# Speaker Recognition Summary
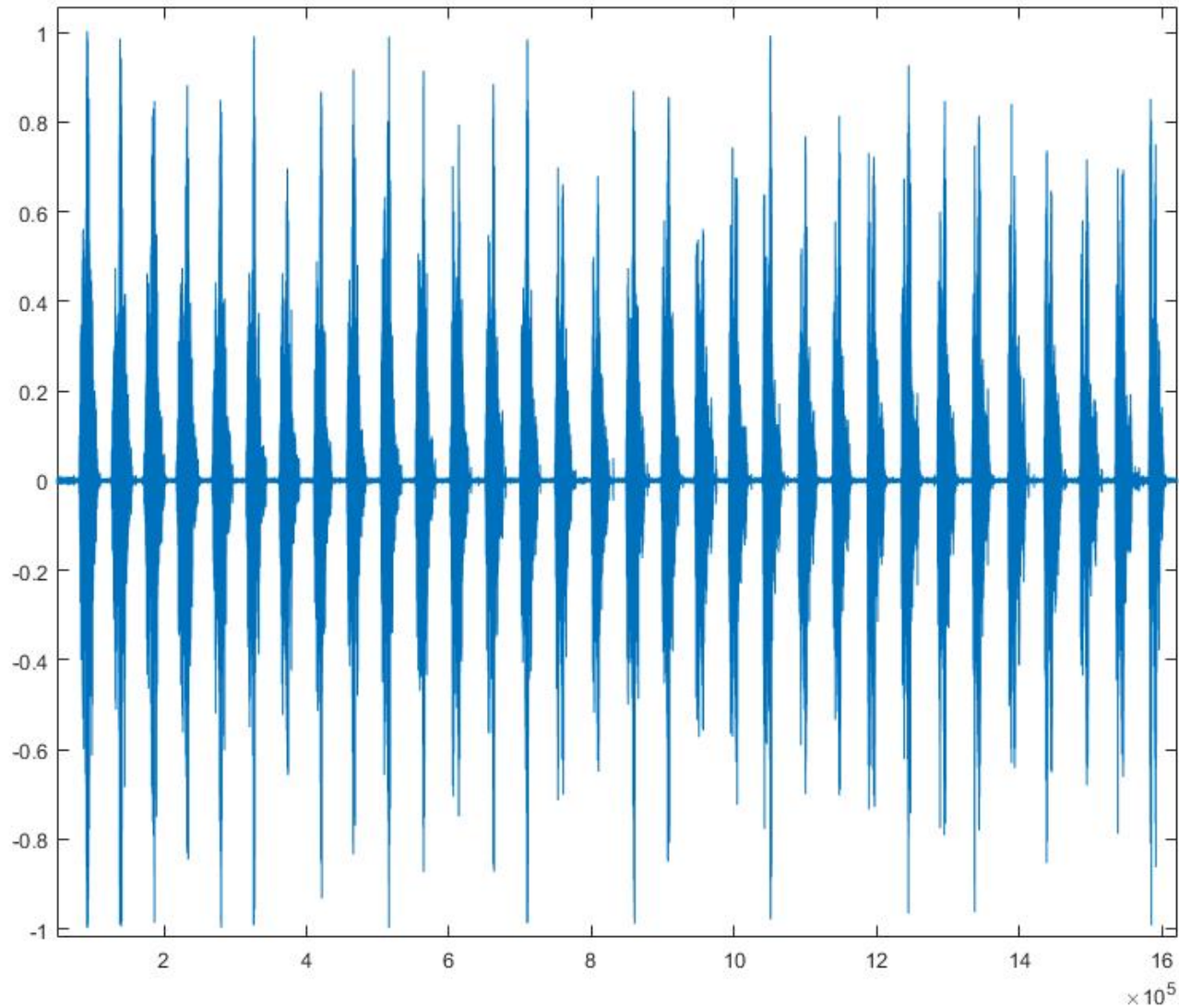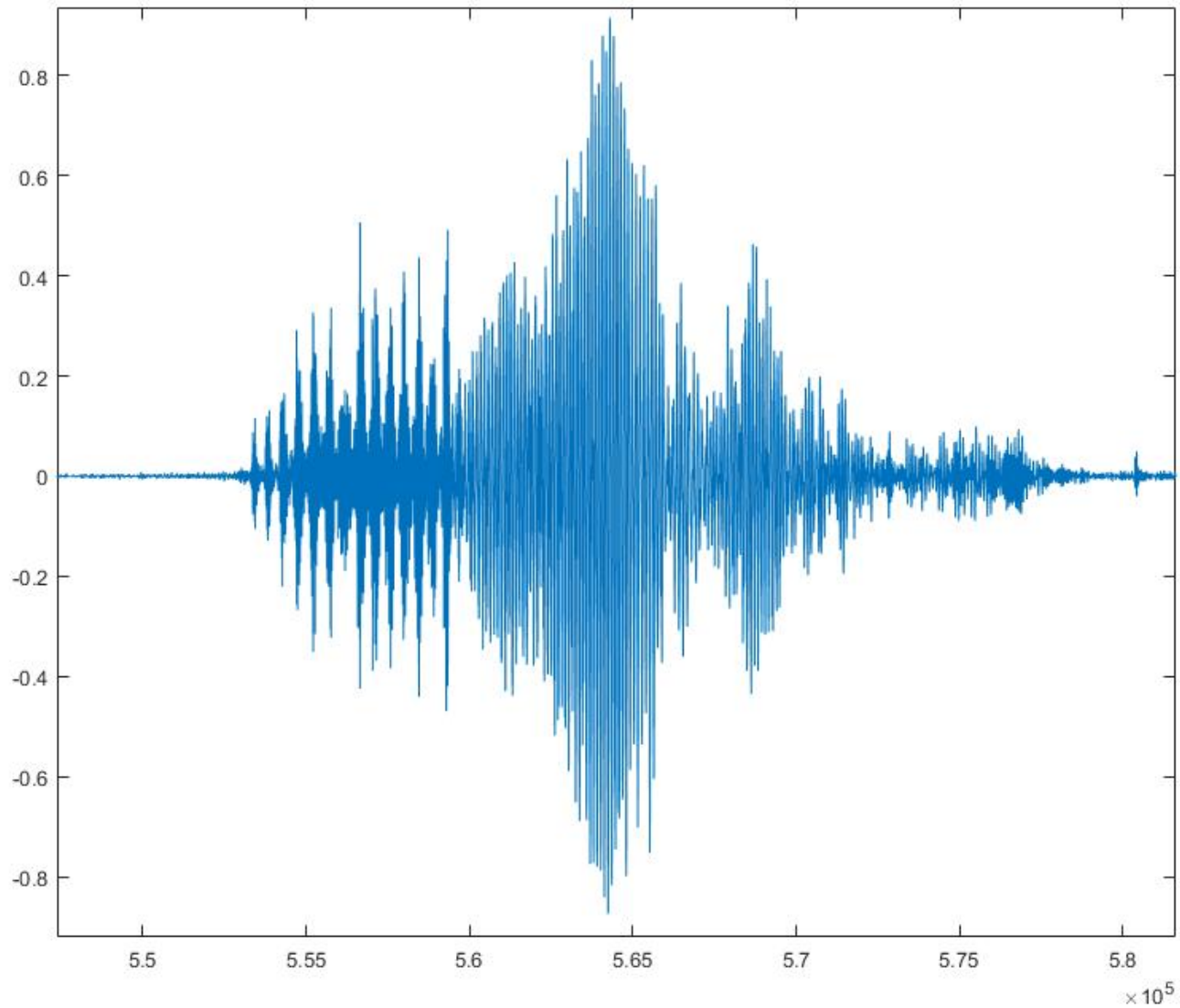
User 2 speech waveform
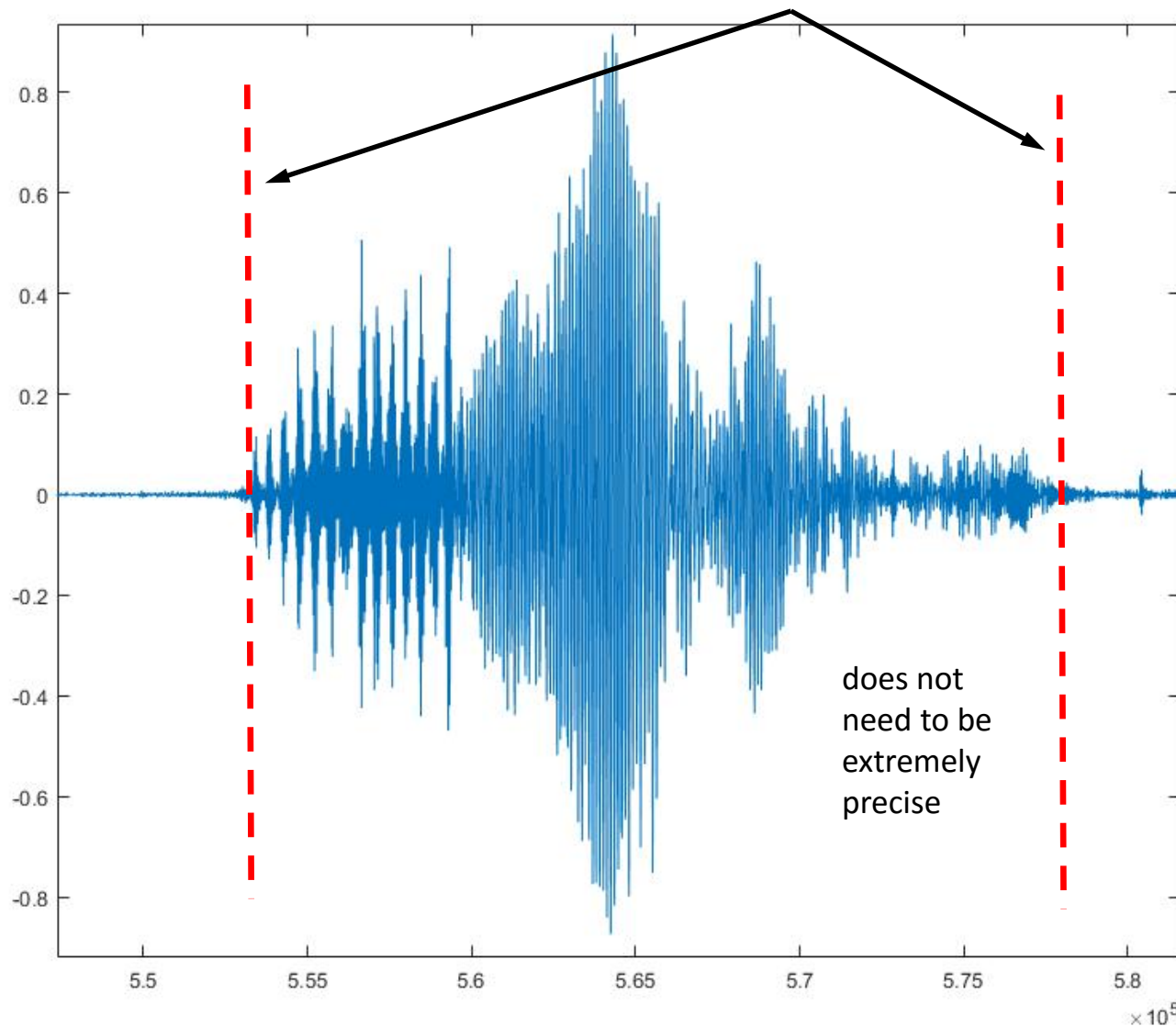
# Word Isolation

zoom in on one word
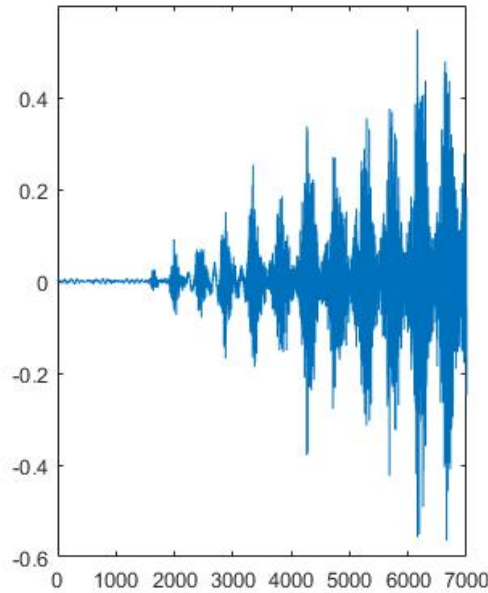
Need to find these boundaries for each spoken word

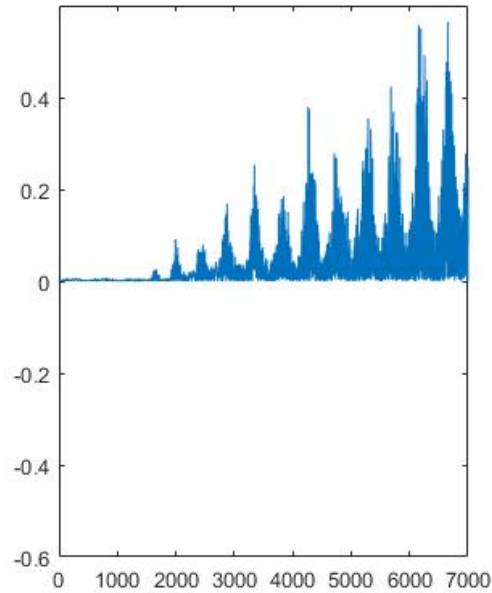does not need to be extremely precise
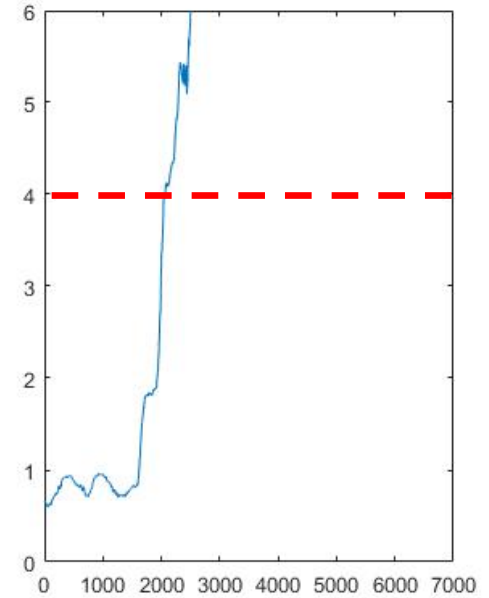
# Word Isolation – A Simple Approach



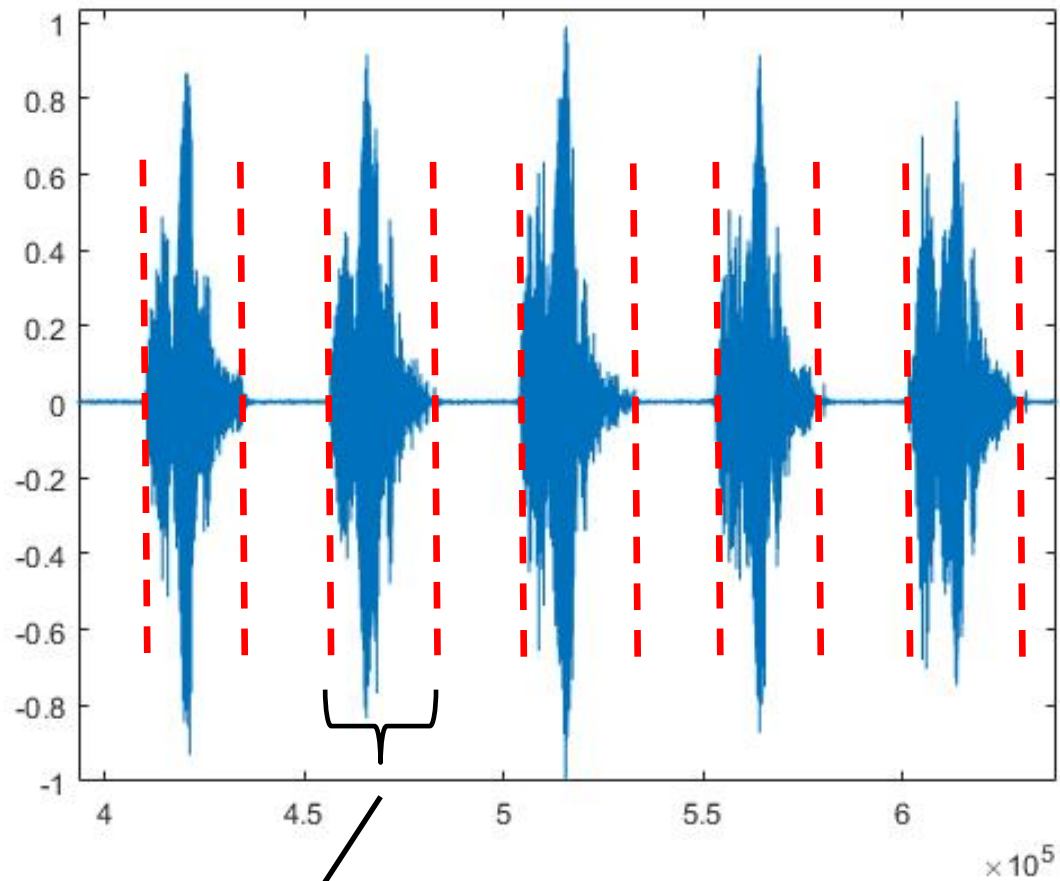Original signal            Rectified signal           Moving average
length = 400

Moving average = convolution with sequence of ones

$$= s_n * \{\cdots\ 1\ 1\ 1\ \cdots\}$$

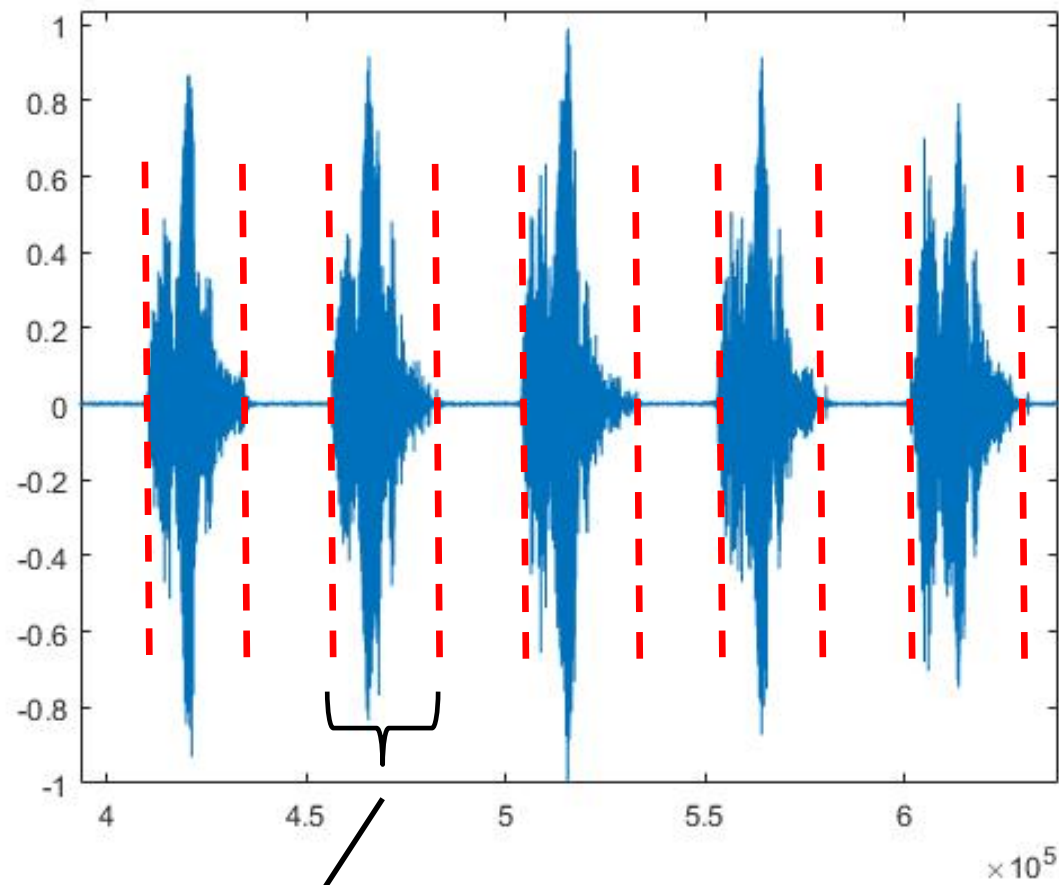# Training Phase



$$\{s_n\} \quad n = 1, \cdots, N$$

isolate sequence for each word
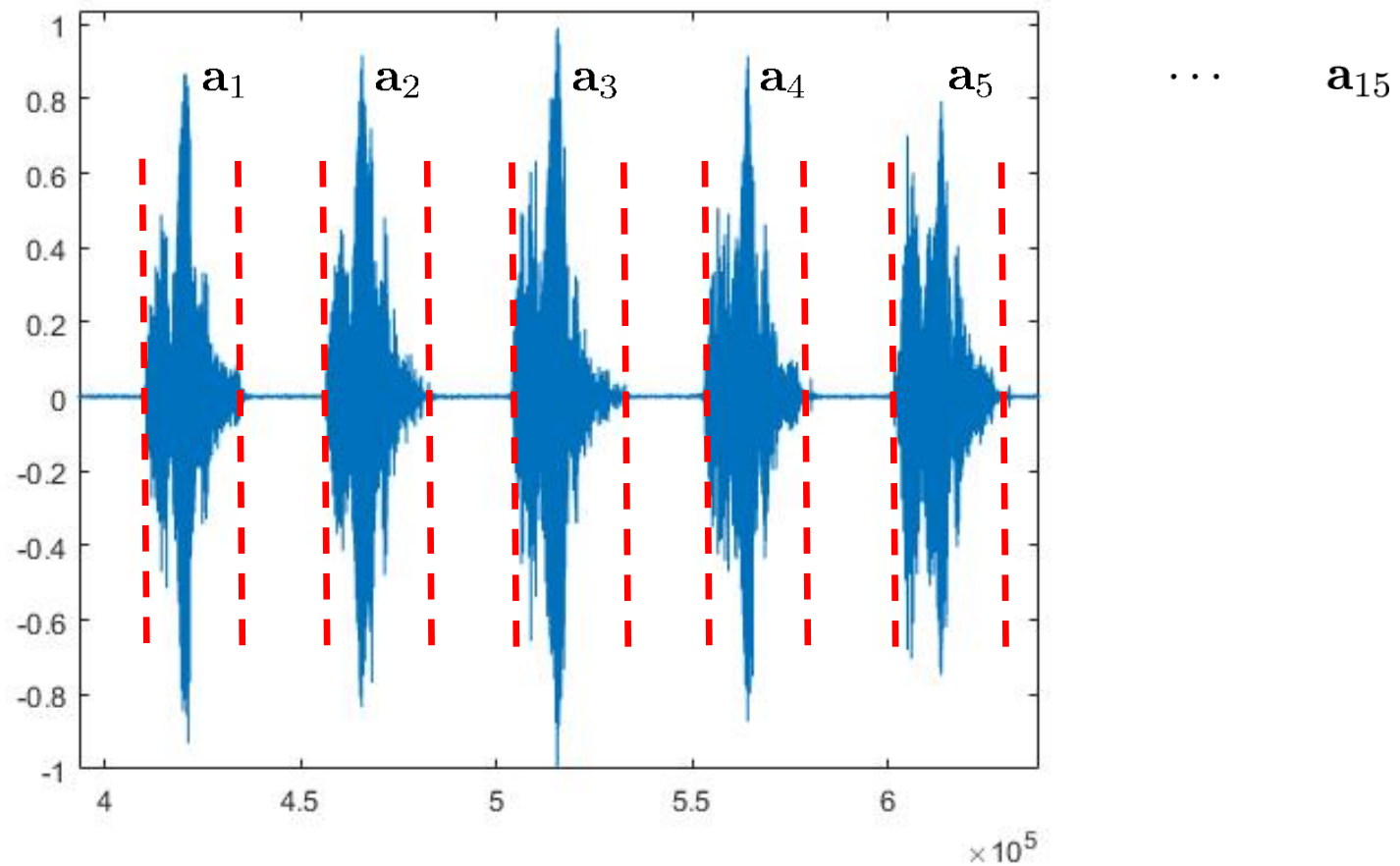
$$\mathbf{a} = \mathbf{R}^{-1}\mathbf{r}$$

calculate LP coefficients for each word

$$\mathbf{r} = \sum_{p+1}^{N} s_n \mathbf{s}_{n-1}$$

$$\mathbf{R} = \sum_{p+1}^{N} \mathbf{s}_{n-1}\mathbf{s}_{n-1}^{T}$$

$$\mathbf{s}_{n-1}^{T} = [s_{n-1} \ s_{n-2} \ \cdots \ s_{n-p}]$$
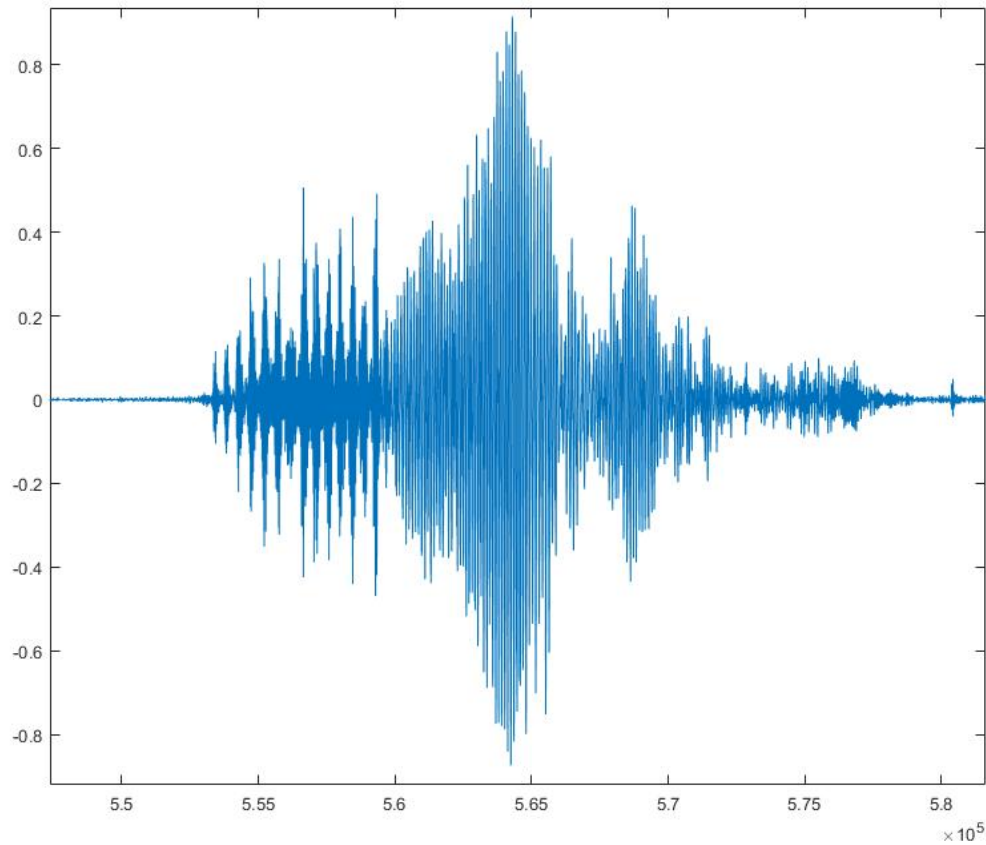
# Training Phase (cont.)



calculate mean and covariance of features:

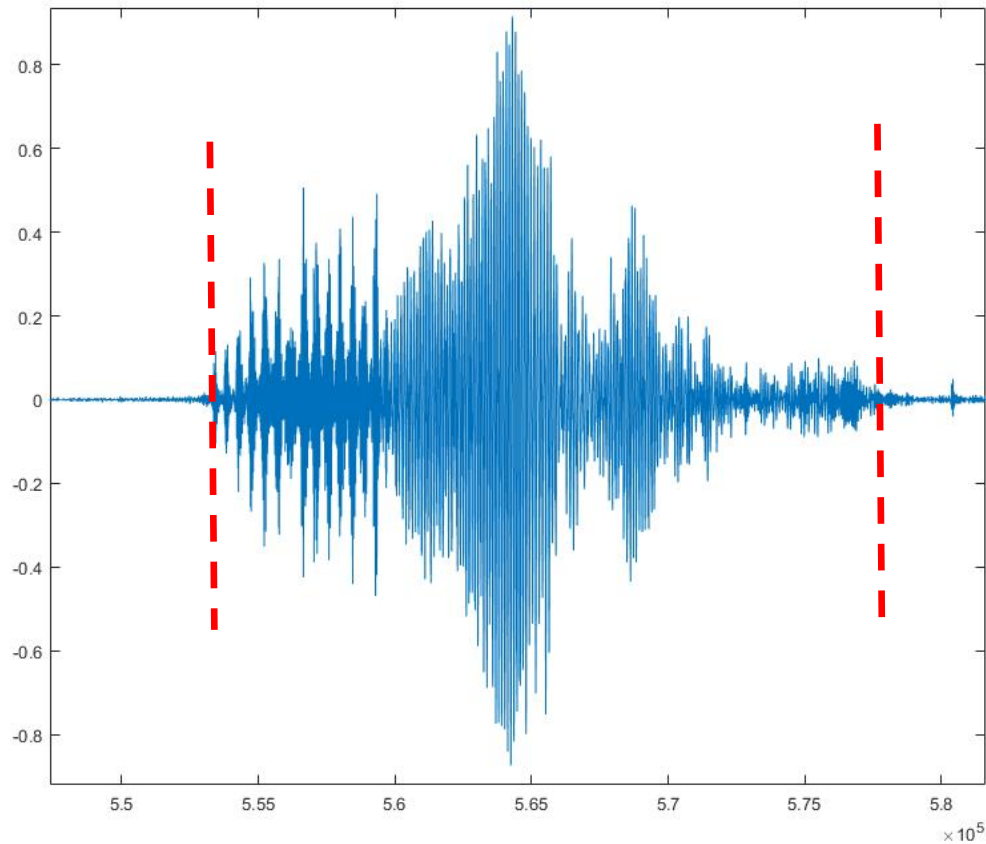$$\boldsymbol{\mu} = \frac{1}{15} \sum_{m=1}^{15} \mathbf{a}_m$$

$$\mathbf{C} = \frac{1}{15} \sum_{m=1}^{15} (\mathbf{a}_m - \boldsymbol{\mu})(\mathbf{a}_m - \boldsymbol{\mu})^T$$

# Testing Phase



have data for 3 users:
$\{\boldsymbol{\mu}_1, \mathbf{C}_1\}, \{\boldsymbol{\mu}_2, \mathbf{C}_2\}, \{\boldsymbol{\mu}_3, \mathbf{C}_3\}$

# Testing Phase



have data for 3 users:
$\{\boldsymbol{\mu}_1, \mathbf{C}_1\}, \{\boldsymbol{\mu}_2, \mathbf{C}_2\}, \{\boldsymbol{\mu}_3, \mathbf{C}_3\}$

calculate LP coefficients for word
not used for training: $\mathbf{x}$

# Testing Phase



have data for 3 users:
$\{\boldsymbol{\mu}_1, \mathbf{C}_1\}, \{\boldsymbol{\mu}_2, \mathbf{C}_2\}, \{\boldsymbol{\mu}_3, \mathbf{C}_3\}$

calculate LP coefficents for word
not used for training:   $\mathbf{x}$

find distance to determine speaker:     $\min_i (\mathbf{x} - \boldsymbol{\mu}_i)^T \mathbf{C}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)$