

低信噪比下无线声传感网络采样率偏移估计方法

石 擎^{1,2}, 杨飞然^{1,2*}, 陈先梅¹, 杨 军^{1,2}

(1. 中国科学院噪声与振动重点实验室(声学研究所), 北京 100190; 2. 中国科学院大学, 北京 100049)

摘要: 现有的采样率偏移(Sampling Rate Offset, SRO)估计算法在低信噪比条件下性能严重下降. 为了解决这个问题, 本文提出一种基于子带二次互相关函数的频率滑动窗二次互相关(Frequency Sliding Double-Cross Correlation Processing, FS-DXCP)算法. 该方法使用频域滑动窗口构建无线节点观测信号间的子带二次互相关函数矩阵, 进而利用奇异值分解来自适应地消除低信噪比频段对二次互相关函数估计的影响, 最后搜寻二次互相关函数的极大值点获得SRO的估计. 计算机仿真实验表明: 在信噪比为-5 dB时, 所提方法的采样率频偏平均估计误差为4.21百万分率(part per million, ppm), 这比现有的DXCP-PHAT算法的估计误差降低了约8.17 ppm. 所提算法有效提升了低信噪比条件下采样率频偏的估计精度.

关键词: 无线声传感网络; 采样率偏移估计; 子带处理

基金项目: 国家自然科学基金(No.62171438); 北京市自然科学基金-小米创新联合基金(No.L223032); 中国科学院声学研究所自主部署“前沿探索”类项目(No.QYTS202111)

中图分类号: TN912

文献标识码: A

文章编号: 0372-2112(2024)06-2131-10

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20230339

Sampling-Rate Offset Estimation for Wireless Acoustic Sensor Networks in Low SNR Environments

SHI Qing^{1,2}, YANG Fei-ran^{1,2*}, CHEN Xian-mei¹, YANG Jun^{1,2}

(1. Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The performance of existing sampling rate offset (SRO) estimation algorithms can be degraded significantly in low signal-to-noise ratio (SNR) conditions. To address this problem, we propose the frequency-sliding double-cross correlation processing (FS-DXCP) algorithm based on the subband secondary generalized cross-correlation function to estimate SRO. The proposed algorithm adopts a frequency-domain sliding window to construct the subband SGCC function matrix of the sensor signals. Then, by utilizing the singular value decomposition (SVD), we adaptively mitigate the influence of low SNR frequency bins on estimating secondary generalized cross-correlation functions. Finally, a higher precision SRO estimation is achieved by tracking the maximum point of the estimated SGCC function. Computer simulations show that the root mean squared error of the proposed method for sampling rate offset is 4.21 ppm when the SNR is -5 dB, which is about 8.17 ppm lower than that of the double-cross correlation processing with phase transform (DXCP-PHAT) algorithm. The proposed algorithm effectively improves the estimation accuracy of the SRO in low SNR conditions.

Key words: Wireless acoustic sensor networks; Sampling-rate offset estimation; Sub-band processing

Foundation Item(s): National Natural Science Foundation of China (No.62171438); Beijing Natural Science Foundation (No.L223032); IACAS Frontier Exploration Project (No.QYTS202111)

1 引言

无线声传感网络(Wireless Acoustic Sensor Networks, WASNs)能够覆盖到比传统集中式处理系统更

广阔的物理区域,因而受到广泛关注^[1]. WASNs中每个节点采用各自独立的模数转换器(Analog-to-Digital Converter, ADC)和时钟源对信号进行采样. 然而,振荡

器的实际采样频率与标称频率之间存在误差(不同材料晶振稳定度与准确度见表1). 直接使用异步采样信号进行处理会导致诸如回声抵消、盲源分离和波束形成等算法性能下降. 因此,在 WASNs 信号处理前进行采样率偏移(Sampling-Rate Offset, SRO)估计是一项基础工作^[2].

表1 不同晶振的性能

	稳定度	准确度	成本
铯原子钟	$10^{-12} \sim 10^{-11}$	$10^{-12} \sim 10^{-11}$	高
铷原子钟	$5 \times 10^{-10} \sim 5 \times 10^{-9}$	10^{-9}	中
石英钟	$10^{-5} \sim 2 \times 10^{-5}$	$10^{-5} \sim 10^{-4}$	低

传统的 SRO 估计方法主要是通过交换无线节点间的时间戳来实现的^[3,4]. 然而,这种有监督估计方法需要无线节点具有收发时间戳的能力并且估计精度不高. 无监督估计方法由于不需要无线节点进行额外数据交换而备受青睐. 文献[5]最早提出了无监督 SRO 估计方法,该方法通过最小化 WASNs 盲源分离的能量相关得分来进行 SRO 估计. 然而,该方法要求声源和无线节点的数量相同,算法复杂度高且估计精确度低. 文献[6]通过最大化 WASNs 节点信号的短时傅里叶变换(Short-time Fourier Transform, STFT)系数的二维高斯似然函数来估计 SRO,这种方法在可用信号时长较长时性能较好. 文献[7]改进了文献[6]的方法,通过寻找间歇性移动的声源在静止时对应的信号帧来进行 SRO 估计,以应对声源间歇移动的场景. 然而,文献[6]和[7]中方法都需要预先使用一组选定的 SRO 值对信号进行重采样,导致这类方法计算量大且不具备在线处理能力. 文献[8]利用失配节点观测信号之间的互相干函数的相位随时间线性变化这一性质来估计 SRO,并提出了一个基于信噪比的加权方案来提高 SRO 的估计精度,但是,该方法需多阶段处理才能实现高精度的 SRO 估计. 文献[9]通过穷举所有可能的 SRO 值对无线节点信号进行相位补偿,使得节点信号之间相关系数最大化来确定 SRO 值. 文献[10]利用无线节点时域观测信号的二次互相关函数实现了高精度的 SRO 估计. 文献[11]提出了一种基于多信号分类算法的主动 SRO 估计方法. 然而,上述方法都只适用于离线 SRO 估计.

文献[12]提出了一种在线 SRO 估计法,该方法使用牛顿-拉尔夫算法最大化与 SRO 有关的模糊函数来估计 SRO. 但是,牛顿-拉尔夫寻峰方法容易陷入局部极大值,因此该方法对初值比较敏感. 目前先进的无监督 SRO 估计方法大都是基于无线节点观测信号之间的互相干或互相关函数实现的^[13-16]. 文献[13]利用相位变换二次互相关(Double-Cross Correlation Processing with Phase Transform, DXCP-PHAT)估计器进行在线 SRO 估计. 这个方法利用相位变换广义互相关(Gener-

alized Cross-Correlation PHAT, GCC-PHAT)算法获得观测信号之间的二次广义互相关函数,实现了更高精度的 SRO 估计. 文献[14]提出了在线的动态加权平均相干漂移法(Dynamic Weighted Average Coherence Drift, DWACD),它使用较短的信号段估计相干漂移,可以更好地处理声源位置和 SRO 变化的情况. DXCP-PHAT 算法与 DWACD 算法是目前性能最好的在线 SRO 估计方法,已有多个工作对这两种算法针对不同的应用场景进行了拓展. 文献[15]将 DWACD 方法与自适应信号估计算法相结合,提出了一种适用于非完全同步的 WASNs 网络中的自适应信号估计算法. 文献[16]以 DXCP-PHAT 方法为基础,使用根-树拓扑方法对 WASNs 中的多个无线节点进行时钟同步,提高了 DXCP-PHAT 方法在多节点 WASNs 中的计算效率和估计精度. 然而,这两种算法在低信噪比时性能急剧下降. 例如:当信噪比为 -5 dB 时,DXCP-PHAT 算法和 DWACD 算法的 SRO 估计误差都在 10 ppm 以上,且异常估计率超过 50%.

本文提出了一种频率滑动窗二次互相关(Frequency Sliding Double-Cross Correlation Processing, FS-DXCP)算法,用来解决低信噪条件下无线声传感网络中的采样率频偏估计问题. 我们将文献[17]中的子带处理思想引入到 DXCP-PHAT 方法中. 我们首先计算出无线节点观测信号之间二次互谱密度,并对其施加滑动窗来获得不同子带上的二次互相关函数. 这些子带二次互相关函数分别表征了不同子带上信号与噪声的信息. 我们证明子带二次互相关函数矩阵的最大奇异值仅与信号分量有关. 这允许我们利用奇异值分解消除低信噪比频段对二次互相关函数估计的影响,获得了稳定的二次互相关函数估计. 相比于 DXCP-PHAT 算法直接使用观测信号之间二次互谱密度函数全频段的相位信息,所提 FS-DXCP 算法能够消除低信噪比频段对 SRO 估计的影响. 仿真实验表明,所提方法在 -5 dB 信噪比环境下,4 种不同声源和混响时间设置的平均 SRO 估计误差为 4.21 ppm,平均异常估计率为 6.5%,与现有的算法相比具有显著的性能优势.

2 信号模型

考虑一个双节点分布式无线声传感网络. 传声器 1 和 2 分别来自不同的无线节点,它们具有相同的标称采样率 f_s . 传声器 2 的 ADC 真实采样率发生了轻微的偏移,两个 ADC 的实际采样率分别为 $f_{s1} = f_s$ 和 $f_{s2} = (1 + \varepsilon)f_s$,其中 ε 为它们之间的 SRO. 记 $s_i(t)$ 为声源在第 i 个节点处产生的连续信号, $v_i(t)$ 为节点 i 处的噪声信号,其中, t 表示连续时间. 则第 i 个节点采样得到的时域离散信号 $z_i[n]$ 为

$$z_i[n] = \tilde{z}_i\left(n/f_{s_i}\right) = s_i(n/f_{s_i}) + v_i(n/f_{s_i}) \quad (1)$$

其中, $\tilde{z}_i(t)$ 是第 i 个节点处的连续观测信号. 定义标准采样的采样间隔为 $T_s = 1/f_s$, 则失配采样的实际采样间隔为 $T_2 \approx (1 - \varepsilon)T_s$. 传声器 2 处的采样结果可表示为

$$z_2[n] = \tilde{z}_2(T_s(n + \tau[n])) \quad (2)$$

其中, $\tau[n] = -\varepsilon n$ 是由 SRO 在采样点 n 处造成的累积时延. 对第 2 节点处的观测信号做 STFT 变换得到

$$Z_2(k, l) = \sum_{n=0}^{N-1} h(n) \tilde{z}_2 \left(\frac{(l-1)N_s}{f_s + \varepsilon f_s} + \frac{n}{f_s + \varepsilon f_s} \right) \exp \left(-j \frac{2\pi kn}{N} \right) \quad (3)$$

其中, $h(n)$ 为窗函数; N, N_s 分别为窗长和帧移; k 代表频率索引; l 表示帧索引. 一般情况下 $\varepsilon \ll 1$, 因此可忽略 SRO 在一帧内的影响. 根据泰勒展开公式 $1/(1 + \varepsilon) = 1 - \varepsilon + \varepsilon^2 - \dots$, 对式 (3) 中的 $(l-1)N_s/(f_s + \varepsilon f_s)$ 项和 $n/(f_s + \varepsilon f_s)$ 项在 $\varepsilon = 0$ 处做泰勒级数展开, 我们忽略 SRO 在一帧内的影响以及 ε^2 以上高阶误差项得到

$$\begin{aligned} Z_2(k, l) &= \sum_{n=0}^N h(n) \tilde{z}_2 \left(\frac{(l-1)N_s}{f_s} (1 - \varepsilon + \varepsilon^2 - \dots) \right. \\ &\quad \left. + \frac{n}{f_s} (1 - \varepsilon + \varepsilon^2 - \dots) \right) \exp \left(-j \frac{2\pi kn}{N} \right) \\ &\approx \sum_{n=0}^N h(n) \tilde{z}_2 \left(\frac{(l-1)N_s}{f_s} (1 - \varepsilon) + \frac{n}{f_s} \right) \exp \left(-j \frac{2\pi kn}{N} \right) \\ &= \sum_{n=0}^N h(n) \tilde{z}_2 \left(\frac{(l-1)N_s}{f_s} + \frac{n}{f_s} - \tau_a(l) \right) \exp \left(-j \frac{2\pi kn}{N} \right) \\ &\approx \exp \left(-j \frac{2\pi k \tau_a(l)}{N} \right) \bar{Z}_2(k, l) \end{aligned} \quad (4)$$

其中,

$$\bar{Z}_2(k, l) = \sum_{n=0}^{N-1} h(n) \tilde{z}_2 \left(\frac{(l-1)N_s}{f_s} + \frac{n}{f_s} \right) \exp \left(-j \frac{2\pi kn}{N} \right) \quad (5)$$

是节点 2 处无失配采样结果 $\tilde{z}_2[n]$ 的 STFT 变换, $\tau_a(l) = -\varepsilon(l-1)N_s$ 是从第 1 帧至第 l 帧由 SRO 造成的累计时延. $Z_2(k, l)$ 只比 $\bar{Z}_2(k, l)$ 多了一个与偏移率有关的相位差 $\exp(-j2\pi k \tau_a(l)/N)$, 这一性质可用来估计 SRO 值.

3 在线 DXCP-PHAT 算法

Chinaev 等人在文献 [10] 中提出使用不同时间帧上观测信号之间的一次互相关函数 (Cross Correlation Function, CCF) 来计算二次互相关函数, 并寻找二次互相关函数的极大值点对应的的时间延迟获得 SRO 的估计. 他们在文献 [13] 中使用广义互相关 (Generalized Cross-Correlation, GCC) 代替 CCF 实现了在线的高精度 SRO 估计. 本节简要介绍这一方法.

存在采样率偏移的情况下, 无线声传感网络中两路观测信号之间第 l 帧的广义互谱密度函数 (Generalized Cross Spectral Density, GCSD) 为 [18]

$$\begin{aligned} \Phi_{z_1, z_2}(k, l) &= E \left\{ \frac{Z_1(k, l) Z_2^*(k, l)}{|Z_1(k, l) Z_2^*(k, l)|} \right\} \\ &= E \left\{ \frac{Z_1(k, l) \bar{Z}_2^*(k, l)}{|Z_1(k, l) \bar{Z}_2^*(k, l)|} \right\} \exp \left(-j \frac{2\pi k(l-1)N_s \varepsilon}{N} \right) \\ &= \Phi_{z_1, \bar{z}_2}(k) \exp \left(-j \frac{2\pi k(l-1)N_s \varepsilon}{N} \right) \\ &= \Phi_{s_1, s_2}(k) \exp \left(-j \frac{2\pi k(l-1)N_s \varepsilon}{N} \right) \end{aligned} \quad (6)$$

其中, $\Phi_{z_1, \bar{z}_2}(k) = E \{ Z_1(k, l) \bar{Z}_2^*(k, l) | Z_1(k, l) \bar{Z}_2^*(k, l) \}$ 表示无失配采样时观测信号之间的 GCSD 函数; $\Phi_{s_1, s_2}(k) = E \{ S_1(k) S_2^*(k) | S_1(k) S_2^*(k) \}$ 表示声源在节点 1 和节点 2 处产生的连续信号之间的 GCSD 函数; 上标 * 表示共轭转置. 两路观测信号之间的 GCC 函数 $\varphi_{z_1, z_2}(l, \tau)$ 是 GCSD 函数 $\Phi_{z_1, z_2}(k, l)$ 的 IFFT 变换, 表示为

$$\begin{aligned} \varphi_{z_1, z_2}(l, \tau) &= \sum_{k=0}^{N-1} \Phi_{z_1, z_2}(k, l) \exp \left(j2\pi \frac{k\tau}{N} \right) \\ &= \sum_{k=0}^{N-1} \Phi_{z_1, \bar{z}_2}(k) \exp \left(-j \frac{2\pi k(l-1)N_s \varepsilon}{N} \right) \\ &\quad \times \exp \left(j2\pi \frac{k\tau}{N} \right) \\ &= \varphi_{z_1, \bar{z}_2}(\tau) * \delta(\tau - \tau_a(l)) \\ &= \varphi_{s_1, s_2}(\tau) * \delta(\tau - \tau_a(l)) \end{aligned} \quad (7)$$

其中, $\delta(\tau)$ 为冲激函数; $\varphi_{z_1, \bar{z}_2}(l, \tau) = \sum_{k=0}^{N-1} \Phi_{z_1, \bar{z}_2}(k) \exp(j2\pi k\tau/N)$ 是无失配采样时两节点观测信号的广义互相关函数; $\varphi_{s_1, s_2}(\tau) = \sum_{k=0}^{N-1} \Phi_{s_1, s_2}(k) \exp(j2\pi k\tau/N)$ 为声源在节点 1 和节点 2 处产生的连续信号的广义互相关函数, 运算符 * 表示卷积运算. 从式 (7) 可以看出, 观测信号 GCC 函数 $\varphi_{z_1, z_2}(l, \tau)$ 的极大值点所对应的 τ 值随时间帧 l 变化, 且该值仅与 $\tau_a(l)$ 有关. 因此, 我们可通过寻找每一帧 GCC 函数 $\varphi_{z_1, z_2}(l, \tau)$ 极大值点对应的 τ 值来估计 SRO. 图 1(a) 为两个存在 62.5 ppm 固定采样率偏移的无线节点信号间 GCC 函数 $\varphi_{z_1, z_2}(l, \tau)$ 的实例. 可以看出两节点观测信号的 GCC 函数 $\varphi_{z_1, z_2}(l, \tau)$ 极大值点所对应的 τ 值随时间线性变化.

在实际中由于噪声和混响等因素的影响, 我们很难准确估计两节点观测信号每一帧的 GCC. 为此, 考虑无线节点观测信号间的二次广义互谱密度 (Secondary Generalized Cross Spectral Density, SGCS) 函数

$$\begin{aligned} \Psi_{z_1, z_2}(k) &= E \left\{ \Phi_{z_1, z_2}(k, l) \Phi_{z_1, z_2}^*(k, l - L_b) \right\} \\ &= E \left\{ \Phi_{z_1, \bar{z}_2}(k) \Phi_{z_1, \bar{z}_2}^*(k) \right\} \exp \left(j \frac{2\pi}{N} k \varepsilon N_s L_b \right) \\ &= \Psi_{s_1, s_2}(k) \exp \left(j \frac{2\pi}{N} k \varepsilon N_s L_b \right) \end{aligned} \quad (8)$$

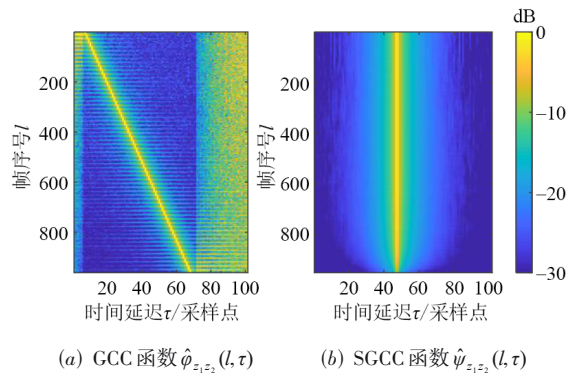


图1 GCC函数和SGCC函数对比

其中, L_b 为用于计算 SGCCSD 函数的帧之间的距离; $\Psi_{s_1, s_2}(k) = E\{\Phi_{s_1, s_2}(k)\Phi_{s_1, s_2}^*(k)\}$ 为声源在节点 1 和节点 2 处产生的连续信号之间的二次广义互谱密度函数. 对 $\Psi_{s_1, s_2}(k, l)$ 进行 IFFT 变换得到观测信号之间的二次广义互相关 (Secondary Generalized Cross-Correlation, SGCC) 函数

$$\begin{aligned} \psi_{s_1, s_2}(\tau) &= \sum_{k=0}^{N-1} \Psi_{s_1, s_2}(k) \exp\left(-j \frac{2\pi k N_s \varepsilon L_b}{N}\right) \exp\left(j 2\pi \frac{k\tau}{N}\right) \quad (9) \\ &= \psi_{s_1, s_2}(\tau) * \delta(\tau - \varepsilon L_b N_s) \end{aligned}$$

其中,

$$\begin{aligned} \psi_{s_1, s_2}(\tau) &= \sum_{k=0}^{N-1} \Psi_{s_1, s_2}(k) \exp\left(j 2\pi \frac{k\tau}{N}\right) \quad (10) \\ &= \sum_{k=0}^{N-1} |\Phi_{s_1, s_2}(k)|^2 \exp\left(j 2\pi \frac{k\tau}{N}\right) \end{aligned}$$

为声源在节点 1 和节点 2 处产生的连续信号的二次互相关函数. 由于 $\Psi_{s_1, s_2}(k)$ 在所有频率点上均为正实数, $\psi_{s_1, s_2}(\tau)$ 在 $\tau=0$ 处有极大值. 失配节点观测信号的 SGCC 函数 $\psi_{s_1, s_2}(\tau)$ 在 $\tau = \varepsilon l_b N_s$ 处有一个恒定的极大值点, 我们可以通过寻找 SGCC 函数 $\psi_{s_1, s_2}(\tau)$ 的极大值点来估计 SRO. 寻找 SGCC 函数 $\hat{\psi}_{s_1, s_2}(l_a, \tau)$ 的极大值点得到第 l_a 帧的 SRO 值:

$$\hat{\tau}_d(l_a) = \arg \max_{\tau} \hat{\psi}_{s_1, s_2}(l_a, \tau) \quad (11)$$

$$\hat{\varepsilon}(l_a) = \frac{\hat{\tau}_d(l_a)}{N_s L_b} \quad (12)$$

其中, $\hat{\tau}_d$ 是二次插值后的第 l_a 帧 SGCC 函数 $\hat{\psi}_{s_1, s_2}(l_a, \tau)$ 的极大值点.

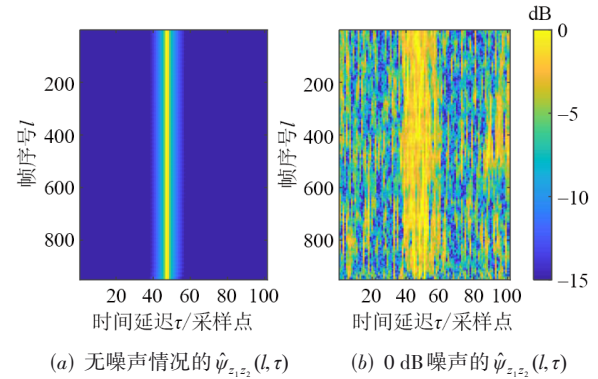
观测信号间的二次广义互谱密度函数 $\Psi_{s_1, s_2}(k)$ 可能是时变的. 我们使用递归平均方法对式(7)及式(9)中的统计期望进行在线估计:

$$\hat{\Phi}_{s_1, s_2}(k, l) = \alpha_1 \hat{\Phi}_{s_1, s_2}(k, l-1) + (1 - \alpha_1) \frac{Z_1(k, l) Z_2^*(k, l)}{|Z_1(k, l) Z_2^*(k, l)|} \quad (13)$$

$$\begin{aligned} \hat{\Psi}_{s_1, s_2}(k, l) &= \alpha_2 \hat{\Psi}_{s_1, s_2}(k, l-1) \\ &+ (1 - \alpha_2) \hat{\Phi}_{s_1, s_2}(k, l) \hat{\Phi}_{s_1, s_2}^*(k, l - l_b) \end{aligned} \quad (14)$$

其中, α_1 和 α_2 为平滑因子. 图 1(b) 是使用图 1(a) 中的信号计算 SGCC 函数 $\hat{\psi}_{s_1, s_2}(l, \tau)$ 的结果图例, 可以看到 SGCC 函数 $\hat{\psi}_{s_1, s_2}(l, \tau)$ 的极大值点对应的 τ 值在所有时间帧 l 上是相同的.

DXCP-PHAT 算法利用 GCC-PHAT 估计无线节点观测信号之间的 SGCC 函数 $\hat{\psi}_{s_1, s_2}(l, \tau)$. 当信噪比较低时, DXCP-PHAT 方法的估计误差较大. 图 2 展示了在无噪声条件与 0 dB 不相关白噪声条件下估计得到的 SGCC 函数 $\hat{\psi}_{s_1, s_2}(l, \tau)$ 的对比实例. 可以看出当信噪比较高时, SGCC 函数 $\hat{\psi}_{s_1, s_2}(l, \tau)$ 的极大值点所对应的 τ 值在所有时间帧 l 上都十分清晰; 当信噪比较低时, SGCC 函数 $\hat{\psi}_{s_1, s_2}(l, \tau)$ 的极大值点对应的 τ 值则不够明显且在某些时间帧内出现了伪峰.

图2 噪声对 SGCC 函数 $\hat{\psi}_{s_1, s_2}(l, \tau)$ 估计的影响

4 所提 FS-DXCP 算法

在这一节, 我们引入滑动子带方法来提高 DXCP-PHAT 算法的性能. 我们的核心思想是利用基于滑动频率子带的广义互相关方法代替 GCC-PHAT 来估计二次广义互相关函数. 在所提 SRO 估计算法中, 首先使用频率滑动窗口构建子带二次相关 (Frequency Sub-band Secondary Generalized Cross Correlation, FS-SGCC) 矩阵, 然后通过奇异值分解从 FS-SGCC 矩阵中获得更精确的 SGCC 估计, 从而提高 DXCP-PHAT 算法在低信噪比时的性能.

4.1 子带 SGCC 矩阵构建

DXCP-PHAT 方法直接使用二次互谱函数的全部频段信息得到 SGCC 函数的估计, 这在高信噪比条件下是有效的. 然而, 当信噪比较低时, 在信号能量较弱频段内的相位信息由噪声主导, 使用这些频段内的相位信息反而会降低 SRO 估计精度. 因此, 我们提出采用子带 SGCCSD 函数, 自适应选择高信噪比频段的信息来估

计 SGCC 函数.

利用窗长为 $2B$ 的窗函数 $W_{2B} = [W_{2B}(0), \dots, W_{2B}(2B-1)]$ 对 SGCSO 函数 $\hat{\Psi}_{z_1, z_2}(k, l)$ 进行加窗. 第 l 帧的第 c 个子带 SGCSO 函数可表示为

$$\hat{\Psi}_c(k, l) = \hat{\Psi}_{z_1, z_2}(k + cM, l)W(k) \quad (15)$$

其中,

$$W(k) = \begin{cases} W_{2B}(k+B), & 0 \leq k \leq B-1 \\ 0, & B \leq k \leq N-2-B \\ W_{2B}(k-N+B+1), & N-1-B \leq k \leq N \end{cases} \quad (16)$$

为频域窗函数; M 为频率窗之间滑动的频点数; $c \in \{0, 1, \dots, C-1\}$ 为子带序号; $C = \lfloor (N/2 - 2B + M)/M \rfloor$ 为子带总数. 频率滑动窗口工作示意图如图 3 所示, 其中黑色实线对应 $c=0$ 时的 SGCSO 函数 $\hat{\Psi}_{z_1, z_2}(k, l)$, 蓝色虚线对应 $c=1$ 时的 SGCSO 函数 $\hat{\Psi}_{z_1, z_2}(k+M, l)$, 红色实线为频率窗函数 $W(k)$. 对子带 SGCSO 函数 $\hat{\Psi}_c(k, l)$ 进行 IFFT 变换得子带 SGCC 函数:

$$r_c(\tau) = \sum_{k=0}^{B-1} \hat{\Psi}_c(k, l) \exp\left(j2\pi \frac{k\tau}{B}\right) \quad (17)$$

将每个子带的 SGCC 函数 $r_c = [r_c(0), r_c(1), \dots, r_c(B-1)]^T$ 作为列向量堆叠在一起得到 FS-SGCC 矩阵:

$$\mathbf{R} = [r_0, r_1, \dots, r_{L-1}] \in \mathbf{C}^{N \times L} \quad (18)$$

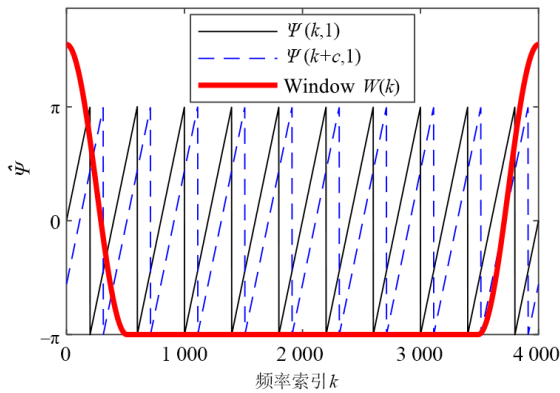


图 3 4096 点 FFT 中 1024 点频率滑动窗示意图

4.2 SGCC 函数提取

理想情况下, FS-SGCC 矩阵的元素为

$$\begin{aligned} \tilde{r}_c(\tau) &= \sum_{k=0}^{B-1} \hat{\Psi}_c(k, l) \exp\left(j2\pi \frac{k\tau}{B}\right) \\ &= \exp\left(-j2\pi c \frac{M}{B} \tau\right) \hat{\Psi}_{z_1, z_2}(l, \tau) * w(\tau) \\ &= \exp\left(-j2\pi c \frac{M}{B} \tau\right) \psi_{s_1, s_2}(\tau) * \delta\left(\tau - \varepsilon l_b N_s \frac{B}{N}\right) * w(\tau) \end{aligned} \quad (19)$$

其中, $\hat{\Psi}_{z_1, z_2}(k, l)$ 表示第 l 帧无噪声情况下两个节点之间的 SGCSO 函数, $w(\tau)$ 为频域上窗函数 $W(k)$ 的逆傅里叶

变换. 记 $\boldsymbol{\psi}_0 = [\psi_0(0), \psi_0(1), \dots, \psi_0(B-1)]^T$, $\boldsymbol{\psi}_0(\tau) = \psi_{s_1, s_2}(\tau) * \delta(\tau - \varepsilon l_b N_s B/N) * w(\tau)$, 那么无噪声情况下的 FS-SGCC 矩阵为

$$\tilde{\mathbf{R}} = \boldsymbol{\psi}_0 \mathbf{e}^* \quad (20)$$

其中, $\mathbf{e} = [1, \exp(-j2\pi M \varepsilon l_b N_s / B), \dots, \exp(-j2\pi M \varepsilon l_b N_s (L-1)/B)]^T$. 可以看出此时矩阵 $\tilde{\mathbf{R}}$ 是秩一的.

而在存在噪声的情况下, 第 l 帧上两个节点之间的第 c 个子带 SGCSO 函数为

$$\hat{\Psi}_c(k, l) = \tilde{\Psi}_c(k, l) + V_c(k, l) \quad (21)$$

其中, $V_c(k, l) = V_{z_1, z_2}(k + cM, l)W(k)$ 表示第 c 个子带 SGCSO 函数的加性噪声; $V_{z_1, z_2}(k, l)$ 是信号间 SGCSO 函数的加性噪声. 纯噪声情况下第 c 个子带的 SGCC 函数为 $\mathbf{n}_c = [n_c(0), n_c(1), \dots, n_c(B-1)]^T$, 则

$$n_c(\tau) = \sum_{k=0}^{B-1} V_c(k, l) \exp\left(j2\pi \frac{k\tau}{B}\right) \quad (22)$$

由于使用了相位变换加权, 根据帕塞瓦尔定理, 纯噪声子带的 SGCC 函数 \mathbf{n}_c 与纯信号子带的 SGCC 函数能量是相等的, 满足关系:

$$\mathbf{n}_c^* \mathbf{n}_c = \tilde{\mathbf{r}}_c^* \tilde{\mathbf{r}}_c \quad (23)$$

其中, $\tilde{\mathbf{r}}_c = [\tilde{r}_c(0), \tilde{r}_c(1), \dots, \tilde{r}_c(B-1)]^T$, $\tilde{\mathbf{r}}_c(\tau) = \sum_{k=0}^{B-1} \hat{\Psi}_c(k, l) \exp\left(j2\pi \frac{k\tau}{B}\right)$ 是纯信号子带的 SGCC 函数.

估计得到的第 c 个子带 SGCC 函数为^[17]

$$\hat{\mathbf{r}}_c = \alpha_c \tilde{\mathbf{r}}_c + (1 - \alpha_c) \mathbf{n} \quad (24)$$

其中, \mathbf{n}_c 为第 c 列中噪声分量; $1 - \alpha_c$ 代表第 c 列中噪声的比重. 因此, FS-SGCC 矩阵为

$$\hat{\mathbf{R}} = \tilde{\mathbf{R}} \mathbf{G} + \mathbf{N}(\mathbf{I} - \mathbf{G}) \quad (25)$$

其中, $\mathbf{N} = [\mathbf{n}_0, \mathbf{n}_1, \dots, \mathbf{n}_{L-1}]$ 是包含各列噪声分量的矩阵, $\mathbf{G} = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_{L-1})$. 由于引入了噪声项, 式 (25) 中的 FS-SGCC 矩阵 $\hat{\mathbf{R}}$ 是满秩的.

为了从矩阵 $\hat{\mathbf{R}}$ 中提取出 SGCC 函数, 我们对该矩阵进行进一步分析, 考虑:

$$\begin{aligned} \hat{\mathbf{R}} \hat{\mathbf{R}}^* &= \left(\sum_{c=0}^{L-1} \alpha_c^2 \right) \boldsymbol{\psi}_0 \boldsymbol{\psi}_0^* \\ &+ \sum_{c=0}^{L-1} (\alpha_c - \alpha_c^2) \exp(-j2\pi c \frac{M}{N} \varepsilon l_b N_s) \mathbf{n}_c \boldsymbol{\psi}_0^* \\ &+ \sum_{c=0}^{L-1} (\alpha_c - \alpha_c^2) \exp(-j2\pi c \frac{M}{N} \varepsilon l_b N_s) \boldsymbol{\psi}_0 \mathbf{n}_c^* \\ &+ \sum_{c=0}^{L-1} (1 - \alpha_c)^2 \mathbf{n}_c \mathbf{n}_c^* \end{aligned} \quad (26)$$

矩阵 $\hat{\mathbf{R}} \hat{\mathbf{R}}^*$ 可分解成四个矩阵之和. 前三个矩阵都是秩一的. 第一个矩阵的非零特征值是 $\sum_{c=0}^{L-1} \alpha_c^2$, 其对应的特征向量是 $\boldsymbol{\psi}_0$. 第二个矩阵的特征值与对应的特征向量由噪

声分量决定. 第三个矩阵唯一的非零特征值是 $\sum_{c=0}^{C-1} (\alpha_c - \alpha_c^2) \exp(-j2\pi c M \varepsilon L_b N_s / N) \mathbf{n}_c^* \boldsymbol{\psi}_0$, 对应的特征向量为 $\boldsymbol{\psi}_0$. 第四个矩阵则是由各个频段上的噪声分量构建的满秩矩阵. 一般情况下, $\hat{\mathbf{R}}\hat{\mathbf{R}}^*$ 的最大特征值由第一个矩阵决定, 其对应的特征向量由 $\boldsymbol{\psi}_0$ 主导, 而噪声分量对该特征值的形成影响较小. 因此可以将矩阵 $\hat{\mathbf{R}}\hat{\mathbf{R}}^*$ 最大特征值所对应的特征向量作为 $\boldsymbol{\psi}_0$ 的估计. 而 $\hat{\mathbf{R}}\hat{\mathbf{R}}^*$ 的特征向量对应于 $\hat{\mathbf{R}}$ 的左奇异矢量, 对矩阵 $\hat{\mathbf{R}}$ 进行奇异值分解:

$$\hat{\mathbf{R}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^* \quad (27)$$

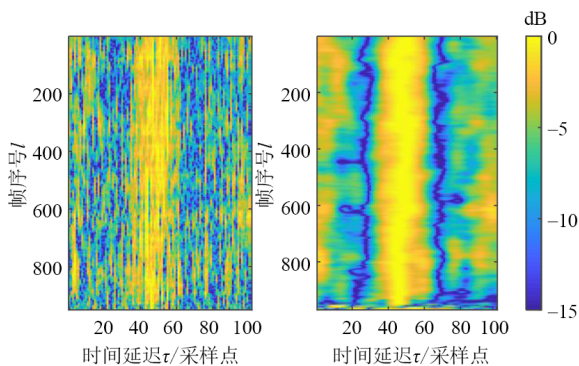
其中, \mathbf{U} 为左奇异矢量矩阵; $\boldsymbol{\Sigma}$ 为将奇异值从大到小排列组成的对角矩阵; \mathbf{V} 为右奇异矢量矩阵. 考虑到 $\pm\boldsymbol{\psi}_0$ 均为左奇异矢量, 我们使用下式对 $\boldsymbol{\psi}_0$ 进行估计:

$$\hat{\boldsymbol{\psi}}_0 = \text{Re}\{\mathbf{u}_1\} \text{sign}(\text{Re}\{u_1[\gamma]\}) \quad (28)$$

其中, \mathbf{u}_i 为左奇异矢量矩阵 \mathbf{U} 的第 i 列; $\text{Re}\{u_1[\gamma]\}$ 为 $\text{Re}\{\mathbf{u}_1\}$ 中绝对值最大的元素. 由于 $w(k)$ 是频域上对称的窗函数, $\omega(\tau)$ 的极大值在零点出现, 因此 $\boldsymbol{\psi}_0(\tau)$ 与 $\boldsymbol{\psi}_{s_1, s_2}(\tau) * \delta(\tau - \varepsilon l_b N_s)$ 具有相同的极大值点:

$$\hat{\tau}_\Delta = \arg \max_{\tau} \boldsymbol{\psi}_0[\tau] \quad (29)$$

我们利用二次插值进一步提高 $\hat{\tau}_\Delta$ 的估计精度. 图 4 展示了所提算法与 DXCP-PHAT 算法得到的 SGCC 函数 $\hat{\boldsymbol{\psi}}_{z_1, z_2}(l, \tau)$ 估计结果, 实验条件与图 3(b) 相同. 可以看出当信噪比较低时, DXCP-PHAT 算法的估计性能明显降低, 在大部分时间帧上无法提供有效的 SRO 估计, 但是所提 FS-DXCP 算法在几乎所有时间帧上都提供了稳定且精确的 SRO 估计.



(a) 文献[13]中方法的 $\hat{\boldsymbol{\psi}}_{z_1, z_2}(l, \tau)$ (b) 本文方法的 $\hat{\boldsymbol{\psi}}_{z_1, z_2}(l, \tau)$

图4 噪声条件下所提方法与 DXCP-PHAT^[13]方法估计的 SGCC 函数对比

5 仿真与分析

本节对所提 FS-DXCP 方法的性能进行验证. 我们在各种实验条件下对比了所提方法与现有两种在线 SRO 估计方法 DXCP-PHAT^[13] 与 DWACD^[14] 的性能.

5.1 仿真设置

为了进行公平的性能对比, 本文房间参数的设置与文献[13]相似. 房间尺寸为 $5 \text{ m} \times 4 \text{ m} \times 3 \text{ m}$, 混响时间为 $T_{60} = 200 \text{ ms}$ 与 $T_{60} = 600 \text{ ms}$. 节点与墙壁距离大于 0.5 m , 节点与声源距离大于临界距离, 声源位置与节点位置随机选择. 为了评估算法在不同数量声源情况下的性能, 设置声源数量为 1 个或 2 个. 纯净声源信号是从 TIMIT 数据库中提取的采样率为 16 kHz 且长度为 1 min 的语音信号. 混响信号中包含 20 dB 的扩散 Babble 噪声以及不同程度的空间不相关白噪声. 对第二个节点进行插值来模拟采样频率偏移, 其中 SRO 为 $\varepsilon \in \pm\{0, 20, 40, 60, 80, 100\} \text{ ppm}$. 各个节点拾取到的信号由纯净语音与房间脉冲响应卷积获得, 房间脉冲响应由镜像法^[19]得到.

算法在进行 SRO 估计前均使用 GCC-PHAT 方法进行粗略同步. 在 DXCP-PHAT 算法以及所提 FS-DXCP 算法中, STFT 窗长设置为 8192 , 帧移设置为 2048 , 窗的类型为 hann 窗. DWACD 算法每次更新输入信号长度为 8192 , 更新步长为 2048 , STFT 窗长为 4096 , Welch 法的帧移为 512 , 平滑参数设置为 0.95 . 在 DXCP-PHAT 与 FS-DXCP 方法中均设置 $\alpha_1 = 0.5$, $\alpha_2 = 0.99$ 和 $L_b = 19$. 在所提 FS-DXCP 方法中, 频域窗长设置为 $B = N/4 = 2048$, 相邻窗之间的间隔为 $B_s = B/2 = 1024$. 性能测试指标为异常估计率 (Percentage of Anomalous Estimates) 与估计的平均均方根误差 (Root-Mean-Square Error, RMSE), 其中估计误差大于 10 ppm 的结果记为异常结果. RMSE 计算公式为

$$\text{RMSE} = \sqrt{\frac{1}{L} \sum_{l_a=L_c+1}^L |(\hat{\varepsilon}(l_a) - \varepsilon(l_a))|^2} \quad (30)$$

其中, L 表示总帧数; $\hat{\varepsilon}(l_a)$ 表示第 l_a 帧的估计 SRO 值; $\varepsilon(l_a)$ 表示真实 SRO 值; L_c 表示建立时间. RMSE 的单位为 ppm. 同时, 我们在固定 SRO 数据集中将估计结果与计算的克拉美劳界 (Cramér-Rao Lower Bound, CRLB) 进行了比较. CRLB 计算过程如下.

根据式 (8), 无线节点信号间第 l 帧的二次互谱密度函数在存在噪声的情况下表示为

$$\begin{aligned} \hat{\Psi}_{z_1, z_2}(k, l) &= \hat{\Psi}_{z_1, z_2}(k, l) + V_{z_1, z_2}(k, l) \\ &= \left| E \left(\frac{S_1(k, l) S_2^H(k, l)}{|S_1(k, l) S_2^H(k, l)|} \right) \right|^2 \exp \left(j \frac{2\pi}{N} k \varepsilon N_s L_b \right) \\ &\quad + V_{z_1, z_2}(k, l) \\ &= \exp \left(j \frac{2\pi}{N} k \varepsilon N_s L_b \right) + V_{z_1, z_2}(k, l) \end{aligned} \quad (31)$$

我们假设噪声的实部与虚部为独立同分布零均

值的高斯变量,其能量在所有频带上平均分布,则有 $1/2\sigma^2 = \text{SNR}$, 其中 σ^2 为 $V(k, l)$ 实部或虚部的方差. $V(k)$ 的实部与虚部服从 $(\text{Re}(V(k)), \text{Im}(V(k))) \sim N(0, 0, \sigma^2, \sigma^2, 0)$ 的二维正态分布. 因此, 我们得到噪声变量的似然分布为^[20]

$$p_k(\Psi_{z_1, z_2}(k), \varepsilon) = (2\pi\sigma^2)^{-\frac{NL}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{l=1}^L \sum_{k=0}^{N-1} \left((\text{Re}(\Psi_{z_1, z_2}(k)) - \cos\left(\frac{2\pi}{N} k\varepsilon N_s L_b\right))^2 + (\text{Im}(\Psi_{z_1, z_2}(k)) - \sin\left(\frac{2\pi}{N} k\varepsilon N_s L_b\right))^2 \right)\right] \quad (32)$$

因此, Fisher 信息矩阵为^[21]

$$I(\varepsilon) = -E\left[\frac{\partial^2 \ln p_k(\Psi_{z_1, z_2}(k), \varepsilon)}{\partial \varepsilon^2}\right] \approx \frac{1}{\sigma^2} \frac{(2\pi N_s L_b)^2}{N^2} \sum_{l=1}^L \sum_{k=0}^{N-1} k^2 \quad (33)$$

我们得到 SRO 估计的 CRLB 为

$$\text{var}(\varepsilon) \geq \frac{\sigma^2}{\frac{(2\pi N_s L_b)^2}{N^2} \sum_{l=1}^L \sum_{k=0}^{N-1} k^2} = \frac{6N\sigma^2}{(N-1)(2N-1)L(2\pi N_s L_b)^2} \quad (34)$$

5.2 固定 SRO 场景下性能评估

图 5 给出了三种算法的 RMSE 随时间的变化曲线, 以及基于二次互相关方法的 CRLB 随时间的变化曲线. 本实验中, 我们选择双声源, 混响时间 $T_{60} = 200$ ms, 空间不相关白噪声为 10 dB. 所展示结果是 100 次实验取平均得到的. 由于估计原理类似, FS-DXCP 方法与 DXCP-PHAT 方法的收敛速度相似. DWACD 方法虽然能够快速收敛, 但是在收敛后稳态误差比较大.

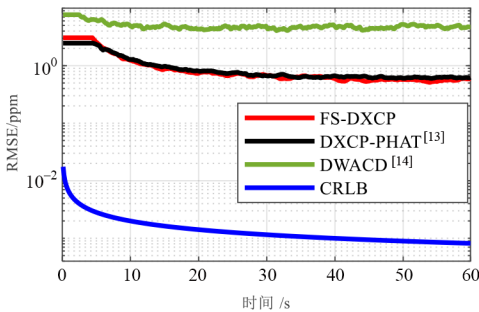


图 5 三种算法收敛性能比较

图 6~9 展示了三种在线的 SRO 估计算法在不同信噪比和不同混响时间时的估计误差和异常估计率. 基于二次互相关方法的 CRLB 也画在相应的图中. 图 6 和图 7 采用单声源, 图 8 和图 9 采用双声源. 图 6 和图 8

中混响时间设置为 $T_{60} = 200$ ms, 图 7 和图 9 中混响时间设置为 $T_{60} = 600$ ms. 信噪比设置为 -10 dB、-5 dB、0 dB、5 dB、10 dB. 最大失配率为 200 ppm.

从图 6(a) 看出在低信噪比情况下, 所提 FS-DXCP 比其他两种算法具有更低的 RMSE; 随着信噪比降低, 所提算法比其他两种算法的性能优越性更显著. 基于二次相关函数的 FS-DXCP 和 DXCP-PHAT 算法的估计精度均优于 DWACD 算法. 这是因为 DWACD 算法使用较短的信号来估计节点信号间的相干漂移, 导致算法对噪声干扰敏感. 当信噪比小于 -5 dB 时, 所提算法估计精度明显优于 DXCP-PHAT 算法, 并且结果更靠近 CRLB. 从图 6(b) 可以看出, 所提 FS-DXCP 算法在异常估计比这个指标的表现也是最好的. 即使信噪比低至 0 dB, 所提 FS-DXCP 算法的异常估计比依然低于 5%. 对比图 6 和图 7, 我们看出: 当混响时间增加时, 所有方法的估计误差都会增大, 但所提算法受到混响的影响比其他两种算法要小. 在 SNR = -10 dB、-5 dB、0 dB 时, FS-DXCP 算法的异常估计率都远低于 DXCP-PHAT 算法.

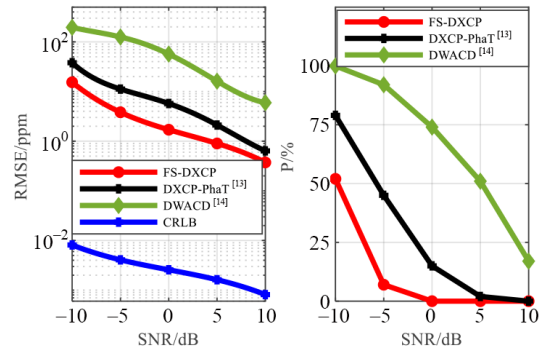


图 6 单声源、混响时间 200 ms 采样率频偏估计性能

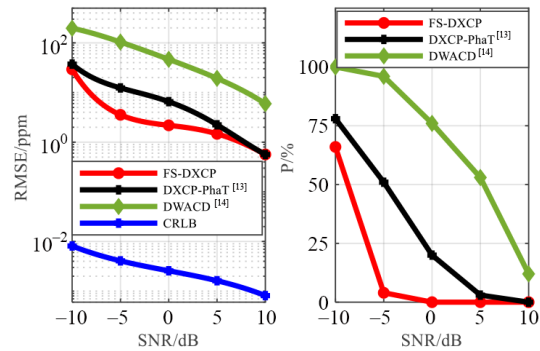


图 7 单声源、混响时间 600 ms 采样率频偏估计性能

在更为复杂的双声源数据集上仿真结果如图 8 和图 9 所示. 相对于单声源的场景, 所有算法在双声源场

景下性能都有下降,但所提 FS-DXCP 方法在低信噪比时仍然取得了最好的性能,并且估计结果更加靠近理论的 CRLB. 但我们注意到,三种算法的 SRO 估计精度都远低于 CRLB. 造成这一现象的原因可能是这三种算法都是在线实时实现的并且采用式(13)和式(14)这样的递归平均来代替统计平均,导致这些算法无法避免瞬态误差的影响.

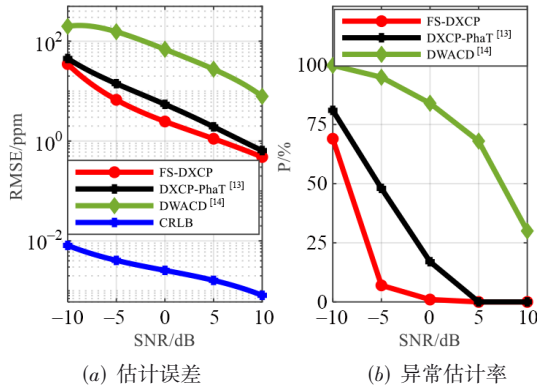


图8 双声源、混响时间 200 ms 采样率频偏估计性能

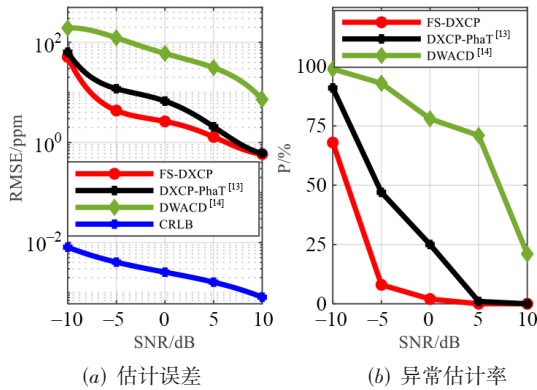


图9 双声源、混响时间 600 ms 采样率频偏估计性能

5.3 时变 SRO 场景性能评估

本节比较三种 SRO 估计算法对时变 SRO 的追踪能力. 我们采用文献[14]中包含声源位置变化和讲话停顿的时变 SRO 数据集. 混响时间为 $T_{60}=300$ ms, 声源与节点距离为 3.2 m. 我们利用 Ornstein-Uhlenbeck 过程模拟时变 SRO, 每一帧的偏移量为

$$\varepsilon[l] = \varepsilon[l-1] + \theta(\mu_{\infty} - \varepsilon[l-1]) + x_{\varepsilon}[l] \quad (35)$$

其中, θ 为平滑参数; $x_{\varepsilon}[l] \sim N(0; \sigma_{OU}^2)$ 服从方差为 $\sigma_{OU}^2=0.05$ ppm 的零均值正态分布; μ_{∞} 为最终达到的 SRO 值. 起始 SRO 值 $\varepsilon[0]=\mu_{\infty} + \Delta_{\text{start}}$, Δ_{start} 的范围为 $-10 \sim 10$ ppm. 图 10 展示了三种在线 SRO 估计算法在时变 SRO 场景下的 RMSE. 当信噪比较高时(20 dB), 三种 SRO 估计算法的性能相当, 其估计误差均在 1 ppm 以内. 随着信噪比降低, 所提 FS-DXCP 算法性能优于其

他两种算法.

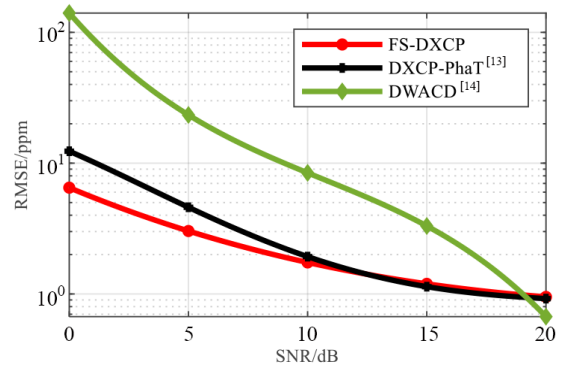


图10 时变 SRO 场景下各算法采样率频偏估计误差

5.4 计算复杂度以及计算时间对比

在这一节中, 我们对比了所提算法与现有的 DXCP-PHAT 算法以及 DWACD 算法单次迭代的计算复杂度. 在计算复杂度分析中, 我们只考虑实值乘法计算. 复值乘法是实值乘法的 4 倍, 除法运算和平方根运算为实值乘法的 20 倍和 40 倍. 分析结果如表 2 所示. 表 2 中 N_1 表示 DWACD 方法中使用 Welch 方法计算互功率谱的 FFT 窗长, $A=(N-N_1)/N_s$, N_s 为 Welch 方法的帧.

表2 各算法进行一次迭代所需的乘法次数

算法	乘法次数
DWACD ^[14]	$(3A+1)N_1 \log_2 N_1 + (6A+36)N_1$
FS-DXCP	$(2 \log_2 N + 16 + 2C)N + (2C \log_2 2B + 2C^2)B$
DXCP-PHAT ^[13]	$(14 + 3 \log_2 N)N$

DWACD 算法中使用了 Welch 平均的方法计算功率谱密度, 并且在计算互相关函数时需要进行开方运算, 因此 DWACD 算法所需的乘法次数比所提 FS-DXCP 算法和 DXCP-PHAT 都要高. 而相比 DXCP-PHAT 算法, 所提 FS-DXCP 算法在计算信号间二次互谱密度后还要进行加窗以及奇异值分解运算. 因此所提 FS-DXCP 算法的所需实值乘法次数比 DXCP-PHAT 算法稍高.

我们在 CPU 为 R7-5800H 的平台上利用 Matlab2018b 软件测试了三种方法的运行时间. DWACD 算法、DXCP-PHAT 算法与所提 FS-DXCP 算法进行一次迭代所需的平均时间分别为 12.25 ms、5.63 ms 以及 8.38 ms. 该结果与运算复杂度分析相符.

6 结论

本文提出了一种适用于低信噪比条件下的 FS-DXCP 算法来估计无线声传感网络中的采样率偏移. 该方法中使用频域滑动窗对无线节点观测信号之间的二

次互谱密度函数进行加窗,构建了子带二次互相关函数矩阵,对子带二次互相关函数矩阵进行奇异值分解获得了二次互相关函数估计,最终计算得到采样率偏移值.仿真实验对比了不同信噪比和不同混响时间条件下所提算法和现有的两种 SRO 估计算法的性能,实验结果证明所提 FS-DXCP 算法在低信噪比条件下取得了比现有的在线算法更加精确的 SRO 估计和更低的异常估计率.

参考文献

- [1] 王睿,陈喆,殷福亮.分布式声传感器阵列校准方法综述[J].电子学报,2021,49(12):2468-2478.
WANG R, CHEN Z, YIN F L. Review on calibration of distributed acoustic sensor arrays[J]. Acta Electronica Sinica, 2021, 49(12): 2468-2478. (in Chinese)
- [2] LIU Z C, ZHANG Z Y, HE L W, et al. Energy-based sound source localization and gain normalization for ad hoc microphone arrays[C]//2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) ICASSP'07. Piscataway: IEEE, 2007: II-761-II-764.
- [3] CHAUDHARI Q M. A simple and robust clock synchronization scheme[J]. IEEE Transactions on Communications, 2012, 60(2): 328-332.
- [4] SCHMALENSTROEER J, HÄB-UMBACH R. Sampling rate synchronization in acoustic sensor networks with a pre-trained clock skew error model[C]//21st European Signal Processing Conference (EUSIPCO). Piscataway: IEEE, 2013: 1-5.
- [5] LIU Z. Sound source separation with distributed microphone arrays in the presence of clocks synchronization errors[C/OL]. (2008-09-17). <https://iwaenc.org/proceedings/2008/contents/papers/9030.pdf>.
- [6] MIYABE S, ONO N, MAKINO S. Blind compensation of interchannel sampling frequency mismatch for ad hoc microphone array based on maximum likelihood estimation[J]. Signal Processing, 2015, 107: 185-196.
- [7] ARAKI S, ONO N, KINOSHITA K, et al. Estimation of sampling frequency mismatch between distributed asynchronous microphones under existence of source movements with stationary time periods detection[C]//2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2019: 785-789.
- [8] SCHMALENSTROEER J, HEYMANN J, DRUDE L, et al. Multi-stage coherence drift based sampling rate synchronization for acoustic beamforming[C]//2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP). Piscataway: IEEE, 2017: 1-6.
- [9] WANG L, DOCCLO S. Correlation maximization-based sampling rate offset estimation for distributed microphone arrays[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(3): 571-582.
- [10] CHINAEV A, THÜNE P, ENZNER G. A double-cross-correlation processor for blind sampling rate offset estimation in acoustic sensor networks[C]//2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2019: 641-645.
- [11] WANG R, CHEN Z, YIN F L. Active sampling rate calibration method for acoustic sensor networks[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28: 3095-3107.
- [12] CHERKASSKY D, GANNOT S. Blind synchronization in wireless acoustic sensor networks[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(3): 651-661.
- [13] CHINAEV A, THÜNE P, ENZNER G. Double-cross-correlation processing for blind sampling-rate and time-offset estimation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29: 1881-1896.
- [14] GBURREK T, SCHMALENSTROEER J, HAEB-UMBACH R. On synchronization of wireless acoustic sensor networks in the presence of time-varying sampling rate offsets and speaker changes[C]//2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2022: 916-920.
- [15] DIDIER P, VAN WATERSCHOOT T, DOCCLO S, et al. Sampling rate offset estimation and compensation for distributed adaptive node-specific signal estimation in wireless acoustic sensor networks[J]. IEEE Open Journal of Signal Processing, 2023, 4: 71-79.
- [16] CHINAEV A, ENZNER G. Distributed synchronization for ad-hoc acoustic sensor networks using closed-loop double-cross-correlation processing[C]//2022 International Workshop on Acoustic Signal Enhancement (IWAENC). Piscataway: IEEE, 2022: 1-5.
- [17] COBOS M, ANTONACCI F, COMANDUCCI L, et al. Frequency-sliding generalized cross-correlation: A sub-band time delay estimation approach[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28: 1270-1281.
- [18] KNAPP C, CARTER G. The generalized correlation

- method for estimation of time delay[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1976, 24(4): 320-327.
- [19] ALLEN J B, BERKLEY D A. Image method for efficiently simulating small-room acoustics[J]. The Journal of the Acoustical Society of America, 1979, 65(4): 943-950.
- [20] ZHANG J, WU P P. Joint sampling synchronization and source localization for wireless acoustic sensor networks[J]. IEEE Communications Letters, 2020, 24(5): 1020-1023.
- [21] KAY S M. Fundamentals of Statistical Signal Processing[M]. Englewood Cliffs: Prentice-Hall PTR, 1993.

作者简介



石 擎 男,1998年生,湖南株洲人.中国科学院声学研究所博士研究生.主要研究方向为音频信号处理.

E-mail: shiqing161@mails.ucas.ac.cn



杨飞然 男,1982年生,山东泰安人.中国科学院声学研究所研究员.主要研究方向为音频信号处理.

E-mail: feiran@mail.ioa.ac.cn



陈先梅 女,1988年生,四川成都人.中国科学院声学研究所助理研究员.主要研究方向为先进功能材料与器件的设计与工艺研究制备.

E-mail: chenxianmeizu@163.com



杨 军 男,1968年生,安徽安庆人.中国科学院声学研究所研究员.主要研究方向为通信声学、3D音频系统、音频信号处理、声场控制和非线性声学.

E-mail: jyang@mail.ioa.ac.cn