受限于半长度滤波器流形的独立向量提取

Zbyn˘ek Koldovsk´y*、Jaroslav ˘Cmejla*、T¨ulay Adali† 和 Stephen O'Regan‡
* 声学信号分析与处理组，机电、信息学与跨学科学院，利贝雷茨技术大学，捷克共和国。

†马里兰大学巴尔的摩县分校电气与计算机工程系，美国马里兰州巴尔的摩市，邮编 21250
‡美国海军水面作战中心卡德罗克分部，美国马里兰州西贝塞斯达

arXiv:2304.01778v1 [eess.AS] 4 Apr 2023

摘要—**独立向量分析（IVA）是独立成分分析（ICA）的一种常用扩展，可用于联合分离一组瞬时线性混合信号，直接应用于频域说话人分离或提取。混合信号由混合矩阵参数化，每个混合信号对应一个矩阵。这意味着 IVA 混合模型没有考虑不同混合信号／频率之间参数的任何关联。分离过程仅通过源模型进行联合建模，其中考虑了声源在不同混合信号之间的统计依赖性。本文提出一种用于联合盲声源提取的混合模型，其中不同频率的混合模型参数彼此关联。实现方法是将可行参数集约束在半长度分离滤波器的流形上；该流形具有明确的解释，并可应用于频域说话人提取。**

索引词—**盲源分离、盲源提取、独立成分分析、独立向量分析、说话人提取**

I. 引言

独立成分分析与独立向量分析（ICA/IVA）是基于这样一种假设的常用盲源分离（BSS）方法：观测到的信号混合由相互独立的原始信号（声源）组成 [1]。在 ICA 中，只考虑一个独立信号混合；当存在$K > 1$ 个混合信号时，每个混合信号都独立处理。在 IVA 中，通过利用不同混合信号中声源之间的依赖关系，联合处理 $K > 1$ 个混合信号 [2]。独立成分/向量提取（ICE/IVE）是与 ICA/IVA 相关的盲源提取（BSE）方法，其目标是从每个混合信号中仅提取特定的目标声源（SOI），即将其与其他信号分离[3]–[5]。

BSS 和 BSE 任务中固有的不确定性，在实践中对应着一些关键问题，例如在说话人提取问题中确定哪位说话人是目标说话人。单混合信号问题可以在原始信号的顺序和尺度不确定的情况下得到解决。考虑 $K > 1$ 个混合信号时，未知的顺序会导致排列问题，也就是说，不同混合信号中分离出的声源顺序不同。IVA 的一个基本思想是，如果联合分离混合信号，并利用不同混合信号中声源之间的依赖关系，就有助于解决排列问题 [7], [8]。

具体而言，在独立向量分析（IVA）中，我们考虑 K 线性瞬时混合

$$\mathbf{x}^k = \mathbf{A}^k \mathbf{s}^k, \qquad k = 1, \ldots, K, \tag{1}$$

其中 $\mathbf{x}^k$ 是表示混合（观测）信号的随机变量（RVs）向量；$\mathbf{A}^k$ 是非奇异混合矩阵；而$\mathbf{s}^k$ 是表示原始声源的独立随机变量向量；k 是混合索引。假设信号样本是相应随机变量的独立实现（独立同分布，即 i.i.d.）。目标是估计解混矩阵$\mathbf{W}^k$，使得$\mathbf{y}^k = \mathbf{W}^k\mathbf{x}^k$ 与$\mathbf{s}^k$相对应，但尺度和顺序存在不可确定性。寻求$\mathbf{W}^k$ 使分离信号$\mathbf{y}^k$ 相互独立。$\mathbf{s}^1, \ldots, \mathbf{s}^K$ 中的第 i 个声源（由向量随机变量 $\mathbf{s}_i = [s_i^1, \ldots, s_i^K]^T$、 $i = 1, \ldots, d$的元素表示）通过联合分布建模为相互依赖，这有助于解决排列问题 [7]。

在说话人提取中，IVA 或 IVE 按如下方式在频域中部署：时域中通过麦克风观测到的混合信号经短时傅里叶变换 (STFT) [9] 转换。随后，模型 (1) 描述第 k 个频带中的观测信号，其样本对应于各帧。混合矩阵$\mathbf{A}^k$ 表示说话人与麦克风之间的声学传递函数，对应于房间脉冲响应；解混矩阵$\mathbf{W}^k$ 表示分离滤波器的传递函数。必须解决排列问题，以对齐每个说话人的分离频率分量。为此，IVA 利用 $s_i^1, \ldots, s_i^K$ 之间的依赖关系，这些依赖关系可解释为同一说话人各频率分量之间的依赖关系 [2]。

尽管 IVA 和 IVE 取得了成功，但至少存在两个值得质疑的方面。回顾（1）可以看出，依赖关系模型可能并不稳健，因为它完全依赖于高阶统计关系 [2]。众所周知，IVA/IVE 对排列问题的解决并没有得到保证，并且会出现许多不完善之处； [2]。其次，随着 STFT 频率分辨率的提高，模型参数量会迅速增加。显然，从某个

本研究得到海军部海军研究局全球项目的资助，项目编号为 N62909-19-1-2105。

点 K 开始，分离精度将不再提高，而计算需求和问题难度却会增加 [11]。

本文提出了一种改进的 IVE 算法，该算法在式 (1) 中一组受约束的可行（解）混合矩阵上运行。具体而言，我们考虑该集合对应于半长度分离滤波器流形这一特殊情形。我们的方法为推导其他类似的 IVE（及 IVA）改进方案提供了指导，并且在频域说话人提取以及将 IVE 和 IVA 作为通用联合盲源提取/盲源分离方法这两个方面都具有意义。实验表明，所提出的（梯度）算法在计算开销、收敛速度和提取精度方面具有若干优势。

本文结构安排如下。第二节对问题进行表述，第三节推导算法。第四节介绍两项实验评估，第五节总结全文。

符号说明：普通字母、粗体小写字母和粗体大写字母分别表示标量、向量和矩阵。上标 $\cdot^T$, $\cdot^H$ 或 $\cdot^*$ 分别表示转置、共轭转置或复共轭。矩阵/向量拼接采用 Matlab 约定，例如 $[1;\ \mathbf{g}] = [1, \mathbf{g}^T]^T$。$\hat{}\,\mathrm{E}[\cdot]$ 表示期望算子，而 $\mathrm{E}[\cdot]$ 表示对该参数所有可用样本取样本均值。

II. 问题表述

A. 独立向量提取

我们首先说明针对联合提取问题对混合模型 (1) 所做的修改；更多详情请参见 [3]。不失一般性，设每个混合信号中的目标信号都是第一个源信号 s_1^k, $k = 1, \ldots, K$。我们可以将 (1) 重写为

$$\mathbf{x}^k = \mathbf{A}^k \mathbf{u}^k, \tag{2}$$

其中 $\mathbf{A}^k = [\mathbf{a}^k, \mathbf{Q}^k]$ 和 $\mathbf{u}^k = [s_1^k;\ \mathbf{z}^k]$。当只需从 $\mathbf{x}^k$ 中提取 s_1^k 时，无需识别和分离 $\mathbf{Q}$ 和 $\mathbf{z}[n]$，只需识别和分离它们对应的子空间即可。它们实际上对应于 $\mathbf{x}^k$ 中的其他“背景”信号。文献 [3],指出，对于这项任务，混合矩阵 $\mathbf{A}^k$ 和解混矩阵 $\mathbf{W}^k = (\mathbf{A}^k)^{-1}$ 可以通过混合向量 $\mathbf{a}^k$（即 (1) 中 $\mathbf{A}^k$ 的第一列）以及分离向量 分离向量$(\mathbf{w}^k)^H$（即 $\mathbf{W}$的第一行）参数化，如下所示:

$$\mathbf{A}^k = \begin{pmatrix} \gamma^k & (\mathbf{h}^k)^H \\ \mathbf{g}^k & \frac{1}{\gamma^k}(\mathbf{g}^k(\mathbf{h}^k)^H - \mathbf{I}_{d-1}) \end{pmatrix}, \tag{3}$$

$$\mathbf{W}^k = \begin{pmatrix} (\mathbf{w}^k)^H \\ \mathbf{B}^k \end{pmatrix} = \begin{pmatrix} (\beta^k)^* & (\mathbf{h}^k)^H \\ \mathbf{g}^k & -\gamma^k \mathbf{I}_{d-1} \end{pmatrix}, \tag{4}$$

其中 $\mathbf{a}^k = [\gamma^k;\ \mathbf{g}^k]$ 和 $\mathbf{w} = [\beta;\ \mathbf{h}^k]$；参数向量$\mathbf{a}^k$ 和 $\mathbf{w}^k$ 通过条件 $(\mathbf{a}^k)^H \mathbf{w}_k = 1$相关联；$\mathbf{I}_d$ 表示 $d \times d$ 单位矩阵。提取的信号为 $s_1^k = (\mathbf{w}^k)^H \mathbf{x}^k$，而背景信号为 $\mathbf{z}^k = \mathbf{B}^k \mathbf{x}^k$。为简化符号，我们进一步将与 SOI（感兴趣信号）对应的向量分量记作 $\mathbf{s} = [s^1, \ldots, s^K]^T$，其中 $s^k = s_1^k$、$k = 1$、$\ldots$、K。

B. 源模型、对比函数和梯度算法

IVE 中使用的基本源模型假设$\mathbf{s}$ 服从联合非高斯概率密度函数 $p(\mathbf{s})$，而 $\mathbf{z}^k, k = 1, \ldots, K$服从圆对称高斯分布；假设$\mathbf{z}^{k_1}$和$\mathbf{z}^{k_2}$对于 $k_1 \neq k_2$ [3]不相关。最重要的是，$\mathbf{s}$ 与 $\mathbf{z}^k$相互独立。由于 $p(\mathbf{s})$ 未知，必须用合适的模型密度替代。在 [12] (第三节.B) 中，建议用以下函数替代 $p(\mathbf{s})$:

$$p(\mathbf{s}) \approx f(\bar{\mathbf{s}}) \left(\prod_{k=1}^{K} \sigma_k \right)^{-2}, \tag{5}$$

其中，$f(\cdot)$ 是合适的归一化非高斯概率密度函数，σ_k^2 是对 s^k 估计值的基于样本的方差，且$\mathbf{s}$由$\mathbf{s}$的归一化元素组成。值得注意的是，在大多数与独立成分分析/独立向量分析相关的方法中，假设 σ_k^2 等于 1，以此应对 s^k 的尺度歧义。本文稍后将说明，在我们的情形下需要采用不同的尺度歧义处理方法。

用于估计模型参数的对比函数为 [12]

$$\mathcal{C}\left(\{\mathbf{w}^k, \mathbf{a}^k\}_{k=1,\ldots,K}\right) = \hat{\mathrm{E}}\left[\log f(\bar{\mathbf{s}})\right] - \sum_{k=1}^{K} \log \sigma_k^2 - \sum_{k=1}^{K} \hat{\mathrm{E}}\left[(\mathbf{z}^k)^H (\mathbf{C}_{\mathbf{z}}^k)^{-1} \mathbf{z}^k\right] + (d-2) \sum_{k=1}^{K} \log |\gamma^k|^2 + C, \tag{6}$$

其中$\mathbf{C}_{\mathbf{z}}^k$ 表示$\mathbf{z}^k$的协方差矩阵；C 是一个与参数向量无关的常数。该函数由似然函数推导而来，其中未知的目标信号概率密度函数由式 (5) 替代，$\mathbf{C}_{\mathbf{z}}^k$ 则由当前的样本估计值 [12]替代。

通过在正交约束下最大化式 (6)，求取$\mathbf{a}^k$ 和$\mathbf{w}^k$ 的估计值；该约束确保 s^k 与$\mathbf{z}^k$ 的当前估计值之间的样本相关性为零。这使$\mathbf{a}^k$ 通过$\mathbf{a}^k = \mathbf{C}^k \mathbf{w}^k / \sigma_k^2$完全依赖于$\mathbf{w}^k$，其中$\mathbf{C}^k$ 是$\mathbf{x}^k$的协方差矩阵。由于用式 (5) 替代了真实概率密度函数，经过归一化步骤后，式 (6) 在正交约束下关于$\mathbf{w}^k$的梯度为

$$\frac{\partial \mathcal{C}}{\partial (\mathbf{w}^k)^*} = \nabla^k = \mathbf{a}^k - \nu_k^{-1} \hat{\mathrm{E}}\left[\phi_k(\bar{\mathbf{s}})\, \bar{\mathbf{x}}^k\right], \tag{7}$$

其中$\mathbf{x}^k = \mathbf{x}^k / \sigma_k$，$\nu_k = \hat{}\,\mathrm{E}[\phi_k(\mathbf{s})\, s^k]$ 是模型概率密度函数的归一化因子，而 $\phi_k(\cdot) = -\frac{\partial}{\partial s_k} \log f(\cdot)$是模型概率密度函数 $f(\cdot)$的第 k个得分函数；详细计算过程见 [12]。

基于梯度的算法通过沿梯度方向更新$\mathbf{w}^k$，来求取式（6）的最大值，即

$$\mathbf{w}^k \leftarrow \mathbf{w}^k + \mu \nabla^k, \quad k = 1, \ldots, K, \tag{8}$$

其中 μ 是步长参数。在 [3], 中，这种算法称为 OGIVE$_{\mathbf{w}}$。由式（7）和式（8）可知，$\mathbf{w}^1$、.....、$\mathbf{w}^K$的更新几乎相互独立。这些“并行”算法通过得分函数 ϕ_k、k =1、.....、K相互影响，而这些得分函数都取决于向量分量$\mathbf{s}$。这种解耦正是大多数盲语音分离与提取方法在短时傅里叶变换域[9], [13]中运行的主要优势。

C. 跨频率关联参数

引入一个矩阵

$$\mathbf{W} = [\mathbf{w}^1, \dots, \mathbf{w}^K], \tag{9}$$

其各列对应于所有频率的分离向量。到目前为止，$\mathbf{W}$ 的各个元素是自由变量；在正交约束下，对比函数 (6) 以这些变量为优化对象，因为其他参数均由该约束决定。一般而言，我们的目标是定义一组新的参数，由一个 $d \times L$ 矩阵的各个元素给出

$$\mathbf{V} = [\mathbf{v}^1, \dots, \mathbf{v}^L], \tag{10}$$

其中，$\mathbf{v}^\ell$ 表示该矩阵的第 ℓ 列，而 $v_{\ell,i}$ 表示 $\mathbf{v}^\ell$的第 i 个元素，即$\mathbf{V}$的第 $i\ell$ 个元素。此外，我们假设存在一个光滑映射

$$\mathbf{W} = \mathcal{F}(\mathbf{V}) \tag{11}$$

且该映射已知。我们希望 $L < K$，因为这样的映射可以减少模型中的自由参数数量。

给定这样的映射后，我们可以应用链式法则，计算对 $\mathbf{V}$ 求导得到的式 (6) 的导数，并推导出在新的流形上执行优化的算法。根据 (7)，并使用复值链式法则 [14],

$$\begin{aligned}\frac{\partial \mathcal{C}}{\partial v^*_{\ell,i}} &= \mathtt{tr}\left[\left(\frac{\partial \mathcal{C}}{\partial \mathbf{W}}\right)^T \frac{\partial \mathbf{W}}{\partial v^*_{\ell,i}}\right] + \mathtt{tr}\left[\left(\frac{\partial \mathcal{C}}{\partial \mathbf{W}^*}\right)^T \frac{\partial \mathbf{W}^*}{\partial v^*_{\ell,i}}\right] \\ &= \mathtt{tr}\left[\mathbf{\Omega}^H \frac{\partial \mathcal{F}(\mathbf{V})}{\partial v^*_{\ell,i}}\right] + \mathtt{tr}\left[\mathbf{\Omega}^T \frac{\partial \mathcal{F}(\mathbf{V})^*}{\partial v^*_{\ell,i}}\right], \end{aligned} \tag{12}$$

其中，tr$[\cdot]$ 表示矩阵参数的迹，且

$$\mathbf{\Omega} = [\nabla^1, \dots, \nabla^K] \tag{13}$$

是由式（7）给出的梯度组成的矩阵。将式（8）替换为两个步骤，即可得到一种新的梯度算法

$$\mathbf{v}^\ell \leftarrow \mathbf{v}^\ell + \mu \frac{\partial \mathcal{C}}{\partial \mathbf{v}^\ell}, \quad \ell = 1, \dots, L, \tag{14}$$

$$\mathbf{W} \leftarrow \mathcal{F}(\mathbf{V}). \tag{15}$$

计算复杂度的增加主要取决于式 (12) 中映射 $\mathcal{F}(\cdot)$ 的导数形式。

III. 所提出的算法

A. 半长度滤波器流形

现在，我们介绍第二节-C中提出的一般方法的一种变体，该变体专为短时傅里叶变换域中的说话人提取方法而设计。在这里，$\mathbf{W}$ 的各行对应于时域分离滤波器在分辨率为 K的离散傅里叶变换 (DFT) 域中的表示。其思路是将$\mathbf{W}$ 约束在长度仅为$K/2$的滤波器流形上。为简单起见，我们假设 $K/4$ 是一个整数。

为求得映射（11），考虑一个 FIR 滤波器 $w(n)$,$n = 0, \dots, K-1$，其长度为 $K/2$，即 $w(n) = 0$ for $n = K/2, \dots, K-1$。其传递函数在分辨率为 K 的 DFT 域中的第 k 个系数为

$$\tilde{w}_K(k) = \sum_{n=0}^{K-1} w(n) \mathsf{W}_K^{kn} = \sum_{n=0}^{\frac{K}{2}-1} w(n) \mathsf{W}_K^{kn}, \tag{16}$$

其中 $\mathsf{W}_K = e^{-i2\pi}{}_K$。现在我们分别针对奇数和偶数 k 表示 $\tilde{w}_K(k)$。

$$\begin{aligned}\tilde{w}_K(2\ell) &= \sum_{n=0}^{\frac{K}{2}-1} w(n) \mathsf{W}_K^{2\ell n} = \sum_{n=0}^{\frac{K}{2}-1} w(n) \mathsf{W}_{\frac{K}{2}}^{\ell n} \\ &= \tilde{w}_{K/2}(\ell), \end{aligned} \tag{17}$$

$$\begin{aligned}\tilde{w}_K(2\ell+1) &= \sum_{n=0}^{\frac{K}{2}-1} w(n) \mathsf{W}_K^{(2\ell+1)n} = \sum_{n=0}^{\frac{K}{2}-1} w(n) \mathsf{W}_K^n \mathsf{W}_{\frac{K}{2}}^{\ell n} \\ &= \{\tilde{w}_{K/2} \circledast v\}(\ell), \end{aligned} \tag{18}$$

其中 $\ell = 0, \dots, K/2-1$, $v(k)$ 表示序列 W_K^n, $n = 0, \dots, K/2-1$ 的离散傅里叶变换（长度为$K/2$），而 $\circledast$ 表示循环卷积。式 (18) 中的后一关系源于离散傅里叶变换的众所周知的性质 [15]。

方程 (17)-(18) 描述了对于每个最大长度为$K/2$的 FIR 滤波器 $w(n)$，$\tilde{w}_K$ 与 $\tilde{w}_{K/2}$ 之间的映射。由于我们希望将 $\mathbf{W}$ 约束为此类滤波器，因此当$\mathbf{V}$ 满足$\mathbf{v}^\ell = \mathbf{w}^{2\ell-1}$（其中 $\ell = 1, \dots, L$，且 $L = K/2$）时，可根据式 (11) 对 $\mathbf{W}$ 进行参数化；这是由式 (17) 得出的。该映射是线性的，可由一个 $L \times 2L$ 矩阵$\mathbf{F}$ 表示，使得[1]

$$\mathbf{W} = \mathbf{V}\mathbf{F}^*. \tag{19}$$

根据（17），$\mathbf{F}$ 的奇数列构成 $L \times L$ 单位矩阵。根据（18），$\mathbf{F}$的偶数列构成循环矩阵，其首行为 $v(k)$,$k = 0, \dots, K/2$ 中的值。

B. 共轭对称性处理

在语音提取中，我们实际上是用实值滤波器处理实值信号。因此，在短时傅里叶变换（离散傅里叶变换）域中存在共轭对称性，只需处理频率 bin $k = 1, \dots, K/2+1$，其中第一个

[1]注意式（19）中$\mathbf{F}$ 的共轭值。这是因为$\mathbf{W}$的列，即分离向量$\mathbf{w}^k$，作用于共轭后的观测信号，即作为 $s^k = (\mathbf{w}^k)^H \mathbf{x}^k$。

末尾的频点为实数，而其余频点为复数。必须考虑这一点，才能使新的参数化尽可能有效。

因此，原参数矩阵 $\mathbf{W}$ 可以替换为截断矩阵 $\mathbf{W}_t = [\mathbf{w}^1, \ldots, \mathbf{w}^{K/2+1}]$；类似地，$\mathbf{V}$ 可以替换为 $\mathbf{V}_t = [\mathbf{v}^1, \ldots, \mathbf{v}^{L/2+1}]$。提取算法只处理频率 bin $k = 1, \ldots, K/2+1$的数据。

现在，式 (19) 的对应形式为

$$\mathbf{W}_t = \mathbf{V}_t \mathbf{F}_1^* + \mathbf{V}_t^* \mathbf{J} \mathbf{F}_2^*, \tag{20}$$

其中，$\mathbf{F}_1$ 和 $\mathbf{F}_2$ 分别是$\mathbf{F}$ 中包含其第 $1, \ldots, L/2+1$ 行和第 $L/2+2, \ldots, L$行的部分；$\mathbf{J}$ 表示去掉首列和末列后的反对角单位矩阵。将式 (20) 代入式 (12) 并经过一些变换，可以证明

$$[\nabla_{\mathbf{v}}^1, \ldots, \nabla_{\mathbf{v}}^{L/2+1}] = \mathbf{\Omega}_1 \mathbf{F}_1^T + \mathbf{\Omega}_1^* \mathbf{F}_2^H, \tag{21}$$

其中

$$\nabla_{\mathbf{v}}^{\ell} = \frac{\partial \mathcal{C}}{\partial (\mathbf{v}^{\ell})^*}, \quad \ell = 1, \ldots, L/2+1, \tag{22}$$

以及 $\mathbf{\Omega}_1 = [\nabla^1, \ldots, \nabla^{K/2+1}]$。

C. 处理尺度歧义

由于存在尺度歧义，提取信号 $s^k = (\mathbf{w}^k)^H \mathbf{x}^k$ 的尺度可以任意变化。这种自由度可能导致收敛问题。因此，通常会对 $\mathbf{w}^k$ 进行缩放，使得 σ_k^2（即 s^k的方差）为 1。然而，我们的方法无法这样做，因为分别对每个$\mathbf{w}^k$, $k = 1, \ldots, K$重新缩放，通常会使它们不再对应半长度滤波器。只有新的参数向量$\mathbf{v}^{\ell}$, $\ell = 1, \ldots, L/2+1$可以重新缩放；我们选择在每次迭代后对这些向量进行归一化。需要注意的是，式（5）给出的尺度不变概率密度函数模型在这方面起着关键作用。对向量$\mathbf{v}^{\ell}$ 进行重新缩放，使得 $\sigma_k^2 = 1$ 对于奇数 k 成立，并不能保证 $\sigma_k^2 = 1$对于偶数 k也成立。采用尺度不变概率密度函数模型后，这个问题便不再出现。

IV. 实验

A. 语音提取

使用由 E. Habets[2]，实现的房间脉冲响应生成器 [16]，我们模拟了这样一种情形：两位说话人由两支相距 20cm 的麦克风录制。房间尺寸为 3×4m，层高为 3m。麦克风位于 $[1.5\ 2\ 1]$m 处，两位说话人分别位于 $[1\ 3\ 1]$m （男性）和 $[2\ 3\ 1]$m（女性）处。混响时间 T_{60} 设为 100ms 和 200ms；采样频率为 16 kHz。生成一段时长为 2.5s 的说话人混合信号，并将其转换到短时傅里叶变换域，离散傅里叶变换长度为 512 个采样点，帧移为 128 个采样点。

对女性说话人的盲提取进行了实验，并以信干比（SIR）进行评估。将所提出的算法与梯度 OGIVE$_{\mathbf{W}}$ 算法进行了比较，后者来自 [3]。OGIVE$_{\mathbf{W}}$ 分为两种变体：使用数据白化和不使用数据白化。使用白化（记作“wh.”）时，该算法涉及矩阵求逆；而所提出的算法和不使用白化的 OGIVE$_{\mathbf{W}}$ 均无需进行矩阵求逆操作。所有方法的步长参数 μ 均设为相同的值 0.05。

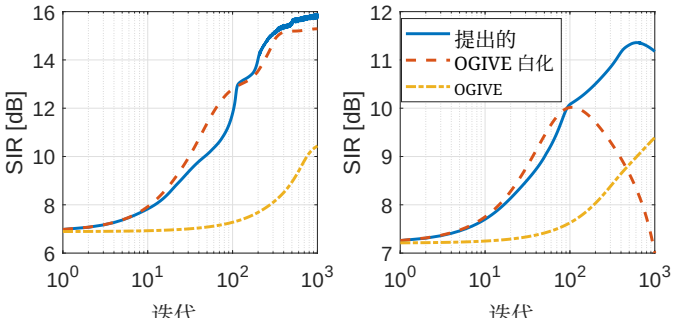

图1. 所提出的算法与 OGIVE 的信干比收敛情况$_{\mathbf{W}}$，当 [3] T_{60} 为 100ms（左）和 200ms（右）时。

图1显示了各算法达到的输出信干比随迭代次数的变化。所提出的算法具有与采用白化的 OGIVE$_{\mathbf{W}}$相当的收敛速度；未采用白化的 OGIVE$_{\mathbf{W}}$ 收敛明显较慢。所提出的算法最终达到的信干比与 OGIVE$_{\mathbf{W}}$ 相当，适用于 $T_{60} = 100$ms。对于 $T_{60} = 200$ms，采用 wh. 的 OGIVE$_{\mathbf{W}}$ 性能在 100 次迭代后下降，这可能是由于排列问题导致收敛到局部极值。图2中 $T_{60} = 300$ms 时也观察到了类似现象。

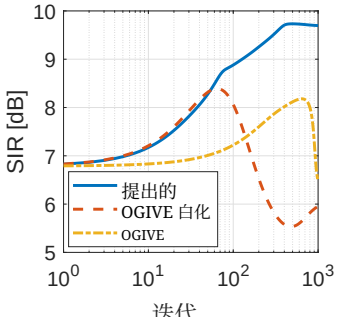

图2. 当 T_{60} 为 300ms 时，信干比随迭代次数的收敛情况。

B. 密集麦克风阵列

这里，我们介绍一项真实世界实验的结果。实验在一间开放式办公室中的阁楼房间内进行，房间尺寸为 $13 \times 8 \times 2.5$ m 和 $T_{60} \approx 500$ ms。播放女性和男性语音的扬声器分别位于距麦克风阵列≈ 0.7 m 处，角度位置分别为 $40°$和 $-40°$；实验装置的照片见图3。说话人由 JBL GO 2 扬声器模拟。

该麦克风阵列是一个密集平面阵列，包含 64 个 MEMS麦克风，按 8×8 规则网格排列，水平和垂直间距均为 20 mm。一个嵌入式

[2]https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-生成器

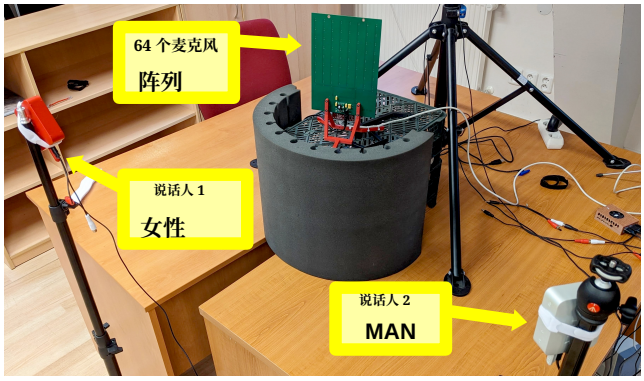

图 3. 实验装置照片。

FPGA 控制器通过网络发送麦克风采集到的所有同步声音数据，随后将其保存并在 Matlab 中处理。录音时长为 3.75 s，采样频率为 48 kHz，之后降采样至 16 kHz。在实验中，我们使用麦克风阵列的第一列和最后一列，即第 1 至 8 个和第 57 至 64 个麦克风。

表 I 显示了提取女性说话人时，各比较算法在迭代 200 次后达到的信干比；初始信干比约为 0.5 dB。实验针对不同的离散傅里叶变换长度重复进行；移位长度始终为离散傅里叶变换长度的四分之一。所提出的算法和采用白化的 OGIVE 在离散傅里叶变换长度为 128 时达到最佳性能；未采用白化的 OGIVE 在长度为 256 时表现最佳，不过这可能是迭代次数有限所致。结果表明，当麦克风数量较多且排列密集时，选择中等长度的分离滤波器是值得的。

表 I 密集麦克风阵列实验中达到的信干比 [DB]

离散傅里叶变换长度	提出的	OGIVE wh.	OGIVE
32	6.17	0.40	-1.17
64	9.11	2.88	-0.70
128	9.51	7.22	4.23
256	6.28	6.23	5.55
512	5.68	3.88	3.72
1024	3.55	3.51	1.72

V. 结论

我们提出了一种改进的独立向量提取算法，该算法在半长度分离滤波器的流形上运行。该算法不需要执行任何矩阵求逆运算，但仍能实现与采用数据预白化的梯度算法相当快的收敛速度。然而，该方法仍有很大的进一步发展和实验验证空间。将这一思想应用于四分之一长度、八分之一长度或十六分之一长度的滤波器，尤其是在零填充的短时傅里叶变换域中实现，将会很有意义。公式（21）有一种有趣的解释，即任何参数向量$\mathbf{v}^{\ell}$ 的变化都会直接影响所有参数向量的变化 $\mathbf{w}^{k}$。这可能会对排列问题的求解产生有趣的影响，因为这意味着，算法在选定频率处向某一给定说话人的收敛，会显著影响其他频率处的梯度；与传统的频域独立向量提取相比，这种影响更为明显。

参考文献

[1] P. Comon 和 C. Jutten，《盲源分离手册：独立成分分析与应用》。独立成分分析及其应用丛书，Elsevier Science，2010年。[2] T. Kim、H. T. Attias、S.-Y. Lee 和 T.-W. Lee，“利用高阶频率依赖进行盲源分离”，*IEEE* 音频、语音与语言处理汇刊，第 70–79 页，2007年1月。[3] Z. Koldovsk´y 和 P. Tichavsk´y，“复数非高斯独立成分/向量提取的梯度算法：收敛性问题”，*IEEE* 信号处理汇刊，第 67 卷，第 1050–1064 页，2019年2月。[4] R. Scheibler 和 N. Ono，“声源数量少于麦克风数量时的独立向量分析”，载于*2019*年 *IEEE* 音频与声学信号处理应用研讨会（*WASPAA*），第 185–189 页，2019年。[5] R. Ikeshita 和 T. Nakatani，“用于联合盲源分离与去混响的独立向量提取”，2021年，arXiv: 2102.04696。[6] H. Sawada、S. Araki、R. Mukai 和 S. Makino，“利用独立成分分析和时频掩蔽盲提取主要目标声源”，*IEEE* 音频、语音与语言处理汇刊，2006年11月。（已接收）。[7] T. Kim、I. Lee 和 T. Lee，“独立向量分析：定义与算法”，载于*2006*年第四十届阿西洛马会议：信号、系统与计算机，第 1393–1396 页，2006年10月。[8] I. Lee、T. Kim 和 T.-W. Lee，“用于卷积混合盲源分离的快速定点独立向量分析算法”，《信号处理》，第 87 卷，第 8 期，第 1859–1871 页，2007年。[9] E. Vincent、T. Virtanen 和 S. Gannot，《音频源分离与语音增强》。Wiley 出版社，第 1 版，2018年。[10] Z. Koldovsk´y、V. Kautsk´y 和 P. Tichavsk´y，“双重非平稳性：从非平稳混合信号中盲提取独立非平稳向量/成分——算法”，*IEEE* 信号处理汇刊，第 70 卷，第 5102–5116 页，2022年。[11] S. Araki、R. Mukai、S. Makino、T. Nishikawa 和 H. Saruwatari，“语音卷积混合信号频域盲源分离的基本局限”，*IEEE* 音频、语音与语言处理汇刊，第 11 卷，第 2 期，第 109–116 页，2003年。[12] Z. Koldovsk´y、V. Kautsk´y、P. Tichavsk´y、J. ˇCmejla 和 J. M´alek，“动态独立成分/向量分析：可由时不变波束形成器分离的时变线性混合信号”，*IEEE* 信号处理汇刊，第 69 卷，第 2158–2173 页，2021年。[13] P. Smaragdis，“频域中卷积混合信号的盲分离”，《神经计算》，第 22 卷，第 21–34 页，1998年。[14] K. B. Petersen 和 M. S. Pedersen，《矩阵手册》，2008年10月。版本 20081110。[15] B. Porat，《数字信号处理教程》。Wiley，1997年。[16] J. B. Allen 和 D. A. Berkley，“高效模拟小房间声学的镜像法”，《美国声学学会杂志》，第 65 卷，第 4 期，第 943–950 页，1979年。