技术流程详细介绍技术框架的每个过程以及特点。

2.3.1.1 声波信号选择

声音信号的发射和捕捉是实现手机音频主动感知系统的第一步,该原理是利用手机内置的扬声器播放特殊的音频信号,声波接触反射物后,由智能手机的麦克风对声波进行捕获,从而赋予手机声呐搜索的功能,使其能够对近场目标进行探测感知。因此,合理的声波设计是非常重要的。通常,主动式声呐设计采取的声波频率区间为 18~23 kHz,因为这个区间的声波信号刚好超过人类的听力范畴,不易被人们所察觉,不会影响人们的日常使用,同时还可以与环境中的绝大部分中低频杂声信号区分开,有利于音频的后期处理,并且对于这个区间的声波频率,基本上现有的手机都支持。随着通信技术的发展,提出的声波模型越来越多,根据使用目的的不同,我们将其分为生命体征检测和手势行为判断两大类进行介绍。

(1) 生命体征的检测

生命体征检测是当下的研究热点之一,因为大多医疗检测设备往往昂贵且不便携带,当人们有需求时,只能去相关医疗单位检查,这样非常不便且无法做到实时检测。移动设备的频繁使用、超声波信号的隐蔽性及高精度感知的特点,使得科研人员聚焦于这种非接触式高精度实时检测方案。众所周知,呼吸和心跳会引起胸腔轻微震动,通过分析声波信号就可以得到呼吸、心跳等生物信息。文献[20]提出了一种基于 FMCW 声波的非接触检测呼吸特征的系统——ApneaApp。FMCW 即调频连续波,它传输的是一个啁啾信号,该信号的频率会在预定的周期内呈线性增加,信号触碰到环境的反射器上,经过一段时间的延迟后返回声呐,这样可以通过比较接受和发射信号的频率来确定时延。如图6所示, f 为频率差,时延可由式(1)求得, k 代表线条的斜率,可以由式(2)得到, F_2 和 F_1 分别代表声波频率的上限和下限, T 是声波信号一个周期的时间,因此目标单位的位移如式(3)所示。

$$t = \frac{f}{k} \quad (1)$$

$$\frac{F_2 - F_1}{T} \quad (2)$$

$$s = \frac{vt}{2} \quad (3)$$

式(3)中, v 代表声音在空气中的传播速度。当多个物体位于发射单位不同的位置时,它们拥有不同的反射频移,因此 FMCW 能够很好地应对多径效应,其更是一种空间滤波技术。在时域中,FMCW 的使用公式为:

$$b = A \cos 2\pi \left(\frac{F_1 + F_2}{2} t + \frac{(F_2 - F_1)(t - NT)^2}{2T} \right) \quad (4)$$

其中, A 代表声波的振幅大小。

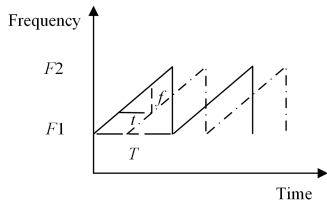


图6 fmcw 声波原理图

Fig. 6 Schematic diagram of FMCW sound wave

(2) 人机交互手势跟踪

手机的计算能力以及配置的传感器的能力越来越强,为手势追踪的实现带来了新的可能。现在的移动设备都是用手指触摸来操作系统,但是当手上沾有水或者屏幕上残留有污渍时,就会严重影响操作的准确性,同时,操作手机页面时手指在一定程度上也会对视线造成遮蔽,而声音不会受此影响,因此利用声波去感应手势行为成为了一种科学的解决方法。文献[21-23]利用声波的 TOA 和 TDOA 技术来测距,可惜精度只能达到厘米级。目前来看,大多数基于音频信号的无设备手势追踪系统都是利用手指反射声音的多普勒频移来测量,但是多普勒频移只能提供粗粒度的识别,如文献[24-26]所述。

Wang 等^[27]提出了无设备手指追踪系统 LLAP,在连续追踪中,利用相位法进行长度测量。在 LLAP 中,科研人员采用连续波(CW)来追踪识别手势的位移,同样地,文献[28]也采用了 CW 来破解手机模式锁,CW 的数学公式相比 FM-CW 更简单,其公式为:

$$A \sin(2\pi ft) \quad (5)$$

或

$$A \cos(2\pi ft) \quad (6)$$

其中, A 同样代表振幅, f 则是所选用声波的频率。一般情况下,声波频率的选择区间为 17~23 kHz。根据音频采集的规则,采集频率只要大于所采集音频信号本身频率的 2 倍,就可保证采集音频的质量。

Nandakumar 等^[29]设计的手势追踪系统则采用的是基于 OFDM 的脉冲信号,该系统的设计精度达到了 8 mm,当用户移动手指时,对应的回声到达的时间会随之发生变化。通过比较两种姿势的回声剖面,科研人员可以提取出与移动手指对应的回声。相比 LLAP 中所采用的 CW 信号,OFDM 信号具有较强的抗环境干扰的能力。OFDM 这项技术的原理是将带宽分割成正交的多个子载波形式,并且每个独立的子载波均可以传输不同的数据。它生成的数学公式为:

$$x_k = \sum_{n=0}^{N-1} X_n e^{2\pi j k n / N}, k=0, \dots, N-1 \quad (7)$$

其中, N 代表带宽被切割成 N 份,在 n^{th} 信道上传输数据位 X_n 。

文献[30]提到的手势感知系统 VSkin 没有选择连续声波,而是选择了 ZC 序列这种周期性的脉冲信号作为感知手势的声波信号。在 Vskin 系统中,利用了声音传播过程中所经过的 3 种路径,即结构路径、LOS 路径和反射路径,虽然短时脉冲信号可以很好地区分各类反射路径,但由于所含能量低,因此无法确切感知手势行为。而 ZC 序列是一种周期性

的高能信号,它依赖信号的自关联性来区分反射路径。ZC 序列在频域上的数学公式为:

$$ZC[n] = e^{-j \frac{2\pi u n (n+1+2q)}{N_{ZC}}} \quad (8)$$

其中, q 是一个整数常量, N_{ZC} 代表 ZC 序列的长度, u 是 0 和 N_{ZC} 之间的一个整数。

综上所述是几种比较先进的声波设计方式,都有各自的独特优势,例如 CW 可以非常清晰地定位声波变化之处,有助于提高信噪比等。上面介绍了声波设计的几种具体理论,为了更直观地展示声波的使用情况,表 1 列出了系统与相应的声波。

表 1 系统与其使用的声波

Table 1 Systems and sound waves it uses

系统名称	声波类型	声波频率/kHz	所使用的传感器
ApneaApp ^[20]	调频连续波	18~20	一个扬声器,两个麦克风
LLAP ^[27]	连续波	17~23	一个扬声器,两个麦克风
Vskin ^[30]	ZC 序列	17~23	一个扬声器,两个麦克风
FingerIO ^[29]	OFDM 信号	18~20	一个扬声器,两个麦克风
PatternListener ^[28]	连续波	18~20	一个扬声器,两个麦克风
ACG ^[31]	调频连续波	17~19	一个扬声器,两个麦克风
AudioGest ^[29]	正弦声波	19	一个扬声器,一个麦克风
EchoTrack ^[30]	线性调频信号	16~23	两个扬声器,一个麦克风
Dolphin ^[34]	Continues tone	21	一个扬声器,一个麦克风
SteerTrack ^[38]	正弦声波	20	一个扬声器,两个麦克风

2.3.1.2 信号预处理

声音预处理环节是整个声音信号处理过程中非常重要的一环,近场感知的重点就是音频信号是否纯粹。我们知道,在音频信号的采集过程中,所录入的音源包括周围环境的自然音以及由多径效应所产生的干扰音,而且所录制音频皆为时域信号,在时域中很难通过观察分析出生物特征。因此,如何消除干扰以及时域信号的转换就成为了这一部分的要点。下面会根据以上两个要点来详细介绍主流的处理方案。

(1) 消除噪声干扰

我们的目的是分析有用音频信号的特征,但是环境中充斥着各种各样的音频信号,且音频信号的传播是多方向性的,除了目标反射回的有用信号,还有多径效应以及扬声器向麦克风的强直功率传播产生的干扰。如何尽可能地去除这些干扰信号成为研究的一个热点。

1) 双麦降噪技术

双麦降噪技术的原理是基于现在智能手机上存在的两个麦克风来实现的,对于环境噪音以及扬声器的强直功率泄露,智能手机无法通过硬件调节来消除这些干扰,通过双麦降噪技术则可以很好地解决这些问题。

假设 A 为手机的主麦克风, B 为摄像头附近的辅助麦克风, A 主要录入扬声器所发出的声波信号 S_A , B 则负责收取背景环境音 S_B ,当然 B 也会顺带捕获扬声器发出的声波信号,但由于位置等因素,其收录的信号远远少于 A 捕获的有效声波信号,所得到的音频信号为:

$$S_M = S_A - S_B \quad (9)$$

通过这样的处理,可以有效地消除噪声的影响。

2) 滤波器的使用

滤波器是一种常见的信号处理方法,简单来说,它是一种频率选择装置,可以让信号中特定的频率成分通过,非特定阶

段的频率则会大幅衰减。滤波器的这种特性可被用来滤除环境噪声,从而更好地进行频谱分析。

按照允许通过的信号频率,滤波器有低通、高通、带通、带阻和全通滤波器 5 种。文献[20]在对呼吸信号进行采集之后,由于所用的音频信号为 18~20 kHz 的超声波信号,为了滤除自然环境中的中低频信号,直接采用了高通滤波器滤除中低频信号。而文献[31]为了获得更高的空间分辨率,将音频信号下传基带信号后,采用低通滤波器进一步过滤所捕获的音频信号。LLAP^[27]则是创新性地采用了 CIC 滤波器,CIC 滤波器是一种三段滤波器,而且其运算只涉及到加减法,没有其他的复杂运算,因此所耗费的运算开销极小。

3) 相干检测

相干检测技术其实是一种信号的解调机制,通常需要和滤波器配套使用,简单举例就是运用调制信号的载波和接收到的已调信号相乘,会得到一个低频信号以及镜像信号,借用低通滤波器将镜像信号去除,得到调制信号,这样的检测方式就是相干检测。这里,我们认为解调信号也是一种去除噪声干扰的手段,文献[28]运用相干检测技术去解调声波信号,与直接检测相比,相干检测技术可以获得更大的信噪比,且更适合应用于密集波分复用系统,但其也会增加系统的复杂性。

4) 向下采样

文献[27-28,31]都提到了下采样,基于超声波的手机近场感知所使用的声波信号都是 17 kHz 以上的高频信号,后续分析会相对复杂,向下采样过后,高频声信号会变成中低频信号,配合低通滤波器就能起到很好的除噪作用。为了更好地理解,我们以 LLAP^[27]中采用的声波 $A\cos(2\pi ft)$ 为例,被捕获的音频信号路径 p 的表达式如下:

$$R_p(t) = 2A_p' \cos(2\pi ft - 2\pi f d_p(t)/c - \theta_p) \quad (10)$$

其中, $2A_p'$ 代表回波的振幅, f 代表声波信号的频率, $d_p(t)$ 为时变产生的路径距离, c 则是声速,而 θ_p 是由硬件引起的延迟和反射引起的相位倒置造成的。获取这些信号后,让信号分别乘以 $\cos(2\pi ft)$ 和 $-\sin(2\pi ft)$,借助滤波器去除高频分量,就可分别得到 I/O 分量。

$$I = A_p' \cos(-2\pi f d_p(t)/c - \theta_p) \quad (11)$$

$$O = A_p' \sin(-2\pi f d_p(t)/c - \theta_p) \quad (12)$$

将这两种分量结合起来,可得到复信号的相位信息:

$$\varphi_p(t) = -\left(\frac{2\pi f d_p(t)}{c} + \theta_p\right) \quad (13)$$

上述为各个预处理过程中的节点详细介绍,为了更直观地展示这些预处理应用过程,我们以文献[28]为例来详细展示预处理过程。文献[28]采取了一样的处理办法,为了获取手指的移动信息,首先将录入的音频信号分成完全一样的两份,利用相干检测对音频信号进行解调,接着利用低通滤波器过滤信号,然后对信号进行下采样,同时去除静态分量,从而获得真实有效的音频信号。在这个处理过程中,扬声器所播放的音频信号被当作载波信号,与手指移动轨迹相关的信号被当作基带信号,因此录入的音频信号是载波信号和基带信号的组合。上述过程的数学公式为:

$$C = F_{ds}(F_{LP}(R(t) * A \sin 2\pi ft)) \quad (14)$$

$$O = F_{ds}(F_{LP}(R(t) * A \cos 2\pi ft)) \quad (15)$$

其中, F_{LP} 代表的是低通滤波器的使用, $R(t)$ 是麦克风所录入的音频信号, F_{ds} 是下采样函数, C/O 分别代表同向和正交。系统结构如图 7 所示。

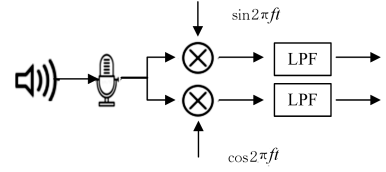


图 7 向下采样系统结构图

Fig. 7 Downsampling system structure diagram

(2) 时域和频域

从人们有时间意识开始,人们所看到的事物都是随着时间在悄然变化,人们也习惯以时间轴来衡量事物的进程,这种观察方法其实就是时域分析^[32]。频域分析则描述的是信号与频率的一个坐标系,它不是真实存在的,而是一种数学构造。时域和频域分析是两种对信号的分析方式,我们捕获音频信号抓到的就是时域信号,但是时域信号是以时间轴来变换的,很难分析出其中的信号特征,但是转换成频域信号后,就可以找到某一行行为所对应的频率特征,便于我们对信号的分析。

傅里叶变换就是我们实现时域信号向频域信号转变的一种重要手段。傅里叶原理表明:任何连续测量的时域信号都可以表示为不同频率的正弦波信号叠加。因此,傅里叶变换算法就是利用原始的时域信号,以累加方式来计算信号中不同正弦波信号的频率振幅等信息。

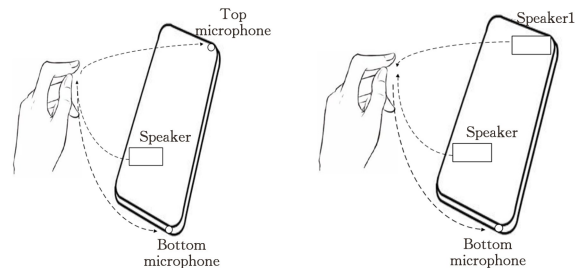
2.3.1.3 声音识别方法

声波经过预处理后,大部分的噪声干扰已经被除去,剩下的就是我们需要的有效的音频信号。为了更清楚地展示这部分思路,我们会在这部分先介绍最新成果的解析方案,最后进行最新成果的详细展示。

从感知目标上来讲,基于超声波的手机近场感知基本上是集中于对生命体征和手势动作的追踪,其感知手段大致可分为基于时间差、基于相位差和机器学习 3 个方向。它们各自均有自己的优缺点,在某些功能应用上有自己独特的优势。

(1) 基于时间差的方法

与声音阵列方向类似,基于时间差的方法是利用音频信号到达麦克风的时间差来建立特定的追踪模型。鉴于手机的硬件条件,一般有两种追踪模型:一种利用一个扬声器发声,两个麦克风收声,如图 8(a) 所示;另一种利用两个扬声器发声,一个麦克风收声来确认目标位置,如图 8(b) 所示。



(a) 一个扬声器两个麦克风

(b) 两个扬声器一个麦克风

图 8 时间差结构原理图

Fig. 8 Schematic diagram of time difference structure

(2) 基于相位差的方法

在声波信号的传播过程中,目标物体的移动等会引起回波信号的相位变化,因此回波信号的相位变化可以被解析为目标物体的位置变化信息。如图7所示,从音频数据中提取相位信息的常用方法就是C/O解调,通过该方法将数据分解为同向和正交两个分量,前面提到的LLAP^[27]也是利用此方法来推断手部的位置。

(3) 机器学习的方法

机器学习是当今分类计算的热门课题,它是人工智能的核心。根据感知目标的不同,基于超声波的手机近场感知可以将其数据特征看成分类解析问题。目前已有部分文献采用机器学习的方式来进行手势行为的分类识别,如文献^[33]采用HMM算法进行手势分类识别。HMM算法是用单一随机变量来描述过程状态的时序概率模型,它对每一种手势都做了HMM运算,计算了其中的相似程度,这种方法可以很好地识别各种定义好的手势,但训练过程相对复杂。文献^[34-35]则利用了支持向量机(SVM)算法来实现手势识别。SVM是一种基于统计学习理论的算法,属于机器学习中有监督学习模型的一种,在小样本环境中,SVM被认为是分类效果最好的分类器之一,而且其基本上不涉及概率测度及大数定律,避开了从归纳到演绎的传统过程,实现了从训练样本到预报样本的“转导推理”,大大简化了分类的复杂度。

2.3.2 应用系统介绍

鉴于手机的普适性、移动性、超强的计算感知能力以及无需额外的硬件支撑特性,科研工作者提出了很多有趣的系统。2.3.1节介绍了基于超声波手机近场感知的相关技术的细节以及框架流程,为了更好地宏观展现最新的科研成果,接下来会详细介绍部分有代表性的系统的情况。

(1) LLAP

Wang等^[27]于2016年提出了无设备的手势跟踪方案——LLAP系统^[27],该系统可被部署于智能手机或者其他可穿戴智能设备上。利用移动设备内置的麦克风抓捕扬声器发出的高频(17~23kHz)连续波声波信号,在除噪过后,Wang等率先使用了相位法识别距离的变化,相比传统多普勒频移,相位法的精确度更高。然后,将信号进行下采样获得I/O分量,将复杂的复信号变为静态矢量和动态矢量两种矢量。静态矢量大多是由多路效应或者静态物体的反射引起的,因此他们采用LEVD^[36]和频率分级来缓解静态干扰,同时利用动态矢量实现空间一维距离测量。在此基础上,结合粗粒度的延迟测量,LLAP实现了二维的收拾追踪。

(2) PatternListener

针对手机PIN码的使用安全问题^[37],Zhou等^[28]于2018年发表了破解手机模式锁的应用系统——PatternListener^[28],根据大多数人使用手机前进行模式锁认证这个过程,在这个时间段进行声波探测。与LLAP相似,PatternListener也采用了连续波作为探测信号,采用相干检测的方法对信号进行解调,在采用局部极值检测算法和线性插值算法除噪之后,利用自创TPI算法,将信号进行拐点切割,随后借助相位变化

来计算路径变化的长度,最后根据概率分布做出一个模式树来推断解锁图案。

(3) ApneaApp

2015年,Nandakumar等^[20]通过使用手机扬声器发射调频连续波(FMCW)来检测用户呼吸引起的胸腔震动。该系统可以同时近距离地对多人进行呼吸检测,扬声器接收到信号后,直接通过高通滤波器进行过滤除噪。为了提升空间分辨率,ApneaApp采用了10组信号一起做傅里叶变换,在此基础上,ApneaApp开发了各种呼吸暂停检测的算法,包括阻塞性呼吸暂停、中央呼吸暂停和声纳反射引起的低呼吸暂停。经临床验证,呼吸暂停和低呼吸事件的平均误差为1.9个事件每小时。

(4) ACG

传统的心电检测设备往往体积庞大且操作复杂,人们需要去医院做定期检测,而可穿戴医学设备穿戴非常不便,并且通常价格昂贵。2018年,Kun等研发了对应于ECG的手机应用ACG^[31]。ACG发射频率为17~19kHz的调频连续波,经双麦降噪后被麦克风捕获,受限于声波频率的带宽,导致空间分辨率不够,ACG将信号下传为基带信号,经低通滤波器和PCA降维主成分分析后,将心跳信号剥离出来,达到检测心跳信号的目的。

(5) FingerIO

2016年,Nandakumar等开发了FingerIO系统^[29],该系统传输18~20kHz的基于OFDM的脉冲信号,该信号具有较强的抗干扰能力,信号经物体反射后被麦克风捕获,当用户移动手指时,对应的回声到达的时间会随之发生变化。通过比较两种姿势的回声剖面,科研人员可以提取出与移动手指对应的回声,OFDM这项技术的原理是将带宽分割成正交的多个子载波形式,并且每个独立的子载波均可以传输不同的数据。通过分析不同的回波信号,计算手与麦克风的距离,实现手势追踪。

(6) VSkin

目前使用手机时,一切操作都需要通过点触屏幕来实现,人机交互的方式过于单一。基于此,Sun等^[30]于2018年研发出了Vskin系统^[30],该系统借助ZC序列实现了细粒度的手指运动检测。这些手势行为可以在手机背面进行,大大增加了人机交互的乐趣。ZC序列选择的频率区间为17~23kHz,收回的回波信号包括通过敲击手机的结构传声以及正常的空中传声。科研人员首先对声波的路径进行分离,测量每种信号路径的相位,为了更好地识别用户手势行为,将空中路径声波信号剥离出来,利用路径长度的变化来估计用户的手势行为。

(7) Dolphin

2014年,Yang等研发了手势识别系统Dolphin^[34],该系统同样使用手机内置的扬声器发声,麦克风捕获声音。该系统利用的是多普勒频移来分析回波信号。与别的时延或者相位差不同,科研人员利用了机器学习的方式来对预定义的手势行为进行分类。实验结果表明,该系统能够对24种手势进行识别,准确率达到94%。

(8)SteerTrack

在驾驶安全方面,大多都是通过图像分析或者可穿戴设备来检测驾驶动作或者驾驶状态。2019年,Xu等研发了基于超声信号的SteerTrack^[38]。该系统借助智能手机发出的正弦波超声信号,在驾驶室内构筑了微小的信号场。利用几何变换将方向盘映射到一个椭圆上,并通过追踪驾驶员双手操作方向盘的行为来判断车辆的转向角度。实验结果表明,该系统追踪方向盘的平均误差为4.61°。

(9)mHealth

2019年,Castillo等研发了mHealth^[39]。该系统与ApneaApp^[20]相似,都是对呼吸暂停信号进行检测。mHealth利用陷波滤波器过滤电力线路的噪声,为了降低心跳所带来的高频噪声,运用了8阶巴特沃斯带通滤波器进行向前向后滤波。同时,mHealth创新性地考虑了口腔呼吸事件,由于传统鼻气流检测装置的局限性,会将口腔呼吸误认为呼吸暂停,这对未来呼吸检测以及疾病的研究具有重要的参考意义。经验证表明,呼吸暂停检测率为89%。

3 音频感知的讨论

本文介绍了基于声波感知的传统声呐原理、声音传感器阵列以及基于智能手机的轻量级音频感知。这3种技术都有自身适宜的应用场景:传统声呐具有功率极强、在水下不易衰减等优点,因此往往被应用于国防安全、水下测绘等大型系统;声音传感器阵列由于其独特的传感器排布方式和数量,往往被应用于小空间的空间定位,如机器人方向判断等;基于智能手机的轻量级音频感知则是当下的研究热点,智能手机的普及以及超声波不易被听到的特性,使其具有极高的普适性和隐蔽性,为近场感知提供了新的科研思路,也为潜在的新型人机交互模式带来了新的方向。本文只对基于智能手机的音频感知来做对比,由于感知目的不同,我们会从声波选择、声音预处理和声音识别的方法这3个方面来做对比。

(1)声波选择

目前基于智能手机的音频感知系统在安卓系统上得到了广泛的验证,少部分应用在IOS上也得到了实现。声波通常选择17~23kHz区间的频率,因为这个区间的频率不易被人类听到,且可以与大多数环境音区别开。相比传统的正弦声波,调频连续波、ZC序列等具备更高的空间解析能力,信号的变化也更加明显。就检测目标的表现而言,连续波对目标最敏感,可以很清楚地看到声波的变化,但其极易受到环境的干扰,对音频分析有一定的难度,且其空间分辨率稍低;调频连续波相比连续波更适合测距和测速,由于其声波特性,使得回波与原声波只有一定的时间差,因此在此基础上有实现多目标探测的可能。ZC序列具有很好的自相关性和很低的互相关性,这种特性被用来产生同步信号,作为对频率和时间的运送。

(2)声音预处理

在采取音频信号样本之后,对音频样本的处理方法基本是一致的,通常是用带通滤波器、向下采样等手段。滤波器可

以很好地将我们不需要的频率段的音频去除,向下采样则是因为智能手机的带宽区间问题,如式(16)所示,科研人员选择的声波带宽容易造成分辨率不够。

$$D = v/2B \quad (16)$$

其中, D 代表分辨率, v 是声音在空中的传播速度, B 是音频的频率区间。这一阶段的目的是消除环境噪声和多径效应带来的影响。

(3)声音识别的方法

声音识别的方法有基于时间差、基于相位差和机器学习这3种。基于时间差和相位差的方法借助特殊的物理模型可以实时追踪目标的运动轨迹,缺点是计算开销较大;机器学习的方法大多利用多普勒频移来确定频移和目标运动的映射关系,然后借助SVM等分类方法来区分确定目标的运动状态,缺点在于只能识别预先设定好的动作行为。

4 未来的发展与工作

随着科技的发展,声学感知领域已经有了长足的进步,尤其是手机性能的高速发展以及手机的大范围应用,为声呐感知带来了新的活力。研究人员利用手机内置的扬声器和麦克风完成了对声音信号的发射和采集,并利用滤波器等手段进行了音频预处理,然后采用基于时延分析法、相位法或者机器学习的方法对音频信号进行了细粒度划分,基于此出现的手机声呐系统包括了对生命状态检测、手势行为分析等各种热点的研究,丰富了潜在的人机交互模式。

本文对声呐感知的发展历程进行了遍历,并对其中的技术细节进行了详细的介绍和分析,包括系统架构、技术细节等。目前来看,基于超声波的手机声呐感知具有高隐蔽性和细粒度探测等多种优点,但是依然存在许多不足和隐患,这些都可以作为我们未来工作的突破点:

(1)传感器的数量。目前发射和采集音频依赖的是手机内置的扬声器与麦克风,内置扬声器与麦克风数量的多少在很大程度上决定了音频的精准性,同时多扬声器麦克风的存在还会丰富时延或者相位法上的物理模型,为多维手势带来更好的选择。

(2)多种传感器配合。文献[28]利用陀螺仪等来探查屏幕唤醒时刻。手机中有丰富的传感器,未来可以使用更多的其他内置传感器(如加速计、陀螺仪等),使得数据采集识别变得更加细粒度且多样化。

(3)除噪能力。自然环境中存在各式各样的噪声干扰,如环境噪声、多径效应^[40]等。这些干扰会导致我们对音频的分析产生巨大的误差,现有的技术不能很好地去除这些影响,如何更好地去除噪声将会是未来的一大挑战。

(4)安全性。新的科研思路^[41]被别有用心之人利用可能会产生新的安全问题,超声波探测的高隐蔽性可能会带来他人攻击手机等安全隐患。

结束语 物联网时代的来临,使得各行各业都有了不同的变化,声波探测领域也经历了从传统声呐设备到声音阵列再到智能手机探测的历程。由于感知设备所面对的多种复杂

场景需求,感知设备在向微小化、细粒度化发展。

目前来看,智能手机具有很强的发展前景,它已经渗透到了人们生活的各方面,其配置的传感器也愈加丰富和强大,为实现更复杂的感知功能提供了强大的技术支持。最新的基于超声波的手机近场感知就是利用手机内置的扬声器发射超声波信号,同时通过手机的麦克风捕获信号,对回波信号加以数学分析,从而进行生命状态的感知或者手势识别等。由于超声波信号不易被听到的特性,使得这种应用可以隐蔽而持续地进行空间感知,丰富了潜在的人机交互的可能性。

本文将声呐、声音阵列和基于手机音频信号的近场感知进行了串联,对这个时间轴的物联网声呐感知进行了描述。首先介绍了声呐的起源和原理;然后介绍了声音阵列方面的知识和应用;接着详细介绍了基于智能手机的主动声呐感知系统的原理,包括系统框架的组成以及其中的技术原理和细节,并按照声波的发射、声音的预处理和声音识别这3部分来全面梳理系统的工作流程;最后,根据现阶段的发展情况对该类应用系统所面临的挑战进行了总结,并对未来工作做出了简单的设想。

参 考 文 献

- [1] CAI L, CHEN H. TouchLogger: inferring keystrokes on touch screen from smartphone motion[C]// Proceedings of the 6th USENIX Conference on Hot Topics in Security. USENIX Association, 2011: 9-9.
- [2] EMILIANO M, ALEXANDER V, SUHRID B, et al. Tapprints: your finger taps have fingerprints[C]// Proceedings of the 10th International Conference on Mobile Systems, 2012: 323-336.
- [3] GUIXIN Y, ZHAN Y T, DING Y F, et al. Cracking Android pattern lock in five attempts[C]// The Network and Distributed System Security Symposium. Internet Society, 2017.
- [4] SONG J, GÁBOR S, PECE F, et al. Inair Gestures Around Unmodified Mobile Devices[C]// Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, 2014: 319-329.
- [5] ZHANG C, TABOR J, ZHANG J, et al. Extending Mobile Interaction Through Near-Field Visible Light Sensing[C]// Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, 2015.
- [6] LI M, MENG Y, LIU J, et al. When CSI Meets Public WiFi: Inferring Your Mobile Phone Password via WiFi Signals[C]// Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016: 1068-1079.
- [7] JIE Z, XIAOLONG Z, ZHANYONG T, et al. Privacy Leakage in Mobile Sensing: Your Unlock Passwords Can Be Leaked through Wireless Hotspot Functionality[J]. Mobile Information Systems, 2016(2): 1-14.
- [8] RODRÍGUEZ V A, TRINIDAD A, GARCÍA B J R, et al. Extended high-frequency (9-20 kHz) audiometry reference thresholds in 645 healthy subjects[J]. International Journal of Audiology, 2014, 53(8): 531-545.
- [9] BUCKLEY K M, GRIFFITHS L J. Broad-band signalsubspace spatial-spectrum (BASS-ALE) estimation[J]. IEEE Transactions on Acoustics Speech & Signal Processing, 1988, 36(7): 953-964.
- [10] LYU Y, LIU J Y, GAO S G, et al. Study on robust adaptive beamforming based on acoustic vector sensor array[C]// Oceans. IEEE, 2017: 1-4.
- [11] AVOKH A, ABUTALEBI H R. Speech enhancement using linearly constrained adaptive constant directivity beam-formers[J]. Applied Acoustics, 2010, 71(3): 262-268.
- [12] LEI Z, YANG K, DUAN R, et al. Localization of low-frequency coherent sound sources with compressive beamforming-based passive synthetic aperture[J]. The Journal of the Acoustical Society of America, 2015, 137(4): EL255-EL260.
- [13] COLLOBERT M, FÉRAUD R, TOURNEUR G L, et al. LISTE-N: a system for locating and tracking individual speakers[C]// Proceedings of the Second International Conference on Automatic Face and Gesture Recognition. IEEE, 1996: 283-288.
- [14] GUSTAFSSON T, RAO B D. Source Localization in Reverberant Environments: Modeling and Statistical Analysis[C]// IEEE Transactions on Speech and Audio Processing. 2003: 791-803.
- [15] DUAN L P. Research on three-dimensional sound source localization of 4-microphone array based on TDOA algorithm [D]. Lanzhou: Lanzhou University of Technology, 2014.
- [16] LIU Y F. Research on sound source localization based on microphone array[D]. Hangzhou: Hangzhou University of Electronic Science and Technology, 2018.
- [17] ZHANG M, YANG P, TIAN C, et al. SoundWrite: Text Input on Surfaces through Mobile Acoustic Sensing[C]// International Workshop on Experiences with the Design & Implementation of Smart Objects. ACM, 2015: 13-17.
- [18] LUO G, CHEN M S, LI P, et al. SoundWrite II: Ambient acoustic sensing for noise tolerant device-free gesture recognition [C]// Proceedings of the International Conference on Parallel and Distributed Systems, 2017: 121-126.
- [19] CAO S, YANG P, LI X, et al. iPand: Accurate Gesture Input with Smart Acoustic Sensing on Hand[C]// 2018 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON). 2018: 1-3.
- [20] NANDAKUMAR R, GOLLAKOTA S, WATSON N. Contact-less Sleep Apnea Detection on Smartphones[C]// the 13th Annual International Conference. ACM, 2015: 22-24.
- [21] TUNG Y C, SHIN K G. EchoTag: Accurate Infrastructure-Free Indoor Location Tagging with Smartphones[C]// Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, 2015: 525-536.
- [22] PENG C, SHEN G, ZHANG Y, et al. BeepBeep: A high accuracy acoustic ranging system using COTS mobile devices[C]// Proceedings of the 5th International Conference on Embedded Networked Sensor Systems. 2007: 1-14.

- [23] ZHANG Z, CHU D, CHEN X, et al. SwordFight; Enabling a New Class of Phone-to-Phone Action Games on Commodity Phones[C]//Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services. ACM, 2012: 1-14.
- [24] GUPTA S, MORRIS D, PATEL S, et al. SoundWave; using the doppler effect to sense gestures[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2012: 1911-1914.
- [25] MD T, SIDHANT G, MAYANK G, et al. Doplink; Using the doppler effect for multi-device interaction. [C]//Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2013: 583-586.
- [26] CHEN K Y, ASHBROOK D, GOEL M, et al. AirLink; sharing files between multiple devices using in-air gestures[C]//Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2014: 565-569.
- [27] WANG W, LIU A X, SUN K. Device-free gesture tracking using acoustic signals[C]//Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking. ACM, 2016: 82-94.
- [28] ZHOU M, WANG Q, YANG J, et al. PatternListener; Cracking Android pattern lock using acoustic signals [C]//Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. 2018: 1775-1787.
- [29] NANDAKUMAR R, IYER V, TAN D, et al. FingerIO; Using Active Sonar for Fine-Grained Finger Tracking[C]//Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016: 1515-1525.
- [30] SUN K, ZHAO T, WANG W, et al. VSkin; Sensing Touch Gestures on Surfaces of Mobile Devices Using Acoustic Signals [C]//Proceedings of the 24th Annual International Conference on Mobile Computing and Networking. 2018: 591-605.
- [31] KUN Q, CHEN S W, FU X, et al. Acousticcar-diagram; Monitoring Heartbeats using Acoustic Signals on Smart Devices[C]//Proceedings of the IEEE Conference on Computer Communications. 2018: 1574-1582.
- [32] COHEN L. Time-frequency analysis; theory and applications [M]. Prentice-Hall, 1995.
- [33] XU Z C, WU K J, HU P. Ultrasonicwaves based gesture recognition method for smartphone platform[J]. Computer Engineering and Applications, 2018(2): 239-245.
- [34] YANG Q F, TANG H, ZHAO X B, et al. Dolphin; Ultra-sonic-Based Gesture Recognition on Smartphone Platform[C]//Proceedings of the 2014 IEEE 17th International Conference on Computational Science and Engineering (CSE). IEEE, 2014: 1461-1468.
- [35] LI X Y, DAI H J, CUI L Z, et al. SonicOperator; Ultrasonic gesture recognition with deep neural network on mobiles[C]//Proceedings of the 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld / SCALCOM/UIC/ATC/CBDCom/IOP/SCI). 2017: 1-7.
- [36] HUANG N E, SHEN Z, LONG S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 1998, 454(1971): 903-995.
- [37] LEUNG X, TANFORD S, JIANG L. Is a picture really worth a thousand words? [J]. Journal of Hospitality and Tourism Technology, 2017, 8(1): 19-38.
- [38] XU X, YU J, CHEN Y, et al. Leveraging Acoustic Signals for Vehicle Steering Tracking with Smartphones[J]. IEEE Transactions on Mobile Computing, 2020, 19(4): 865-879.
- [39] CASTILLO-ESCARIO Y, FERRER-LLUIS I, MONTSERRAT J M, et al. Entropy Analysis of Acoustic Signals Recorded With a Smartphone for Detecting Apneas and Hypopneas; A Comparison With a Commercial System for Home Sleep Apnea Diagnosis[J]. IEEE Access, 2019(99): 1-1.
- [40] LING K, DAI H, LIU Y, et al. UltraGesture; Fine-grained gesture sensing and recognition[C]//Proceedings of the 2018 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON). 2018: 1-9.
- [41] CHENG P, BAGCI E, ROEDIG U, et al. SonarSnoop; Active acoustic side-channel attacks[C]//Proceedings of the International Journal of Information Security. 2018: 1-16.



CHEN Chao, born in 1991, master student, is a member of China Computer Federation. His main research interests include audio perceptual and so on.



ZHAO Chun-lei, born in 1979, Ph.D, associate professor, is a member of China Computer Federation. Her main research interests include cyber security and so on.