arXiv:2301.08925v1 [eess.AS] 21 Jan 2023

语音与音频内容隐私面临的新挑战

Jennifer Williams [1], *Karla Pizzi* [2, 3], *Shuvayanti Das* [4], *Paul-Gauthier No´e* [5]

[1]南安普顿大学电子与计算机科学系，英国[2]弗劳恩霍夫 AISEC，德国加兴；[3]慕尼黑工业大学，德国[4]Aflorithmic Labs Ltd.，英国[5]阿维尼翁大学阿维尼翁计算机科学实验室（LIA）´，法国j.williams@soton.ac.uk, karla.pizzi@aisec.fraunhofer.de, shuvayanti@gmail.com, paul-gauthier.noe@univ-avignon.fr

摘要

语音与音频中的隐私涉及多个方面。该领域中一个尤其尚未得到充分研究的隐私问题，是对与内容和语境相关信息的考量。语音内容可能包括词语及其含义，甚至风格标记、病理性语音、语调模式或情绪。更广泛地说，在真实环境中采集的音频可能包含背景语音，或暴露位置标记、房间特征、副语言声音或其他可听事件等语境信息。音频录制设备和语音技术在日常生活中正变得越来越普遍。与此同时，商业化的语音与音频技术并未为消费者提供多样的隐私选择。即使隐私受到法规约束或法律保护，隐私保障与执行方面的技术解决方案仍不完善。本文立场论文提出了语音与音频内容隐私领域三个重要且亟待研究的挑战。我们着重指出当前的不足与机遇，并确定若干重点领域，这些领域可能对开发合乎伦理且更安全的语音技术产生重大影响。

索引词：语音隐私、内容屏蔽、隐私评估、语音识别、说话人识别

1. 引言

尽管隐私与安全都是数字时代个人保护的核心，但音频数据中的隐私概念并非泾渭分明 [1]。隐私与安全常常被混为一谈，尤其是在法律领域 [2]。本文采用的立场是，**隐私**与控制信息访问权限有关，而**安全**则涉及这些信息可能如何被使用（或滥用）。如何确保任何数据集的数据隐私，尤其是语音与音频数据集的数据隐私，仍是一个开放问题。音频数据中可能包含许多不同类型的私人信息。背景噪声可能透露地理位置，录音中可能录下未经同意分享其语音数据的人的声音，甚至还可能包含与个人相关的信息。本文尤其关注语音内容隐私，不过我们也认识到，语音隐私与更广义的音频处理密不可分。

早期针对语音隐私的努力，采用了使用无意义的掩蔽噪声来模糊语音信息的技术，从而有效地使人类窃听者无法听懂语音 [3]。这一噪声掩蔽的基础理念通常假设人类说话者位于房间内（且已知房间特征），并且希望隐藏整段对话。噪声掩蔽这一理念延续至今，近来还被用于保护智能手机采集的音频的隐私 [4]。从概念上说，噪声掩蔽与通过麦克风干扰实现隐私的方式有关 [5]。然而，这两种技术（噪声掩蔽和干扰）都对音频采集条件以及预期的窃听者（更准确地说，是攻击向量）作出了假设。这些技术还旨在提供一种全面隐私，既不允许精细控制，也不给对话者选择隐私级别的机会。



图1：通过技术掩蔽方案隐藏或掩盖音频内容的概念图。左侧：未使用掩蔽器的语音识别；右侧：使用掩蔽器的语音识别。包含出生日期信息的音频部分经过遮蔽，使自动语音识别系统无法将其转录出来。

本文关注语音与音频中的内容隐私。我们使用内容 一词，以区别于较狭义的说话人声音特征定义，即内容/语境信息。这样做旨在区分内容隐私与 VoicePrivacy Initiative [6] 中所说的声音隐私。VoicePrivacy Initiative 于 2020年成立，目前主要关注说话人身份的匿名化。相反，我们关注的是语音与音频中与语义相关的信息。例如，如果某人通过言语透露了自己的姓名、生日、地址或银行账号，那么仅仅将其声学声音特征匿名化，并不足以隐藏这类私人信息。请看图1中的概念示意图。在这个例子中，掩蔽器作用于语音，以隐藏话语中的一个特定词语（出生日期）。我们将在第 3.1 节更详细地讨论技术掩蔽方案。更广泛地说，内容信息还可能包括副语言音频事件，例如不同风格的笑声，或口吃等言语障碍模式。

在撰写本文时，从消费者视角来看，语音技术已日益普及。欧盟 [7] 和英国 [8] 等管理机构已通过立法，或正在进一步调查是否需要为公民提供隐私保护。此外，已有案例显示，具备音频功能的设备会在未经授权的情况下录音，并将数据提供给未经授权的人，这引发了人们对隐私保护在法律和技术层面应达到何种范围的担忧 [7] 。

表1: 人机交互中不同类型的信息拦截。在此，数据在参与者之间传输（例如，在对话过程中）。这些数据通过信息泄露被截取。

No	数据传输	隐私泄露	示例
1	人与人之间	人	在任何公共场所，人们都可能听到许多对话（例如在公交车上或餐馆里）。
2	人与人	机器	在智能设备附近进行对话，可能会使其中一方的隐私受到威胁。备附近进行对话，可能会让设备听到谈话内容，前提是某些提及唤醒词，或（有意图）在后台录制语音。
3	人与人（通过机器）	机器	使用实时翻译程序时，设备可能会窃听或截取某人的私人信息人与人之间的对话。
4	人对机器	人	随着智能助手的普及，越来越多的操作是基于语音指令执行的。窃听它们也可能泄露私人信息（例如，家庭购物清单或人们的姓名和电话号码）。
5	人对机器	机器	案例 *1*：用于激活设备的唤醒词检测设备通常无法区分命令由其他自然对话直接传送至设备。这些其他对话可能是人与人之间的对话，也可能是人对机器的对话。情况 *2*：如果包含私人信息的音频数据属于训练数据集的一部分，那么机器学习算法可能会将这些私人信息输入到模型中（这些信息之后可能会从模型中提取出来（之后从模型中得到）。
6	机器到人	人	使用语音助手时，通常会将部分信息以合成语音的形式朗读出来（例如，使用导航程序或使用屏幕阅读辅助设备时）。

设备上的语音数据也曾作为调查严重暴力犯罪的一部分而受到关注 [10]。如果此类录音可在法院作为证据采纳，那么人们可能希望只采纳对话的部分片段。扣押设备中的数据也会对数据治理产生影响，因为这些数据（或其副本）随后会被存储在其他位置。这可能影响对个人被遗忘权的执行 [11]。

语音与音频内容隐私问题所面临的威胁范围极为广泛，涉及数据的传输方式以及信息可能遭泄露的途径。每种情况都可以视为独特的，因此很难对语音隐私的要求形成统一观点。表1概述了与语音隐私相关的不同情形。在某些场景中，造成伤害的风险较高（例如案例2和3，发言者本身就有隐私预期），即使信息泄露需要高深的知识或技术技能才能实现。

如果语音与音频研究界不开始推进内容隐私相关工作的成熟发展，那么要落实赋予个人的法律保护或保护商业机密信息，将会变得非常困难。这在一定程度上是因为能够侵犯隐私的技术日益复杂 [12], , 此外还存在无意的隐私侵犯 [13]。弱势群体以及依赖语音技术作为辅助设备的人所受到的伤害尤为严重。归根结底，内容隐私对每个人都有益。

本文提出了内容隐私研究的三个方向，并阐述了其中的挑战以及一些建议重点关注的领域。这三个方向是：(1) 内容隐私的技术解决方案，(2) 对下游任务的影响，以及 (3) 内容隐私评估的标准化。我们鼓励语音与音频研究界将这些研究方向的相关内容纳入当前和未来的工作中。如此重要的议题不应任由行业驱动的利益左右，也不应因缺乏创新而停滞不前。

2. 背景

关于语音隐私的早期研究中，有些关注如何防止人工窃听者通过偷听对话截取信息。在这类情形下，隐藏对话内容的方法着重于确保所有可听语音都变得无法辨认。例如，在 [3] 中，研究人员生成了掩蔽噪声，并结合声压级对其进行评估，以此表示隐私级别。在 [14], 的研究中，研究人员采用语音重合成技术，应用声道变换函数，并用重新录制的元音替换语音片段。这些实现内容隐私的技术可以进一步发展，以应对表 1 中隐私场景 1 的有限几种变化。对于大量语音转录文本，研究表明，语言模型有助于检测仇恨言论中的敏感词类别 [15], ；这一思路也可以进一步用于检测其他需要遮蔽的敏感私人信息类别。事实上，BERT 模型 [16] 已经能够从掩码数据中学习。从

掩码数据中学习的概念或许可以拓展，从而创造出新型隐私保护语言模型。

内容隐私的通用方法通常采用衡量人类视角下可懂度的指标进行评估 [3], , 近年来也会报告自动语音识别（ ASR）的词错误率 [14, 17]。但隐私评估取决于数据传输模式和隐私泄露类型。第 3.3 节将更详细地讨论通过可懂度衡量隐私的现有标准。文献[18] 从行为科学视角提出了符合伦理的音频数据收集指南。他们呼吁研究界探索更多隐私技术解决方案，例如开发在存储音频前提取特征的技术，从而避免存储录音。

音频隐私方面的工作可能涉及开发算法，在隐藏私人信息的同时保留其他有用信息，以用于环境声音分类等任务[14, 19]。类似的工作也已应用于咳嗽检测[20]： 研究人员会录制音频，但可以将意外录入的语音处理至不可懂。近期， [21]调查了人们在音频劣化前后对隐私的看法。劣化的目的是让语音信息变得不可懂，同时在音频中保留足够的信息，以便识别常见家务活动（洗碗、做饭、洗澡等）。研究发现，有意劣化音频显著改善了人们对隐私的看法，同时分类任务的表现仍然良好。然而，该研究没有评估语音是否仍能被自动语音识别系统识别，也没有考察如果人们看到劣化音频的自动语音识别转录文本，他们对隐私的看法会如何变化。

尽管并非所有研究人员都会积极处理与音频采集相关的隐私问题，但非语音分类任务仍日益普遍。例如， [22]开发了一种利用可穿戴音频传感器（如麦克风）检测活动的技术，并使用采集到的音频将声音分类为进食、饮水、说话和大笑等事件，这就是“BodyScope”。要在一般音频采集中平衡对私人信息（如语音）的遮蔽，难度很大。即使是麦克风干扰的最新进展，也无法实现对遮蔽的精细控制 [5]。干扰或许可以掩盖人类语音，但有时也会彻底破坏原本可用于其他任务的音频。

3. 挑战

本节将介绍实现语音与音频内容隐私进一步发展的三个最重要挑战。要拓展当前的技术现状并弥补持续存在的技术差距，包括改变人们对隐私重要性的态度，就必须应对所有这些挑战。对于每项挑战，我们都会先提出一个挑战性问题，再介绍建议的重点领域和讨论。

3.1. 内容隐私的技术解决方案

对于第一个挑战，我们笼统地指表1中的第2至6种情况。我们需要更精确的定义来界定攻击场景，以便制定掩蔽算法的技术要求。

挑战 1：哪些类型的算法有助于屏蔽语音与音频中的内容？具体采用何种算法，如何随待屏蔽的内容信息类型而变化？

重点领域：首先，我们建议建立内容隐私分类体系，以便为攻击场景定义合适的保护措施。基于此，理想情况下，我们可以开发新的屏蔽算法，使其能够适用于多种内容隐私场景（例如，屏蔽敏感细粒度内容，如口述出生日期，以及在背景中意外录下的短暂第三方对话）。

下文中，我们用遮蔽 这一术语指代任何用于掩盖信息的方法（例如，施加特殊噪声或修改信号级特征）。遮蔽可以是可听见的（用于对人类听众隐藏语音），也可以是次声的（用于机器听觉）；它既可以遮蔽语音内容，也可以遮蔽包含私人信息的非语音内容。其主要目的是让人类和机器窃听者无法理解这些内容。为了将掩蔽算法应用于语音数据，我们认为合适的分类体系应涵盖以下几个方面:

- 与非语音音频相比，语音的隐私有何不同？ → 人类语音可能包含各种不同类型的私人信息，包括健康信息或个人内容。因此，语音的隐私要求可能比一般音频数据更为严格。敏感音频信息可能包含地理标识，这些标识可能局限于语音信号中的某个局部（如汽车喇叭声），也可能是全局性的（如房间特征）。
- 可遮蔽的内容是否有限？ → 似乎不可能同时遮蔽所有私人内容。例如，如果我们尝试同时遮蔽私人语音内容、背景噪声和情绪，可能会损失过多数据。目前还没有办法量化或描述这类数据损失。
- 我们希望对控制粒度达到什么程度，原因是什么？控制粒度需求会如何影响设计决策？ → 遮蔽完整的词语似乎是非常直观的解决方案。不过，有时只遮蔽部分词语就足以确保适当的隐私保护水平（例如，对于出生于20世纪的人，只遮蔽其出生的年代和确切年份）。
- 我们能否利用近期的一些机器学习（ML）趋势？ → 在机器学习领域，隐私已成为一个重要议题，涵盖了模型反演 [23] 和成员推理 [24]等不同的攻击场景。

这些问题可以从不同角度来考虑，包括软件方案（基于算法的隐私保护）、硬件方案或边缘解决方案（例如，嵌入音频采集设备中）。在某些情况下，解决方案需要结合软件与硬件，例如进行边缘计算（即与云端断开连接），或在设备上采用数字信号处理（DSP）技术。更具体地说，让我们考虑自动语音识别（ASR）中的内容屏蔽。如果数据中包含某人提及其生日的内容，对于某些应用来说，只需在揭示日期的那段音频中添加额外噪声即可。如果 ASR 模型随后识别出另一个生日，这是因为底层语言模型预测词序列中会出现一个日期；这仍可视为可以接受（因为原始私人信息已被隐藏）。

在其他情况下，我们可能希望模型不要转录任何虚假数据，而是省略特定词语。在这种情况下，完全静音遮蔽可能会有所帮助。因此，需要根据要求考虑不同的掩蔽类型

。如果需要可逆遮蔽，那么技术掩蔽方案还需要一个“密钥”，以便在应用遮蔽后将其撤销。可逆遮蔽可用于实时语音传输，或需要与授权方共享高度敏感语音数据库的场景。

显然，音频数据并不存在一种适用于所有情况的遮蔽方案。选择哪种掩蔽器，取决于我们要遮蔽的是数据库中的语音、实时语音（如电话通信或视频电话），还是压缩状态下的语音等。此外，选择还可能取决于要遮蔽的信息类型，例如语音中的词语、音频事件噪声或背景说话者。我们还需要考虑遮蔽后的音频/语音是否需要能够恢复或还原。

3.2. 对下游任务的影响

第二个挑战主要涉及表 1 中的情形 5，但也可能推广到情形 2 和 3。我们主要讨论在其他模型基础上构建模型（或基于经过显著变换的数据构建模型）所带来的问题。如果隐私保护方案通过遮蔽来修改音频，我们就需要考虑这对基于这些预处理数据开发其他任务的后续模型有何影响。

挑战 2： 在隐藏语音或一般音频中不同类型的信息内容后，我们如何预判并为其对数据库共享、语音识别、说话人验证或音频场景分析等下游应用的影响做好准备？

重点领域： 为了解决下游任务中的问题，不应将流水线中的不同模型视为彼此独立的模块。因此，我们提议创建共享任务或挑战任务，提供一套与示例下游任务相对应的基线工具，用于基准测试。

这一领域的问题在于，如何找到最优的隐私政策，同时保留足够的信息，使音频能够用于特定任务。在以往关于这一主题的研究中，最常用的下游任务是自动语音识别（ASR）和自动说话人验证（ASV）。除此之外，还有许多其他以音频数据为输入的下游任务，例如健康分析或活动识别。对于模型流水线，还会出现新的问题：

- 如果部分音频被屏蔽（例如词语或其他内容特征），会对某些任务产生什么影响？ →例如，如果我们移除房间或位置信息的标记，是否仍能进行音频场景分析？又或者，如果我们从电话通信或视频电话通话中移除儿童的背景语音，这是否会使语音压缩或语音到语音翻译等其他任务更难完成？一个主要挑战在于，我们并不总是知道被屏蔽的语音/音频之后会用于什么用途。
- 有哪些策略可用于评估技术掩蔽方案的影响，包括在处理流程的其他阶段，之后还能对音频进行哪些操作？ →我们认为，选择掩蔽器时应始终考虑至少一项下游任务。不过，未来研究若能确定是否存在一种更通用的方法，先优先保护隐私（即需要遮蔽的内容），再在敏感内容被隐藏后执行下游任务，同时不损失性能，那将很有价值。
- 有没有办法创建足够可靠的内容隐私保护方案，使其既适用于高度敏感的语音（例如用于医学研究的语音），又能用于其他任务，或至少能在机构之间共享，同时无需将其存储在专用服务器上，并保护个人的权利？ → 这对于大型语音数据库尤为重要，特别是那些可能被视为敏感数据的数据库（例如儿童医学语音 [25]）。如果以恰当的方式开展数据共享，就能让最有望从语音与音频技术中受益的人获得巨大益处。

对于上述部分问题，一种潜在的解决方案是将隐私保护机制直接整合到下游任务或模型中，而不是仅仅依赖数据转换。例如，可以开发一种自动语音识别系统，并使用已屏蔽受保护内容的数据进行训练；屏蔽操作或许可以通过在信号中添加声学标记来实现，提示自动语音识别系统跳过或忽略敏感语音内容。此类系统的表现可能优于使用未屏蔽语音训练的系统。例如，可以训练自动语音识别系统，使其学会绝不识别或转录信用卡信息。另一方面，还是以此为例，如果要进行基于语音的在线购物，信用卡详细信息就是必不可少的。我们认为，这应由语音技术的使用者自行选择。此外，要将具有隐私保护设计的自动语音识别系统纳入商业产品，就需要行业采用该系统。

3.3. 内容隐私标准化评估

第三项挑战泛指音频数据中的隐私概念，因此，应考虑表 1中的所有情形。

挑战 3： 我们如何制定高效、客观的内容隐私衡量方法，用于隐私保障和合规？

重点领域： 我们建议语音与音频研究界成立一个新的国际工作组，协作制定一套新的国际标准和测量方法，以应对现有标准尚未涵盖的细粒度内容隐私问题。

在此，我们进一步探讨制定标准时需要考虑的一些因素。这些想法可以作为一项独立研究议程的组成部分进一步发展，旨在解决语音内容隐私问题；也可以纳入现有及正在开展的研究工作（例如，数据集创建、训练新的自动语音识别系统等）。

- 内容隐私度量指标应具备哪些理想特征？ → 其他语音标准易于用于评估语音。算法运行速度快、结果可靠，而且度量指标通常易于解释。本文提出的首个挑战表明，建立内容隐私分类体系会很有帮助，因为不同类型的内容可能需要不同的度量指标。此外，除了数据转换之外，能够评估可确保隐私的算法和机器学习系统的度量指标或许也很有用。

• 在什么特殊情况下，使用标准度量指标评估音频隐私会特别困难？ →某些类型的内容可能被视为敏感信息，但要为其建立真实标注，以便比较信息是否已被有效移除，会十分困难。例如，副语言事件（如大笑、哭泣、尖叫等）的表现差异很大，尽管这些信息很容易被用来识别音频中的人。对于自然变化很大的语音和音频事件，建立标准度量指标来评估其遮蔽效果可能颇具挑战。

•内容隐私与说话人匿名化的标准隐私度量指标有哪些相似之处和差异？ → 在另一种隐私形式中，重点是隐藏说话人的声学声音特征，其中一个目标是衡量两个说话人声音在声学方面有多接近或相似。或许可以借鉴这一思路来处理内容隐私，例如开发一种新型音频嵌入，用于比较遮蔽前后的差异。这种方法面临的一项挑战是，内容隐私追求细粒度的遮蔽（在时域或频域内），而音频嵌入通常是基于较大的时间语境学习得到的。

语音领域已经存在许多隐私度量指标。 [26]列出了这些隐私度量指标。现有指标已被证明非常有用，但它们往往仅基于感知可懂度（例如，在表1中的场景1、4和6里，我们作为人类窃听者），并且是在对整段对话进行测量时采用非常宏观的尺度。它们无法处理非常细粒度的隐私问题，例如隐藏特定词语或短语，或敏感非语音事件。 [27], 指出，凡是有音频的地方就有标准，因此标准无处不在。 [27] 中介绍的具体标准针对音频相关硬件和设备，与本文并无直接关联，但其指出现有标准众多，这一点值得重视。

隐私指数（PI）用于开放式办公室声学环境 [28]。对于封闭房间，语音传输指数（STI） [29] 更为适用，并且在噪声、混响和说话风格等多种条件下都适用。语音隐私等级（SPC） [30] 是另一种标准隐私指标。语音可懂度指数（SII）是美国国家标准协会的标准（例如，美国国家标准协会） [31]。 [32] 的研究比较了快速语音传输指数（RASTI）、STI 和 SII，并报告称，STI 和 SII 在衡量可懂度方面具有可比性。所有这些指标都无法应用于细粒度场景，而且它们面向的是可懂度，不适用于副语言内容隐私。

近期， [33] 提出了掩蔽错误率（MER），以应对对新型内容隐私度量的需求。对于该度量，他们假设内容隐私仅指词语，且已有语音的真实语音转写（即未遮蔽内容）。基于这些假设，该度量会计算正确或错误遮蔽的词语比例，并得出一个单一数值评分。目前尚不清楚如何将该度量应用于现实场景，因为相关研究只是提出了这一度量，并未在具体任务中使用。此外，MER 仅适用于词语，并未说明如何将其调整（如果可以调整的话），以用于副语言内容隐私或非语音事件。

4. 讨论

我们提出了语音与音频研究领域需要开始探索的三项挑战，以推动隐私这一长期以来探索不足的关键领域向前发展。重要的是，应将隐私级别视为由个人自行选择。为实现这一目标，我们讨论了开发技术解决方案的一些要求。随后，我们探讨了如何从下游任务的角度考量内容隐私。最后，我们呼吁语音与音频领域的研究群体携手合作，开始制定内容隐私的国际标准，并提出了一系列可纳入此类标准考量的问题。

众所周知，科技公司会存储使用其产品的消费者的语音信息 [34, 35]。这些公司也承认会与第三方共享数据，从而损害消费者隐私。因此，消费者往往选择有限：他们可以使用方便、优质且易于获取的产品，同时冒着隐私受损的风险 [36]；也可以不使用这类产品，从而放弃语音技术带来的益处。

即使已经采取了安全和隐私保护机制，我们仍常常在研究与开发领域发现对抗性发展周期。一项保护技术被开发并发表后，黑客（或其他学者）便会尝试指出其中的缺陷。为从一开始就避免危及隐私，除非企业在产品上市销售前评估隐私相关方面，并向消费者公布相关信息，否则不应将产品商业化。理想情况下，学者和其他感兴趣的人随后仍会探查商业产品中未被发现的隐私弱点，并进一步将相关信息公之于众。究竟需要向消费者告知哪些内容，目前尚不明确。由于这会给企业带来高昂成本，立法者开始更加关注这一问题 [37, 38]。除非法律要求，企业很少会主动在其技术中加入隐私功能。

我们建议研究人员采取一种对待语音隐私与安全的立场，类似于近年来应对算法偏见问题的趋势。我们鼓励研究人员说明其工作中如何考虑隐私与安全问题。我们还期待加强对监管与政策的讨论，以追究技术开发者的责任，促使他们在商业产品中提供更多隐私选项。

5. 致谢

本研究得到了巴伐利亚州经济事务、地区发展与能源部的资助。

6. 参考文献

[1] A. Nautsch、C. Jasserand、E. Kindt、M. Todisco、I. Trancoso 和 N. Evans，《GDPR与语音数据：法律与技术界的思考，以及迈向共识的第一步》，载于*Interspeech 2019*，2019年，第3695–3699页。[2] D. E. Bambauer，《隐私与安全》，载于《刑法与犯罪学杂志》，第103卷，第667页，2013年。[3] T. Tamesue、S. Yamaguchi 和 T. Saeki，《利用掩蔽噪声实现语音隐私的研究》，载于《声音与振动杂志》，第297卷，第3–5期，第1088–1096页，2006年。[4] Y.-C. Tung 和 K. G. Shin，《利用声音掩蔽保护智能手机上的音频隐私》，载于《*2019*年*ACM*亚洲计算机与通信安全会议论文集》，2019年，第257–268页。

[5] Y. Chen、M. Gao、Y. Li、L. Zhang、L. Lu、F. Lin、J. Han 和 K. Ren，《老大哥在监听：超声波麦克风干扰器评估框架》，载于*IEEE INFOCOM 2022——IEEE*计算机通信会议，2022年，第1119–1128页。[6] N. Tomashenko、B. M. L. Srivastava、X. Wang、E. Vincent、A. Nautsch、J. Yamagishi、N. Evans、J. Patino、J.-F. Bonastre、P.-G. No´e等，《VoicePrivacy Initiative简介》，载于*INTERSPEECH 2020*，2020年。[7] P. Voigt 和 A. Von dem Bussche，《欧盟通用数据保护条例（GDPR）》，载于《实用指南》，第*1*版，施普林格国际出版：施普林格国际出版，第10卷，第3152676期，第10–5555页，2017年。[8] 《议会议员将审查互联技术日益普及带来的影响》，https://committees.parliament.uk/committee/378/digital-culture-media-and-sport-committee/news/170742/mps-to-examine-implications-of-rising-popularity-of-connected/tech，2022, [在线；访问日期：6月26日2022]。[9] A. Griffin，《Amazon Echo如何录下一个家庭的私人对话，并将其发送给一名陌生人》，https://www.independent.co.uk/tech/amazon-echo-recording-alexa-message-family-security-stop/how-to-a8369311.html，2018, [在线；访问日期：6月24日2022]。[10] A. Cuthbertson，《亚马逊被要求在双重谋杀案中提供Alexa证据》，https://www.zdnet.com/article/apple-stores-your-voice-data-for-two-years，2018，[在线；访问日期：6月24日2022]。[11] J. Rosen，《被遗忘权》，载于《斯坦福法律评论在线》，第64卷，第88页，2011年。[12] C. Shi、T. Zhang、Z. Xu、S. Li、Y. Yuan、A. Petropulu、C. T. M. Wu 和 Y. Chen，《利用相控多输入多输出雷达，通过房间物体振动发起语音隐私攻击》，载于《第*20*届移动系统、应用与服务国际会议论文集》，2022年，第573–574页。[13] A. Kwong、W. Xu 和 K. Fu，《听觉硬盘：利用合成麦克风窃听的磁盘》，载于《*2019*年*IEEE*安全与隐私研讨会（*SP*）》，2019年，第905–919页。[14] F. Chen、J. Adcock 和 S. Krishnagiri，《音频隐私：降低语音可懂度，同时保留环境声音》，载于《第*16*届*ACM*多媒体国际会议论文集》，2008年，第733–736页。[15] M. Mozafari、R. Farahbakhsh 和 N. Crespi，《基于BERT的迁移学习方法：检测在线社交媒体中的仇恨言论》，载于复杂网络及其应用国际会议，2019年，第928–940页。[16] J. D. M.-W. C. Kenton 和 L. K. Toutanova，《BERT：用于语言理解的深度双向变换器预训练》，载于*NAACL-HLT*论文集，2019年，第4171–4186页。[17] M. Dong、D. Yan 和 R. Wang，《语音增强中的对抗式隐私保护》，载于*arXiv*预印本 *arXiv:2206 08170*，2022..[18] M. Cychosz、R. Romeo、M. Soderstrom、C. Scaff、H. Ganek、A. Cristia、M. Casillas、K. De Barbaro、J. Y. Bang 和 A. Weisleder，《日常生活的长时录音：伦理与最佳实践》，载于《行为研究方法》，第52卷，第5期，第1951–1969页，2020年。[19] S. Kumar、L. T. Nguyen、M. Zeng、K. Liu 和 J. Zhang，《声音粉碎：保护隐私的音频感知》，载于《第*16*届移动计算系统与应用国际研讨会论文集》，2015年，第135–140页。[20] E. C. Larson、T. Lee、S. Liu、M. Rosenfeld 和 S. N. Patel，《利用低成本麦克风实现准确且保护隐私的咳嗽感知》，载于《第*13*届普适计算国际会议论文集》，2011年，第375–384页。[21] D. Liang、W. Song 和 E. Thomaz，《音频劣化对隐私感知及基于音频的人类活动识别推断性能的影响》，载于第*22*届移动设备与服务人机交互国际会议

[22] T. Rahman、A. T. Adams、M. Zhang、E. Cherry、B. Zhou、H. Peng 和 T. Choudhury，“BodyBeat：一种用于感知非语音身体声音的移动系统”，载于 *MobiSys*，第 14 卷，编号 10.1145，2014 年，第 2–594 页。

[23] M. Fredrikson、S.Jha 和 T. Ristenpart，《利用置信度信息的模型反演攻击及基本应对措施》，载于第*22*届 *ACM SIGSAC* 计算机与通信安全会议论文集，2015年，第1322–1333页。

[24] R. Shokri、M. Stronati、C. Song 和 V. Shmatikov，《针对机器学习模型的成员推理攻击》，载于*2017*年 *IEEE*安全与隐私研讨会 *(SP)*，2017年，第3–18页。[25] S. Das、N. N. Lønfeldt、A. K. Pagsberg、L. Clemmensen 等，《面向资源匮乏真实环境条件下丹麦语儿童与临床医生对话的语音检测：案例研究》，*arXiv* 预印本 *arXiv:2204.11550*，2022年。[26] J. Stout，《语音隐私标准》，*Cambridge Sound Management, Inc*，2015年。[27] D. A. Bohn，《应对复杂且令人沮丧的音频标准世界》，*Sound & Video Contractor*，第56–64页，2000年。

[28] M. Gover，《隐私指数与语音隐私等级之间的关系》，美国声学学会杂志，第 145 卷，441，2019 年 5 月，第 435 页，EL–EL。

[29] K. L. Bradley 和 I. B，《一种根据语音波形确定语音传输指数的方法》，美国声学学会杂志，第106卷，第6期，3637–3648页，1999年。

[30] J. Bradley 和 G. Gover，《基于语音隐私等级（SPC）值的一套新语音隐私标准》，载于国际声学大会，*ICA*，2010年。

[31] C. Pavlovic，《SII——语音可懂度指数标准：ANSI S3.5 1997》，美国声学学会期刊，第143卷，第3期，第1906–1906页，2018年。

[32] P. Larm 和 V. Hongisto，“语音传输指数、快速语音传输指数与语音可懂度指数的实验比较”，美国声学学会杂志，第 119 卷，第 2 期，第 1106–1117 页，2006年。[33] J. Williams、J. Yamagishi、P.-G. No´e、C. Valentini-Botinhao 和 J.-F. Bonastre，“重新审视语音内容隐私”，载于第 *1* 届 *ISCA* 语音通信安全与隐私研讨会，2021年。[34] Z. Whittaker，“苹果公司将你的语音数据存储两年”，https://www.zdnet.com/article/apple-stores-your-voice-data-for-two-years，2013，[在线；访问日期：6月24日，2022]。[35] E. Field，“Google Home 和 Amazon Echo 可以存储你的语音录音。它们何时可能被用来对付你”，https://www.entrepreneur.com/article/322896, 2018, [在线；访问日期：6月24日，2022]。[36] H. Zhu、Y. Zhang、X. Guo 和 X.-Y. Li，“防泄漏：保护隐藏在语音中的隐私”，载于*2021*年第七届大数据计算与通信国际会议（*BigCom*），2021年，第 114–120 页。[37] R. N. Zaeem 和 K. S. Barber，“《通用数据保护条例》对隐私政策的影响：近期进展与未来展望”，*ACM*管理信息系统汇刊（*TMIS*），第 12 卷，第 1 期，第 1–20 页，2020 年。[38] A. D. Vanberg，“《通用数据保护条例》之后的信息隐私——道路的尽头，还是漫长旅程的起点？”，《国际人权杂志》，第 25 卷，第 1 期，第 52–78 页，2021年。