Solving the cocktail party problem

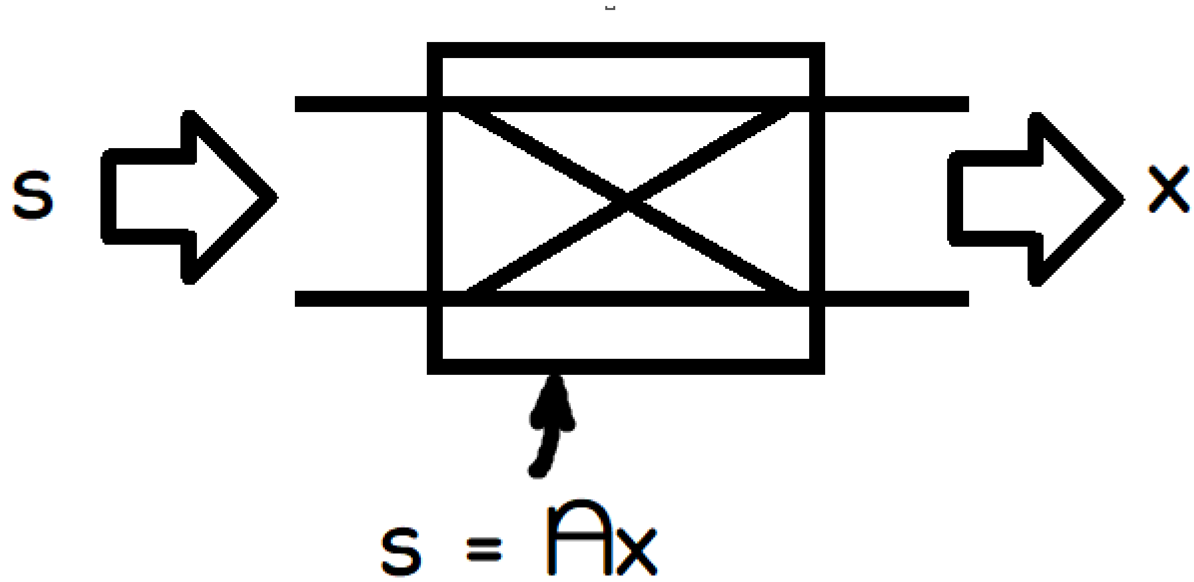
Solving the cocktail party problem

(https://stackoverflow.com/questions/20414667/cocktail-party-algorithm-svd-implementation-in-one-line-of-code)

(Andrew Ng) The original one-liner was:
[W,s,v]=svd(
**(repmat(sum(x.*x,1),size(x,1),1).*x)x'*
);

The matrix being svd'ed is an estimate of the power spectrum of x(t). Actually it is the time-frequency spectrogram. Columns are instantaneous fft's, while rows run in time.

X = repmat(sum(x.*x,1),size(x,1),1).*x)*x'



(https://www.youtube.com/watch?v=P5mnh_xOSAA)
Assumptions:

1. The mixing matrix *is* invertible
2. Sources are statistically independent (ie no correlation)
3. The independent components have non-Gaussian distributions
4. The data has been centered, both x and s have zero mean.

$x = As$
The goal is to estimate an unmixing matrix W:

$\hat{s} = Wx \approx s$

Strategy to find W:
$W = (U\Sigma V)^{-1} = V\Sigma^{-1}U^T$

1. Covariance of the observed (mmicrophone) data gives U and Σ
2. Independence of the *sources* gives V

Covariance of the centered data:
$\langle xx^T \rangle = \langle Ass^T A^T \rangle = \langle U\Sigma V^T ss^T V\Sigma U^T \rangle$

(The U, V, Σ matrices are all constant)

$= U\Sigma V^T \langle ss^T \rangle V\Sigma U^T$

Now, $\langle ss^T \rangle$ is the covariance of the source vector itself. Note that the source vector is structured as $s = s(t) = \begin{bmatrix} s_1(t) \\ s_2(t) \end{bmatrix}$. We are taalking about the covariance of the two sub-processes $s_1(t)$ and $s_2(t)$.

Assumption: The covariance matrix $\langle ss^T \rangle = I$, justify this!

This gives:
$\langle xx^T \rangle = U\Sigma^2 U^T$ (Since $VV^T = I$). This is a spectral decomposition of xx^T

So to obtain U and Σ, perform spectral decomposition of the covariance matrix $\langle xx^T \rangle$ (*after removing the mean, of course*)

The guiding idea here is that a good way to unmix the signals is to use a linear transform that seperates the data out into statistically independent channels (ie if there is a trace of one signal common to both channels, then that woud show up in covariance)

Covariance is $\langle xy \rangle \equiv E[(x - \mu_x)(y - \mu_y)]$, and the expectation operator is realized as an integration over time samples, rather than an instantaneous ensemble average (ie we assume ergodicity)

So do the following:

1. Call the actual signals $s_1(t)$ and $s_2(t)$, and let $s(t) \equiv $
   Similar packaging for the mixed signal, $x(t) \equiv \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = A \begin{bmatrix} s_1(t) \\ s_2(t) \end{bmatrix}$
2. Center the recorded data $x(t) \rightarrow \begin{bmatrix} x_1(t) - \tilde{x}_1 \\ x_2(t) - \tilde{x}_2 \end{bmatrix}$. So from now on x_i means $(x_i(t) - \tilde{x}_i)$
3. Evaluate the covariance:
   1. $\langle xx^T \rangle = \left\langle \begin{bmatrix} x_1^2 & x_1 x_2 \\ x_2 x_1 & x_2^2 \end{bmatrix} \right\rangle \equiv C$ (say).
   2. Diagonalize C to catch U and Σ, as $C \simeq U\Sigma U^T$.

3.