

基于车载分布式声学器件的手势静音控制

范梦瑶^{1, 2}, 苏玉琪^{1, 2}, 张扶桑^{3, 1, 2}, 苏林^{1, 4}, 金蓓弘^{1, 2}

(1. 中国科学院大学, 北京 100190; 2. 中国科学院软件研究所, 北京 100190; 3. 北京航空航天大学, 北京 100191; 4. 中国科学院声学研究所, 北京 100190)

摘要: 随着智能座舱交互技术的发展, 非接触式手势识别已成为车载人机交互的重要补充。针对传统视觉与雷达方案在隐私保护、部署成本及覆盖范围方面的固有局限, 提出了一种新型的手势静音控制系统。该方案复用车载分布式的多扬声器与多麦克风阵列, 构建低成本的近场手势识别系统。该系统通过发送高频声学信号, 并提取手部动作引起的信道冲激响应变化来识别静音手势。具体而言, 为消除系统延迟, 首先创建一个电回环通道以提取参考信号, 该信号可用于动态计算系统延迟, 实现准确的信道冲激响应的对齐。其次, 提出了单通道多维特征判决与多通道协同决策的融合方案, 通过两者的协同优化实现预设手势与环境干扰的精准区分。在多种场景下开展的实验验证表明, 系统静音手势识别准确率超过 98%, 误报率低于 1.5%, 响应时间小于 0.05 秒。

关键词: 车载声学感知; 手势静音控制; 非接触式手势识别; 智能座舱

doi: 10.11959/j.issn.2096-3750.XXXX.

Leveraging in-cabin distributed acoustic devices for gesture-based mute control

FAN Mengyao^{1,2}, SU Yuqi^{1,2}, ZHANG Fusang^{3,1,2}, SU Lin^{1,4}, JIN Beihong^{1,2}

1. University of Chinese Academy of Sciences, Beijing 100190, China

2. Institute of Software, Chinese Academy of Sciences, Beijing 100190, China

3. Beihang University, Beijing 100191, China

4. The Institute of Acoustics of the Chinese Academy of Sciences, Beijing 100190, China

Abstract: With the rapid evolution of smart cabin interaction technology, contactless gesture recognition has become an essential component of in-cabin human-computer interaction. Traditional methods, such as vision-based and radar-based systems, often face challenges like privacy concerns, high deployment costs, and limited coverage. To overcome these limitations, we introduce a novel gesture-based mute control system that leverages the advantages of contactless interaction to enhance user experience and system efficiency. This solution repurposes existing audio hardware to construct an affordable near-field recognition system. By transmitting high-frequency acoustic signals and analyzing the changes in the channel impulse response (CIR) induced by hand movements, the system effectively recognizes mute gestures. Specifically, to eliminate system delay, we first establish an electrical loopback channel to extract a reference signal. This reference signal is then used to dynamically compute the system delay, ensuring accurate CIR time alignment. Additionally, we propose a fusion scheme that combines single-channel multi-dimensional feature judgment with a multi-channel collaborative decision-making mechanism. This integrated approach is designed to accurately distinguish predefined gestures from environmental interferences. Experimental validations across multiple scenarios show that the system achieves a mute gesture recognition accuracy exceeding 98%, a false alarm rate below 1.5%, and a response time within 0.05 sec.

收稿日期: XXXX-XX-XX; 修回日期: XXXX-XX-XX

通信作者: 张扶桑, fusang@iscas.ac.cn

基金项目: 北京市科技新星计划-交叉合作课题 (20240484641), 国家自然科学基金 (62422213, 62172394, 62072450), 北京市自然科学基金 (L223034)

Foundation Items: Beijing Nova Program(20240484641), National Natural Science Foundation of China (62422213, 62172394, 62072450), Beijing Natural Science Foundation(L223034)

onds.

Key words: in-cabin acoustic sensing, gesture mute control, contactless gesture recognition, smart cabin

0 引言

随着智能座舱技术的快速发展,人车交互领域正面临新的挑战与需求,涵盖驾驶安全保障、操作便捷性提升及响应速度优化等需求。传统按键或触摸屏交互方案要求驾驶员“眼手协调”,不仅分散驾驶注意力、危及行车安全,还存在覆盖范围局限——交互组件通常仅支持前排操作,导致后排乘客缺少操控能力,难以满足全车交互需求。而传统基于语音的交互方案又存在易受环境噪声干扰、识别准确率低等问题,此类问题会降低整体用户体验。

非接触式手势交互因便捷性与自然性,成为智能座舱人机交互的重要方案。尤其在音频静音控制等高频操作场景中,其优势显著:抗噪声干扰能力强、多用户适应性好、指令响应迅速,既保障驾驶安全,又为智能座舱交互开辟新路径。

目前车内手势识别方案主要基于视觉或雷达技术,但面临实际挑战。视觉方案通过摄像头捕捉图像,存在隐私泄露风险,且对光照条件变化敏感,座椅布局可能遮挡识别范围形成盲区;雷达方案易受车内多路径反射干扰,且硬件成本高昂,难以支持全覆盖部署。多模态融合技术虽旨在整合模态优势,却面临模态同步难度大、计算资源需求高等问题,导致系统难以快速响应,限制了其在车载场景的广泛应用。

本文提出利用车内声学感知实现手势交互。我们观察到越来越多的汽车配备了大量声学模块如扬声器与麦克风阵列,这些声学模块主要用于语音控制和车内娱乐。通过复用车载分布式的多扬声器与多麦克风阵列,我们开发了支持全覆盖交互、快速响应及多用户操作的手势交互系统,可实现“挥手静音”等自然便捷的交互功能。

尽管前景可期,但要将这一构想转化为一个功能齐全的系统,我们仍需解决多个挑战。

1) 系统延迟导致的手势定位偏差问题。在车载音频系统中,声学信号传输需经过 ADC/DAC 转换、音频缓冲、音频接口及通用操作系统资源竞争

[1]等环节。每个阶段的随机延迟在信号传输过程中累积,形成不可忽视的系统延迟,且该延迟在每次系统启动时随机变化,无法预先测量或消除。这种系统延迟会引发测距误差,导致手势距离对齐失准、特征提取偏差,最终造成手势空间定位偏差。

基于参考信号的系统延迟消除方法。为解决该问题,我们提出一种基于声学系统中电回环通道^①参考信号的系统延迟消除方法。该方法利用声学系统设计中的一路电回环通道来获取参考信号,从而模拟只包含系统延迟的传播路径。具体而言,通过提取参考信号信道冲击响应(Channel Impulse Response, CIR)的主峰位置作为系统延迟的时间偏移参考。随后,对每个麦克风接收通道的 CIR 进行校准以补偿时间偏移。该校准过程可准确重建声学信号在空气中传播的实际距离。因此,后续的手势特征可更精确的提取,从而更准确地确定手势的空间位置。

2) 复杂环境干扰下的识别准确率问题。在实际驾驶场景中,车内环境受到车辆振动、乘客肢体动作、车载音响功率不平衡及内外噪声等因素干扰,严重影响弱手势反射信号的检测。

抗干扰手势识别方法。为了提高在复杂干扰环境下对弱手势信号识别的可靠性,我们开发了一种创新的手势识别方法。该方法将单通道多维特征判决与多通道协同决策相结合,从而提高了在复杂条件下手势检测的整体鲁棒性和准确性。首先,根据每个麦克风通道接收信号的 CIR 进行三维手势特征判决。这涉及三个关键约束: i) 空间局部性约束: 聚焦近场距离以确保空间精度; ii) 能量显著性约束: 检测能量的突变以识别重要手势; iii) 时间瞬时性约束: 匹配短时特征以准确捕捉瞬态手势。其次,基于多麦克风子阵列的空间划分,我们建立了一个多通道协作决策机制。在同一时间域窗口内,如果目标区域内多个麦克风通道满足协作条件,即触发系统响应。通过将多维特征约束与空间阵列协同相结合,该方法有效抑制了环境干扰,并提升了全座舱各座位区域的识别准确性。

^①在一些声卡的 PCB 设计中,会将发送信号通过电路的方式引至接收端,用于为音频监控或回声消除提供参考信号。

通过解决上述挑战，我们在真实车辆上开发了手势静音控制系统。在多乘客、高噪声、颠簸路况等场景进行了大量实车测试，结果表明，我们系统在复杂环境中具有良好的性能。具体而言，其平均识别准确率超过98%，平均误报率低于1.5%。全车各座位区域检测性能稳定，响应时间低于0.05秒。综上所述，通过声学感知技术，我们实现了实时、全覆盖、高鲁棒性的车载手势静音控制系统。

1 相关工作

1.1 基于声波的感知

基于声波的感知技术，通过巧妙利用设备内置的声学器件，不仅实现了功能的复用，还在生命体征监测、距离测量以及精准定位等多个关键应用领域展现出了巨大的发展潜力。在细粒度生命体征感知方面，C-FMCW技术[2]通过普通扬声器实现非接触式呼吸监测；Zhang等人结合智能音箱与经验模态分解算法分离心跳与呼吸信号[3]；基于智能手机双麦克风的Acousticcardiogram[4]通过消除直达波干扰，实现呼吸与心跳的同步监测；BreathJunior[5]、RespTracker[6]实现厘米级呼吸监测；SpeakPrint[7]结合语音与超声信号完成用户认证；Su等人[8]提出的分布式声学感知方案，利用车载扬声器与麦克风阵列实现座舱内无盲区儿童存在检测，为车载复杂空间感知提供了新思路。

在大尺度测距与定位方面，研究主要聚焦于时间特征提取。BeepBeep[9]、RFBeep[10]分别通过双向感知与内核优化实现厘米级至分米级测距精度；BatMapper[11]、SAMS[12]构建声学雷达系统完成室内环境映射；Guoguo[13]、ARABIS[14]借助ToA/TDoA实现厘米级定位。

1.2 非接触式手势识别

非接触式手势识别技术主要通过声学、射频、视觉三类方式实现。

基于声学信号的方法利用设备内置扬声器与麦克风，通过分析手部反射的回波信号特性识别手势。基于信号多普勒效应的SoundWave[15]、AudioGest[16]可识别滑动、点击等基础手势，但对复杂手指动作的分辨率不足。基于信号信道冲激响应分析的EchoTrack[17]、Strata[18]、LLAP[19]及引入深度学习的UltraGesture[20]、RobuCIR[21]，通过CIR的幅度或相位变化实现轨迹跟踪，提升了手

势识别精度，却受限于复杂环境下的频率选择性衰落问题。基于主动声纳的解决方案如FingerIO[22]虽具备毫米级手势定位精度，但仅专注于精细指尖运动追踪，且检测范围仅有 $50 \times 25 \text{ cm}^2$ 。

此外，基于射频信号的方法通过WiFi、毫米波雷达、超宽带（UWB）雷达等来实现手势识别。WiGest[23]、WiSee[24]、WiFinger[25]等利用WiFi信号强度（RSSI）或信道状态信息（CSI）实现非接触式检测，但是由于WiFi的带宽较小且距离分辨率较低，容易受多径干扰。雷达方案如AllSee[26]、Rf-IDraw[27]、RF-IPad[28]虽支持空中手写等复杂动作，却存在硬件成本高昂的问题，难以大规模部署。

相比之下，基于视觉的方法通过摄像头捕获图像信息实现手势识别。深度摄像头设备如Microsoft HoloLens和Sony PlayStation VR支持高精度3D交互，却受限于光照条件与设备成本。基于普通摄像头的方案如DigitEyes[29]成本较低，但也受遮挡和光照变化的影响，鲁棒性不足，且存在隐私泄露风险，因此难以广泛应用。

与现有声波手势识别方法相比，本文的创新主要体现在三方面：其一，硬件层面完全复用车载分布式扬声器与麦克风阵列，无需额外专用感知器件；而许多既有方案依赖定制化声纳或非车载消费级音频设备，难以满足多座位覆盖与车规适配。其二，决策机制采用多麦克风子阵协同判决，通过空间分区与共识规则增强抗干扰，区别于单通道特征匹配或简单阵列叠加。其三，信号处理引入馈线信号实时校准，动态消除车载音频链路的随机延迟，解决静态补偿易致定位偏差的问题，为手势检测提供稳定的时间基准。

此外，与雷达类方案相比，本文方法零新增硬件成本且具备全舱覆盖；与视觉类方案相比，可规避隐私、光照与遮挡等问题。因此，本文基于车载声学器件，以无额外硬件成本的方式实现了支持全舱覆盖、快速响应、可多人操作的声学手势交互系统。

2 系统概述

本文提出的基于车载声学器件的手势交互系统旨在通过使用车上已经部署的分布式的多扬声器与多麦克风，构建一种低成本、全覆盖的近场声学手

势感知方案。其核心目标是实现对不同座位乘客的静音手势的实时检测与可靠识别，以满足紧急静音场景下的即时交互需求。

2.1 场景定义

本系统的核心应用场景为车载环境下的高频静音控制场景，主要为扬声器播放音乐、导航语音等音频内容时的实时手势静音控制，同时也支持在静音状态或未播放娱乐音频时的独立手势识别（如用户自定义控制需求），以适配不同交互场景。系统将核心交互动作定义为“近场单次挥手”——用户在扬声器前方近场自定义距离内自然挥手靠近。当前聚焦“挥手静音”，系统已预留多手势扩展接口。该手势通过手部瞬间移动靠近引起的声学信号反射变化触发静音指令，如图1所示。



图1 基于声学的交互手势

这种手势设计有两个关键特点。首先，在信号检测方面，短时强反射在CIR中表现为能量局部突变，可通过滑动窗口和能量阈值识别。其次，在用户操作方面，该手势无需用户精确对准设备或执行复杂动作，操作自然简洁，能有效适配前排与后排不同座位的乘客。

2.2 系统框架

系统总体框架如图2所示，包含信号调制、预处理和手势识别三个核心模块，具体实现过程如下。

信号生成：系统首先调制生成具有优异自相关特性的高频Zadoff-Chu（ZC）序列[30]信号，然后分两路传输：一路通过电回环通道直接输出参考信号，另一路经扬声器发射至空气，由车内目标反射后被麦克风阵列接收。

信号预处理：多通道麦克风阵列接收多径反射信号，电回环通道捕获参考信号，然后计算这些信号的CIR。我们首先基于参考信号的CIR计算系统延迟，然后利用此延迟对多麦克风通道的CIR进行

时间延迟校准。这一过程恢复了声学信号在空气中的真实传播距离，从而精确定位手势位置，为后续特征提取奠定基础。

特征提取与识别：手势特征识别模块以多通道校准后的CIR数据为输入，提取满足空间局部性、能量显著性和时间瞬时性的多维度手势特征。各通道特征识别结果传输至抗干扰决策模块。该模块通过空间阵列协同决策机制，利用多通道协同过滤单通道噪声及偶发干扰。因此，复杂环境下的识别可靠性得到提升。最终决策结果用于触发静音指令，形成“挥手-感知-响应”的完整闭环。全流程处理时长控制在0.05秒内，满足实时性要求。

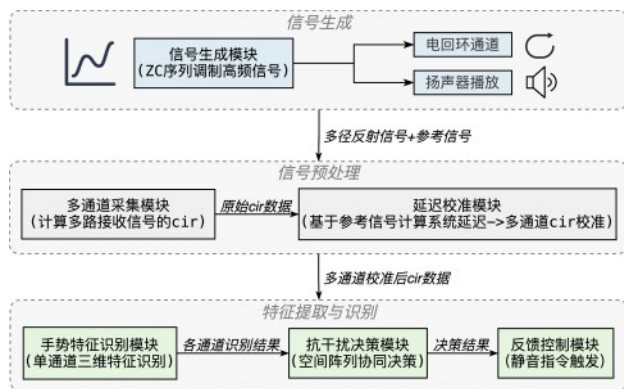


图2 系统框架

3 系统详细设计与实现

本章重点探讨系统模块的详细设计与实现。首先，我们讨论了在车内环境中发送信号的调制过程以及接收信号的建模过程。随后，我们介绍了基于参考信号的系统延迟消除方法。接着，我们详细阐述了在延迟消除后利用CIR进行单通道三维手势特征判决的方法。最后，我们介绍了基于多通道的空间协同决策机制。

3.1 发送信号调制与接收信号建模

本节主要介绍ZC发送信号的设计与车载环境下的接收信号的建模。首先，我们将详细阐述基带ZC序列的定义，和将基带ZC序列调制至超声频段的关键步骤。然后，我们将介绍信号在车载环境中传播与接收的过程及模型，以及系统延迟对信号处理产生的影响。

3.1.1 ZC序列的定义与特性

为减少多径效应的影响，构建高分辨率的距离-时间谱，发送端采用具有强自相关性的ZC序列

作为基带信号。该信号可表示为：

$$zc[n] = \exp\left(-j \frac{\pi u n(n+1+2q)}{N_{zc}}\right)$$

其中， u 是 ZC 序列根参数，决定序列正交性与自相关特性。 q 为相位偏移量，用于调整信号的初始相位。 $n = 0, 1, \dots, N_{zc} - 1$ ，其中 N_{zc} 为 ZC 序列的长度，决定信号的周期长度。

3.1.2 发送信号的调制

为了在识别过程中不对人产生噪音干扰，需将基带 ZC 信号调制至不可听的超声频段进行发送。具体流程如下。

步骤1:首先，对 ZC 序列进行快速傅里叶变换以得到其频域表示。随后通过汉明窗抑制旁瓣并减少频谱泄漏，从而得到频域基带 ZC 信号。

步骤2:将基带 ZC 信号搬移至目标载波中心频率 f_c 位置，并将其共轭信号对称搬移至 $-f_c$ 处。最后通过逆快速傅里叶变换转换回时域，完成频谱搬移。

步骤3:我们重复并拼接单帧 ZC 信号多次，以生成指定持续时间的周期性长信号。随后，我们通过 18kHz - 20kHz 带通滤波器对信号进行滤波，以抑制带外噪声并聚焦能量于目标超声频率带，从而获得发送信号。

如图 3 所示，调制后的发送信号在超声频段具有集中的能量和良好的旁瓣抑制。

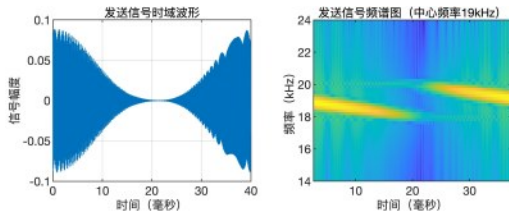


图3 发送信号的时域波形和频谱图

3.1.3 接收信号的建模

发送信号由多扬声器发送，在车载封闭空间中传播并由多麦克风接收。由于车载空间内存在座椅、仪表盘、车门等密集的静态物体，空间结构复杂，因此信号传播时会同时沿直射路径与多条反射路径抵达麦克风阵列。因此，接收信号可表示为直射信号、多路反射信号和背景噪声的叠加[8]，如下所示：

$$y[n] = \sum_{i=1}^P A_i e^{-j\phi_i} \cdot zc_T[n - \tau_i f_s] + \text{noise}[n]$$

其中： $zc_T[n]$ 是发送信号。 $y[n]$ 是接收信号。 A_i 是第 i 条路径的幅度衰减，它与传播距离、反射面材质相关。 $\phi_i = \frac{2\pi d_i}{\lambda}$ 是传播引起的相位延迟。 τ_i 是第 i 条路径的总延迟。 f_s 是采样率。 $\text{noise}[n]$ 是背景噪声。 d_i 是第 i 条路径的路径长度。 λ 是声波波长。 c 是声音在空气中的传播速度。

在车载音频系统中，声学信号需经过 ADC/DAC 信号转换、音频缓冲区、音频接口及通用操作系统资源竞争等多个阶段，各阶段的不确定性延迟在传播路径中累加，形成具有启动时不确定性的系统延迟（该延迟无法通过事先测量确定）。因此，每条路径的总延迟 τ_i 必须综合考虑传播路径的物理延迟及音频系统各阶段引入的不确定性系统延迟，可表示为：

$$\tau_i = \frac{d_i}{c} + \tau_{\text{sys}}$$

其中 $\frac{d_i}{c}$ 表示声学信号在空气中传播所需的时间。

τ_{sys} 表示随机系统延迟。

若不消除系统延迟，将导致信道冲激响应峰值位置偏移。这种偏移会导致距离模糊。此外，这还会使手势特征与车内其他动态干扰难以得到有效区分。例如，若反射路径长度为 0.5 m，则传播时间为约 1.46 ms，若叠加系统延迟 $\tau_{\text{sys}} = 2.9$ ms，手势距离将被误判为 1.5 m 左右，这将严重影响影响手势的特征提取和位置估计。

3.2 基于参考信号的系统延迟消除

3.2.1 参考信号建模

为了获得准确的手势位置，我们提出了一种通过利用声卡中构建的电回环通道的参考信号来消除系统延迟的方法。我们的硬件连接方案如下。

如图 4 所示，计算机与声卡相连，用于驱动扬声器发射信号并接收麦克风阵列采集的信号。声卡的右侧有一个输出端口和一个输入端口。在输出端口，我们采用一分二音频线将发送的信号分为两路（即路径 A 和路径 B）：路径 A 连接到功率放大器，用于驱动扬声器将声学信号传输到空气中。路径 B（即电回环通道）连接到双头 3.5mm 音频线，将信号回传至声卡的输入端口。

通过上述硬件连接方法，我们建立了电回环通道，为随后的系统延迟校准和多径信号分析提供了可靠的参考基准。



图4 硬件连接方法

参考信号包含了系统延迟这一关键信息，其时间表达式为：

$$y_{\text{ref}}[n] = A \cdot zc_T[n - \tau_{\text{sys}} f_s] = h_{\text{sys}}[n] \cdot zc_T[n]$$

其中 A 表示信号传输过程中的幅度衰减，在电回环通道中近似为1。参考信号 $y_{\text{ref}}[n]$ 也可以视为循环卷积 $h_{\text{sys}}[n] \otimes zc_T[n]$ 。

由于信号在音频线中的传播时间相较于系统延迟可忽略不计，因此参考信号的 CIR 的主峰位置就对应于总系统延迟。参考信号的 CIR 可表示为：

$$h_{\text{sys}}[n] = \delta[n - \tau_{\text{sys}} f_s]$$

其中 $\tau_{\text{sys}} f_s$ 为系统延迟采样点的索引， δ 是单位冲激函数。

3.2.2 系统延迟计算

在参考信号建模的基础上，我们通过以下步骤动态计算系统延迟。

1) **参考信号分帧**：将参考信号按 ZC 信号单周期长度 L 分帧，得到多帧信号 $y_{\text{ref}}[m, n]$ （其中 m 为帧序号）。这便于后续以帧为单位进行互相关处理。

2) **互相关计算**：对于参考信号的每个帧，与发送信号帧 zc_T 进行互相关计算，以获得 CIR 帧：

$$h_{\text{sys}}[m, n] = \text{IFFT}(\text{FFT}(zc_T)^* \cdot \text{FFT}(y_{\text{ref}}[m, n]))$$

其中 $\text{FFT}(\cdot)$ 表示快速傅立叶变换操作， $\text{IFFT}(\cdot)$ 表示逆快速傅立叶变换操作， $(\cdot)^*$ 为发送信号 zc_T 经傅里叶变换后的共轭。通过采用在频域中计算互相关的方法，计算复杂度从 $O(L^2)$ 降至 $O(L \cdot \log L)$ ，满足实时需求。

我们将多帧互相关结果沿时间维度排列。排列后的结果对应于距离-时间谱（如图 6(a)所示）。来自麦克风的接收信号的 CIR 以相同方式计算，并表示为 $h[m, n]$ 。

3) **峰值检测与延迟估计**：如图 5 所示，取互相关幅度最大位置作为延迟估计：

$$P_{\text{sys}} = \max_n |h_{\text{sys}}[m, n]|$$

其中 P_{sys} 对应强互相关主峰的采样点值， $P_{\text{sys}} = \tau_{\text{sys}} \cdot f_s$ 。

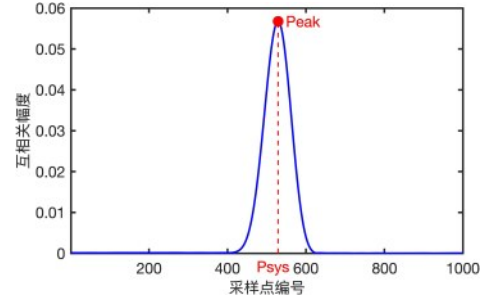


图5 参考信号的互相关结果

3.2.3 系统延迟消除

为消除系统延迟对各接收通道 CIR 距离估计的影响，需对原始 CIR 进行动态校准。具体来说，将原始 CIR 设为 $h[m, n]$ （其中 $n = 0, 1, \dots, L - 1$, L 为 CIR 序列长度）。我们沿采样点维度将 CIR 循环移位 P_{sys} 个采样点，以获得校准后 CIR $h_{\text{true}}[m, n]$ ：

$$h_{\text{true}}[m, n] = \text{circshift}(h[m, n], P_{\text{sys}})$$

其中， $\text{circshift}(\cdot)$ 表示循环移位操作。通过这样的操作，原始 CIR 中的采样点被重新对齐到其真实位置。例如，原本对应于采样点 $n + P_{\text{sys}}$ 的飞行时间，被调整回采样点 n 。此调整消除了系统延迟对距离测量的影响。图 6(b) 和图 6(c) 展示了信号校准前后 CIR 的对比。这表明采样点索引与实际距离之间的对应关系已得到修正，验证了该方法的有效性。

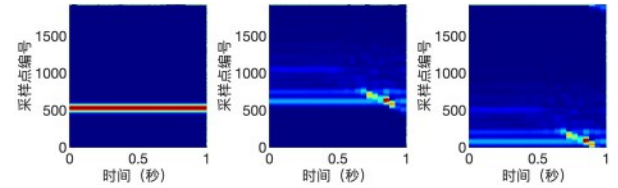


图6 系统延迟消除

(a)参考通道的距离 (b)消除系统延迟前 (c)消除系统延迟后时间谱的距离-时间谱

在固定距离标定试验中（采样率 48 kHz），我们统计参考通道主峰对齐的采样偏移：未校准时均值/标准差为 512/12 个采样点，对应时间偏差约 10.67 ms，距离等效偏差约 3.66m；校准后降至 0.9/0.4 个采样点，对应时间偏差约 18.83 μs ，距离等效偏差约 6.43mm。在相同阈值与评测协议下，整体识别准确率由 25% 提升至 99.05%，误报率由 35% 降至 0.5%，表明所提延迟校准显著抑制相对时基偏差并降低由此引发的定位漂移。

3.3 单通道三维特征判决方法

为精准提取手势特征并抑制环境干扰，本节基

于CIR构建了一个包含空间局部性、能量显著性和时间瞬时性的三维特征模型。通过层级化约束逐步筛选出有效的手势信息。

3.3.1 空间局部性约束

标准手势操作集中于扬声器近场区域（定义为0 - 0.3 m）。为抑制由环境噪声、随机车辆颠簸或乘客体动引入的远场类手势特征，需对分析范围进行约束。

我们定义空间约束函数如下，即仅保留近场区域的CIR信息：

$$\mathcal{S}_D(\mathbf{h}) = \begin{cases} h[m, n] = h[m, n], & n \in [0, L_w - 1] \\ h[m, n] = 0, & \text{otherwise} \end{cases}$$

其中 n 为采样点索引， L_w 是与规定的近场区域边界对应的采样点。该函数在 n 属于 $[0, L_w - 1]$ 时取原值，表示近场区域；否则设置 $h[m, n] = 0$ ，表示远场区域。

对消除系统延迟后的CI ($h_{true}[m, n]$) 应用该函数，得到近场约束的CIR：

$$h_D[m, n] = \mathcal{S}_D(h_{true}[m, n])$$

图7(a)到图7(b)即展示了单个时间窗口的空间约束前后的距离时间谱。此操作可抑制远场干扰，仅保留手势集中区域的CIR信息。

3.3.2 能量显著性约束

由手势触发的高能反射脉冲在时域幅度上显著区别于稳态直射和反射路径，需进一步筛选能量显著点。

首先，我们定义一个动态能量阈值以自适应场景特异性能量变化：

$$\theta = \alpha \cdot \max(h_D[m, n])$$

其中 $\alpha \in (0, 1)$ 是灵敏度系数，以平衡检测率与抗干扰能力，而 $\max(h_D[m, n])$ 表示矩阵的最大值，即近场CIR的最大能量。然后，通过能量约束函数对CIR进行二值化：

$$\mathcal{S}_E(\mathbf{h}) = \begin{cases} h[m, n] = 1, & h[m, n] \geq \theta \\ h[m, n] = 0, & \text{otherwise} \end{cases}$$

应用能量约束后，得到二值化能量显著矩阵：

$$h_E[m, n] = \mathcal{S}_E(h_D[m, n])$$

当 $h_D[m, n] \geq \theta$ 时， $h_E[m, n] = 1$ ，否则 $h_E[m, n] = 0$ 。这筛选出了能量显著的点，如图7(b)到图7(c)所示。

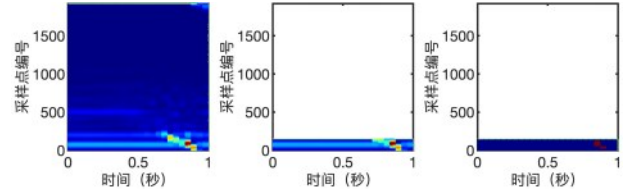


图7 空间约束和能量约束

(a)消除系统延迟后 (b)空间约束后的 (c)能量约束后的的距离时间谱 距离时间谱 距离时间谱

3.3.3 时间瞬时性约束

时间瞬时性约束的物理意义在于，手势反射因动作突发性呈现时间上的瞬态特性（持续时间短，如图8中部分带有手势的窗口的距离时间谱所示）。相比之下，静态背景反射具有跨周期时间稳态特性（持续时间长，如图8中无手势的窗口的距离时间谱所示）。我们通过利用这种时间维度差异来提取手势特征。

我们定义一个时间连续性度量函数，以计算同一距离处高能点的时间占比：

$$\mathcal{T}(n, \mathbf{h}) = \frac{1}{M} \sum_{m=1}^M h[m, n]$$

其中 M 为总帧数，该函数反映了固定距离下高能点的持续时间。

设定阈值 $\tau_h = 0.2$ （经实验验证的鲁棒值），并定义时间约束函数：

$$\mathcal{S}_T(\mathbf{h}) = \begin{cases} 1, & \forall n, \mathcal{T}(n, \mathbf{h}) < \tau_h \\ 0, & \text{otherwise} \end{cases}$$

当所有距离单元的时间占比均小于 τ_h 时，判定为手势特征。

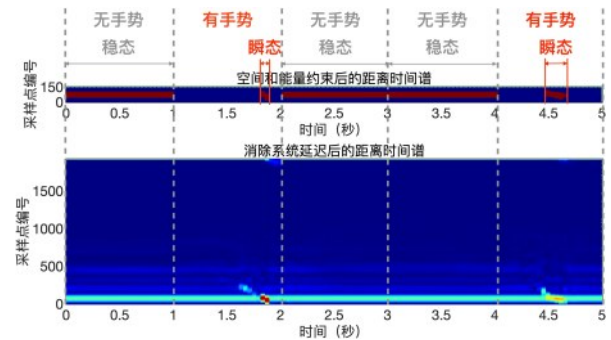


图8 多时间窗口的距离时间谱

3.3.4 最终特征判决

通过空间、能量、时间的三层约束，第 i 个通道的最终特征判决结果为：

$$s_g^i = \mathcal{S}_T(\mathcal{S}_E(\mathcal{S}_D(h_{true}[m, n])))$$

若 $s_g^i = 1$ ，则第 i 个通道检测到有效手势，否

则若 $s_g^i = 0$ ，则第 i 个通道未检测到手势。

该判决逻辑本质上是对三维约束条件的联合筛选过程。一个手势特征仅在同时满足以下三个条件时才被视为有效：i) 近场特征存在，ii) 能量显著性，以及 iii) 能量显著点在时间上的瞬态连续性。

3.4 多通道协同决策机制

根据车载场景多座位交互需求，本系统采用前后排四座位分布式声学感知单元布局。基于对常见车型扬声器和麦克风阵列布局的研究，声学单元安装在 A/B 柱上。

每个单元包含 1 个扬声器与 4 个麦克风通道。对应通道编号为：子阵 1（驾驶座）包含麦克风 1-4，状态向量为 $\mathbf{s}_1 = [s_g^1, \dots, s_g^4]$ ；子阵 2（副驾驶座）包含麦克风 5-8，状态向量为 $\mathbf{s}_2 = [s_g^5, \dots, s_g^8]$ ；子阵 3（后排左座）包含麦克风 9-12，状态向量为 $\mathbf{s}_3 = [s_g^9, \dots, s_g^{12}]$ ；子阵 4（后排右座）包含麦克风 13-16，状态向量为 $\mathbf{s}_4 = [s_g^{13}, \dots, s_g^{16}]$ 。其中 $s_g^i \in \{0, 1\}$ 表示第 i 通道是否检测到手势信号。

各子阵采用一致的多通道协同决策逻辑，以子阵 k 为例，定义子阵决策函数：

$$f_{\text{vote}}(\mathbf{s}_k) = \mathbb{I}(\|\mathbf{s}_k\|_0 \geq 3)$$

其中 $\|\cdot\|_0$ 表示 l_0 范数，用于量化子阵中检测到手势的通道数；若通道数 ≥ 3 则返回 1（表示该子阵检测到手势），否则返回 0。

通过分析四个子阵的决策结果，系统确定最终反馈结果。首先，识别所有输出为 1 的子阵。若仅有一个子阵输出为 1，则最终结果为该子阵对应的座位编号。若多个子阵同时输出为 1，则结果为这些子阵对应的座位编号的集合。若无子阵输出 1，则最终结果标记为 0。这种逻辑设计既保留了单座位精准定位的能力，又支持多座位同时交互的场景。

通过这种多通道协同决策机制，系统能更可靠地检测和定位手势。通道级协作确保单个阵列的多个通道一致检测手势。子阵级仲裁通过多个阵列的逻辑组合实现空间定位。

最终决策结果通过总线反馈至车载信息娱乐系统，触发自定义命令（例如音频静音），实现挥手控车。该机制在复杂车内环境中有效提升手势检测可靠性和定位精度，同时降低误触发率。

4 实验评估

4.1 实验设置



图9 车内声学感知单元部署

我们在一辆真实汽车的音频模块上部署了本系统。在典型部署场景中，共有 4 组声学感知单元，如图 9 所示，分别安装在 A/B 柱上。

每组声学感知单元包含 1 个 Bose 扬声器与 1 个四元麦克风阵列，如图 10 所示。同时，我们在车内后视镜下方部署了一台摄像头，用于记录车内手势事件作为真值。

计算机生成周期 ZC 信号（中心频率 $f_c = 19\text{kHz}$ 、带宽 $B = 2\text{kHz}$ 、信号周期采样长度 $L = 1920$ ）。通过声卡（PXUA216MB-DL2-M/XMOS）驱动扬声器发送声学信号。麦克风阵列以采样率 $f_s = 48\text{kHz}$ 收集反射信号，并通过声卡实时传输至计算机以进行信号处理。信号处理采用滑动窗口机制，窗口长度 1s，步长 0.5s，以平衡实时性与检测准确率。

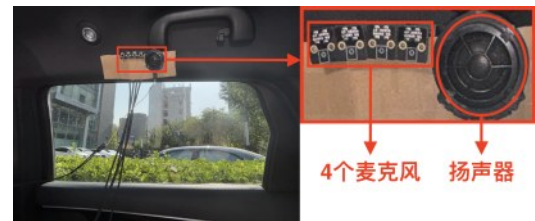


图10 单个声学感知单元的硬件说明

基于摄像头获取的真实数据，定义四类手势事件类型如下。1) 真阳性 (TP)，即成功检测真实手势且方向判定正确，2) 假阴性 (FN)，即未检测到真实手势，3) 方向性假阳性 (FP_1)，即检测到手势但方向判断错误，4) 误报 (FP_2)，即在无真实手势时输出手势事件。

因此，定义系统性能评估的关键指标为：准确率 $= \text{TP} / (\text{TP} + \text{FN} + \text{FP}_1)$ ，误报率 $= \text{FP}_2 / (\text{TP} + \text{FN} + \text{FP}_1)$ 。

同时评估每个窗口的平均处理时长。

4.2 整体性能

我们对12名测试人员在多种设置与车内场景下进行了评估。为了确保实验更接近真实场景，我们未对测试人员的其他活动进行限制，允许测试人员自由进行日常活动，如身体晃动、使用手机、进食等。在我们的实验中，我们考虑了不同座位和乘客数量对系统性能的影响。由于我们基于声学信号进行手势感知，我们还验证了不同环境噪声对系统性能的影响，例如：雨声、音乐、人声和交通噪声。此外，我们还在不同道路条件下评估系统性能，例如：减速带、山路、街道和高速公路。总体而言，我们在各种设置和场景下累计收集12小时的数据，形成20个数据集。此外，为验证个体层面的稳定性，我们在12名测试人员上给出被试级准确率与虚警率的可视化结果（见图11），整体表现稳定，仅个别被试在特定姿态/距离下出现轻微波动。

检测结果显示，系统平均准确率超过99.05%，平均误报率低于0.5%。每个窗口的平均处理时间不超过0.05秒，远低于0.5秒的滑动步长阈值。这充分验证了系统在车内环境中的识别能力和实时处理性能。同时，从被试级观察可见，绝大多数被试的准确率达到100%、虚警率为0%，个别被试出现轻微虚警。进一步分析表明，轻微虚警主要源于两类个体化因素：其一，年龄较小的用户动作幅度与轨迹不够规范，手部难以稳定进入最佳感知区，导致有效回波特征能量偏低、与背景扰动的可分性下降；其二，身材较高的用户在乘坐过程中前后晃动幅度较大，头部短时过于接近发射/接收阵列进入声学近场，产生与目标手势相近的高能反射，从而诱发误判。

在资源占用与能耗方面，系统核心算法的低复杂度特性确保了在嵌入式车载平台的适用性：CIR估计采用频域快速互相关，复杂度 $O(L \cdot \log L)$ ，三维特征判决为线性阈值运算 $O(N)$ ，多通道决策基于子阵投票 $O(M)$ ，总计算复杂度低。理论上在主流车载嵌入式平台（如ARM Cortex-A55）上的单帧处理时间 $< 0.05s$ ，无需专用硬件加速即可满足实时性需求。能耗方面，19kHz超声波信号的持续发射功率 $\leq 5mW$ （远低于车载扬声器额定功率），多

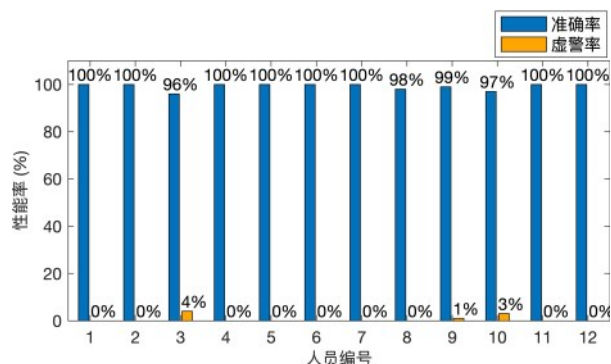


图11 不同测试人员的性能表现

通道数据处理的算力需求约0.5TOPS，对应功耗增加 $\leq 3W$ ，处于车载电子设备的常规功耗预算范围内，不会对车载电池续航产生显著影响。

此外，通过对比不同场景下的性能表现，发现系统即使在复杂声学环境中仍能保持高鲁棒性。这得益于单通道三维特征判决方法与多通道协同决策机制的共同作用。

4.3 不同座位的性能

为了测试不同座位的性能，测试人员轮流坐在1-4号座位（如图12所示），每个座位收集20分钟数据，共收集了约1.3小时的数据。



图12 测试人员在不同座位

如图13所示的实验结果表明，系统在每个座位上的检测准确率均超过98.5%，误报率低于1.5%。这表明系统通过分布式声学感知单元和多维度特征判决方法，实现了车内手势检测的全面覆盖。通过利用系统延迟消除方法和空间约束函数，系统精确划分近场区域，确保每个座位的声学单元仅对本地手势特征作出响应。

4.4 多乘客场景的性能

我们对系统在多乘客场景下的性能进行了评估，通过控制车内乘客总数从1到4，分析其他乘

①我们在实验前已获得所有测试人员的书面同意，且我们的机构审查委员会已批准整个研究活动。

客活动对声学信号传播的影响。随着乘客数量的增加,车内声学信号的传播路径因人体运动和反射而变得更加复杂,可能导致信号衰减或多径效应。

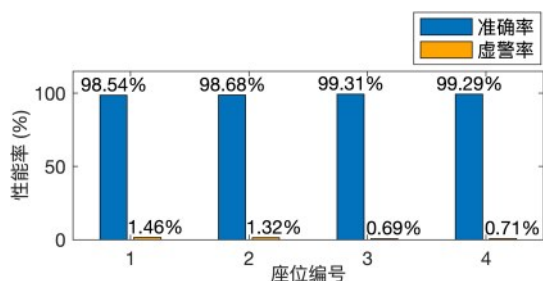


图13 不同座位的性能表现

如图14所示,实验结果表明,当乘客数量从1增加到4时,准确率在99.45%至98.68%之间保持相对稳定。这是因为我们的多通道协作决策机制通过利用空间局部性约束,有效排除了其他乘客的干扰。因此,我们通过移除该模块进行消融实验。如图14中绿色部分所示,随着乘客数量增加,准确率逐渐下降(即从87.80%降至77.92%),这表明该机制在抵御乘客活动干扰方面发挥着重要作用。

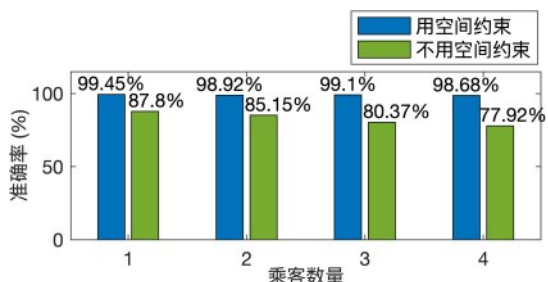


图14 多乘客场景下的性能表现

基于多乘客场景,我们进一步评估了乘客身体动作对系统的干扰效应。实验结果显示,常规身体动作(如挥动手臂、调整姿势或举手触摸头部)不会导致误判。然而,当身高较高的乘客前后大幅度摇晃身体,导致头部过于靠近扬声器时,头部对声信号的反射可能形成类似手势的特征,这可能被误认为有效手势。

为缓解此问题,可将声学单元重新布置至更接近车头方向的位置。此空间调整确保乘客头部(在正常坐姿下)难以侵入扬声器的近场感应区域,从而直接减少因头部引起的信号反射导致的误报,同时平衡车内人体工学与检测可靠性。此外,若因车厢狭窄或需大幅度体动导致头部仍易靠近扬声器,系统还支持用户自定义灵敏度阈值,通过提高对瞬

态手势特征的判定标准,增强“头部靠近”与“挥手静音”的区分度,进一步减少此类场景下的误报,提升用户体验的灵活性。

4.5 不同噪声场景的性能

由于我们基于声学信号进行手势感知,噪声对系统性能的影响是研究的重点,包括来自外部的环境噪声和扬声器播放音频产生的内部噪声。因此,我们选取四种日常噪声进行实验,即雨声、音乐、人声和交通噪声。实验中,我们设置了两种噪声的播放方式:外部噪声通过非车载扬声器如车内的手机、电脑播放;内部噪声通过车载扬声器与ZC信号同步播放。

如图15所示的实验结果表明,环境噪声与内在噪声对系统性能的影响微乎其微。这是因为日常生活中大多数噪声(如雨声、日常对话、背景音乐)的主要能量集中在低频(频率< 18 kHz),而系统的目标手势信号使用18 kHz - 20 kHz的载波频段发送,该频段自然避开了噪声能量集中的频段。即使面对汽车发动机、刹车等产生的全频谱交通噪声干扰,信号能量仍然主要集中在非目标频段。此类干扰的典型时频特性如图15中0.3到0.7 s区间所示。在如图17所示的距离-时间谱中,噪声仅干扰由多径效应引起的低能量区域。因此,通过利用三维特征判决方法中的空间局部性约束,系统能够有效识别并抑制与手势特征重叠的噪声干扰。这种抑制机制消除了非目标频率带和低能量多径区域中的噪声影响。综上所述,日常环境噪声与内部噪声对系统的手势检测性能几乎没有实质性影响。

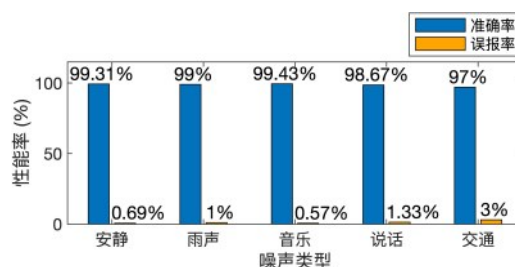


图15 不同噪声场景下的性能表现

4.6 不同路况场景的性能

在道路条件干扰实验中,我们测试了车辆在四种典型道路条件(即减速带、山路、城市街道和高速公路)下行驶时的手势识别性能。我们还进行了一项开窗/闭窗对比实验,以评估高速行驶时强风对声学信道的干扰。在实验中,测试人员从不同座

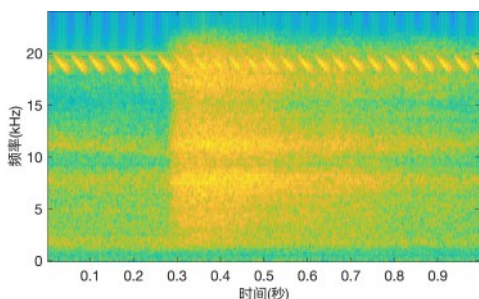


图16 高频噪声干扰场景下接收信号的频谱图

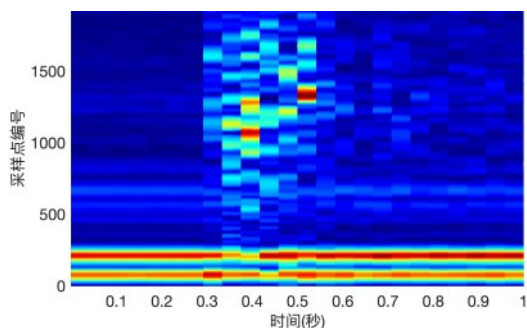


图17 高频噪声干扰场景下接收信号的距离时间谱

位执行手势动作，并在机械振动和风噪声条件下同时收集数据。

如图18所示的实验结果表明，在声学设备牢固固定的条件下，汽车机械振动对手势检测的影响较小，其准确率在所有道路条件下均保持在较高水平。尽管高速强风可能引入信道波动，但这些波动不符合手势特征所需的近场高能瞬态特性，因此不会导致误判。

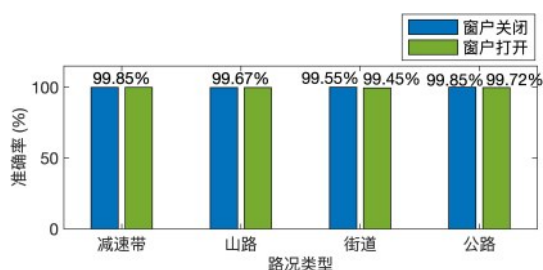


图18 不同路况场景下的性能表现(准确率)

4.7 与其他感知技术的比较

为系统性评估所提声学方案的优势与边界，我们补充开展了与UWB雷达与毫米波雷达的对比实验。两类雷达按行业常见方案布设于车顶中央（见图19、图20），与本系统在同车、同人、同工况条件下对照。测试覆盖四个座位（驾驶/副驾/后排左/后排右），并分别考察两种手势交互姿态：面向设备（正向视场）与面向车窗（自然交互位姿）；

后者与本系统的自然交互位置一致（见图21）。为确保评测公平，我们在三类设备上统一滑窗步长、判决阈值与人工审核流程，并保持同一判决模块；仅对不同传感器实施幅度归一化与带宽对齐等必要预处理，不引入任何传感器特定优化。

在上述统一协议下，我们统计了三类设备在四个座位上的平均识别准确率。如图22、图23所示，结果显示：在正向视场条件下，声学方案、UWB雷达与毫米波雷达分别达到 99.05% / 90% / 95%；而在更贴近真实使用的自然交互位姿（面向车窗）条件下，声学方案仍保持 99.05%，UWB 雷达的平均准确率显著下降至 25%，毫米波雷达则仅为 15%。这表明，当交互由“面向设备”转向更自然的“面向车窗”时，顶置雷达的有效覆盖与可见度显著受限，而声学方案在相同位姿下仍能稳定保持高识别率。

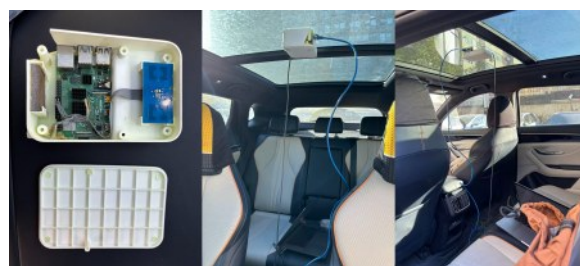


图19 超宽带雷达设备和部署示意图

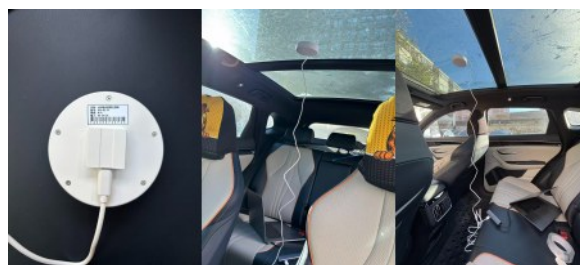


图20 毫米波雷达设备和部署示意图



图21 两种手势交互姿态

机理分析表明，上述差异主要来自视场与遮挡：顶置UWB/毫米波雷达在车内受波束指向性与头枕/躯干遮挡影响，车窗侧手势的有效散射截面与多普勒特征在传感器主瓣内的能量表达不足，导

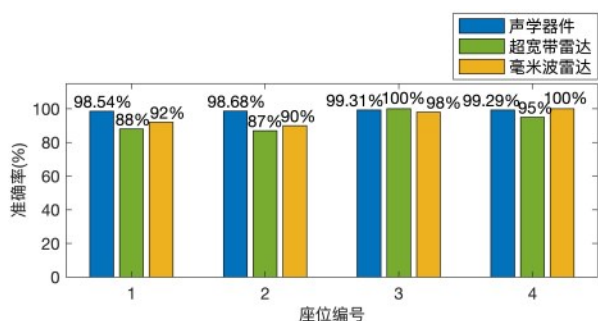


图22 正向视场下不同技术方案在4个座位上的性能表现

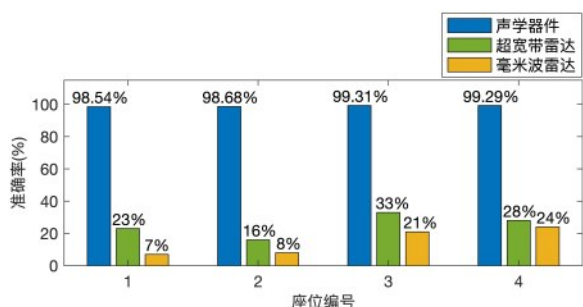


图23 自然交互位姿下不同技术方案在4个座位上的性能表现

致触发与判别性能明显下降；相对地，复用车载扬声器/麦克风的声学方案在该位姿下可依托声路扩散与多径鲁棒性，结合多通道协同决策维持稳定识别。同时，从人机可达性看，雷达在正向视场虽有一定准确率，但前排乘客需向后上方伸手进入其主要视场，使用不便且姿势不规范时波动更大；声学方案无需改变乘坐姿态即可完成交互，符合日常驾驶与乘坐习惯。

从工程实现与成本角度，UWB/毫米波方案在传感器 BOM、线束布置与 EMC 改造等方面的总体成本明显高于声学方案；后者复用量产车载扬声器/麦克风与现有音频链路，集成路径成熟，整车改动小，具备更好的量产可行性与性价比。需要指出的是，若允许对雷达进行位姿/阵列形态重构（例如在车窗侧增设侧向阵列、扩大视场或融合多传感器），其自然位姿下的识别性能有望提升，但相应会带来额外的硬件成本、供电与线束、外饰开模及 EMC 适配，工程复杂度与整车改造成本也将显著增加。

综上，在真实车内交互需求（乘客面向车窗的自然手势）下，所提声学方案在可达性（无需改变乘坐姿态）、覆盖范围与性能稳定性方面均优于车顶中央布置的 UWB/毫米波雷达方案，并兼具更低成本与更易量产集成的优势；在正向视场场景下，

雷达具备一定识别能力，但其人机可达性与自然使用一致性不足。上述结论与工程取舍为后续车规级落地提供了明确参考。

5 结论

本文提出了一种基于车内声学感知技术的手势静音控制系统，旨在克服高成本、操作范围有限及隐私保护等关键挑战。通过利用车内现有扬声器和麦克风，我们开发了一种高频声学感知技术，并结合自校准延迟机制。这些技术突破有效克服了传统解决方案中固有的成本和隐私限制。最终形成了一种低成本、轻量化、响应迅速且覆盖全面的智能车内交互解决方案，显著提升用户体验。

尽管本系统在单一手势识别与特定场景中表现优异，其功能拓展性与硬件适配性仍有提升空间。未来工作将聚焦：在手势类型与功能扩展方面，从单一静音控制向多模态手势交互延伸，支持音量调节、媒体切换等复杂操作，构建更丰富的人车交互体系；在多模态融合与自适应算法方面，结合视觉、惯性测量单元（IMU）等多源数据，提升复杂场景下的识别精度；引入机器学习模型动态优化检测阈值，增强系统对不同车型与用户习惯的适应性；在泛化验证方面，计划选取超过3款声学布局与内饰材料差异显著的车型，采用统一协议复现实验，以量化评估跨车型鲁棒性。

综上所述，本系统为车载声学手势交互提供了一种低成本、高可靠的解决方案；随着功能扩展与算法优化的推进，系统有望在智能座舱交互领域获得更广泛的应用。

参考文献：

- [1] TUNG Y-C, BUI D, SHIN K G. Cross-platform support for rapid development of mobile acoustic sensing applications[C]//Proceedings of the 16th ACM MobiSys (Annual International Conference on Mobile Systems, Applications, and Services). New York: ACM, 2018: 455-467.
- [2] WANG T B, ZHANG D Q, ZHENG Y Q, GU T, ZHOU X S, DORIZZI B. C-FMCW based contactless respiration detection using acoustic signal[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 1(4): 170.
- [3] ZHANG F S, WANG Z, JIN B H, XIONG J, et al. Your smart speaker can "hear" your heartbeat![J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2020, 4(4): 161.

- [4] QIAN K, WU C S, XIAO F, et al. Acousticcardiogram: monitoring heartbeats using acoustic signals on smart devices[C]//IEEE INFOCOM 2018-IEEE Conference on Computer Communications. Piscataway, NJ: IEEE, 2018: 1574-1582.
- [5] WANG A R, SUNSHINE J E, GOLLAKOTA S. Contactless infant monitoring using white noise[C]//Proceedings of the 25th Annual International Conference on Mobile Computing and Networking. New York: ACM, 2019: 1-16.
- [6] WAN H R, SHI S Y, CAO W Y, et al. RespTracker: multi-user room-scale respiration tracking with commercial acoustic devices [C]//IEEE INFOCOM 2021-IEEE Conference on Computer Communications. Piscataway, NJ: IEEE, 2021: 1-10.
- [7] DAI H P, WANG W, LIU A X, et al. Speech based human authentication on smartphones[C]//Proceedings of the 2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON). Piscataway, NJ: IEEE, 2019: 1-9.
- [8] SU Y Q, ZHANG F S, NIU K, et al. Embracing distributed acoustic sensing in car cabin for children presence detection[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2024, 8(1): 16.
- [9] PENG C Y, SHEN G B, ZHANG Y G, 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 (SenSys '07). New York: ACM, 2007: 1-14.
- [10] UDDIN M, NADEEM T. RF-Beep: a light ranging scheme for smart devices[C]//2013 IEEE International Conference on Pervasive Computing and Communications (PerCom). Piscataway, NJ: IEEE, 2013: 114-122.
- [11] ZHOU B, ELBADRY M, GAO R, et al. BatMapper: acoustic sensing based indoor floor plan construction using smartphones[C]//Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services. New York: ACM, 2017: 42-55.
- [12] PRADHAN S, BAIG G, MAO W G, QIU L L, CHEN G H, YANG B. Smartphone-based acoustic indoor space mapping[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(2): 75.
- [13] LIU K K, LIU X X, LI X L. Guoguo: enabling fine-grained indoor localization via smartphone[C]//Proceedings of the 11th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '13). New York: ACM, 2013: 235-248.
- [14] WANG Y T, LI J, ZHENG R, et al. ARABIS: an asynchronous acoustic indoor positioning system for mobile devices[C]//Proceedings of the 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN). Piscataway, NJ: IEEE, 2017: 1-8.
- [15] 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 (CHI '12). New York: ACM, 2012: 1911-1914.
- [16] RUAN W J, SHENG Q Z, YANG L, et al. AudioGest: enabling fine-grained hand gesture detection by decoding echo signal[C]//Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York: ACM, 2016: 474-485.
- [17] CHEN H J, LI F, WANG Y. EchoTrack: Acoustic device-free hand tracking on smart phones[C]//IEEE INFOCOM 2017-IEEE Conference on Computer Communications. Piscataway: IEEE, 2017: 1-9.
- [18] YUN S, CHEN Y - C, ZHENG H, et al. Strata: fine - grained acoustic - based device - free tracking[C]//Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services. New York: ACM, 2017: 15-28.
- [19] 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. New York: ACM, 2016: 82-94.
- [20] LING K, DAI H P, LIU Y T, et al. UltraGesture: fine - grained gesture sensing and recognition[C]//2018 15th Annual IEEE International Conference on Sensing, Communication, and Networking. Piscataway, NJ: IEEE, 2018: 1-9.
- [21] WANG Y W, SHEN J X, ZHENG Y Q. Push the Limit of Acoustic Gesture Recognition[C]//IEEE INFOCOM 2020 - IEEE Conference on Computer Communications. Piscataway: IEEE, 2020: 566 - 575.
- [22] NANDAKUMAR R, IYER V, TAN D, et al. FingerIO: using active so-nar for fine - grained finger tracking[C]//Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. New York: ACM, 2016: 1515-1525.
- [23] ABDELNASSER H, YOUSSEF M, HARRAS K A. WiGest: a ubiquitous WiFi - based gesture recognition system[C]//2015 IEEE Conference on Computer Communications (INFOCOM). Piscataway, NJ: IEEE, 2015: 1472-1480.
- [24] PU Q F, GUPTA S, GOLLAKOTA S, et al. Whole-home gesture recognition using wireless signals[C]//Proceedings of the 19th Annual International Conference on Mobile Computing & Networking (MobiCom '13). New York: ACM, 2013: 27-38.
- [25] TAN S, YANG J. WiFinger: leveraging commodity WiFi for fine-grained finger gesture recognition[C]//Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing. New York: Association for Computing Machinery, 2016: 201-210.
- [26] KELLOGG B, TALLA V, GOLLAKOTA S. Bringing gesture recognition to all devices[C]//Proceedings of the 11th USENIX Conference on Networked Systems Design and Implementation. Berkeley, CA: USENIX Association, 2014: 303-316.
- [27] WANG J, VASISHT D, KATABI D. RF-IDraw: virtual touch screen in the air using RF signals[C]//Proceedings of the 2014 ACM Conference on SIGCOMM. New York: ACM, 2014:

- 235-246.
- [28] DING H, QIAN C, HAN J S, WANG G, XI W, ZHAO K, ZHAO J Z. RFIPad: Enabling Cost-Efficient and Device-Free In-air Handwriting Using Passive Tags[C]//2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS). Piscataway: IEEE, 2017: 447-457.
- [29] REHG J M, KANADE T. Visual tracking of high DOF articulated structures: an application to human hand tracking[C]//Proceedings of the Third European Conference on Computer Vision (ECCV '94). Berlin: Springer-Verlag, 1994: 35-46.
- [30] POPOVIC B M. Generalized chirp-like polyphase sequences with optimum correlation properties[J]. IEEE Transactions on Information Theory, 1992, 38(4): 1406-1409.