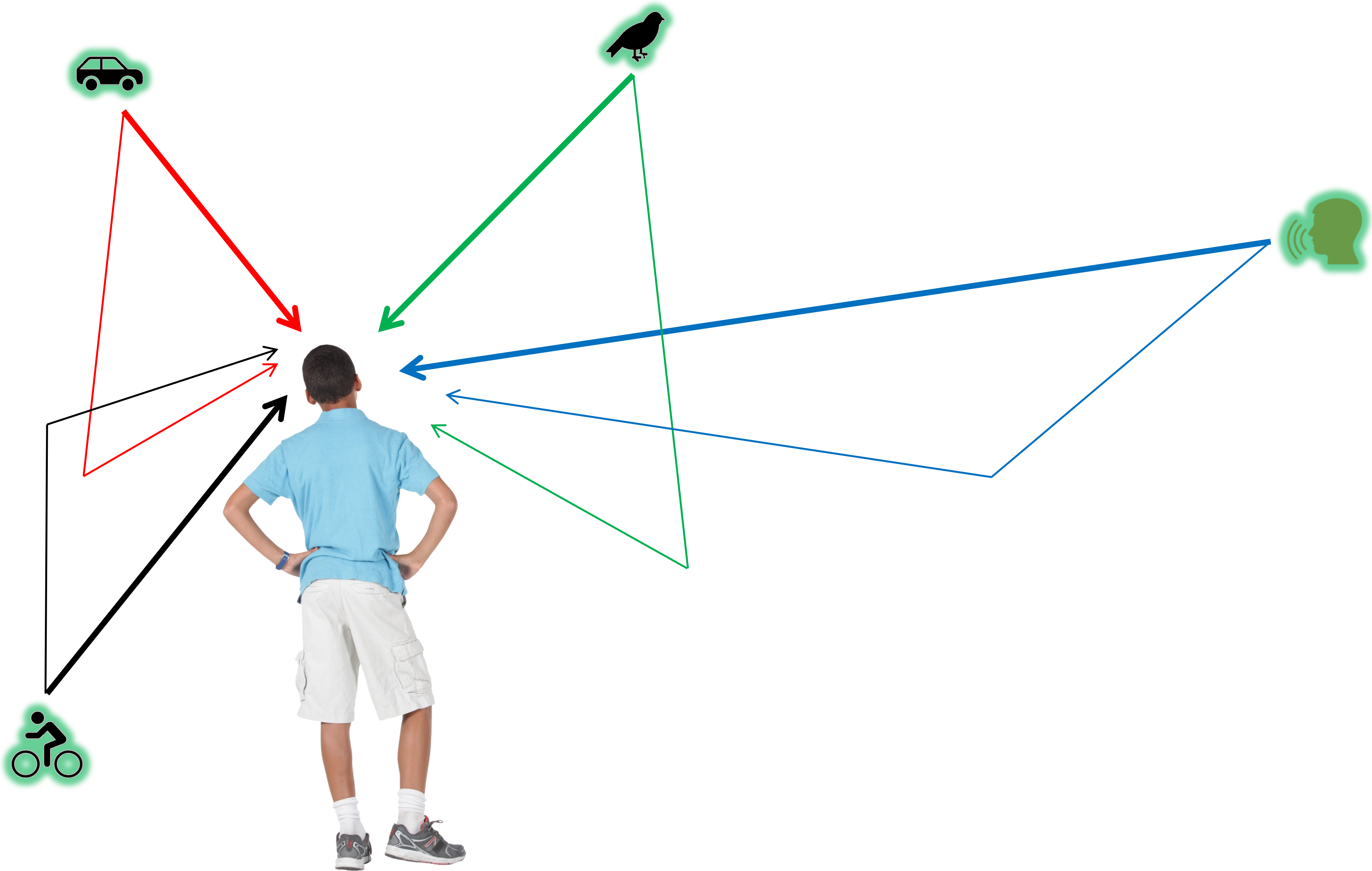
Source separation

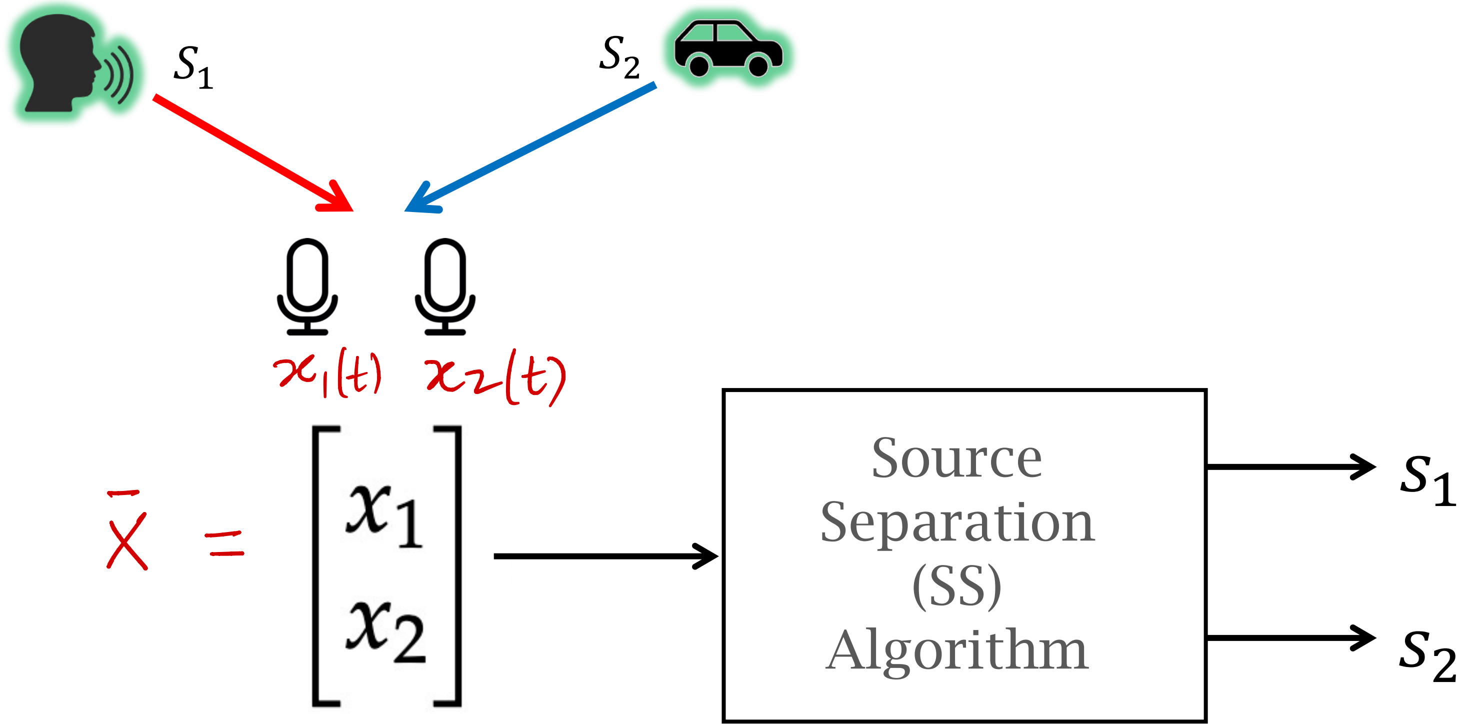
Source separation



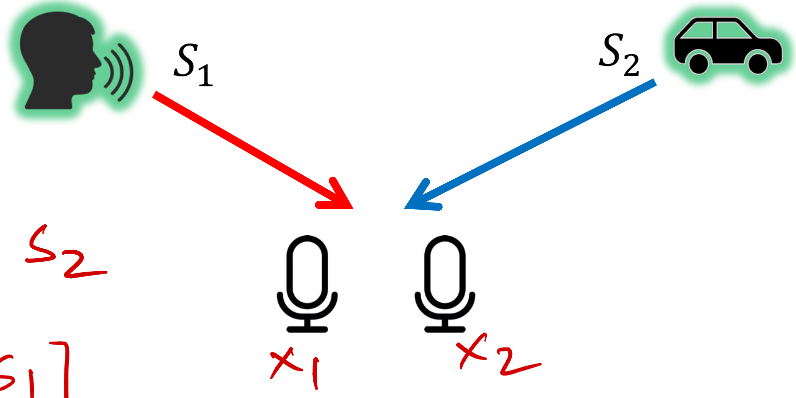
Source separation

Source separation preliminaries

Source separation: The general problem statement



Source separation: The general problem statement



$\bar{x}$ is a mixture of s_1 and s_2

$x_1 = a_1 s_1 + b_1 s_2$
$x_2 = a_2 s_1 + b_2 s_2$

$x_1 = a_1 s_1 + b_1 s_2 = [a_1 \;\; b_1] \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

Mixing matrix

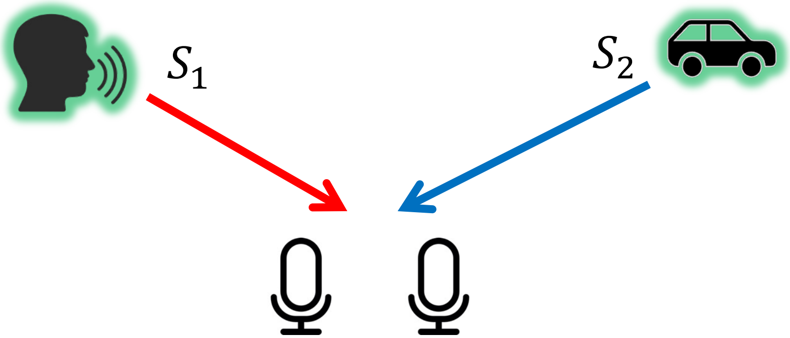
$\bar{s} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$

$x = A\bar{s}$

Unknown mixing matrix, unknown source signals → heavily under-determined

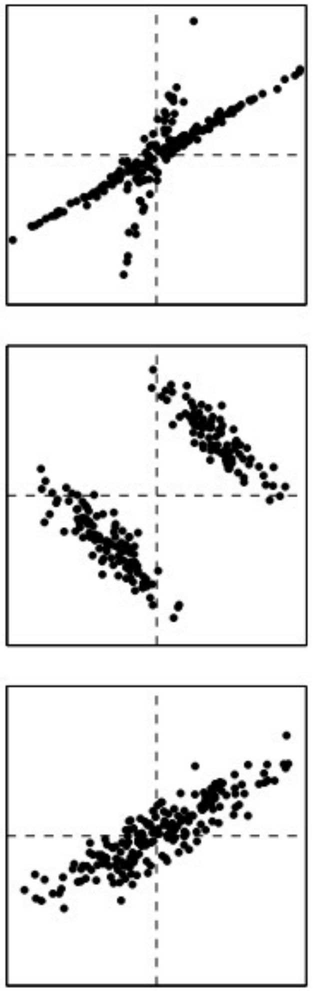
Source separation preliminaries

Source separation: The general problem statement



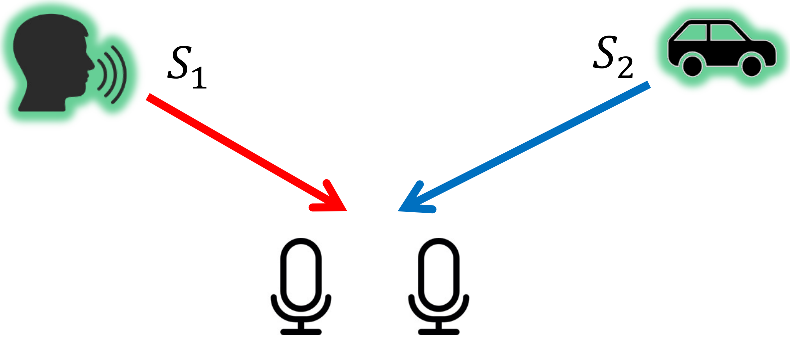
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}}_{\text{Mixing matrix}} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} =$$



Source separation preliminaries

Source separation: The general problem statement



overdetermined system

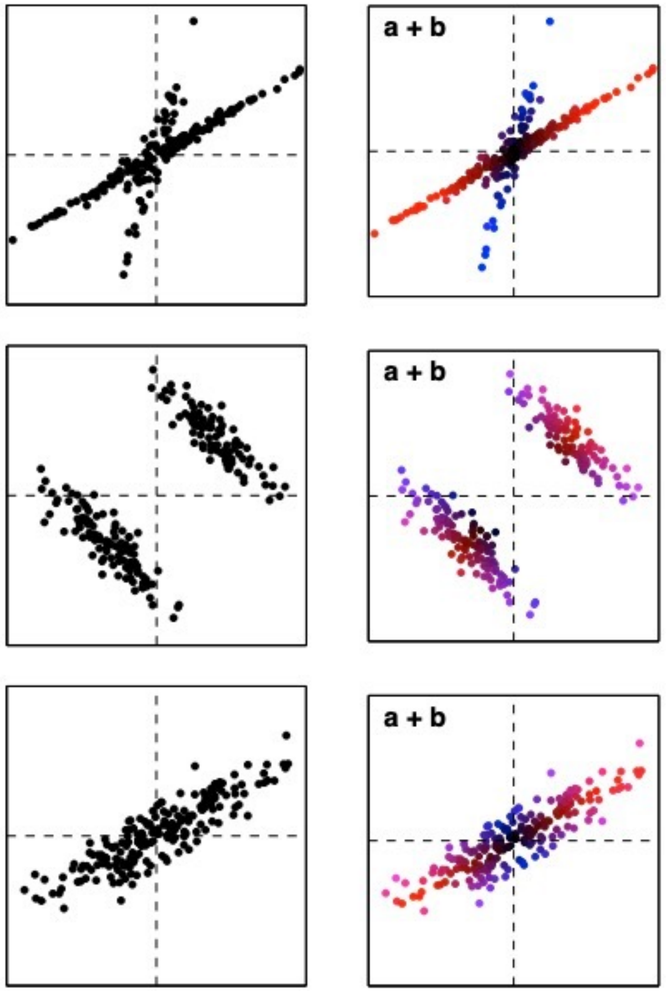
$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} | & | \\ c_1 & c_2 \\ | & | \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

Mixing matrix

underdetermined system

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} =$$

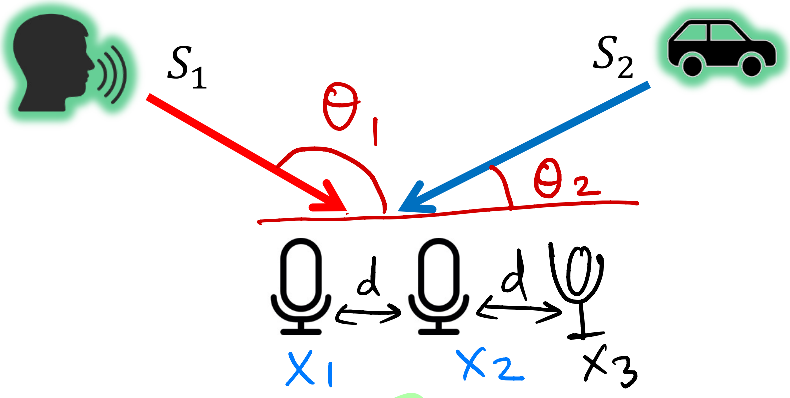


Hard to separate the sources even visually

When can we solve SS?

Let's make some simplifications: Mixing matrix is known based on **Angle of Arrival (AOA)**

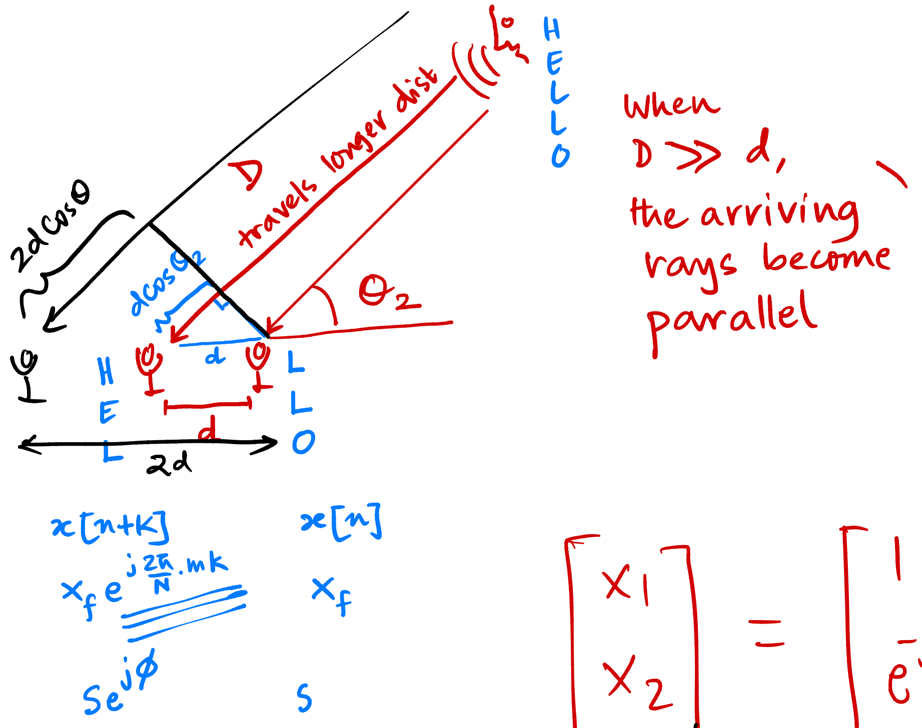
- What is angle of arrival (AoA)? How do you quantify it?
- Relation between AoA and FFT
- How do you get AoA? From camera or from audio itself?
- How to solve X = A S + N even when A is known



$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} | & | \\ \vec{a}_{\theta_1} & \vec{a}_{\theta_2} \\ | & | \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

AoA matrix A

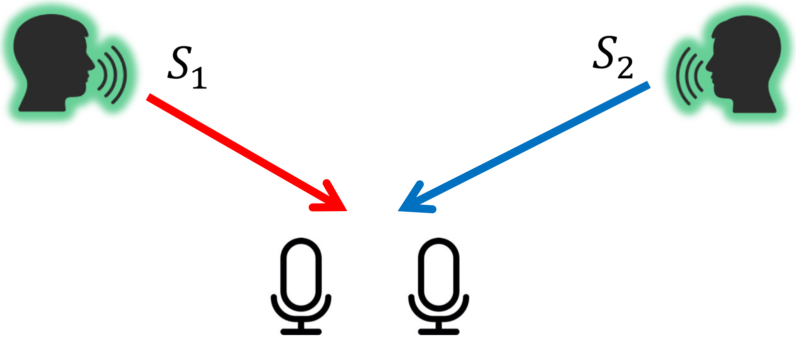
$$X = A.S + N$$



$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}_f = \begin{bmatrix} 1 & 1 \\ e^{-j\phi_1} & e^{-j\phi_2} \\ e^{-j2\phi_1} & e^{-j2\phi_2} \end{bmatrix}_f \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}_f$$

Source = Speech

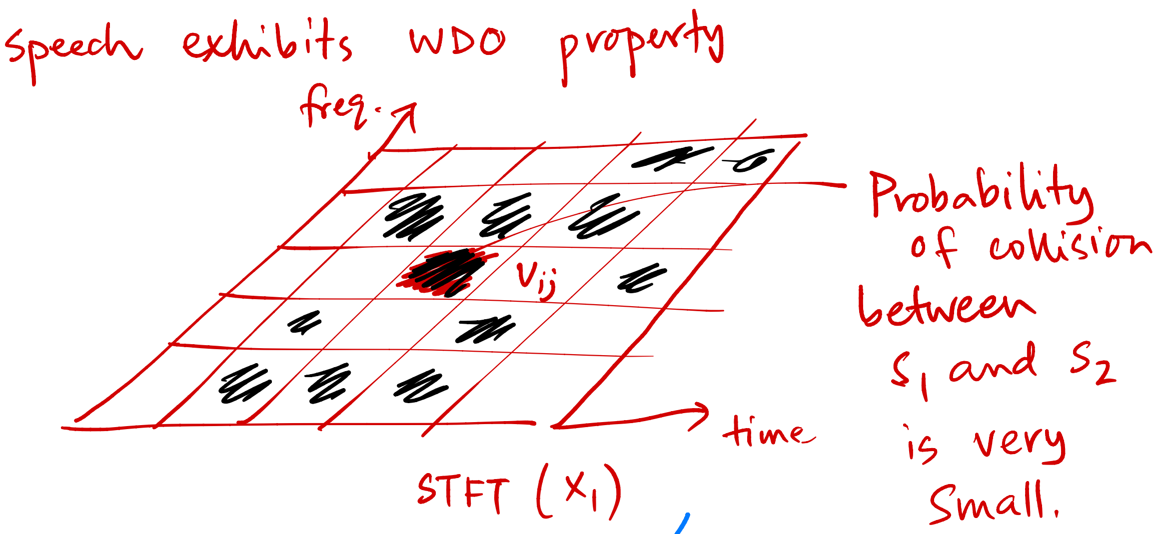
But what if AoA unknown? **It's hard to solve for S … but what if S is speech signals?**

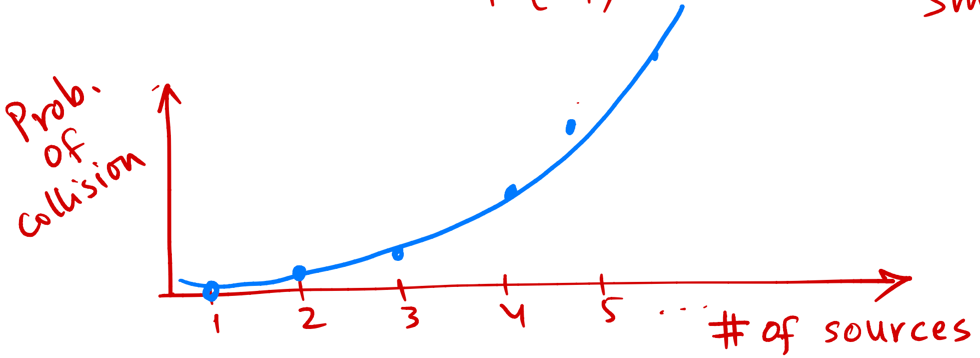


$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} | & | \\ \vec{a}_{\theta_1} & \vec{a}_{\theta_2} \\ | & | \end{bmatrix}}_{\text{AoA matrix A}} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

AoA matrix A

$$X = A.S + N$$

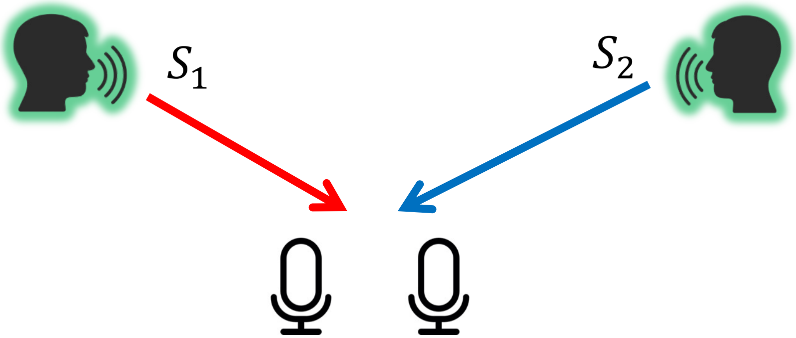




Source = Speech

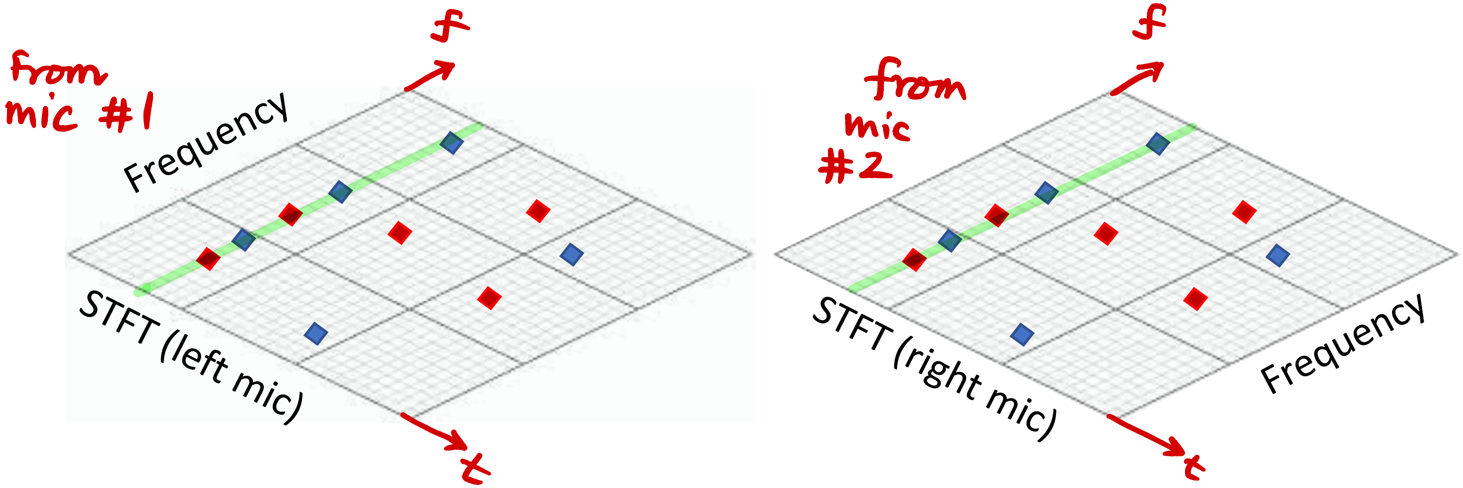
But what if AoA unknown? **It's hard to solve for S … but what if S is speech signals?**

When the source signal is speech, exploit TF-disjointness



$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} | & | \\ \vec{a}_{\theta_1} & \vec{a}_{\theta_2} \\ | & | \end{bmatrix}}_{\text{AoA matrix A}} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

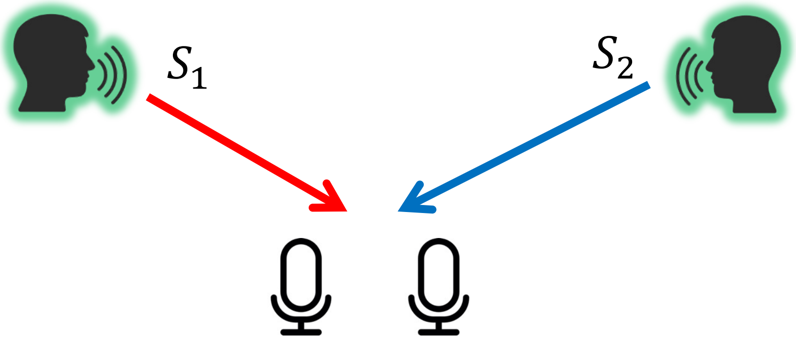
$$X = A.S + N$$



Source = Speech

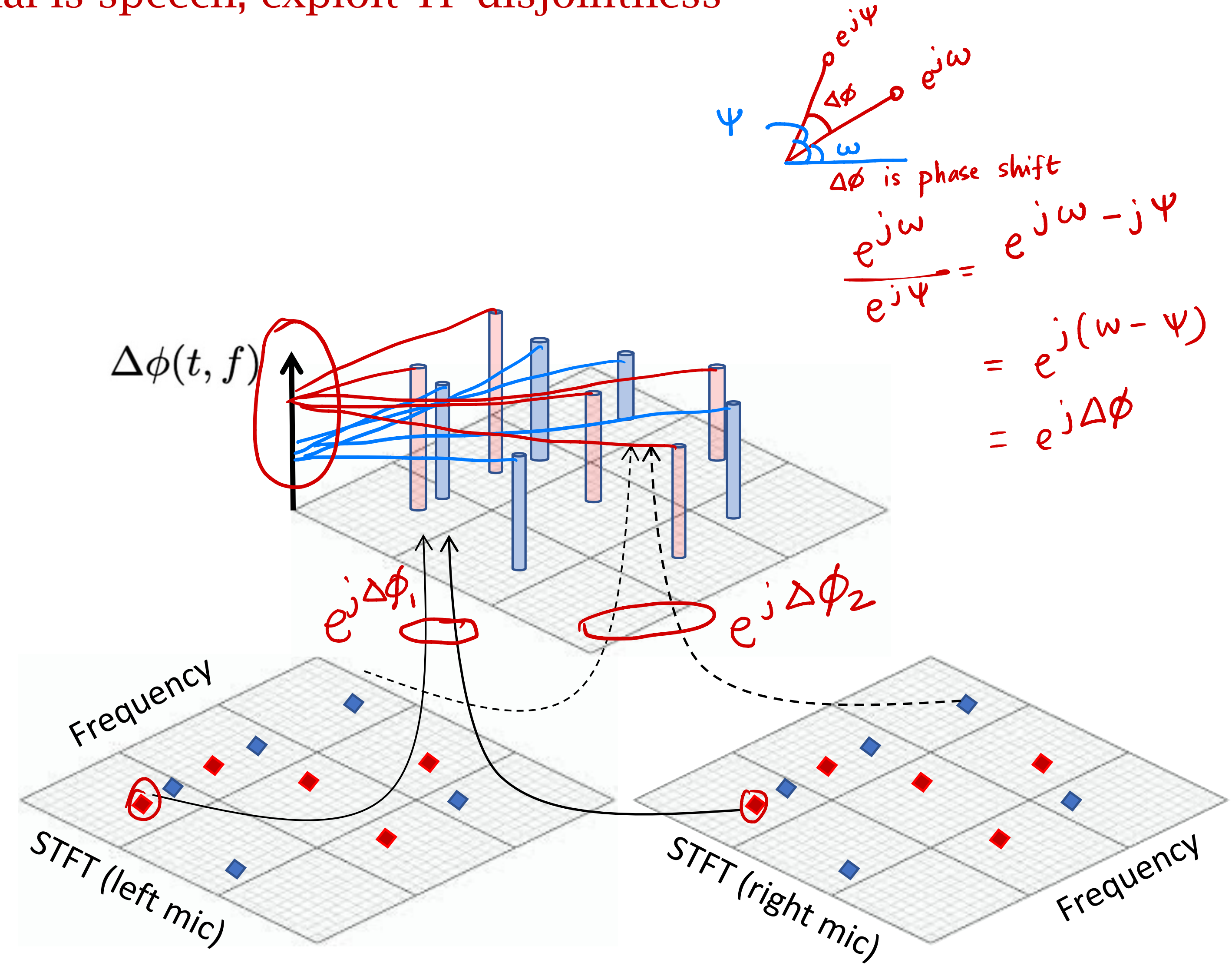
But what if AoA unknown? **It's hard to solve for S … but what if S is speech signals?**

When the source signal is speech, exploit TF-disjointness



$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} | & | \\ \vec{a}_{\theta_1} & \vec{a}_{\theta_2} \\ | & | \end{bmatrix}}_{\text{AoA matrix A}} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

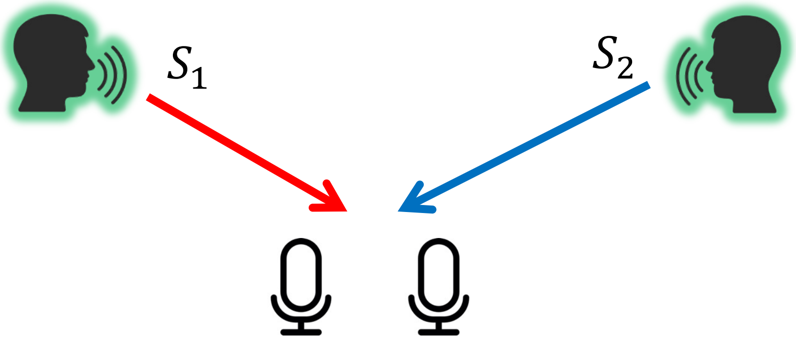
$$X = A.S + N$$

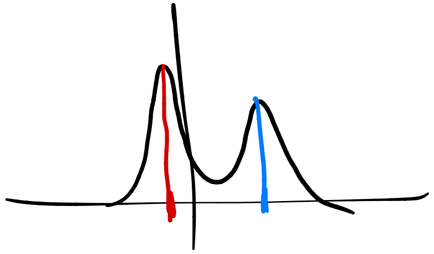


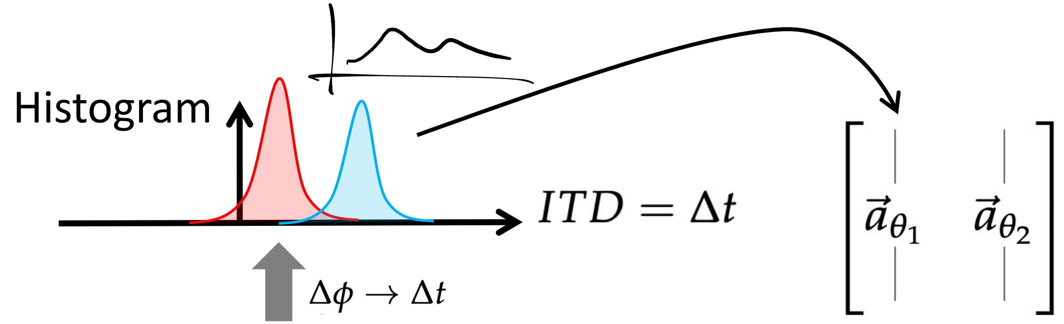
Source = Speech

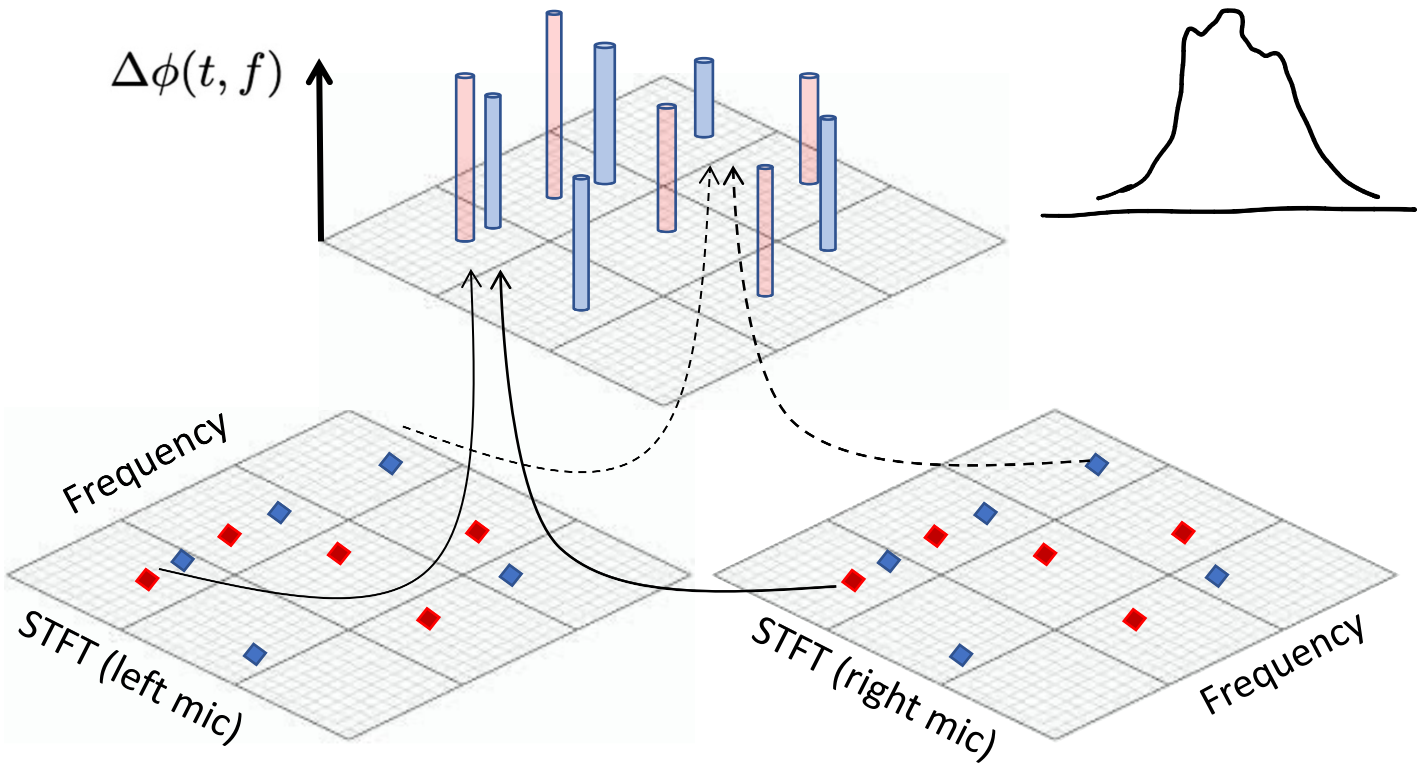
But what if AoA unknown? **It's hard to solve for S … but what if S is speech signals?**

When the source signal is speech, exploit TF-disjointness







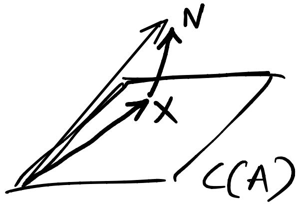


$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} | & | \\ \vec{a}_{\theta_1} & \vec{a}_{\theta_2} \\ | & | \end{bmatrix}}_{\text{AoA matrix A}} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

AoA matrix A

$$X = A.S + N$$

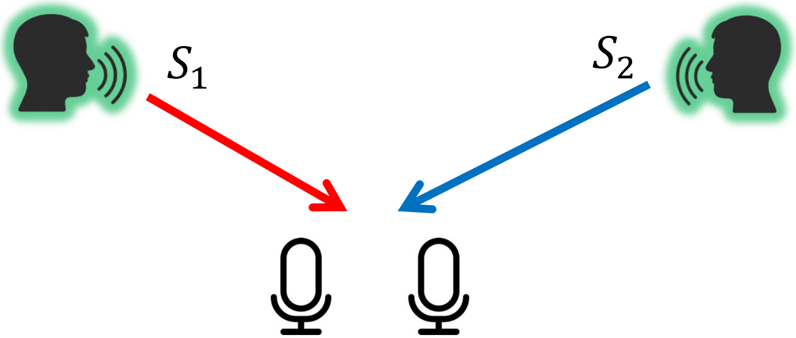
$$(X - N) = AS$$



Source = Speech

But what if AoA unknown? **It's hard to solve for S … but what if S is speech signals?**

When the source signal is speech, exploit TF-disjointness



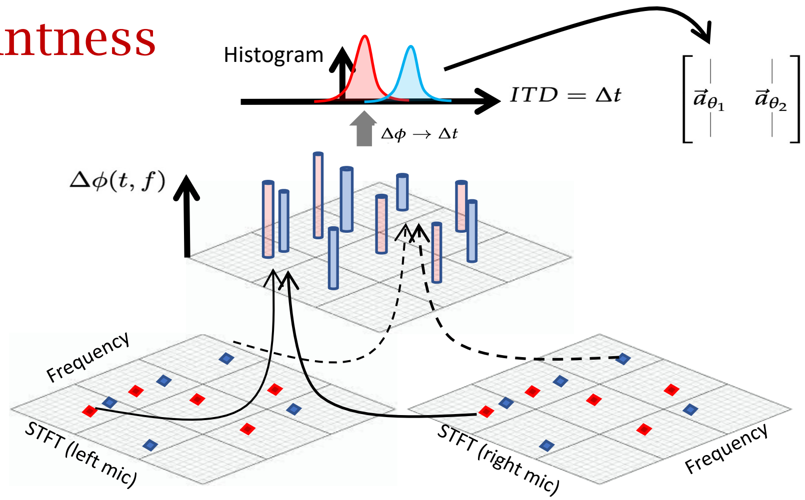
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \vec{a}_{\theta_1} & \vec{a}_{\theta_2} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

AoA matrix A

$$X = A.S + N$$

Reverberation

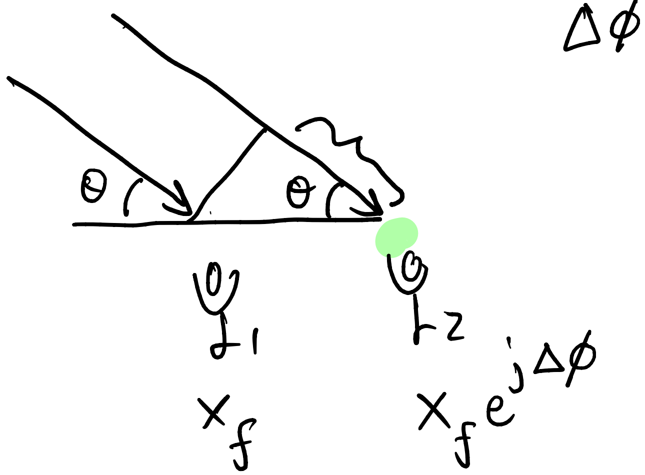
$$X = \begin{bmatrix} a^1_{\theta_1} & a^1_{\theta_2} & a^1_{\theta_3} & a^2_{\theta_4} & a^2_{\theta_-} \end{bmatrix} \begin{bmatrix} s_1 \\ s_1 \\ s_1 \\ s_2 \\ s_2 \end{bmatrix}$$



DUET Algorithm Steps:

1. Take STFT of both mics
2. Divide $STFT_1$ by $STFT_2$ bin by bin ⇒ gives phase shift $\Delta\phi$ for each TF bin
3. Cluster along $\Delta\phi_{TF}$ ⇒ should see 2 clusters (assuming low noise, AoAs not too close, low reverb.)
4. Assign each TF bin to the closest centroid in the clusters
5. Take TF bins belonging to the same cluster and take the IFFT to reconstruct source signal.

$$\Delta\phi = f(\theta)$$



$$\frac{29.3\, e^{j(32+\Delta\phi)}}{29.3\, e^{j32}} = 1 \cdot e^{j\Delta\phi}$$