مقدمه

در یک مهمانی، ما می‌توانیم به یک صدای خاص توجه ویژه داشته باشیم و دیگر صداهای تداخلی موجود در محیط اطراف خود را فیلتر نماییم. این قابلیت ادراکی، باعث ایجاد انگیزه‌ای برای پدید آوردن یک زمینه مطالعاتی جدید گردید. هدف این زمینه مطالعاتی، طراحی سیستم‌های جداسازی گفتار بر اساس اصول سیستم شنوایی انسان است. در بسیاری از کاربردها نظیر بازشناسی گفتار اتوماتیک و مخابرات راه دور به یک سیستم موثر که توانایی جداسازی سیگنال گفتار هدف از سیگنال تداخلی را در شرایط تک‌میکروفونه داشته باشد، نیاز می‌باشد.

جداسازی کور منابع یکی از موضوعات مورد بررسی در زمینه پردازش سیگنال است که توجه به آن در دو دهه اخیر افزایش یافته است. جداسازی سیگنال ها در کاربردهای متنوعی از پردازش سیگنال از جمله پردازش سیگنال های صحبت تا تحلیل تصویرهای پزشکی به کار می رود .

سیستم جداسازی گفتار تک‌میکروفونه پیشنهادی بر اساس ویژگی فرکانس گام در حوزه فرکانس مدولاسیون طراحی گردیده است. جداسازی بر اساس فیلتر نمودن سیگنال نویزی با استفاده از ماسک تخمین زده شده در حوزه طیف مدولاسیون با به‌کارگیری محدوده فرکانس گام تخمین زده شده، انجام می‌گیرد. برای بهبود عملکرد سیستم جداسازی پیشنهادی، یک سیستم جداسازی تک‌میکروفونه ترکیبی نیز پیشنهاد می‌گردد. در این سیستم، از فیلتر نمودن مدولاسیون همدوس وفقی برای جداسازی زیرباندهای با فرکانس پایین و از سیستم جداسازی تک‌میکروفونه ناهمدوس بازگشتی برای جداسازی زیرباندهای با فرکانس بالا استفاده می‌گردد. در فیلتر نمودن مدولاسیون همدوس وفقی، برای حذف سیگنال تداخلی از فیلتر وفقی Affine Projection استفاده می‌گردد که این فیلتر بر روی سیگنال مدولاتور بدست آمده از تبدیل مدولاسیون همدوس، اعمال می‌گردد.

همچنین با به‌کارگیری یک میکروفون اضافه یک سیستم جداسازی دومیکروفونه بر اساس ویژگی‌های اختلاف زمانی برای زیرباندهای با فرکانس پایین و اختلاف چگالی برای زیرباندهای با فرکانس بالا، به منظور افزایش کیفیت سیگنال جدا شده پیشنهاد شده است . در سیستم دومیکروفونه ، جداسازی سیگنال هدف از تداخل بر مبنای ماسک باینری زمان- فرکانس تخمین زده شده بر اساس دو ویژگی مکانی اختلاف زمانی و اختلاف چگالی انجام می‌گیرد. نتایج ارزیابی نشان می‌دهد که سیستم‌های پیشنهادی تک‌میکروفونه در مقابل تداخل مقاوم است و در شرایطی که انرژی سیگنال تداخلی زیاد باشد نیز قادر به جداسازی گفتار هدف با کیفیت خوب می‌باشد. همچنین نتایج بدست آمده از سیستم جداسازی دومیکروفونه نشان دهنده جداسازی قسمت های واکدار و بی‌واک سیگنال گفتار هدف از سیگنال تداخلی با کیفیتی مورد قبول است.

هدف از جداسازی منابع، تخمین سیگنال N منبع ناشناخته مختلف با استفاده از مخلوط سیگنال های دریافتی توسط P سنسور است. به دلیل اینکه اطلاعات اولیه ای راجع به منابع و چگونگی ترکیب آنها وجود ندارد. مسئله جداسازی، جداسازی کور نامیده می شود .

به طور کلی در مسئله جداسازی کور منابع، Pمخلوط خطی از N منبع داریم که تابع تبدیل بین منابع و سنسورها، ماتریس مجهول A به ابعاد N*P می باشد و در رابطه x=As بردار s شامل منابع، s=[s1,s2,…SN]T و x=[x1,x2,…xP]T هم مخلوط سیگنال های دریافتی توسط P سنسور است. بلوک دیاگرام کلی مسئله BSS در شکل 1-1 نشان داده شده است .

شرایط محیطی و نوع مخلوط روی پیچیدگی مسئله BSS تاثیر می گذارند. در یک محیط طبیعی سیگنال های با انعکاس توسط سنسورها دریافت می شوند و بنابراین تخمین ماتریس A به شناسایی جهت منبع در زمان های مختلف نیاز دارد. عموما برای ساده تر شدن مسئله، فرضیاتی برای محیط در نظر گرفته می شود که عبارتند از:

الف) مخلوط لحظه ای: فرض ابتدایی که برای محیط در نظر گرفته می شود این است که سیگنال ها به صورت همزمان ولی با تضعیف های متفاوت به سنسورها برسند. در این محیط رابطه خطی ثابتی بین منابع و سنسورها برقرار است. (ماتریس A یک ماتریس اسکالر به ابعاد N*P با مقادیر ثابت است

$$x(t)=As(t)$$

ب) مخلوط بدون اکو: در این محیط فرض می شود، سیگنال هر منبع با یک تضعیف و تاخیر منحصر به فرد به هر سنسور برسد. در این حالت بین منابع و سنسورها رابطه کانولوشنی برقرار است

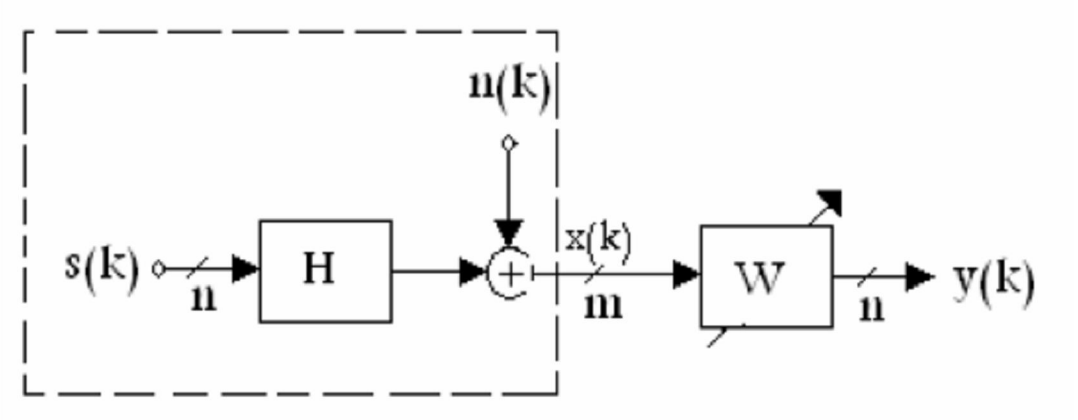
$$x(t)=A*s(t)$$

ج) مخلوط اکودار: این محیط کامل ترین حالت است که در آن بین هر منبع و هر سنسور چند مسیر در نظر گرفته می شود. رابطه بین منابع و سنسور یک رابطه کانولوشنی می باشد که ماتریس Aنسبت به حالت قبل پیچیدگی بیشتری دارد $x(t)=A(z)*s(t)$.

همچنین در مورد منابعی که در مسئله جداسازی کور سیگنال وجود دارند می توان فرضیاتی در نظر گرفت. این فرضیات اساس کار بیشتر الگوریتم های جداسازی منابع را تشکیل می دهند که شامل مشخصات آماری نظیر استقلال، غیرگوسی بودن و... می باشد .

یکی از فرضیات قدرتمند معروف این است که منابع در یک حوزه تبدیل (مانند تبدیل فوریه، تبدیل زمان – فرکانس و...) روی هم افتادگی نداشته باشند. روش هایی که از این فرض استفاده می نمایند. به عنوان روش های اسپارس شناخته می شوند. مزیت این فرض این است که احتمال اینکه دو یا تعداد بیشتری از منابع همزمان در یک نقطه از فضای اسپارس فعال باشند بسیار کم است .

بنابراین در یک فضای اسپارس می توان با تخمین ضریب مربوط به هر منبع به تنهایی، سهم منبع مورد نظر را از ترکیبات حذف کرد. این فرض در شرایطی که تعداد منابع بیشتر از سنسورها می باشد (حالت نامعین) کاربرد دارد. برای نمایش اسپارس یک سیگنال آکوستیک اغلب از تبدیل فوریه، تبدیل گابر و تبدیل موجک استفاده می شود. شکل 1 بلوک دیاگرام این مدل را نشان میدهد.



شکل1 بلوک دیاگرام مربوط به BSS

مدلسازی جداسازی کورکورانه سیگنال به صورت زیر می باشد:

سیگنال اصلی:

$$S(t) = [s_1(t), s_2(t) \ldots s_n(t)]^T$$

بردار مشاهده:

$$X(t) = AS(t) = [x_1(t), x_2(t) \ldots x_m(t)]^T$$

ماتریس جداساز کننده W :

$$Y(t) = WX(t)$$

هدف محاسبه ماتریس جداساز W می باشد.

در اینجا سه نوع روش مختلف برای جداسازی سیگنال های مخلوط شده آورده شده است. ریاضیات مربوط به هر یک به طور خلاصه آورده شده است.

روش اول EASI می باشد.

این روش مبتنی بر روش LMS بوده و روابط آن به شرح زیر است:

$$W(0) = I$$

مقدار دهی اولیه W

$$y(t) = W(t)S(t)$$

$$g(y(t)) = y^3(t)$$

تابع غیرخطی G به منظور مدلسازی ماهیت غیر خطی سیگنال صوت

$$W(t+1) = W(t) + u[y(t)y^T(t) - I + g(y(t))y^T(t) - y(t)g^T(y(t))]W(t)$$

روش دوم gradient RLS می باشد.

ابتدا عملیات سفید کننده انجام می شود. بدین منظور میانگین و واریانس محاسبه می شود

$$\bar{x}(t) = x(t) - E\{x(t)\}$$

$$v(t) = E\{\bar{x}(t)\bar{x}^T(t)\}^{-1/2} x(t)$$

سپس تابع هزینه تعریف شده در زیر مینیمم می شود.

$$J(W) = \sum_{i-1}^{t} \beta^{t-i} \left\| v(i) - W(i) g(W^T(i-1)v(i)) \right\|^2$$

$$y(t) = W(t-1)v(t)$$
$$z(t) = g(y(t))$$
$$h(t) = P(t-1)z(t)$$
$$m(t) = \frac{h(t)}{\beta + z^T(t)h(t)}$$
$$P(t) = \frac{1}{\beta} Tri\ [P(t-1) - m(t)h^T(t)]$$
$$W(t) = W(t-1) + m(t)[v^T(t) - z^T(t)W(t-1)]$$

در روابط Tri ماتریس بالا مثلثی است. به طور کلی همگرایی روش RLS از روش LMS سریع تر است.

اسم روش سوم natural gradient of the RLS method است. این روش توسعه یافته روش RLS است و هدف آن سرعت همگرایی و دقت بیشتر در ازای پردازش بیشتر می باشد. خلاصه روند آن به شرح زیر می باشد:

$$y(t) = W(t-1)v(t)$$
$$z(t) = g(y(t))$$
$$Q(t) = \frac{P(t-1)}{\beta + z^T(t)P(t-1)y(t)}$$
$$P(t) = \frac{1}{\beta}[P(t-1) - Q(t)y(t)z^T(t)P(t-1)]$$
$$W(t) = W(t-1) + [P(t)z(t)v^T(t) - Q(t)y(t)z^T(t)W(t-1)]$$

نتایج شبیه سازی

سیگنال اصلی به صورت زیر در نظر گرفته شده است. که ترکیبی از 5 سیگنال زیر می باشد

$$s(t) = [\text{sgn}(\cos(2\pi 155t)), \sin(2\pi 800t), \sin(2\pi 300t + 6\cos(2\pi 60t)), \sin(2\pi 90t), v(t)]^T$$

Where v (t) [-1, 1] is uniformly distributed noise,

the sampling frequency of 10kHz,

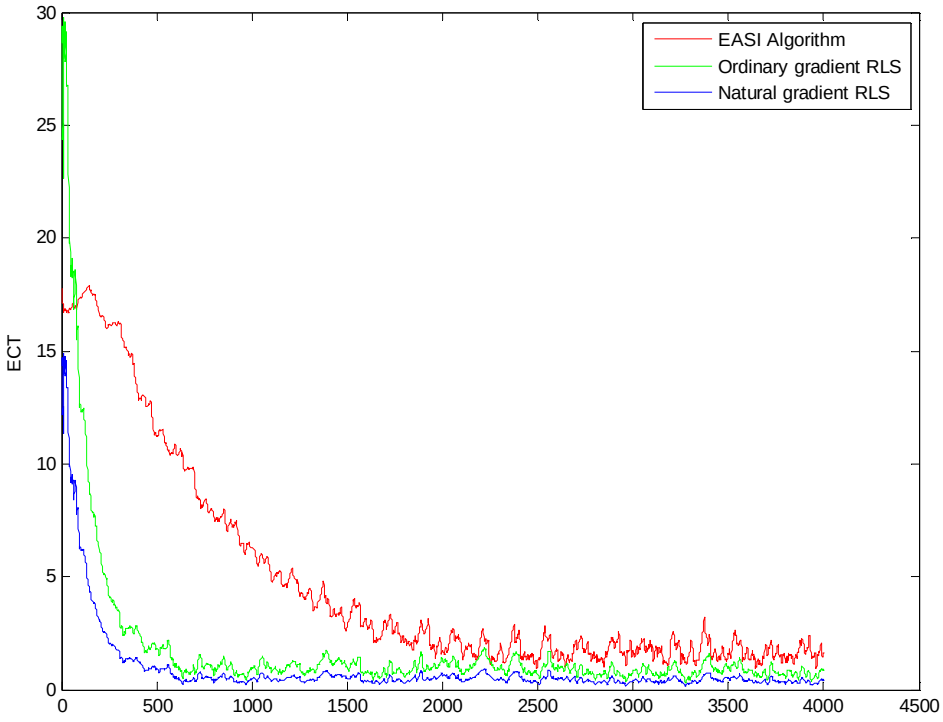
the sample size was 4000.

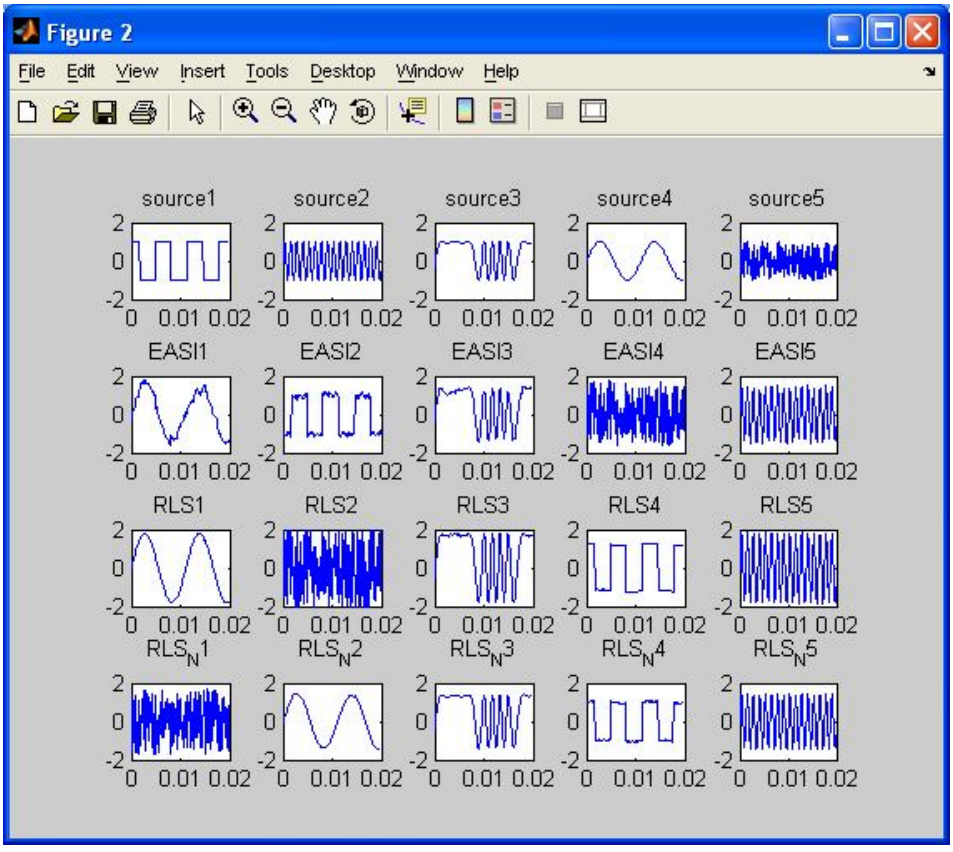
Mixing matrix A is [0,1] uniform distribution of 5 * 5 random matrix.

EASI method parameters: non-linear function g (y) = y^3

RLS method parameters are: nonlinear function g (y) = Tanh - (y)

Forgetting factor β = 0.99

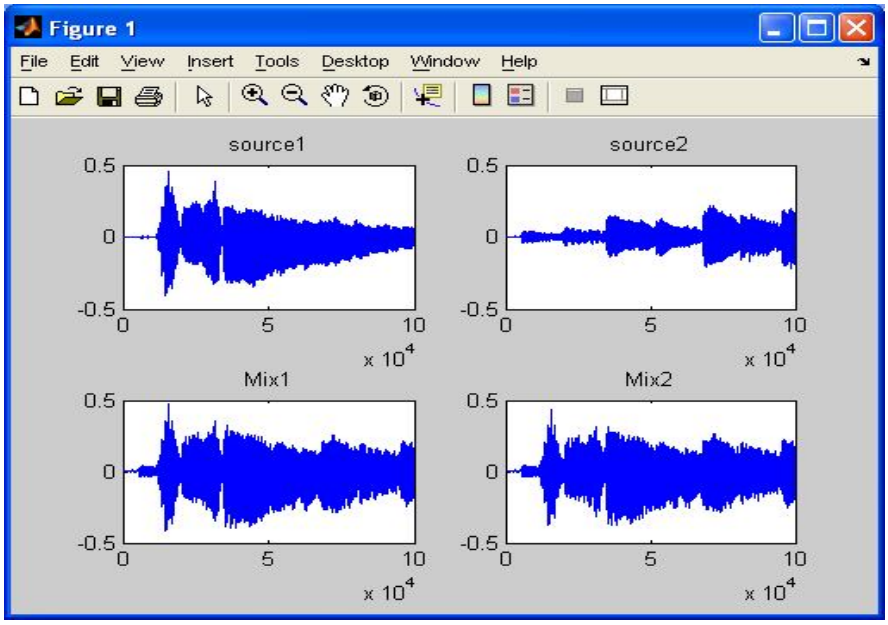




نتیجه جداسازی 5 سیگنال مختلف مخلوط شده با 3 روش بیان شده

$$A=\begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

نمایش سیگنال های صوتی اصلی و ترکیب آنها با یکدیگر



تابع غیرخطی استفاده شده:

g (y) = y-Tanh - (y).

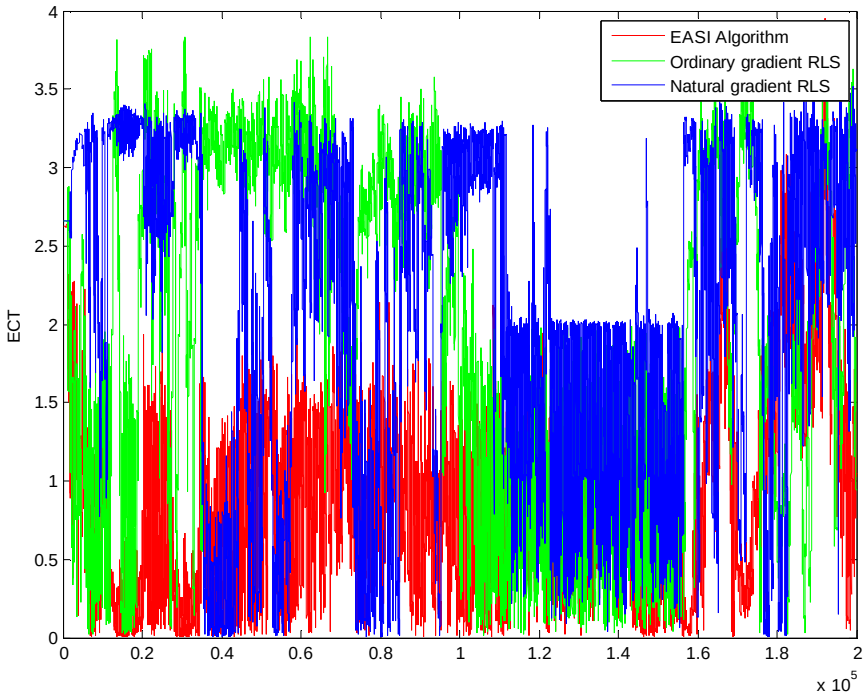
EASI: learning rate u = 0.006

RLS method: forgetting factor β = 0.9 9

Isolated ECT curve as follows:

crosstalk error of ECT:

$$E=\sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{|c_{ij}|}{\max_k|c_{ik}|}-1\right)+\sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{|c_{ij}|}{\max_k|c_{kj}|}-1\right)$$



نتیجه جداسازی سیگنال های مخلوط صوتی به شکل زیر می باشد: