arXiv:2512.16518v1 [cs.HC] 18 Dec 2025

海报：利用多任务学习识别耳内隐藏式私钥，实现可靠的无声语音接口

Xuefu Dong
Liqiang Xu
东京大学
日本东京

Lixing He
香港中文大学
香港
中国香港特别行政区，中国

Zengyi Han*
大连海事大学
中国大连
zyhan@dlmu.edu.cn

Ken Christofferson
多伦多大学
加拿大安大略省多伦多

Yifei Chen
清华大学
中国北京

Akihito Taya
东京大学
日本东京

Yuuki Nishiyama
东京大学
日本千叶

Kaoru Sezaki
东京大学
日本千叶

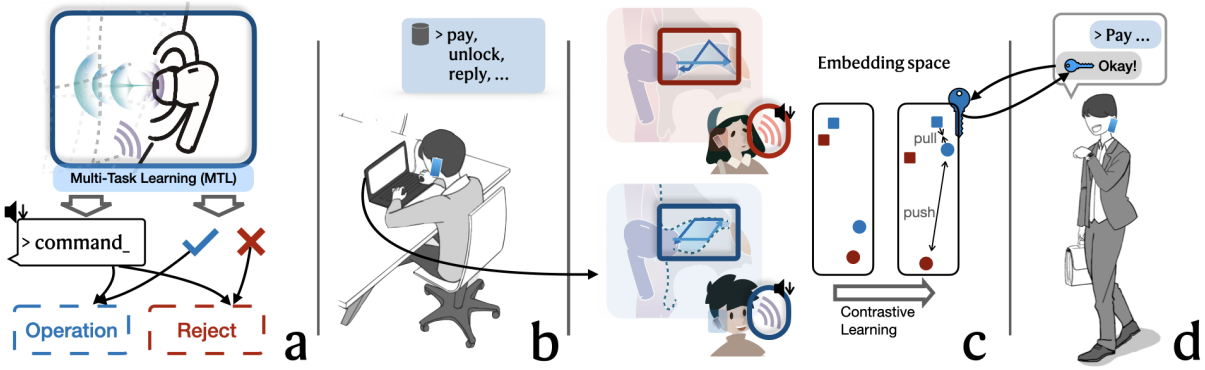


图 1：（a）HEar-ID 使用多任务学习（MTL）来验证用户身份并可靠地推断无声语音；（b）新用户通过录制少量无声指令样本在 HEar-ID 中注册；（c）HEar-ID 利用对比学习，拉近耳语与 ECDM 投影特征之间的余弦相似度；（d）用户可以通过自然且安全的方式完成验证。

摘要

无声语音接口（SSI）可实现免提输入，无需通过听觉发声，但大多数无声语音接口系统无法验证说话人身份。我们提出 HEar-ID，它利用消费级主动降噪耳塞采集低频“耳语”音频和高频超声反射。两种信号流的特征均经过共享编码器，生成的嵌入向量分别输入用于用户认证的对比学习分支和用于静默拼写识别的 SSI头。该设计可解码 50 个单词，同时可靠地拒绝冒充者；整个系统仅需在普通耳塞上运行单一模型。实验表明，HEar-ID 具有较高的拼写准确率和稳健的认证能力。

CCS 概念

• 安全与隐私 → 生物识别；• 以人为本的计算 → 文本输入；基于声音的输入／输出。

关键词

无声语音接口、用户认证、对比学习、多任务学习、耳戴式计算、声学感知

*Zengyi Han 为通讯作者。仅供个人使用；不得再分发。正式版本将发表于 UbiComp Companion '25。https://doi.org/10.1145/3714394.3754429

ACM 引用格式：
Xuefu Dong、Liqiang Xu、Lixing He、Zengyi Han、Ken Christofferson、Yifei Chen、Akihito Taya、Yuuki Nishiyama 和 Kaoru Sezaki。2025。海报：使用多任务学习识别隐藏于耳内的私钥，实现可靠的无声语音接口。载于 *2025*年 *ACM*普适与泛在计算国际联合会议（*UbiComp Companion '25*）伴随会议，*2025*年*10*月*12*日至*16*日，埃斯波，芬兰。ACM，纽约，纽约州，美国，5页。https://doi.org/10.1145/3714394.3754429

1 引言

尽管基于语音的交互为手动或视觉控制提供了一种自然的替代方式，但在公共场所并非总是可行：在安静的环境（如图书馆）中大声说话可能会打扰他人或泄露私人信息（例如口述短信）。无声语音接口（SSI）通过依靠口部和面部动作而非可听见的发声来应对这些问题，使用户在嘈杂和安静的场所都能使用，且不会打扰旁人。因此，在公共场合，用户往往更倾向于使用 SSI，而非传统语音识别 [26]。然而，先前与 SSI 相关的研究 [4, 30, 37, 39] 几乎没有考虑安全问题。虽然常规的基于语音的身份验证会在授予访问权限前验证说话者身份，从而降低这些风险，但它仍容易受到重放攻击和注入攻击（例如通过扬声器触发 Siri），因此亟须开发可靠的无声语音接口系统。

我们分析发现，无声语音识别任务与说话人认证任务彼此相关，而非相互独立。每个人耳道固有结构的独特性会形成不同的声传播路径，因此细微的耳道形变同时编码了话语内容和说话人身份。幸运的是，近年来耳戴式感知取得了显著进展，涵盖活动识别 [17, 23, 28], 、姿势或网格重建 [6, 10, 21, 25], 、语音增强 [16, 18, 19], 以及用户认证 [7, 8, 11, 31]。其中，有若干研究采用带入耳式麦克风的普通耳塞，通过超声感知检测耳道动态形变（ECDM），并将耳道用作反射腔。因此，我们在耳戴式平台上采用多任务学习（MTL）方法，利用这种相关性，同时实现静默拼写识别和说话人认证。

在这项工作中，我们提出 HEar-ID，实现可靠的 SSI。该系统仅利用一副市售主动降噪耳塞式耳机发射不可听的 OFDM信号，并记录超声反射和耳语音频，从而支持静默拼写输入（例如 /iː eɪ ɑr/ 表示单词"ear"）和基于单个机器学习模型的用户验证。我们将无声语音视为等同于耳语，因为我们发现，大多数实验参与者由于缺乏训练，仍会发出微弱的语音。如图 1 所示，HEar-ID 从超声波形（17.5–23kHz）和耳语（0-11kHz）中提取特征，再将这两组特征输入共享编码器（时序卷积网络→双向门控循环单元→多层感知机），生成 d维嵌入向量。对比头对齐真实用户耳语—超声配对样本，并排斥攻击者样本；而针对拼接后嵌入向量的 SSI头则支持对静默拼写进行词级解码。在初步实验中，HEar-ID 始终取得了有希望的结果：对于 11 名参与者，系统以 3.2% 的假阳性率（FPR）和真正例率（TPR）可靠地拒绝了冒充者；其中 8 名参与者的 Top-1词语识别准确率达到 90.25%。

我们的贡献如下：(1) 一种多任务 SSI 框架，将超声特征与耳语特征融合，用于联合拼写和认证；(2) 一种对比学习方法（CLWUM），将耳语与超声波信号嵌入"私钥"空间，以实现鲁棒验证；(3) 评估结果表明，该方法可准确拼写 50 个单词，并具备出色的认证性能。

2 相关工作

2.1 超声无声语音接口

通过多普勒效应 [36] 或其他机器学习技术 [2, 3, 24] 解读的超声感知，可以检测到与射频- [22, 35], 光学- [29, 32] 和基于运动的传感器 [14, 15] 捕捉到的相似动作，从而实现无声语音接口（SSI）技术。可穿戴方案包括 Jin等人 [20]的 EarCommand，该方案使用定制耳塞发射调频连续波啁啾信号，并通过耳道内的麦克风记录反射信号，在 12 名参与者无声说出的 32 个词语上达到 89.9% 的准确率。

2.2 基于耳戴设备的用户认证

可穿戴设备认证可以采用密码——例如基于语音或手势的方案——在耳塞式耳机 [1, 33]上实现；但口述密码可能泄露个人数据，也容易遭受重放攻击，因此促使人们寻求生物特征识别替代方案。

然而，几乎所有现有系统都要求用户使用正常语音或做出可检测动作，这限制了它们在公共场合或安静环境中的实用性。据我们所知，HEar-ID 是首个可通过无声语音或耳语进行认证的耳戴设备方案。与仅依赖超声反射的 EarDynamic [31], 不同，我们学习超声波信号与耳语信号之间的联合映射，即使没有可听见的语音，也能捕捉用户身份。

3 对比多任务学习（CMTL）

本节将介绍 HEar-ID 用于词语推断和用户验证的方法，如图 2 所示。

3.1 信号处理与特征提取

HEar-ID 使用采样率为 48000 Hz、包含 2046 个采样点的 OFDM信号，持续约 42 ms。接收到来自耳道的反射信号后，HEar-ID 首先使用高通滤波器分离出超声频带（17.5–23 kHz），然后将每个接收片段对齐到一个精确的 OFDM周期。粗对齐会寻找时延t_{coarse}，使一个发射符号与两个接收符号之间的互相关达到最大值。随后，精细对齐通过最小化与发射信号之间的相位差，校正亚采样延迟（±25 个采样点）。原始音频则根据结果索引进行对齐。

我们将数据切分为 426 ms（10 帧）的滑动窗口，步长为 85 ms（2 帧），并从中提取特征。每个窗口的超声频带包含耳道动态运动（ECDM）的信息；为补偿商用耳塞式耳机在 16 kHz 以上频率的快速衰减，我们进一步对其进行二阶微分增强，随后用 200 阶自回归（AR）系数表示。另一方面，我们对对齐后的原始音频再应用一个 12 kHz 低通滤波器，以获取耳语音频。为降低计算复杂度，我们将音频降采样至原采样率的一半，然后将其输入广泛采用的梅尔频谱图提取器（$n_fft = hop_length \approx 42ms$）。随后将配对特征输入我们的神经网络。

3.2 耳语—超声映射的对比学习（CLWUM）

当用户无声发音时，面部动作会同时在低频耳语信号（由耳塞式耳机麦克风采集）和耳道内的高频超声反射中引发相关扰动。CLWUM 利用这种内在联系，将配对的耳语与超声信号映射到共享嵌入空间，从而有效充当"私钥"，仅对齐真实用户嵌入。

如图 2 所示，每个批次为每种模态提供 $2N$ 个片段：前 N 个来自真实 用户，接下来的 N 个来自攻击者 （实验中将 N 设为 50 个词）。经过模态专用编码器 $f_w(\cdot)$（耳语）和 $f_u(\cdot)$（超声波）以及小型

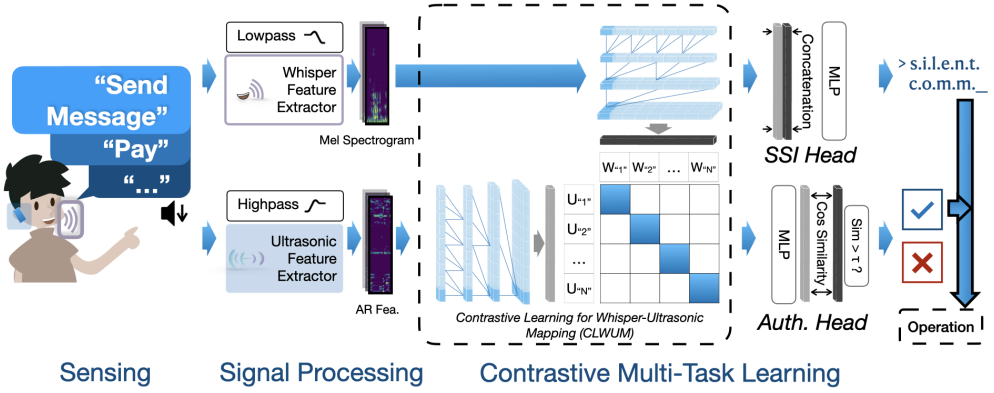


图 2：HEar-ID 的三阶段工作流程。

投影头 $h_w(\cdot)$、$h_u(\cdot)$，我们得到嵌入向量

$$w_i = h_w\big(f_w(x_i^{\text{whisper}})\big), \quad u_j = h_u\big(f_u(x_j^{\text{ultra}})\big), \quad i, j = 1, \ldots, 2N,$$

我们将其收集为$W(d\times 2N)$，以及 $U(d\times 2N)$。由于 CLWUM 专注于真实用户映射，因此我们仅提取真实用户的前 N 个嵌入向量：$W^g(d\times N)$ $U^g(d\times N)$。随后，我们计算来自 N真实用户样本 $S_{ij} = sim(w_i,\ u_j)$的耳语嵌入与超声嵌入之间的余弦相似度数组。这里，对角线上的值 S_{ii} 代表真实的耳语—超声波配对。对比学习能促进超声嵌入与耳语嵌入之间稳健的跨模态对齐，以实现静默拼写识别，同时提取支撑可靠用户认证的共享特征。因此，我们定义对比损失：

$$L_{\text{CL}} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp\left(S_{i,i}/\tau\right)}{\sum_{j=1}^{N}\exp\left(S_{i,j}/\tau\right)},$$

其中，$\tau > 0$ 是温度超参数，根据经验设为 0.7。最小化 L_{CL} 会提高真实配对的相似度，并降低不匹配配对的相似度，从而将来自真实用户的耳语与超声信号嵌入到一个连贯的共享空间中。

3.3 基于多任务学习的可靠静默拼写界面

为同时验证用户身份并识别无声语音中的拼写词，我们采用多任务学习（MTL）范式。多任务学习旨在利用相关任务之间的共享表示和有用信息，提升多个任务的整体性能 [38]。我们的架构分为两个任务——用户认证和静默拼写识别，并以 CLWUM 编码器作为共享基础。

3.3.1 用户认证。如图 2 所示，每个耳语或超声波片段首先经过 CLWUM 编码器——由三个堆叠的 TCN模块和两个双向门控循环单元层组成——然后经过嵌入多层感知机，生成一个 d维表示。随后，通过轻量级模态专用头对这一共享嵌入进行进一步投影，得到认证向量。对于每个批次，令 h_i^{whisper} 和 h_i^{ultra}分别表示对应于第 i个真实用户静默拼写的共享维嵌入（其中 h_j^{ultra}表示第 j个攻击者片段的嵌入（其中 $i,\ j = 1,\ \ldots\ldots,\ N$，且每个批次有 N个不同的词）。然后，在认证头中分别对超声波和耳语数据应用两个多层感知机，计算认证向量。接着，系统计算这对向量的余弦相似度，并在相似度高于阈值 thr时判定认证通过。在训练期间，我们从真实用户和攻击者的数据集中构造配对：

$$w_i = \text{MLP}_w\big(h_i^{\text{whisper}}\big), \quad u_i^{+} = \text{MLP}_u\big(h_i^{\text{ultra}}\big), \quad u_j^{-} = \text{MLP}_u\big(h_j^{\text{ultra}}\big),$$

其中 $w_i,\ u_i^{+},\ u_j^{-} \in \mathbb{R}^d$。这里，$(w_i,\ u_i^{+})$ 构成真实用户配对（两者嵌入的是同一个词），而每个 u_j^{-} 都表示该词对应的攻击者嵌入向量。

接着，我们计算余弦相似度 $s_i^{+} = sim(w_i,\ u_i)$ 用于衡量真实配对的相似度，$s_i^{-} = sim(w_i,\ u_{i^*}^{-})$ 则用于衡量真实用户与攻击者配对的相似度。训练过程通过带间隔 m 的角度三元组损失进行优化（其中 m 设为 11.45）：

$$L_{\text{auth}} = \frac{1}{N}\sum_{i=1}^{N}\max\Big(0, m + arccos(s_i^{+}) - arccos(s_i^{-})\Big).$$

该损失会缩小真实配对之间的夹角（提高相似度），同时增大最难负样本（攻击者）之间的夹角，从而形成清晰的决策边界。我们使用 Youden 的 J 统计量 [34]，对 thr进行校准，并采用经验值 m = 30 度。

3.3.2 静默拼写识别。如图 2 所示，对于静默拼写，首先将真实用户耳语和超声波片段输入共享的 CLWUM 编码器（时序卷积网络 → 双向门控循环单元 → 嵌入向量 MLP），得到两个由 d 维向量组成的序列：$h^{\text{whisper}} = \left[h_1^{w}, \ldots, h_N^{w}\right]$、$h^{\text{ultra}} = \left[h_1^{u}, \ldots, h_N^{u}\right]$.。随后，我们在每个时间步拼接对应的耳语嵌入向量和超声波嵌入向量，得到 $c_i = \left[h_i^{w} \,\|\, h_i^{u}\right]$，并将拼接后的序列 $\{c_i\}$ 输入一个小型投影多层感知机，再接一个 Softmax 分类器，以生成字母表（26 个字母 ＋空白符）上的帧级逻辑值：$\ell_i = \text{MLP}^{\text{spell}}(c_i)$，其中 $K = 27$。将整个逻辑值序列记为 $\mathbf{L} = [\ell_1, \ldots, \ell_N]$。

我们使用联结主义时序分类（CTC）损失 [13] 来对齐$\mathbf{L}$与目标字母序列$\mathbf{y}$，无需帧级标签：$L_{\text{CTC}} = -\log\ P(\mathbf{y}\ |\mathbf{L})$. CTC使模型能够学习

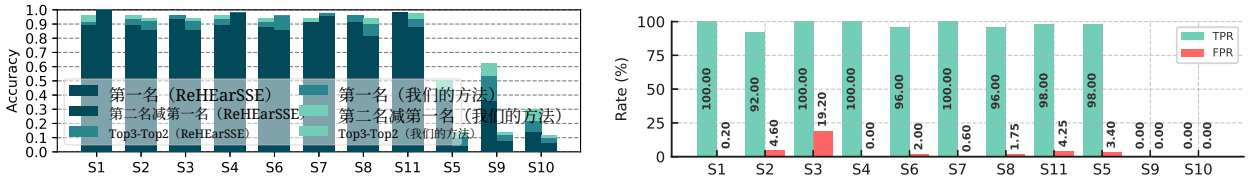


图 3：（a）词语推断准确率，以及（b）用户身份验证性能。

字母身份及其时间转移，这对于解码无声口述的单词至关重要。

3.4 身份验证流程

身份验证处理流程分为两个阶段：

注册阶段。 每位用户佩戴耳塞式耳机，先默默拼读一小组单词样本。随后，HEar-ID 使用该用户的耳语—超声波配对样本以及其他用户的数据来训练网络。我们认为，制造商可以随产品一并提供这些额外的训练数据。HEar-ID 通过最小化组合损失进行端到端训练

$$L_{\text{total}} = \alpha L_{\text{CL}} + \beta L_{\text{auth}} + \gamma L_{\text{CTC}},$$

超参数调优后，任务权重设为 $\alpha = 0.1$、$\beta = 0.5$ 和 $\gamma = 0.3$。随后校准验证阈值。

验证阶段。 用户默默拼写单词时，系统会捕获相应的耳语和超声波片段，计算它们的嵌入向量，并将其与存储的参考集进行比较。如果与至少一个参考嵌入向量的相似度超过阈值，则接受该用户；否则，拒绝其访问权限。同时，拼接后的嵌入向量会发送至预测头，由其生成逻辑值，再通过解码（例如，使用束搜索）得出最终拼写的单词。

4 初步结果

我们使用了 ReHEarSSE [5] 数据集的一个子集（10 名参与者），并另外招募了一名参与者，使研究总人数达到 11 人（9 名男性、2 名女性；平均年龄 22 岁）。所有数据均使用市售主动降噪耳塞式耳机（漫步者 W380NB）采集；我们将耳塞式耳机中的入耳式麦克风从电路中拆下，并通过 3.5毫米插孔直接连接到智能手机（Google Pixel 3a），以确保获得未经处理的原始音频。所有数据采集程序均遵循机构审查委员会（IRB）的指导方针。所有参与者都使用微弱语音（例如，/iː eɪ ɑr/，表示单词"ear"）对同一词汇表进行了 4 轮拼写；该词汇表包含从《牛津英语词典》最常用的 1000 个单词中随机选出的 50 个词 [27]。

4.1 整体性能

4.1.1 实验设计：我们轮流将一名参与者视为真实用户，其余参与者视为冒充者。对于每位参与者，我们随机选取一个重新佩戴会话作为测试集，并将除留出会话之外的所有其他会话作为训练集。认证头和 CLWUM 均使用所有参与者的数据进行训练和测试。这意味着在测试期间，每位参与者的模型会受到 500 次攻击，并被真实用户访问权限调用 50 次。对于 SSI 头，我们仅使用真实用户的数据，从而得到一个用户相关模型。我们实现了 [5] 的处理流程，作为基线。

4.1.2 指标。对于识别，我们报告*Top–n* 词准确率（n=1，2，3），即真实标签出现在前-n 个预测结果中的比例。对于认证，我们采用标准的**TPR**（真正例率）、**FPR**（假正例率）及由此得出的**EER**。图3展示了我们的**耳语+超声波**静默语音接口（SSI）的识别准确率，并与仅使用超声波的基线 ReHEarSSE [5], 进行了比较；图3则汇总了各用户的认证结果（TPR/FPR）。

静默语音识别： 我们的耳语 +超声波SSI平均 Top-1 准确率达到 **67.3 %**。不过，我们发现，在11名用户中，有8名用户的超声感知性能未出现下降，其 Top-1 准确率达到90.25%，与仅使用超声波的 ReHEarSSE 基线（**91.4%**）相当。3名说话者（S1、S4、S7）的准确率超过90%，证实了耳语流的优势。S5、S9和S10的表现较差（<10 %）——这可能是由于发音不清晰以及个体特异的传感器放置位置所致，而耳语与超声波对齐会放大这些问题。

认证： 平均性能为**真正例率** = **81.76 %**、**假正例率** = **3 2** **5** **99 9** **3 4** **.%**。值得注意的是，尽管 S 的识别效果较差，仍取得.%/.%，这表明即使词汇解码失效，我们的身份嵌入依然稳健。相反，S9 和 S10 的真正例率接近于零，与其较低的识别得分相吻合，说明其在不同会话中的耳机佩戴位置不一致。

总体而言，结果证实，将耳语麦克风与超声感知相结合，能够实现与最先进的声学静默语音接口相当的识别准确率，同时保持较高的认证鲁棒性，从而通过单一可穿戴设备和机器学习模型，兼顾内容检索与用户验证这两个目标。

5 结论

我们推出了 HEar-ID，它结合 AR超声波和耳语特征，可在耳塞式耳机上实现静默拼写和用户认证。CLWUM模块通过对齐真实用户嵌入并排斥冒充者，同时支持两项任务。测试表明，该方法在真实场景下具有稳健的认证性能；在 11 名参与者中的 8 名身上，对 50 个词的拼写准确率达到 90.25%，表现出良好前景。未来工作将扩充词汇表并改进持续验证，同时利用生成模型等方法生成逼真的合成数据 [9, 12]。

致谢

本研究获得了日本学术振兴会科学研究费助成事业项目（编号：JP24K20759）以及中央高校基本科研业务费（3132025240，中国）的支持。

参考文献

[1] Tao Chen、Yongjie Yang、Chonghao Qiu、Xiaoran Fan、Xiuzhen Guo 和 Longfei Shangguan。2024。通过耳机实现免提语音助手激活。载于第 22 届国际移动系统、应用与服务年度会议论文集。155–168。[2] Haiming Cheng 和 Wei Lou。2022。PD-FMCW：利用调频连续波中的相位差，将无设备声学感知推向极限。《IEEE移动计算汇刊》22，8（2022），4865–4880。[3] Haiming Cheng、Wei Lou、Yanni Yang、Yi-pu Chen 和 Xinyu Zhang。2023。TwinkleTwinkle：通过眨眼与智能设备交互。《ACM交互式、移动、可穿戴和普适技术》7，2（2023），1–30。[4] Debadatta Dash、Evan Kittle、Isabel Gerrard、Richard Csaky、Gabriel Gonzalez、David Taylor、Juan Pablo Llinas、Dominic Labanowski、Nishita Deka 和 Richy Yun。2025。利用可穿戴磁力计进行无声语音识别。bioRxiv（2025），2025–08。[5] Xuefu Dong、Yifei Chen、Yuuki Nishiyama、Kaoru Sezaki、Yuntao Wang、Ken Christofferson 和 Alex Mariakakis。2024。ReHEarSSE：识别耳内隐藏的无声拼写表达。载于 2024 年计算机系统人因大会论文集。1–16。[6] Di Duan、Shengzhe Lyu、Mu Yuan、Hongfei Xue、Tianxing Li、Weitao Xu、Kaishun Wu 和 Guoliang Xing。2025。Argus：基于精简型可穿戴毫米波扩展设备的多视角自我中心人体网格重建。载于第 23 届 ACM 嵌入式网络传感器系统会议论文集。1–14。[7] Di Duan、Zehua Sun、Tao Ni、Shuaicheng Li、Xiaohua Jia、Weitao Xu 和 Tianxing Li。2024。F2key：基于商用现成耳机将面部动态转换为私钥，以实现可靠的语音交互。载于第 22 届国际移动系统、应用与服务年度会议论文集。127–140。[8] Yongjian Fu、Ke Sun、Ruyao Wang、Xinyi Li、Ju Ren、Yaoxue Zhang 和 Xinyu Zhang。2025。利用商用现成无线耳塞中的耳内心冲击图实现心脏监测。arXiv 预印本 arXiv:2501.06744（2025）。[9] Yongjian Fu、Shuning Wang、Linghui Zhong、Lili Chen、Ju Ren 和 Yaoxue Zhang。2022。SVoice：通过商用设备上的声学感知实现静默语音通信。载于第 20 届 ACM 嵌入式网络传感器系统会议论文集。计算机协会，纽约州纽约，美国，622–636。[10] Tian Gao、Xuefu Dong、Akihito Taya、Yuuki Nishiyama 和 Kaoru Sezaki。2025。使用耳塞和超声波基于耳道形状检测进行表情识别。载于电子信息通信学会会议档案。电子信息通信学会。[11] Yang Gao、Wei Wang、Vir V Phoha、Wei Sun 和 Zhanpeng Jin。2019。EarEcho：利用耳道回声进行可穿戴设备认证。《ACM交互式、移动、可穿戴和普适技术》3，3（2019），1–24。[12] Chen Gong、Bo Liang、Wei Gao 和 Chenren Xu。2025。数据能够自我发声：无线合成数据的质量引导的利用。arXiv 预印本 arXiv:2506.23174（2025）。[13] Alex Graves。2012。联结主义时序分类。Springer Berlin Heidelberg，柏林，海德堡，61–93。doi:10.1007/978-3-642-24797-2_7[14] Zengyi Han、Xuefu Dong、Liqiang Xu、Zhen Zhu、En Wang、Yuuki Nishiyama 和 Kaoru Sezaki。2024。RideGuard：利用智能手机进行微型交通工具转向动作预测。载于 2024 IEEE 第 44 届分布式计算系统国际会议（ ICDCS ）。IEEE，1039–1049。[15] Lixing He、Yunqi Guo、Haozheng Hou 和 Zhenyu Yan。2025。VibOmni：面向可扩展的耳戴设备骨传导语音增强。arXiv 预印本 arXiv:2512.02515（2025）。[16] Lixing He、Haozheng Hou、Shuyao Shi、Xian Shuai 和 Zhenyu Yan。2023。面向头戴式可穿戴设备的骨传导振动语音增强。载于第 21 届国际移动系统、应用与服务年度会议论文集。计算机协会，纽约州纽约，美国，14–27。[17] Lixing He、Bufang Yang、Di Duan、Zhenyu Yan 和 Guoliang Xing。2025。EmbodiedSense：利用耳机理解具身活动。arXiv 预印本 arXiv:2504.02624（2025）。[18] Hirotaka Hiraki、Shusuke Kanazawa、Takahiro Miura、Manabu Yoshida、Masaaki Mochimaru 和 Jun Rekimoto。2024。用于抑制噪声的导电织物振膜耳机麦克风。载于第 37 届 ACM 用户界面软件与技术年度研讨会附属论文集（美国宾夕法尼亚州匹兹堡）（UIST Adjunct '24）。计算机协会，纽约州纽约，美国，文章 56，3 页。doi:10.1145/3672539.3686768 [19] Hirotaka Hiraki、Shusuke Kanazawa、Takahiro Miura、Manabu Yoshida、Masaaki Mochimaru 和 Jun Rekimoto。2024。WhisperMask：用于耳语语音的抑噪口罩式麦克风。载于 2024 年增强人类国际会议论文集（澳大利亚维多利亚州墨尔本）（AHs '24）。计算机协会，纽约州纽约，美国，

1 14 10 1145 3652920 3652925 –。[20] doi: 。/。Yincheng Jin、Yang Gao、Xuhai Xu、Seokmin Choi、Jiyang Li、Feng Liu、Zhengx-iong Li 和 Zhanpeng Jin。2022。EarCommand：“听见”你的无声语音

Commands In Ear。《ACM交互式、移动、可穿戴与泛在技术》6，2（2022），1–28。[21] Ke Li、Ruidong Zhang、Bo Liang、François Guimbretière 和 Cheng Zhang。2022。EarIO：一种低功耗声学感知耳戴设备，可持续追踪细致的面部动作。《ACM交互式、移动、可穿戴与泛在技术论文集》6，2，文章 62（2022年7月），24 页。doi:10.1145/3534621[22] Wenwei Li、Jiarun Zhou、Jie Xiong、Yuhui Xie、Leye Wang、Duo Zhang 和 Daqing Zhang。2025。重新审视基于 WiFi 的角度估计，以实现稳健的被动室内定位。《ACM交互式、移动、可穿戴与泛在技术论文集》9，4（2025），1–28。[23] Shengzhe Lyu、Yongliang Chen、Di Duan、Renqi Jia 和 Weitao Xu。2024。Earda：实现准确且数据高效的耳戴设备活动感知。载于《2024年IEEE人工智能物联网系统感知与计算耦合会议（CSCAIoT）》。[24] Wenguang Mao、Wei Sun、Mei Wang 和 Lili Qiu。2020。DeepRange：通过深度学习实现声学测距。《ACM交互式、移动、可穿戴与泛在技术论文集》4，4（2020），1–23。[25] Vimal Mollyn、Riku Arakawa、Mayank Goel、Chris Harrison 和 Karan Ahuja。2023。IMUPoser：利用手机、手表和耳塞中的惯性测量单元进行全身姿态估计。载于《2023年计算机系统人因大会论文集》（德国汉堡）（CHI '23）。计算机协会，纽约，美国纽约州，文章 529, 12 10 1145 3544548 3581392 页。

[26] 2021 doi: . / . Laxmi Pandey、Khalad Hasan 和 Ahmed Sabbir Arif。接受私人和公共场合语音输入方法与无声语音输入方法的程度。载于《2021年计算机系统人因大会论文集》。计算机协会，纽约，美国纽约州，1–13。[27] 牛津大学出版社。2023。《牛津英语词典在线版》。牛津大学出版社，牛津。https://www.oed.com/search/advanced/Entries? textTermOpt0=WordPhrase&dateOfUseFirstUse=false&page=1&sortOption=Frequency [28] Kazuki Shimojo、Zengyi Han、Yuuki Nishiyama 和 Kaoru Sezaki。2023。初步研究：使用耳戴设备检测街道步行过程中的视觉搜索行为。载于《2022年ACM普适与泛在计算国际联合会议暨2022年ACM可穿戴计算国际研讨会附属会议论文集》（英国剑桥）（UbiComp/ISWC '22 Adjunct）。计算机协会，纽约，美国纽约州，254–257。[29] Zixiong Su、Shitao Fang 和 Jun Rekimoto。2023。LipLearner：移动设备上的可定制无声语音交互。载于《2023年计算机系统人因大会论文集》（德国汉堡）（CHI '23 696, 21 10 1145 3544548 3581465）。计算机协会，纽约，美国纽约州，文章页码。doi: . / .[30] Xue Sun、Jie Xiong、Chao Feng、Haoyu Li、Yuli Wu、Dingyi Fang 和 Xiaojiang Chen。2024。EarSSR：通过耳机实现无声语音识别。《 IEEE 移动计算汇刊》在线优先发表（

2024 1 17 10 1109 2024 3356719 ），–。[31] 2021 doi: . /TMC. . Zi Wang、Sheng Tan、Linghan Zhang、Yili Ren、Zhi Wang 和 Jie Yang。耳道形变驱动的入耳式可穿戴设备连续用户认证。《ACM交互式、移动、可穿戴与泛在技术论文集》5，1（2021），1–27。[32] Jingwen Wei、Lupeng Zhang、Jingchi Zhang、Bin Han 和 Lei Wang。2024。UniQR：基于设备姿态和环境匹配的安全二维码支付方案。载于《2024年第21届IEEE感知、通信与网络国际年会（SECON）》。IEEE，1–9。[33] Xuhai Xu、Haitian Shi、Xin Yi、WenJia Liu、Yukang Yan、Yuanchun Shi、Alex Mariakakis、Jennifer Mankoff 和 Anind K. Dey。2020。EarBuddy：利用无线耳塞实现面部交互。载于《2020年计算机系统人因大会论文集》（美国夏威夷州檀香山）（CHI '20）。计算机协会，纽约，美国纽约州，

1 14 10 1145 3313831 3376836 –。doi: . /

[34] 1969 300 301 1969 . WJ Youden。统计技术。NBS 特别出版物 - （），421。[35] Shang Zeng、Haoran Wan、Shuyu Shi 和 Wei Wang。2023。mSilent：利用商用毫米波雷达实现通用语料无声语音识别。《ACM交互式、移动、可穿戴与泛在技术论文集》7，1（2023），1–28。[36] Chuanxin Zhang、Xue Jiang 和 Dean Ta。2024。揭示声学旋转多普勒效应中与入射角无关的频移。《物理评论快报》。

132 2024 114001 11 10 1103 132 114001 （3月），。[37] 期号 。doi: . /PhysRevLett. . Ruidong Zhang、Ke Li、Yihong Hao、Yufan Wang、Zhengnan Lai、François Guimbretière 和 Cheng Zhang。2023。EchoSpeech：通过声学感知技术，在低干扰眼镜上实现连续无声语音识别。载于《2023年计算机系统人因大会论文集》。计算机协会，纽约，美国纽约州，1–18。[38] Yu Zhang 和 Qiang Yang。2018。多任务学习综述。《国家科学评论》5，1（2018），30–43。[39] Juntao Zhou、Dian Ding、Yijie Li、Yu Lu、Yida Wang、Yongzhao Zhang、Yi-Chao Chen 和 Guangtao Xue。2025。M2SILENT：通过共享空间中的多方向扬声器实现多用户无声语音交互。载于《2025年计算机系统人因大会论文集》。1–19。