独立向量分析[1]

FanmingL

一、问题提出

盲源分离(Blind Source Separation)问题是指在信号的理论模型和源信号无法精确获知的情况下，如何从观测信号中分离出各源信号的过程。“鸡尾酒会问题”(cocktailparty problem)是一个非常经典的 BSS 问题，对于瞬时混合模型，独立成分分析(Independent Component Analysis, ICA)非常有效，但是实际生活中，声音的混合往往不是瞬时的，而是有延迟的，当前观测到的信号总是直达声和混响声的混合，这种模型下，声音是有延迟的，换句话说，是卷积混合的。由于卷积运算在频域上是一个相对来说更简单的乘法运算，所以对于卷机混合信号，我们理论上可以对每个频率点(frequency bin)应用 ICA 去分离来自不同声源的分量，但是 ICA 存在一个固有的排序问题——我们并不能判断分离的结果究竟来自哪个声源。这样一来，就算我们在每一个频率点都成功将信号分离开来，我们也不能正确的复原声源信号的频谱，这样也就不能在时域复原声源信号。近几年来，一些科研工作者提出了一些方法能够在特定条件下解决排序问题，但是在一般情况下，排序问题还是不能被解决。2006 年，Taesu Kim 提出了独立向量分析(Independent Vector Analysis, IVA)这一方法，这一方法很好的避免的排序问题，它是一种 ICA 的一种拓展算法，同样是在频域对信号进行处理，不同的是 IVA 将每个声源信号的变换成的所有频率点看成一个向量，然后利用这些向量之间的独立性，去求解这个 BSS 问题，也就是从一开始各个频率点就没被分开，而是看作一个整体，所以也不存在排序问题。

二、独立成分分析

我们可以用两种方式去推导独立成分分析的迭代公式。第一种方法参考自[2]，下面主要由这种方法推导迭代公式。

假设我们有 n 个信号源、n 个麦克风，在瞬时混合模型下，存在一个矩阵 A 使得某个时刻

$$\begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = \mathbf{A} \begin{bmatrix} s_1 \\ s_2 \\ \cdots \\ s_n \end{bmatrix} \tag{1.1}$$

其中$x_i(i = 1,2,3\ldots n)$是第 i 个麦克风捕获到的观测信号，$s_j(j = 1,2,3\ldots n)$是第 j 个信号源产生的信号。所以只要解得了矩阵 A 的逆A^{-1}，我们就能由观测信号求得每个声源的信号。我们不妨假设

$$\mathbf{W} = \mathbf{A}^{-1} = \begin{bmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \cdots \\ \mathbf{w}_n^T \end{bmatrix} \tag{1.2}$$

那么

$$\begin{bmatrix} s_1 \\ s_2 \\ \cdots \\ s_n \end{bmatrix} = \mathbf{W} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \cdots \\ \mathbf{w}_n^T \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{w}_1^T\mathbf{x} \\ \mathbf{w}_2^T\mathbf{x} \\ \cdots \\ \mathbf{w}_n^T\mathbf{x} \end{bmatrix} \tag{1.3}$$

即

$$s_j = \mathbf{w}_j^T\mathbf{x} \tag{1.4}$$

假设，$\mathbf{x}$满足分布$p_{\mathbf{x}}(\mathbf{x})$，$s$满足分布$p_s(s)$，那么可以证明如下关系

$$p_{\mathbf{x}}(\mathbf{x}) = p_{\mathbf{s}}(\mathbf{W}\mathbf{x}) \cdot |\mathbf{W}| \tag{1.5}$$

所以，若$\mathbf{x}$每个分量都是独立的，那么显然满足

$$p(\mathbf{x}) = \prod_{i=1}^{n} p_{\mathbf{s}}(\mathbf{w}_i^T \mathbf{x}) \cdot |\mathbf{W}| \tag{1.6}$$

若我们进行了 m 次采样，那么我们可以求出$p(\mathbf{x})$的极大似然函数$L(\mathbf{W})$

$$L(\mathbf{W}) = \prod_{j=1}^{m} (\prod_{i=1}^{n} (p_{\mathbf{s}}(\mathbf{w}_i^T \mathbf{x}^{(j)}) \cdot |\mathbf{W}|)) \tag{1.7}$$

其中$\mathbf{x}^{(j)}(j = 1,2,\ldots,m)$表示第 j 个样本。为了方便计算，我们对$L(\mathbf{W})$求对数

$$l(\mathbf{W}) = log(L(\mathbf{W})) = \sum_{j=1}^{m} (\sum_{i=1}^{n} log p_{\mathbf{s}}(\mathbf{s})(\mathbf{w}_i^T \mathbf{x}^{(j)})) \tag{1.8}$$

这里，我们假设s的概率分布函数为 sigmoid 函数$g(\mathbf{s}) = \frac{1}{1+e^{-\mathbf{s}}}$，那么$p_{\mathbf{s}}(\mathbf{s}) = g^{'}(\mathbf{s})$。所以，

$$l(\mathbf{W}) = log(L(\mathbf{W})) = \sum_{j=1}^{m} (\sum_{i=1}^{n} log g^{'}(\mathbf{s})(\mathbf{w}_i^T \mathbf{x}^{(j)})) \tag{1.9}$$

又根据如下公式

$$\nabla_{\mathbf{W}} |\mathbf{W}| = |\mathbf{W}| (\mathbf{W}^{-1})^T \tag{1.10}$$

我们可以求得$l(\mathbf{W})$对$\mathbf{W}$的梯度

$$\nabla_{\mathbf{W}} l(\mathbf{W}) = \begin{bmatrix} 1 - 2g(\mathbf{w}_1^T \mathbf{x}^j) \\ 1 - 2g(\mathbf{w}_2^T \mathbf{x}^j) \\ \cdots \\ 1 - 2g(\mathbf{w}_n^T \mathbf{x}^j) \end{bmatrix} \mathbf{x}^{(j)^T} + (\mathbf{W}^T)^{-1} \tag{1.11}$$

应用梯度上升法(gradient ascent)使$l(\mathbf{W})$的值最大

$$\mathbf{W} := \mathbf{W} + \alpha \left(\begin{bmatrix} 1 - 2g(\mathbf{w}_1^T \mathbf{x}^j) \\ 1 - 2g(\mathbf{w}_2^T \mathbf{x}^j) \\ \cdots \\ 1 - 2g(\mathbf{w}_n^T \mathbf{x}^j) \end{bmatrix} \mathbf{x}^{(j)^T} + (\mathbf{W}^T)^{-1} \right) \tag{1.12}$$

其中α为一常量。

由此，我们便得到了$\mathbf{W}$的迭代公式。但是如之前所述，这一方法存在排序问题，不能够用于卷积混合的信号。

第二种方法，参考自[3]，主要是利用高阶统计量来衡量变量的非高斯性，这是基于“混合信号的非高斯性必然小于组成该混合信号各个成分的非高斯性”这一假设。而评价一组信号的非高斯性有许多评价函数，如峭度、负熵、互信息等，对这些评价函数应用梯度下降\上升法，即可得到迭代公式。

三、独立向量分析

1、独立向量模型的导出，每个信号源 j 在观测点 i 处产生的信号是卷积的，假设其单位冲激响应为$h_{ij}(\tau)(\tau = 0,1,2\ldots,T-1)$。（这里假设了其单位冲激响应的长度为 T，是有限长的，因为混响声是以指数形式衰减的，所以在τ比较大时，$h_{ij}(\tau)$已经非常小了）我们将 i 处观测到的来自 j 声源的信号表示为x_{ij}

$$x_{ij}(t) = \sum_{\tau=0}^{T-1} h_{ij}(\tau) s_j(t-\tau) \tag{1.13}$$

上式中$s_j(t-\tau)$表示第 j 个声源在$t-\tau$时间产生的信号大小。所以在 i 处的总的信号为所有声源的叠加，即

$$x_i(t)=\sum_{j=1}^{L}\sum_{\tau=0}^{T-1}h_{ij}(\tau)s_j(t-\tau) \tag{1.14}$$

其中 L 表示，我们假设一共有 L 个声源。对$x_i(t)$使用短时傅里叶变换(short-time Fourier transform)

$$x_i^{(k)}[n]=\sum_{t=0}^{K-1}\omega(t)x_i(nJ+t)e^{-j\frac{2\pi(k-1)}{K}t} \tag{1.15}$$

其中 J 表示，窗函数进行 STFT 的滑动距离(shift size)，$\omega(t)$为窗函数，K 表示窗函数长度。若假设窗函数长度 K 远远大于单位抽样响应$h_{ij}(\tau)$的长度 T，我们可以做如下假设

$$x_i^{(k)}[n]\approx\sum_{j=1}^{L}h_{ij}^{(k)}s_j^{(k)}[n] \tag{1.16}$$

所以存在下列关系

$$y_i^{(k)}[n]=\sum_{j=1}^{M}g_{ij}^{(k)}x_j^{(k)}[n]\approx s_i^{(k)}[n] \tag{1.17}$$

其中$g_{ij}^{(k)}$为第 k 个频率点的解混滤波器，M 为观测点的数量。于是我们就得到了对原始信号$s_i^{(k)}[n]$的估计$y_i^{(k)}[n]$的线性表达式。为了方便表示，我们写成矩阵的形式。

$$\begin{bmatrix}y_1^{(k)}[n]\\y_2^{(k)}[n]\\\cdots\\y_L^{(k)}[n]\end{bmatrix}=\begin{bmatrix}g_{11}^{(k)}[n] & g_{12}^{(k)}[n] & \cdots & g_{1M}^{(k)}[n]\\g_{21}^{(k)}[n] & g_{22}^{(k)}[n] & \cdots & g_{2M}^{(k)}[n]\\\cdots & \cdots & \cdots & \cdots\\g_{L1}^{(k)}[n] & g_{L2}^{(k)}[n] & \cdots & g_{LM}^{(k)}[n]\end{bmatrix}\begin{bmatrix}x_1^{(k)}[n]\\x_2^{(k)}[n]\\\cdots\\x_n^{(k)}[n]\end{bmatrix} \tag{1.18}$$

$$\mathbf{y}^{(k)}[n]=\mathbf{G}^{(k)}[n]\mathbf{x}^{(k)}[n] \tag{1.19}$$

按照 IVA 的思想，把一个信号的所有频率点看成一个向量即

$$\mathbf{y}_i[n]=\begin{bmatrix}y_i^{(1)}[n]\\y_i^{(1)}[n]\\\cdots\\y_i^{(K)}[n]\end{bmatrix} \tag{1.20}$$

然后假设每个信号$\mathbf{y}_i$看作一个随机向量，其概率密度函数(probability density function
, PDF)为$q(\mathbf{y}_i)$，若满足$\mathbf{y}_i$之间相互独立，那么$\mathbf{y}_i$之间的联合概率分布函数$p(\mathbf{y}_1,\mathbf{y}_2,\mathbf{y}_3,\ldots,\mathbf{y}_L)$与所有$q(\mathbf{y}_i)$的乘积$\prod_{i=1}^{L}q($的 LK 散度将会达到极小。即

$$\min_{\mathbf{G}}\ \iota=KL\left(p(\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_L)||\prod_{i=1}^{L}q(\mathbf{y}_i)\right)$$
$$\text{s.t.}\ \ \mathbf{y}^{(k)}=\mathbf{G}^{(k)}\mathbf{x}^{(k)}(k=1,2,3,\ldots,K) \tag{1.21}$$

可以将上式化简为

$$\min_{\mathbf{G}}\ \iota=const.-\sum_{k=1}^{K}log|det\mathbf{G}^{(k)}|-\sum_{i=1}^{L}E[logq(\mathbf{y}_i)]$$

$$\text{s.t. } \mathbf{y}^{(k)} = \mathbf{G}^{(k)}\mathbf{x}^{(k)} (k = 1,2,3,\ldots,K) \tag{1.22}$$

其中 E[*]为期望运算符，我们应用梯度下降法来解决这一优化问题，即

$$\mathbf{G}^{(k)} := \mathbf{G}^{(k)} - \eta \nabla_{\mathbf{G}^{(k)}} \iota \tag{1.23}$$

η为一常数，控制梯度下降的速度，$\nabla_{\mathbf{G}^{(k)}}\iota$表示$\iota$对$\mathbf{G}^{(k)}$求梯度，结合$\nabla_{\mathbf{G}}|\mathbf{G}| = |\mathbf{G}|(\mathbf{G}^{-1})^T$，我们可以得到

$$\nabla_{\mathbf{G}^{(k)}}\left(-\sum_{k=1}^{K} log|det\mathbf{G}^{(k)}|\right) = -\frac{\left(\mathbf{G}^{(k)^{-1}}\right)^T}{det\mathbf{G}^{(k)}} det\mathbf{G}^{(k)} = -\left(\mathbf{G}^{(k)^{-1}}\right)^T \tag{1.24}$$

$$\nabla_{\mathbf{G}^{(k)}}\left(-\sum_{i=1}^{L} E[logq(\mathbf{y}_i)]\right) = -E\left[\begin{bmatrix} \frac{\partial logq(\mathbf{y}_1)}{\partial y_1^{(k)}} \\ \frac{\partial logq(\mathbf{y}_2)}{\partial y_2^{(k)}} \\ \cdots \\ \frac{\partial logq(\mathbf{y}_L)}{\partial y_L^{(k)}} \end{bmatrix} \mathbf{x}^{(k)^T}\right] \tag{1.25}$$

令

$$\varphi^{(\mathbf{k})}(\mathbf{y}) = \begin{bmatrix} \frac{\partial logq(\mathbf{y}_1)}{\partial y_1^{(k)}} \\ \frac{\partial logq(\mathbf{y}_2)}{\partial y_2^{(k)}} \\ \cdots \\ \frac{\partial logq(\mathbf{y}_L)}{\partial y_L^{(k)}} \end{bmatrix} \tag{1.26}$$

所以在梯度下降法下，迭代公式为

$$\mathbf{G}^{(k)} := \mathbf{G}^{(k)} + \eta\left(\left(\mathbf{G}^{(k)^{-1}}\right)^T + E\left[\varphi^{(\mathbf{k})}(\mathbf{y})\mathbf{x}^{(\mathbf{k})^{\mathbf{T}}}\right]\right) \tag{1.27}$$

为了加快收敛速度，将梯度下降法改进为自然梯度(Natural Gradient)法，即对梯度因子右乘以一个缩放比例矩阵$\mathbf{G}^{(k)^T}\mathbf{G}^{(k)}$根据$\left(\mathbf{G}^{(k)^{-1}}\right)^T\mathbf{G}^{(k)^T}\mathbf{G}^{(k)} = \mathbf{G}^{(k)}$以及$\mathbf{x}^{(k)}\mathbf{G}^{(k)^T}\mathbf{G}^{(k)} = \left(\mathbf{G}^{(k)}\mathbf{x}^{(k)}\right)^T\mathbf{G}^{(k)} = \mathbf{y}^{(k)^T}\mathbf{G}^{(k)}$，可将迭代公式改进为，

$$\mathbf{G}^{(k)} := \mathbf{G}^{(k)} + \eta\left(I + E\left[\left(\varphi^{(\mathbf{k})}(\mathbf{y})\right)\mathbf{y}^{(k)^T}\right]\right)\mathbf{G}^{(k)} \tag{1.28}$$

那么如何选择$q(\mathbf{y}_i)$呢，$q(\mathbf{y}_i)$是一个多元先验概率密度函数，根据以往的经验，语音信号具有超高斯分布的特征，所以我们可以选择一个多元超高斯分布作为$q(\mathbf{y}_i)$，而拉普拉斯分布是一个典型的超高

斯分布，对于一个多元的超高斯分布，我们有两种选择，一是假设各分量相互独立，二是假设各分量相互依赖，
各分量独立的拉普拉斯分布

$$q(\mathbf{y}_i) = \alpha \prod_{k=1}^{K} exp(-\frac{|y_i^{(k)} - \mu_i^{(k)}|}{\sigma_i^{(k)}}) \tag{1.29}$$

α为归一化因子，$\mu_i^{(k)}$为$\mathbf{y}_i^{(k)}$的均值，$\sigma_i^{(k)}$为$\mathbf{y}_i^{(k)}$的方差，作零均值及单位方差假设，可得

$$\varphi^{(\mathbf{k})}(\mathbf{y}) = -\begin{bmatrix} \frac{y_1^{(k)}}{|y_1^{(k)}|} \\ \frac{y_2^{(k)}}{|y_2^{(k)}|} \\ \cdots \\ \frac{y_L^{(k)}}{|y_L^{(k)}|} \end{bmatrix} \tag{1.30}$$

各分量相互依赖的拉普拉斯分布

$$q(\mathbf{y}_i) = \alpha exp\left(-\sqrt{\mathbf{y}_i - \mu_i^T \Sigma_i^{-1} (\mathbf{y}_i - \mu_i)}\right) \tag{1.31}$$

α为归一化因子，μ_i为$\mathbf{y}_i$的均值，Σ_i为$\mathbf{y}_i$的协方差矩阵，作零均值及单位协方差矩阵假设，可得

$$\varphi^{(\mathbf{k})}(\mathbf{y}) = -\begin{bmatrix} \frac{y_1^{(k)}}{\sqrt{\mathbf{y}_1^T \mathbf{y}_1}} \\ \frac{y_2^{(k)}}{\sqrt{\mathbf{y}_2^T \mathbf{y}_2}} \\ \cdots \\ \frac{y_L^{(k)}}{\sqrt{\mathbf{y}_L^T \mathbf{y}_L}} \end{bmatrix} = -(\mathbf{y}^{(\mathbf{k})})./(\sqrt{diag(\mathbf{y}^{\mathbf{T}}\mathbf{y})}) \tag{1.32}$$

式中，diag(*)表示取矩阵对角线元素作为一个列向量，“./”代表左边矩阵的每一行都除以右边向量的对应行。
相对比可以看出各分量独立的拉普拉斯分布模型对$\mathbf{y}_i$的梯度意义不大，所以我们采用各分量相互依赖的拉普拉斯分布作为$q(\mathbf{y}_i)$。

现在，我们可以得到 IVA 的完整的迭代公式

$$\mathbf{G}^{(k)} := \mathbf{G}^{(k)} + \eta\left(I - E\left[\left(\mathbf{y}^{(k)}./\left(\sqrt{diag(\mathbf{y}^T\mathbf{y})}\right)\right)\mathbf{y}^{(k)^T}\right]\right)\mathbf{G}^{(k)} \tag{1.33}$$

这样经过不断的迭代，我们就能将$\mathbf{G}^{(k)}$都收敛到一个稳定值，这就是离线的 IVA 算法。

四、实时独立成分分析算法

为了将 IVA 能够实时运行，需要对上述算法进行改进，在批量下降法中，我们需要用所有样本的数据去估算$\nabla_{\mathbf{G}^{(k)}}\iota$，于是我们想到随机梯度下降法，用$\left(\mathbf{y}^{(k)}[n]./\left(\sqrt{diag(\mathbf{y}^T[n]\mathbf{y}[n])}\right)\right)\mathbf{y}^{(k)^T}[n]$代替$E\left[\left(\mathbf{y}^{(k)}./\left(\sqrt{diag(\mathbf{y}^T\mathbf{y})}\right)\right)\mathbf{y}^{(k)^T}\right]$这样一来，运算量大大减少，也能够运用于实时运算。

我们令

$$\Re[n]=\left(\mathbf{y}^{(k)}./\left(\sqrt{diag(\mathbf{y}^T\mathbf{y})}\right)\right)\mathbf{y}^{(k)^T} \tag{1.34}$$

上面所列出的迭代公式就可以化简为

$$\mathbf{G}^{(k)}:=\mathbf{G}^{(k)}+\eta(I-\Re[n])\mathbf{G}^{(k)} \tag{1.35}$$

但是当$\Re[n]$趋近于I时，梯度会趋于零，所以我们需要再做些改进。令$\Lambda[n]$的对角线元素与矩阵$\Re[n]$相等，非对角线元素为零，迭代公式化为

$$\mathbf{G}^{(k)}:=\mathbf{G}^{(k)}+\eta(\Lambda[n]-\Re[n])\mathbf{G}^{(k)} \tag{1.36}$$

为了提高收敛速度和算法鲁棒性，我们需要对梯度归一化，迭代公式化为

$$\mathbf{G}^{(k)}:=\mathbf{G}^{(k)}+\eta\sqrt{\xi^{-1(k)}[n]}(\Lambda[n]-\Re[n])\mathbf{G}^{(k)} \tag{1.37}$$

其中$\sqrt{\xi^{-1(k)}[n]}$为 n 时间点的归一化因子，归一化因子迭代公式如下

$$\sqrt{\xi^{-1(k)}[n]}=\beta\xi^{-1(k)}[n-1]+(1-\beta)\mathbf{x}^{(k)^T}\mathbf{x}^{(k)}/L \tag{1.38}$$

每一次迭代，之后，我们用$\mathbf{G}^{(k)}$把估计的信号求出来，然后把每个频率点的信号都除以$|\mathbf{G}^{(k)}|$的模，因为如果各个频率点的值的大小不一样的话，也会影响最后重建的效果。

overlap-add，最后我们求出了解混信号之后，需要将信号变换至频域，此时我们应当用 overlap-add 的方式将不同时间的窗的位置所重叠部分的信号加起来，因为用频域相乘得到的结果是圆周卷积，为了防止信号混叠或丢失，我们一般会用 overlap-add 或者 overlap-save 的方法。

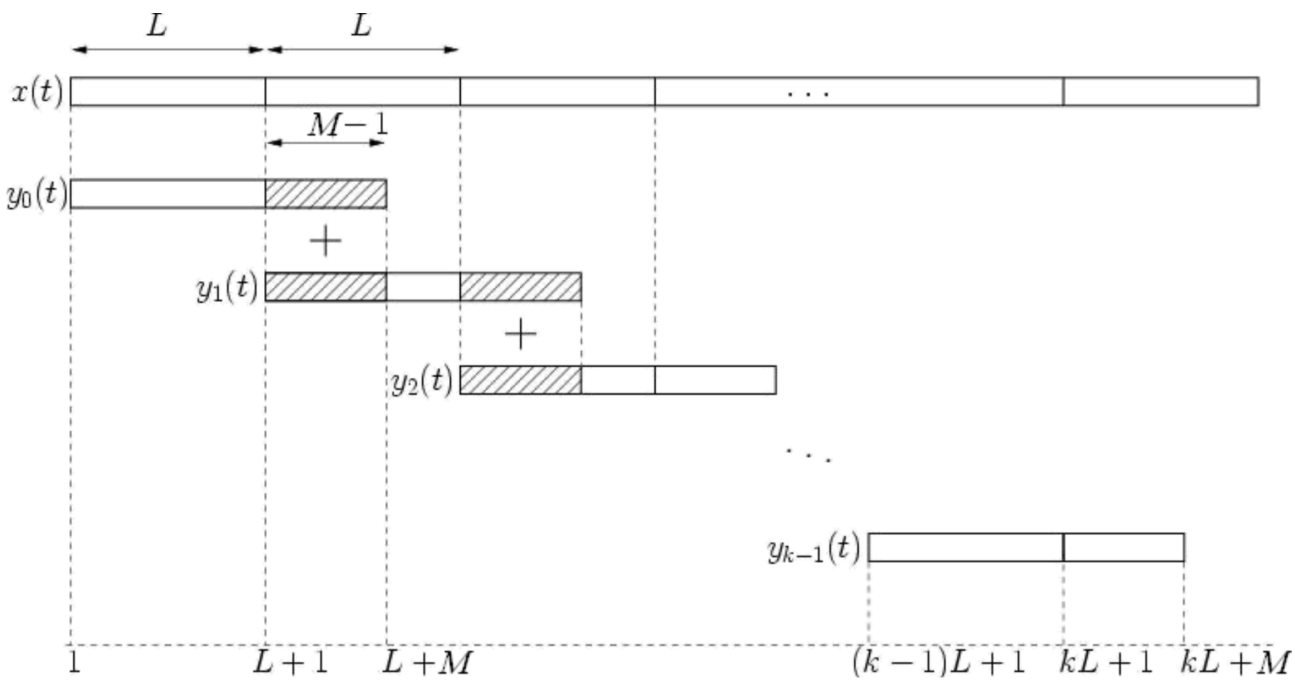

图 1 overlap-add 示意图

五、实时 IVA 仿真

parameters	
N-point FFT	256-point FFT
Window	256-point hanning window
Shift size	64
β (smoothing factor)	0.5
reverberation time(ms)	250
Sample rate(Hz)	8000
η(gradient descent coefficient)	1.4
Audio length(s)	140
Sound velocity(m/s)	344
Reciever1 position(x(m), y(m), z(m))	(2.0, 2.0, 2.4)
Reciever2 position(x(m), y(m), z(m))	(2.0, 2.0, 1.9)
Sound source1 position(x(m), y(m), z(m))	(2.0, 2.4, 2.0)
Sound source2 position(x(m), y(m), z(m))	(2.0, 1.6, 2.7)
Room size(x(m), y(m), z(m))	(5.0, 4.0, 6.0)

表 1 仿真参数列表

仿真中，用 ImageModel[4]去模拟 i 声源在 j 麦克风处的单位抽样响应h_{ij}，然后用两首歌作为声源，并将其与h_{ij}进行卷积，得到麦克风捕获到信号，然后运用上述讨论的实时 IVA 算法，用其迭代公式进行迭代，每经过一个 shift size 可以还原一个 shift size 长度的信号。仿真的具体参数如上表，最后这两段 140s 混合音乐被较成功的分离，耗时约 107s。具体信号图像如下。除了对模拟混合的信号进行仿真外，还对一首音乐的两个通道施行实时 IVA 算法，最后也相当成功的把背景音乐分离了出来。这两个仿真的素材与代码均上传至了 https://github.com/FanmingL/independent-analysis。

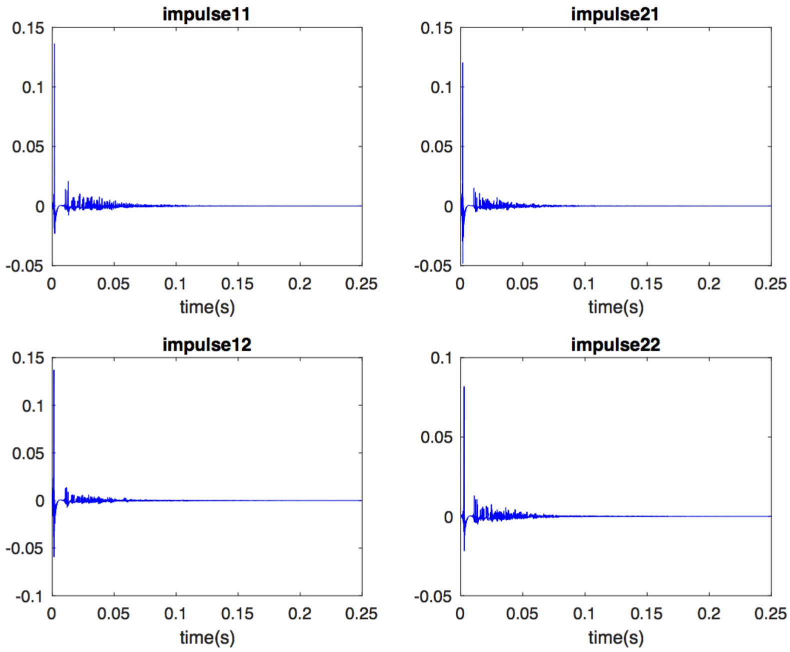

图 2 单位冲激响应

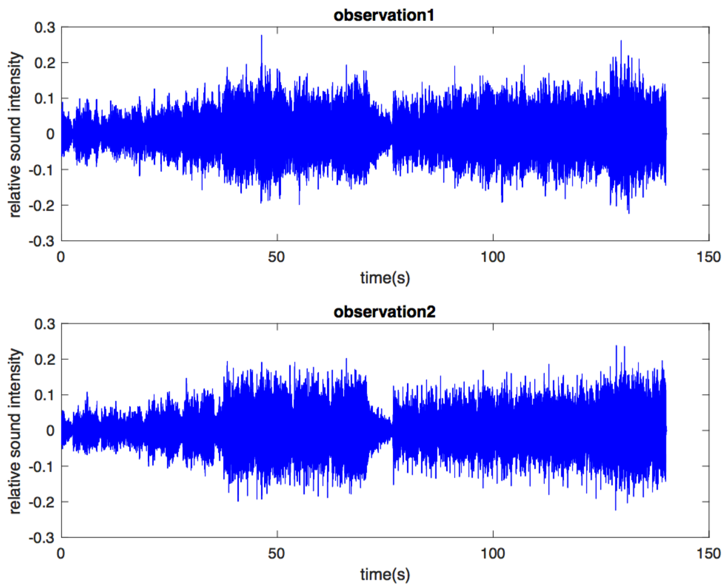

图 3 麦克风捕获的信号

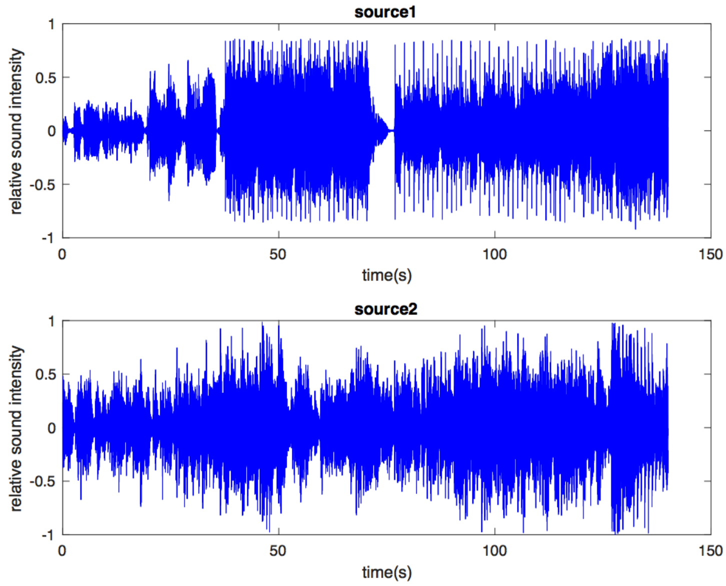

图 4 源信号

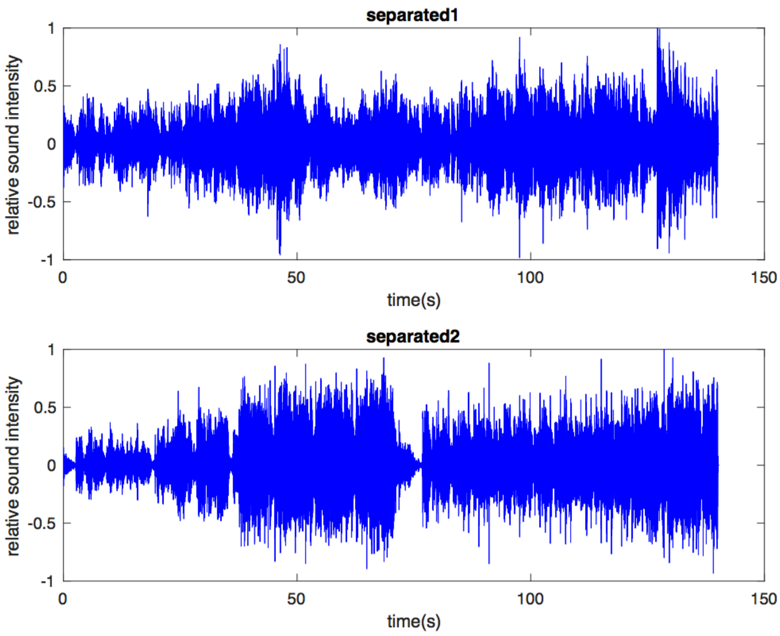

图 5 分离信号

六、总结与展望

在编写好实时 IVA 算法之后，多次尝试发现这种方法包括速度快、鲁棒性强、分离效果好、收敛速度快等优点，但是把结果拿出来仔细听时可以感觉到，分离效果还是有提升空间的。

未来希望首先能在 DSP 系统上运行该算法，然后希望将此算法应用到语音增强的应用上。

七、参考文献：

[1] Kim T. Real-Time Independent Vector Analysis for Convolutive Blind Source Separation[J]. IEEE Transactions on Circuits & Systems I Regular Papers, 2010, 57(7):1431-1438.

[2] http://cs229.stanford.edu/notes/cs229-notes11.pdf

[3] Aapo Hyvarinen, Juha Karhunen, Erkki Oja. 独立成分分析[M]. 电子工业出版社, 2007.
[4]J. B. Allen and D. A. Berkley, “Image method for efficiently simu- lating small room acoustics,” *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.