DOI:10.19651/j.cnki.emt.2415587

基于时频特征的反监听技术研究与应用*

李建勋 王 开
(东南大学信息科学与工程学院 南京 211189)

摘 要:本研究旨在探讨一种基于时频特征设计的反监听技术,重点研究如何通过动态修改时序和频率增强特定频率范围内的人类语音干扰。本文针对现有的语音干扰技术展开研究,并与标准噪声注入方法进行了比较。研究方法包括理论分析和实验验证,通过对实际原型进行测试和验证,评估了基于时频特征提取的干扰信号在干扰语音识别系统方面的有效性。实验结果显示,当信噪比低于 0 dB 时,所提出方法的文本识别错误率超过了 60%;而当信噪比为 0 dB 时,本文算法的文本识别错误率平均比当前干扰算法高出 20%以上。此外,当干扰系统与录音设备保持相同距离时,本文算法在录音设备上产生的信噪比比当前算法低近 2 dB,这说明了所提出算法的高能量利用效率。因此,本研究成果对于提高通信安全和保护隐私具有重要意义,特别是在需要高度保密的通信环境中。

关键词:隐私保护;防窃听;超声波干扰;时频特征提取;语音干扰
中图分类号:TN912.3 **文献标识码**:A **国家标准学科分类代码**:510.4040

Anti-eavesdropping technology research and application based on time-frequency features

Li Jianxun Wang Kai
(School of Information Science and Engineering, Southeast University, Nanjing 211189, China)

Abstract: This study aimed to explore an anti-eavesdropping technique based on time-frequency feature design, focusing on how to dynamically modify the temporal and frequency aspects to enhance human speech interference within specific frequency ranges. The paper conducted research on existing speech interference techniques, comparing them with standard noise injection methods. The research methods included theoretical analysis and experimental validation. By testing and evaluating the interference signals based on time-frequency feature extraction on an actual prototype, the effectiveness of the interference in disrupting speech recognition systems was assessed. The experimental results showed that when the signal-to-noise ratio (SNR) was lower than 0 dB, the proposed method's text recognition error rate (WER) was over 60%. Moreover, when the SNR was 0 dB, the WER of the algorithm in this paper was higher than that of current jamming algorithms by more than 20% on average. Additionally, when the jamming system maintained the same distance from the recording device, the SNR produced by this paper's algorithm on the recording device was lower than that of the current algorithm by almost 2 dB. This demonstrates the high energy utilization efficiency of the proposed algorithm. Therefore, the findings of this research have significant implications for improving communication security and protecting privacy, especially in environments that require a high level of confidentiality.

Keywords: privacy protection; anti-eavesdropping; ultrasonic interference; time-frequency feature extraction; speech interference

0 引 言

在信息时代,电子设备的广泛应用给通信和安全系统带来了巨大挑战,尤其是在当前智能设备中麦克风的广泛使用下,窃听问题变得日益严重。窃听行为可以轻松地利用各种隐秘的录音设备实现,包括小型录音设备和智能手机。Guri 等[1]实现了一个软件可以操纵连接到计算机的耳机,将它们作为麦克风进行窃听攻击。这些隐秘的麦克风在许多间谍活动中发挥着关键作用,引发了对个人音频保密性的担忧。

收稿日期:2024-03-06
* 基金项目:国家自然科学基金(62234012)项目资助

针对这些威胁并确保用户信息和隐私的保护，在反窃听技术领域已经展开了广泛的研究。对于可听噪声干扰的应用，例如使用白噪声播放器或高频噪声源[2]，是一种常见的方法，用于挫败窃听企图并防止记录。然而，这些方法可能会导致用户体验下降，因为它们在通信过程中引入了不必要的干扰，影响了有效的交流。另一种替代的反窃听技术涉及电磁干扰，通过引入电磁信号来扰乱麦克风传感器系统，以防止正常的录音操作[3]。然而，这种方法需要大量的电源，并可能对其他电子设备造成干扰[4]。此外，它可能违反电磁频谱管理法规，因而产生法律风险[5]。

因此，研究重点已经转向利用麦克风的非线性特性产生难以辨别的干扰[6-7]。学者们已经应用了多种技术来应对窃听问题。其中一种方法是通过将低频带限白噪声与超声波调制，引入麦克风中的噪声，并发展了便携式超声波反窃听设备[8]。Chen 等[9-10]提出一种可穿戴式麦克风干扰器 Wearable Jammer 增加便携性的同时，阻止用户周围环境中的麦克风窃听。为了实现区域覆盖，专用的超声波干扰系统已被研发[11]，同时高功率单音信号已被用于录音衰减[12]。通过优化超声波的工作区域，干扰效果的提升主要通过增强实现。

另一种方法是基于人类语音特征设计干扰信号波形，这可能为反窃听工作提供多重优势。汪志成等[13]使用声谱图来提取声音信号的短时频率、能量分布，结果提高了识别的精度，证明可以通过这种方法定制特定信号以有效干扰语音识别系统，从而实现有效的干扰。例如，信息掩蔽可以减小噪声和目标语音之间的差异，使人耳难以过滤掉噪声，降低窃听者识别录音语音的能力。

本研究旨在推动反窃听技术的发展，通过引入一种新的干扰信号设计方法，提出了基于时频特征的干扰(time-frequency feature interference，TFFI)信号反窃听方法。该方法利用语音时频图能够可视化表达语音信号的时、频域联合分布，从中分析并提取相关的图像特征，突破了传统声学特征的单一性，获取了传统声学特征无法表征的语音信息[14]。相较于典型的噪声注入技术，TFFI 不仅调整其频率，还调制其时序，以更真实地模拟和干扰人类语音的微妙之处。

本研究采用了优化超声传输方法，并进一步提高干扰效果，以获得更大的传输能量和减少信号失真。设计了一个原型系统，利用声学参量阵列生成高度定向的超声传输，避免与非窃听设备的无意干扰。

最终的实验结果表明，TFFI 技术在低能量场景下，即信噪比(signal to interference plus noise ratio，SNR)等于 0 时，能够有效阻碍语音识别系统，与传统的噪声注入方法相比，显著降低了语音识别系统的识别准确性。本研究的创新之处在于将 TFFI 定制配置以适应人类语音频率范围，形成了一种更为稳定的防止窃听的方法，从而在反窃听领域具有重要意义。

1　时频特征干扰信号

1.1　麦克风系统中的非线性

麦克风作为广泛应用的传感器系统，在各种设备中得到广泛应用，包括录音设备和智能手机等。其主要功能是将声音信号转换为电信号。麦克风系统使用放大、滤波等技术处理模拟信号。然后，模拟信号被传递到模数转换器(analog to digital converter，ADC)。最终，数字化的音频数据被传输到处理器进行额外处理[15]。

图 1 展示了典型的麦克风系统的结构。麦克风的非线性是指在麦克风工作过程中，其输入和输出之间的关系不是简单的线性关系。随着变换后的信号穿过不同模块，如放大器等，组件的模拟特性导致非线性特征浮现。这些特征导致了互调失真和谐波失真，随后在最终信号中产生了原始输入信号中不存在的频率成分。

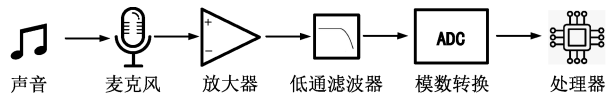


图 1　典型的麦克风系统

对于具有多个频率成分的两个声音信号 a 和 b，系统生成的非线性成分可以表示为：

$$s_{nl}(t)=\sum_i\sum_j \beta A_{a,i}A_{b,j}\cos(2\pi(f_{a,i}\pm f_{b,j})t) \quad (1)$$

在式(1)中，$A_{a,i}$、$f_{a,i}$ 分别表示声音信号 a 在第 i 处的幅度信息以及频率信息，$A_{b,j}$、$f_{b,j}$ 表示声音信号 b 在第 j 处的幅度信息以及频率信息，β 是增益系数，由硬件本身决定，$s_{nl}(t)$ 则表示转换后的信号值。

人类听觉和语音通信的频率范围通常涵盖 20 Hz～20 kHz。因此，大多数麦克风系统在这个范围内运作。即使麦克风能够检测 20 kHz 以上的频率，低通滤波器能有效地限制输入信号中的高频成分，以防止高频噪声或干扰影响到转换后的数字信号质量。例如，当一个 41 kHz 和一个 40 kHz 信号发送到麦克风端时，由于麦克风的非线性特性，会产生两个信号的和频 81 kHz 和差频 1 kHz。这些频率之和和频率之差是由非线性效应引起的，通常被称为互调产物。根据奈奎斯特采样定理，模数转换器(ADC)的典型采样率为 44.1 kHz，该定理规定数字信号频率应低于 22 kHz，所以最终 1 kHz 信号被麦克风采集并保留了下来。

1.2　干扰系统概述

干扰系统的整体结构如图 2 所示，其中包含两个核心模块：分析模块和硬件模块。分析部分的主要功能是制定干扰信号，而硬件部分则根据分析模块提供的干扰数据集来生成必要的输出信号。随后，超声波探头阵列传输这些干扰信号。

干扰系统的第一步要求采集用户声音信息，保存该用户的音频文件。接着，将捕获的声音文件数据在电脑端经

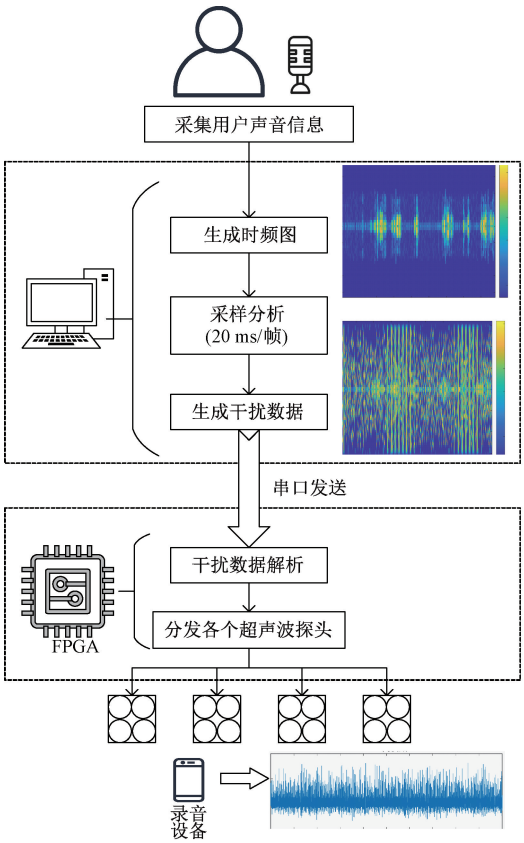


图 2 干扰系统概述

过时频转换和频谱分析,然后进行分帧提取并分析该用户的主要基频特征和共振峰频特征。在这个过程中,构建该用户的时频特征库。随后,基于这个特征库,制定适应该用户的个性化干扰数据。

在干扰信号数据集构建完成后,由电脑端串口发送到硬件模块进行处理,并嵌入到超声波信号中分布给不同的超声波模块处理并进行下一步发射,以实现干扰功能。硬件模块能够根据干扰数据集中的数据创建必要的干扰信号模式,以确保对该用户声音信号的有效干扰。生成的干扰信号随后通过超声波探头阵列进行传输,从而与被用于窃听的麦克风所接收到的信号叠加来实现干扰。

1.3 时频特征干扰信号

传统方法包括使用白噪声调制超声波或采用高幅差分单频超声波[16],这两种方法均属于能量抑制的范畴。白噪声方法主要基于将解调的白噪声完全覆盖在频率上的思路,从而降低语音信号的信噪比并损害有用信息。而差分单频干扰方案则通过互调产生高幅度的单音信号,利用人类听觉的掩蔽效应。单音干扰的高声压降低了听者对其他声音的听觉敏感度。

然而,纯粹的能量抑制方法对干扰系统的功率和组件提出了极高的要求,并且容易受到降噪算法的影响而降低屏蔽效果。设计用于干扰录音和窃听设备的系统必须符合效率、智能、稳定性和实用性的标准,特别强调在有效阻碍人类语音特征的同时抑制不必要的能量消耗。为了解决这些问题,本文提出了一种根据人类语音特征定制的时频特征干扰系统。

个体的声音表现出独特的声学特征。男性的声纹数量大约是女性的两倍,而男性和女性的声音显示出不同的频谱分布[17]。此外,男性的声音通常在 50～250 Hz 之间具有较低的基频,而女性的基频范围在 200～300 Hz 之间[18]。共振峰值频率与语言的语义结构相关,在元音和辅音之间变化。因此,必须提取不同频率以中断语义并将干扰引导到说话者可区分的基频。此外,人类语音特征高度依赖于时间[19],因此使用时频结合而不仅仅依赖频率进行干扰可以更有效地实现干扰。因此,本文提出了时频特征干扰信号的方法。

首先,对说话者的声音频谱进行客观分析,以确定频谱的能量和频率分布,并确定说话者的主频带。这些频带有助于确定处理干扰信号所需的发射通道数量。在完成分析后,确定存在 3 个频率发射间隔,每个间隔分配给负责超声波传输的特定阵列组。

确定了发射阵列组的配置后,接下来需要对相应的时频特征干扰信号进行仔细设计和检查。在进行人类语音的时频分析后,可以明显地发现其属于非平稳信号的范畴。然而,值得注意的是,在短暂的时间窗口内(15～25 ms),人类语音信号可以被视为具有短时平稳性。因此,选择了 20 ms 作为“时频特征干扰信号”的宽度。这一时段足够短,可以有效地干扰并扰乱人类语音的传输和麦克风的接收过程。

然而,仅考虑时频特征信号的宽度是不够的,还需要对其持续时间以及频率范围进行深入的考量。时频特征干扰信号的持续时间至关重要,因为它直接影响了干扰的持久性和效果。同时,频率范围的选择也是至关重要的,因为不同频率的干扰信号可能对目标系统产生不同程度的影响。因此,在设计时频特征干扰信号时,必须综合考虑其持续时间和频率范围,以确保系统能够在特定情况下有效地执行干扰任务。

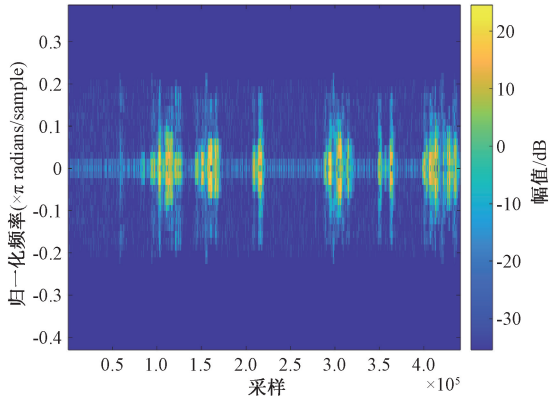
通过基于 20 ms/帧长度对人类语音数据进行分割和提取时频样本,得到了人类语音的短时时频样本。分析每个短时样本的频率能量分布,并使用谱估计方法估算主共振峰频率。这个估算过程涉及最小化预测误差,使用最小均方误差准则,本质上是解决一个正则化方程:

$$\begin{aligned}&\Phi^{k}_{(i,j)}(n)=\sum_{n}s^{k}_{i}(n)s^{k}_{j}(n)\\&\sum_{m=1}^{P}a_{m}\Phi^{k}_{(i,j)}(n)=-\Phi^{k}(n)\end{aligned}\tag{2}$$

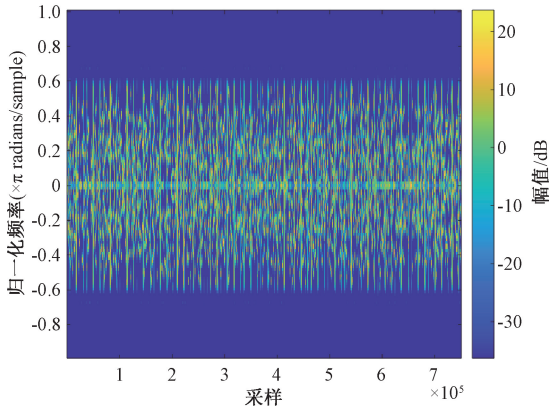
在式(2)中 $\Phi^{k}_{(i,j)}(n)$ 表示由第 k 帧处第 i 个和第 j 个的超声波探头合成的干扰信号值,$s^{k}_{i}(n)$ 表示干扰信号在第 k 帧处第 i 个超声波发射的信号,a_m 则代表干扰信号的幅度

信息，P 的值为超声波探头的组合数 C_m^2。

基于提取的短时共振峰频率特征和相应的个体遗传频率特征，确定了时频特征干扰信号的具体干扰频率 f_i 及其相应的频率范围 B_i，从而构建了时频特征干扰库。在图 3 中呈现的时频图中，图 3(a)代表用户的语音信号，而图 3(b)描述了基于语音信号设计的时频特征干扰信号。



(a) 人声的时频图



(b) TFFI的时频图

图 3　人声和 TFFI 时频图

为了生成与干扰库相对应的时频特征干扰，必须根据时频特征干扰和麦克风传感器的非线性来计算驱动超声波传感器阵列的驱动控制信号。同时需要将干扰信号调制到高频超声载波上来驱动超声探头，以便输出的干扰对人耳不可听见。对于超声探头的驱动控制信号，最常用的调制方法是双边带幅度调制(double sideband amplitude modulation，DSB-AM)，或简单地使用多个单频超声波。多单频信号不能满足时频特征干扰信号的时变特性，而 DSB-AM 调制在非线性解调时的能量转换效率较低。本文采用频移键控调制的思想，使用了一种多频编码调制信号叠加的干扰方案。该方法实现简单，并且作为窄带时变信号，在空间传输过程中损耗较低，解调时能量损失较少，提高了能源利用率。

驱动控制信号可以表示如下：

$$s_{con}^{k}(t)=A_k\cos\left(2\pi\left(f_k\pm\frac{B_k}{T_0}\cdot t\right)t+\varphi_k\right) \tag{3}$$

在式(3)中，A_k 控制振幅，f_k 控制起始频率点，B_k 控制频率带宽，T_0 代表短时帧长度，在本文中设置为 20 ms，而 φ_k 表示相位控制。

2　评估与讨论

2.1　实验设置

TFFI 干扰系统的硬件电路实现如图 4 所示。该系统主要包括干扰信号生成与处理模块以及超声波传输阵列。在干扰信号生成与处理模块中，现场可编程门阵列(field programmable gate array，FPGA)(芯片型号为 ZYNQ7020)充当着控制中心，负责协调和管理各个子模块的工作。四通道数模转换器(digital to analog converter，DAC)则负责将数字信号转换为模拟信号，以供功率放大器进一步处理。这四个功率放大器负责增强干扰信号的强度，确保其能够有效地传播和影响目标系统。在系统的另一端，超声波传输阵列作为干扰信号的传播介质，配备了 16 个 40 kHz 超声波探头。这些探头被分布在不同位置，以达到对目标设备较好的干扰效果，并且每组探头传输两个独特的时频特征干扰信号，从而实现了符合时频特征干扰的干扰效果，提高了系统的干扰性能和适用范围。

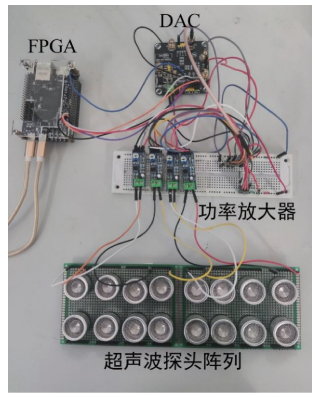


图 4　TFFI 干扰系统测试设备

时频特征干扰系统的软件设计实现如图 5 所示。该程序的实现部分涵盖了多个模块，以有效地实现对干扰信号的控制和调节。在程序的核心设计中，串口模块负责接收并解析外部指令，其中包括指定的干扰库频率信息。随后，通过电脑端将这些信息循环发送给硬件模块以实现干扰信号的生成和调整。超声波发送模块是系统的关键组成部分之一，它负责生成超声波信号，并且具有独立调节四个通道的频率和相位的能力。这种独立调节的设计使得系统能够针对不同的干扰目标进行精准频率干扰。此外，数字模拟转换控制模块则承担着将数字信号转换为模拟信号的任务，以便将超声波信号发送到干扰目标中。整个软件设计旨在实现对时频特征干扰系统的灵活控制，使其能够在复杂的环境下实现高效的干扰效果。

评估语音干扰算法的干扰效果通常涉及使用一系列客观和主观评估指标，以确保干扰算法的性能和实际可用性。

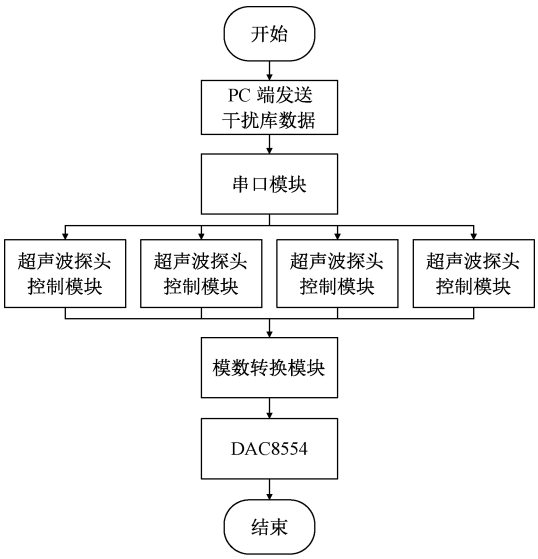


图 5 TFFI 干扰系统软件设计

平均意见得分(mean opinion score,MOS)方法是一种涉及专家(听众)对语音进行评分的主观评估方法。通过使用语音转文本(speech-to-text,STT)系统测试受干扰的语音并获取转录文本,然后与正确文本进行比较,可以计算单词错误率(word error rate,WER)。WER 是评估语音识别系统性能的一项重要指标,用于衡量系统输出文本与参考文本之间的差异程度。计算方法是将系统输出文本转换为参考文本所需的插入、删除和替换操作的数量,然后除以参考文本中的总单词数。其计算基于插入(insertions,I)、删除(deletions,D)和替换(substitutions,S)的操作数量。该指标通过式(4)定义:

$$WER = \frac{S + D + I}{N} \tag{4}$$

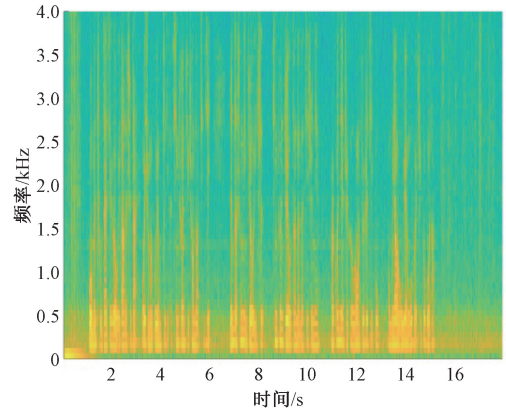
例如,如果参考文本是“机器学习在信息安全中起到重要作用”,而系统输出为“机器学习信息安全中起到重要性作业”,其中“作用”被替换成“作业”,因此 $S=1$;“在”被删除,因此 $D=1$;“重要性”插入了“性”,因此 $I=1$。因为总单词数 N 为 16 个,因此计算得到的 WER 为 3/16,即 18.75%。

本文采用语音转文本测试方法,WER 提供了一个定量的指标,用于评估语音识别系统的准确性,特别适用于比较不同系统的性能或评估单个系统随时间的改进情况。较低的 WER 值表示转录的准确性更高。为了减少评估的主观性,使用 WER 来指示干扰的质量。较高的 WER 表明语音转文本系统对语音的识别较差,从而表示更好的干扰效果。

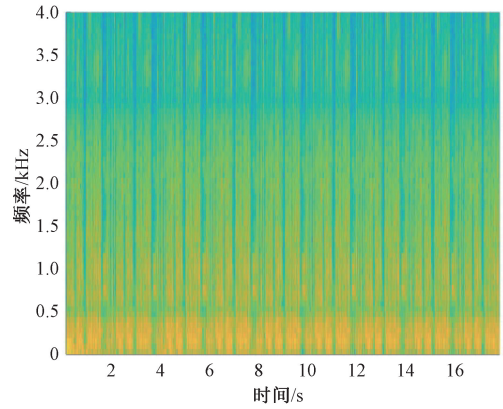
2.2 干扰能量利用效率分析

在能量效率分析中,录音设备和干扰设备保持了相同的距离,距离为 20 cm。随后录制了在有和无白噪声以及 TFFI 干扰情况下的用户语音信号。在录音采集过程中,说话者、文本内容和录音设备都保持不变。随后,导出音频文件并对其进行干扰效率对比分析。

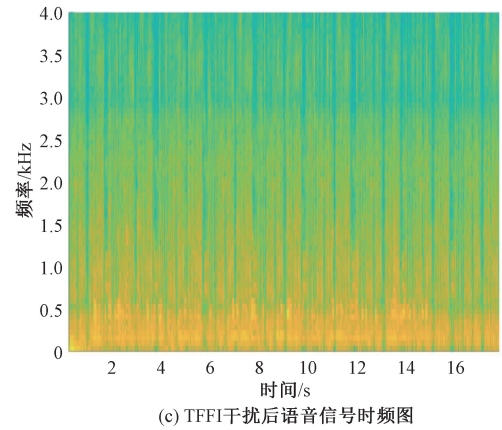
为了比较 TFFI 噪声和白噪声对相同声音的干扰效果,分析并绘制了图 6 中纯声音、白噪声干扰的声音和 TFFI 干扰的声音的时频域波形图。白噪声在时频域中的能量分布均匀,而 TFFI 噪声则可以定位到人类语音的特定成分,以实现更好的掩蔽效果。



(a) 原语音信号时频图



(b) 受到Backdoor[8]中噪声干扰后语音信号时频图



(c) TFFI干扰后语音信号时频图

图 6 不同情况下语音信号时频谱

同时,从图 7 可以观察到,在相同距离下,TFFI 的能量高于白噪声。为了确认干扰信号的能量传递效率,记录了相同距离下的白噪声干扰信号和 TFFI 干扰信号。计算了两个记录的干扰噪声与相同纯人声信号分布片段的信噪比。在不同距离下获得信噪比,如图 7 所示。图 7 显示,在

低距离情况下，TFFI 干扰信号产生的噪声比白噪声干扰信号高 2 dB。在长距离情况下，由于传输设备功率有限和空气中超声波传播的损耗，两条信噪比曲线迅速增加并收敛，导致干扰信号产生的噪声迅速减少。因此，TFFI 干扰信号相比白噪声需要更少的功率才能达到相同的信噪比，表明 TFFI 干扰系统具有更高的能量利用效率。

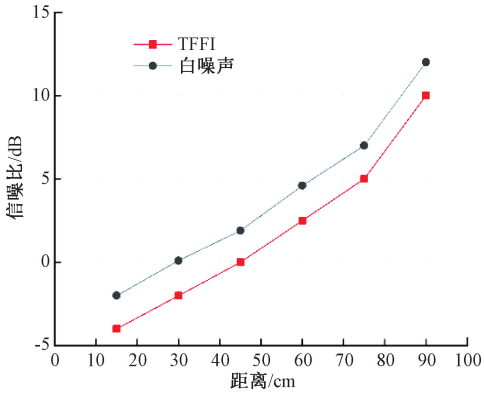


图 7　不同距离下两种干扰方法的信噪比

2.3　实验结果与分析

为了全面评估干扰效果，本次对比实验使用了四种不同的语音转文本系统进行测试：其中由两个主要的商业语音转文本系统：腾讯语音转文本系统和讯飞语音转文本系统，一个免费的语音转文本系统：笛云语音转文本系统，以及一个常用的开源语音转文本系统：DeepSpeech。

对 TFFI 和 Backdoor[8] 中的噪声进行了比较，噪声的参数设置与 Backdoor 中使用的噪声一致，即带限制的白噪声，频率范围从 0～12 kHz，并由 40 kHz 的载波调制。图 8 显示了原语音与受 TFFI 干扰语音的波形对比。图 8(a) 为原始语音波形，图 8(b) 为干扰信号波形。在各种信噪比(SNR)下，具有 TFFI 和白噪声干扰的录音被输入到 STT 系统中。

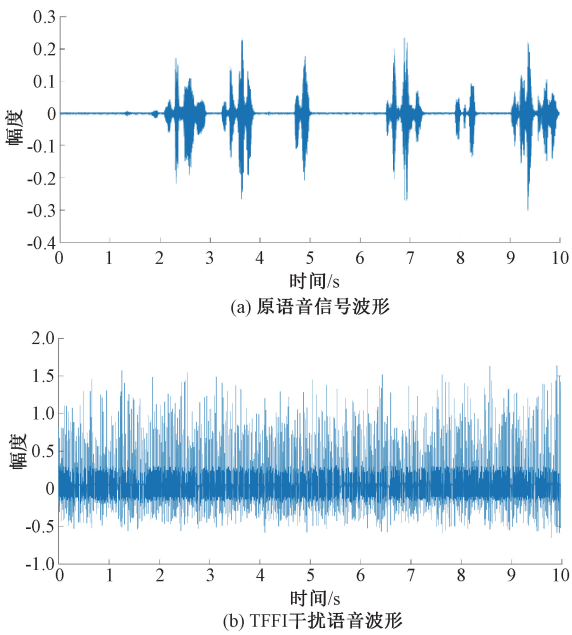
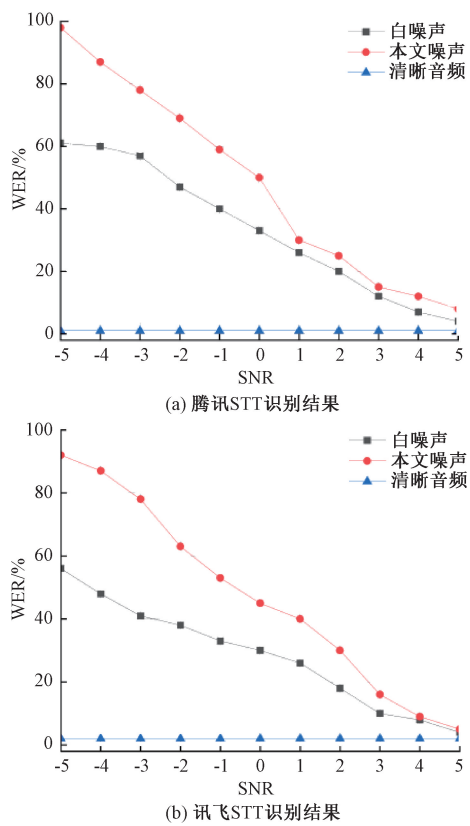
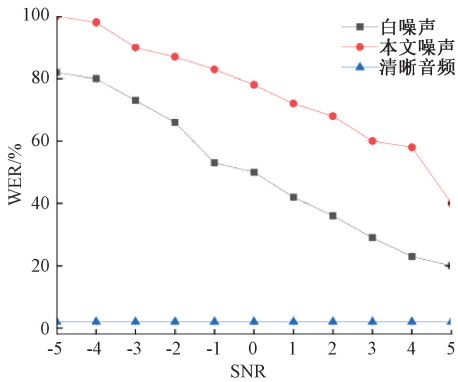


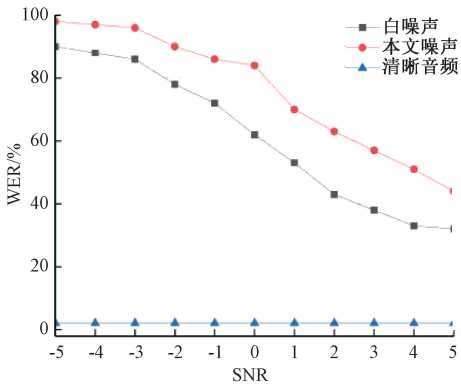
图 8　原语音与 TFFI 干扰语音波形图

图 9 说明了在 SNR 范围为[－5，5] dB 内，每个 STT 系统的词错误率(WER)值。图 9 显示，TFFI 设计在信噪比(SNR)低于 0 dB 时表现优于白噪声干扰，特别是在 0 dB 的 SNR 下。在 0 dB 的 SNR 下，TFFI 的平均词错误率(WER)比白噪声高 20%。对于两个主要的商业 STT 系统，在低 SNR 下，TFFI 和白噪声干扰的干扰性能差异较大，而在开源 STT 系统中差异较小。这种变化可以与商业 STT 系统中噪声降低算法的使用有关，显示当前商业噪声降低方法对 TFFI 的抑制较弱。这说明了 TFFI 的抗干扰能力。此外，在笛云和 DeepSpeech 系统中采用非优化的噪声降低算法的情况下，TFFI 在高于 0 dB 的 SNR 下表现出优于白噪声干扰的性能。这表明 TFFI 具有更高的能量利用效率和更好的干扰效果。为了更好地说明 TFFI 的效果，在使用相同的设备下，传输不同的干扰信号，即 TFFI 中的噪声，Backdoor[8] 中的噪声和 Patronus[16] 中的噪声，并在 10 cm、50 cm、100 cm 处测试实际的干扰效果。从表 1 可以发现，在相同的距离下，TFFI 的干扰效果远高于其他噪声，除了 TFFI 对语义的干扰效果更强外。此外，TFFI 在具有相同传输设备功率的被干扰设备中产生更高功率的干扰，显示了 TFFI 的高能量利用效率。





(c) 笛云STT识别结果



(d) DeepSpeech STT识别结果

图 9 4 个语音转文字系统的识别结果

表 1 在不同距离下的平均 WER 识别结果

距离/cm	平均 WER 识别结果/%		
	Patronus[16] 中的噪声	Backdoor[8] 中的噪声	TFFI 干扰
10	78.6	57.4	93.2
50	29.4	21.6	56.5
100	10.4	4.5	35.8

3 结 论

本文研究的核心思路是通过时频特征分析引入一种利用时频特征干扰的高效防窃听系统。该系统为智能家居及其他敏感环境提供了更加可靠的保护手段。在防窃听过程中,系统通过对时频特征的分析和利用,能够针对不同环境下的威胁进行精准干扰,从而提高了防范窃听的效果和精度。最后,针对干扰技术搭建的实时音频处理硬件系统,使得能够对窃听威胁进行及时响应和干预,有效保护机密信息的安全。此外,从实验结果来看,由于系统功率的限制导致了干扰范围的局限性,从而影响了系统的实际应用效果,为了进一步提升系统的应用价值,未来将提高系统的功率输出,以满足更广泛的应用需求,并增强系统的应用前景。综上所述,系统作为一种实用的音频安全解决方案,有望在保护隐私和信息安全方面发挥积极作用,为智能化生活的安全性推进提供可靠保障。因此,这项研究不仅在技术上具有实用性,还在信息安全方面具有重要的意义。

参考文献

[1] GURI M, SOLEWICZ Y A, DAIDAKULOV A, et al. Speake(a)r: turn speakers to microphones for fun and profit[C]. WOOT'17: Proceedings of the 11th USENIX Conference on Offensive Technologies, 2017: 13-13.

[2] XU D, YAN X H, CHEN B, et al. Energy-constrained confidentiality fusion estimation against eavesdroppers[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2021, 69(2): 624-628.

[3] 吴燕静. 基于电磁干扰麦克风的防窃听技术研究[D]. 杭州:浙江大学,2016.

[4] TUNG Y C, SHIN K G. Exploiting sound masking for audio privacy in smartphones[C]. Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security, 2019: 257-268.

[5] ZHANG G, YAN C, JI X, et al. DolphinAtack: inaudible voice commands[C]. ACM, 2017: 103-117.

[6] 吴燕静,马卓然,徐文渊. 基于电磁干扰模拟传感器的防窃听技术研究[J]. 电子技术,2016,45(4):47-51.

[7] ROY N, SHEN S, HASSANIEH H,et al. Inaudible voice commands: the long-range attack and defense[C]. 15th USENIX Symposium on Networked Systems Design and Implementation(NSDI 18), 2018: 547-560.

[8] ROY N, HASSANIEH H, CHOUDHURY R R. BackDoor: Making microphones hear inaudible sounds[C]. Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, 2017: 2-14.

[9] CHEN Y, LI H, TENG S Y, et al. Wearable microphone jamming[C]. Association for Computing Machinery, 2020: 1-12.

[10] CHEN Y, LI H, NAGELS S, et al. Understanding

the effectiveness of ultrasonic microphone jammer[J]. ArXiv Preprint,2019, ArXiv:1904.08490.

[11] GUI H, YAN C, CHENG Y S, et al. Black hole of sound: constructing an anti-eavesdropping area using ultrasound[C]. 2022 IEEE 6th Conference on Energy Internet and Energy System Integration(EI2), 2022: 1717-1722.

[12] SHEN H, ZHANG W, FANG H, et al. JamSys: Coverage optimization of a microphone jamming system based on ultrasounds[J]. IEEE Access, 2019,(99): 1-1.

[13] 汪志成,王泽旺,朱梦帆,等.基于卷积神经网络的局部放电声音识别研究[J].电子测量技术,2023,46(20):148-155.

[14] 李 响,李国正,邓明君,等.基于语音频谱图像特征的人体疲劳检测方法[J].仪器仪表学报,2021,(2):123-132.

[15] 刘书成,陈颖,陈相宁.管道声通信智能传感系统设计与实现[J].电子测量技术,2015,38(11):106-109.

[16] LI L, LIU M, YAO Y, et al. Patronus: Preventing unauthorized speech recordings with support for selective unscrambling[C]. ACM, 2020: 245-257.

[17] SNYDER D, GHAHREMANI P, POVEY D, et al. Deep neural network-based speaker embeddings for end-to-end speaker verification[C]. 2016 IEEE Spoken Language Technology Workshop(SLT), IEEE, 2016: 165-170.

[18] 张国明. 基于非线性特性的语音安全和声波通信关键技术研究[D].杭州:浙江大学,2022.

[19] 金薛冬,李东新.基于谱减法的语音信号降噪改进算法[J].国外电子测量技术,2018,37(5):63-67.

作者简介

李建勋,硕士研究生,主要研究方向为信号处理。

王开(通信作者),副教授 主要研究方向为信号处理、线性/非线性信号处理和模块化仪器。

E-mail:kaiwang@seu.edu.cn