Feature Mapping

By David Cinciruk

Outline

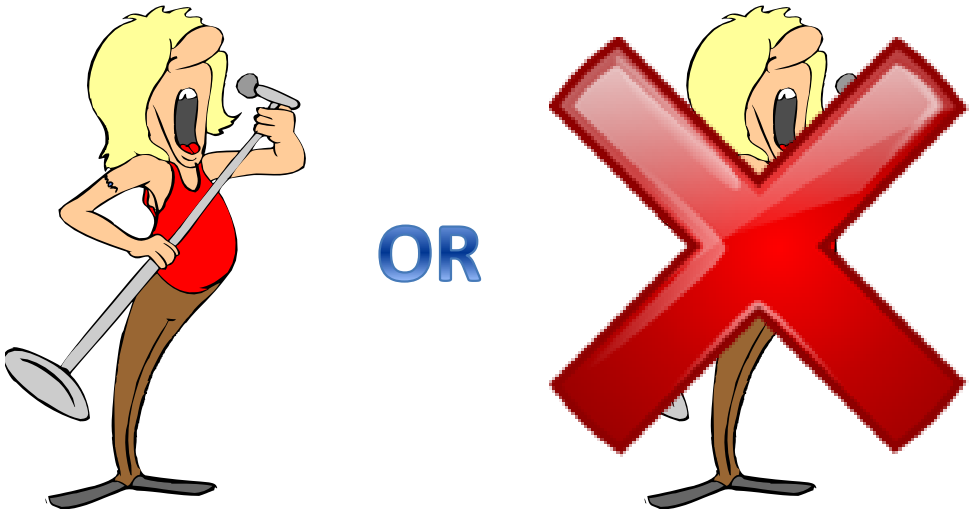
- Introduction
- Overview of GMM/UBM Text Independent Speaker Recognition
- Feature Mapping
- Conclusion

Outline

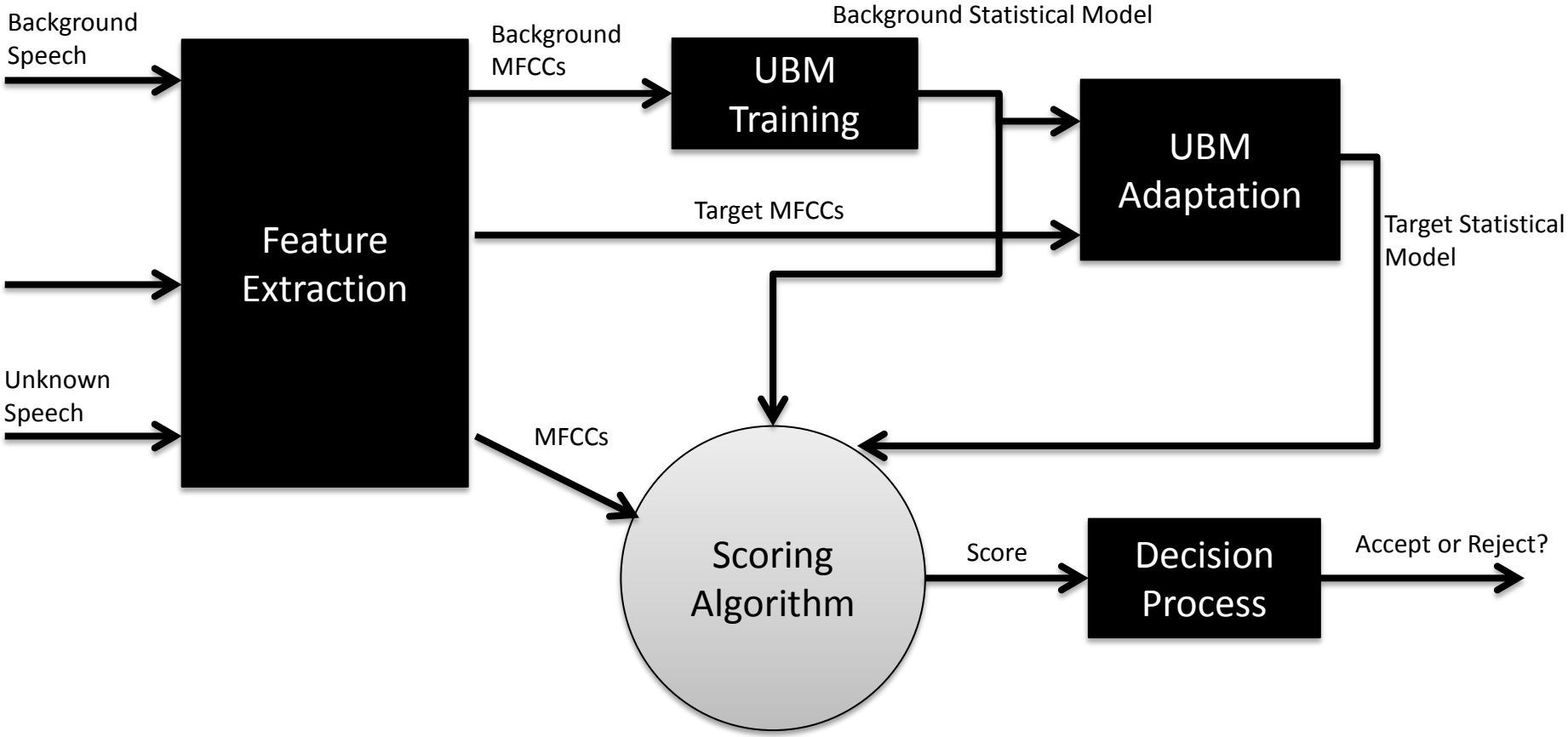
- **Introduction**
- Overview of Speaker Recognition
- Feature Mapping
- Conclusion

What is Speaker Verification

- Process of confirming if an unknown speaker is a certain person
- One way to perform this is by using Gaussian Mixture Models (GMM)



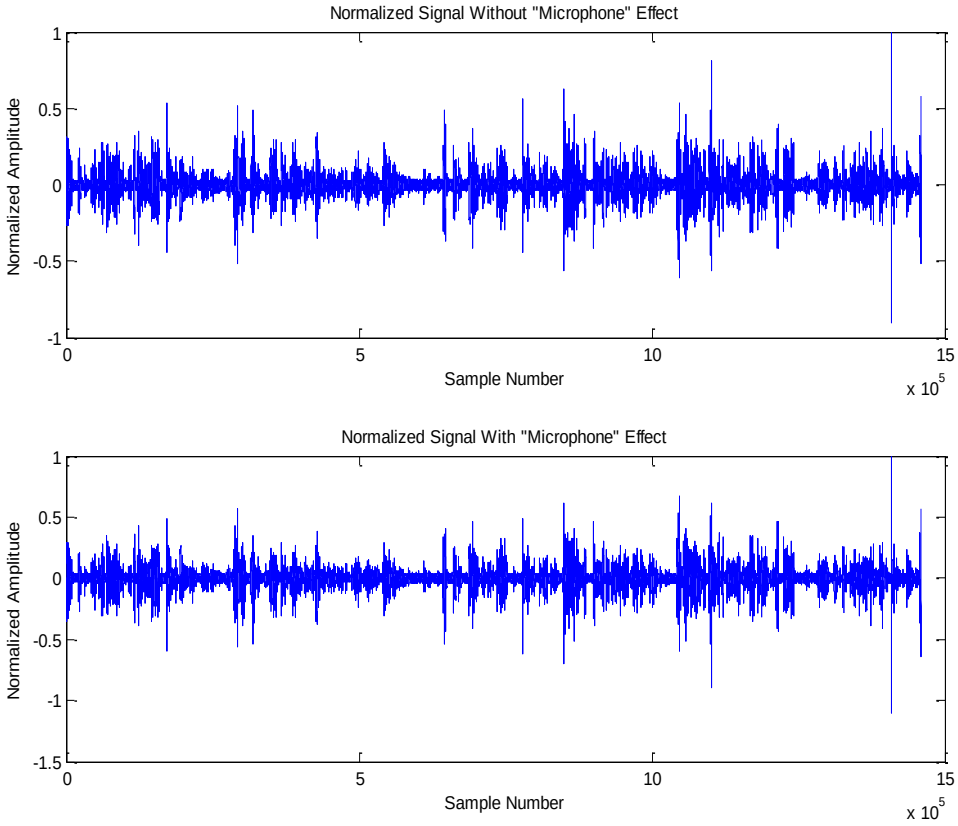
Performing GMM-UBM Based Speaker Verification

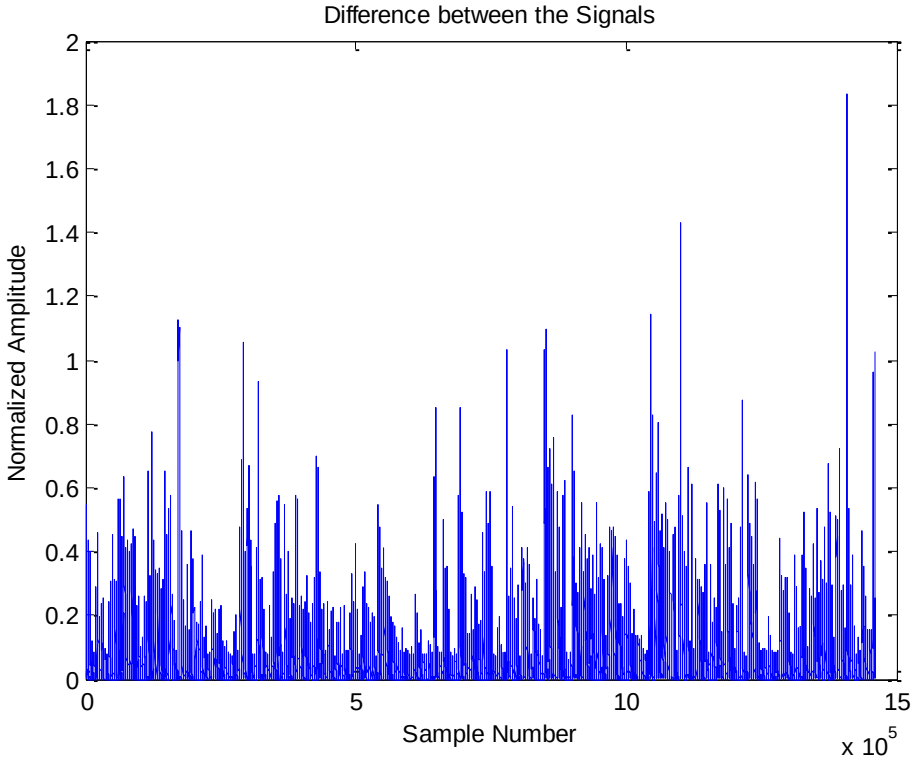


The Effects of Variability

- Speech can be broken down into a part relevant to speaker's identity and a Variability Factor that mixes with Speech.
- Variability Factor: Conditions that speech is performed under

The Effects of Variability



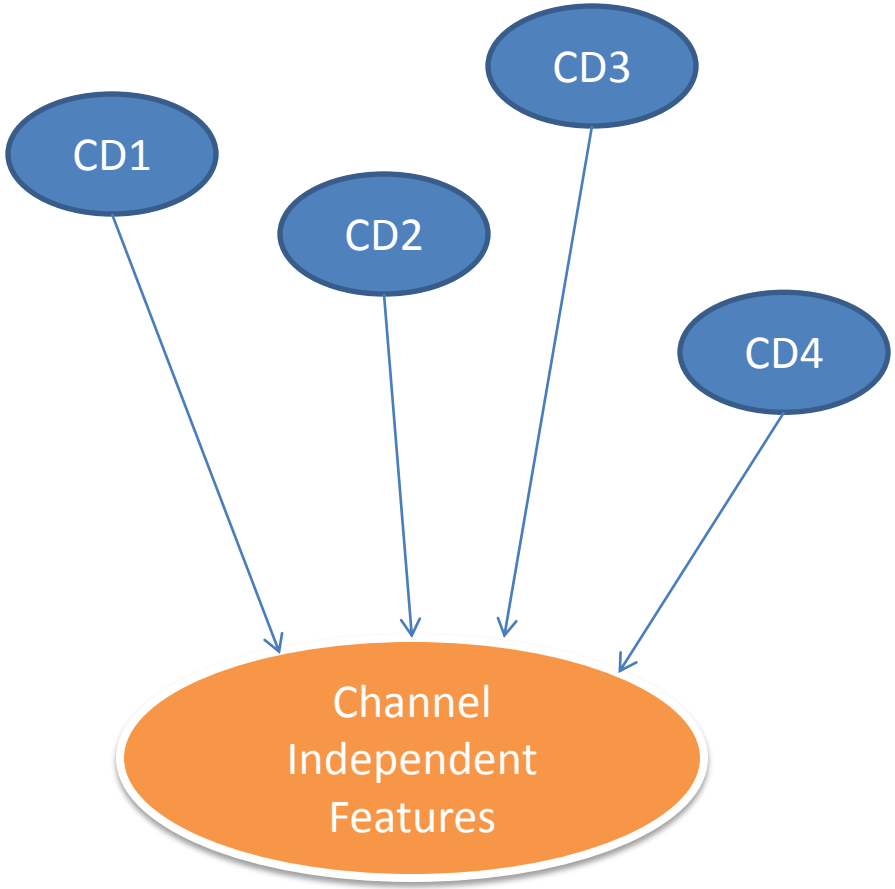


The Effects of Variability

- The effects of these factors mix with the speaker parameters
- Affects how the system “sees” a certain audio file
- Features become a function of both speaker properties and channel properties

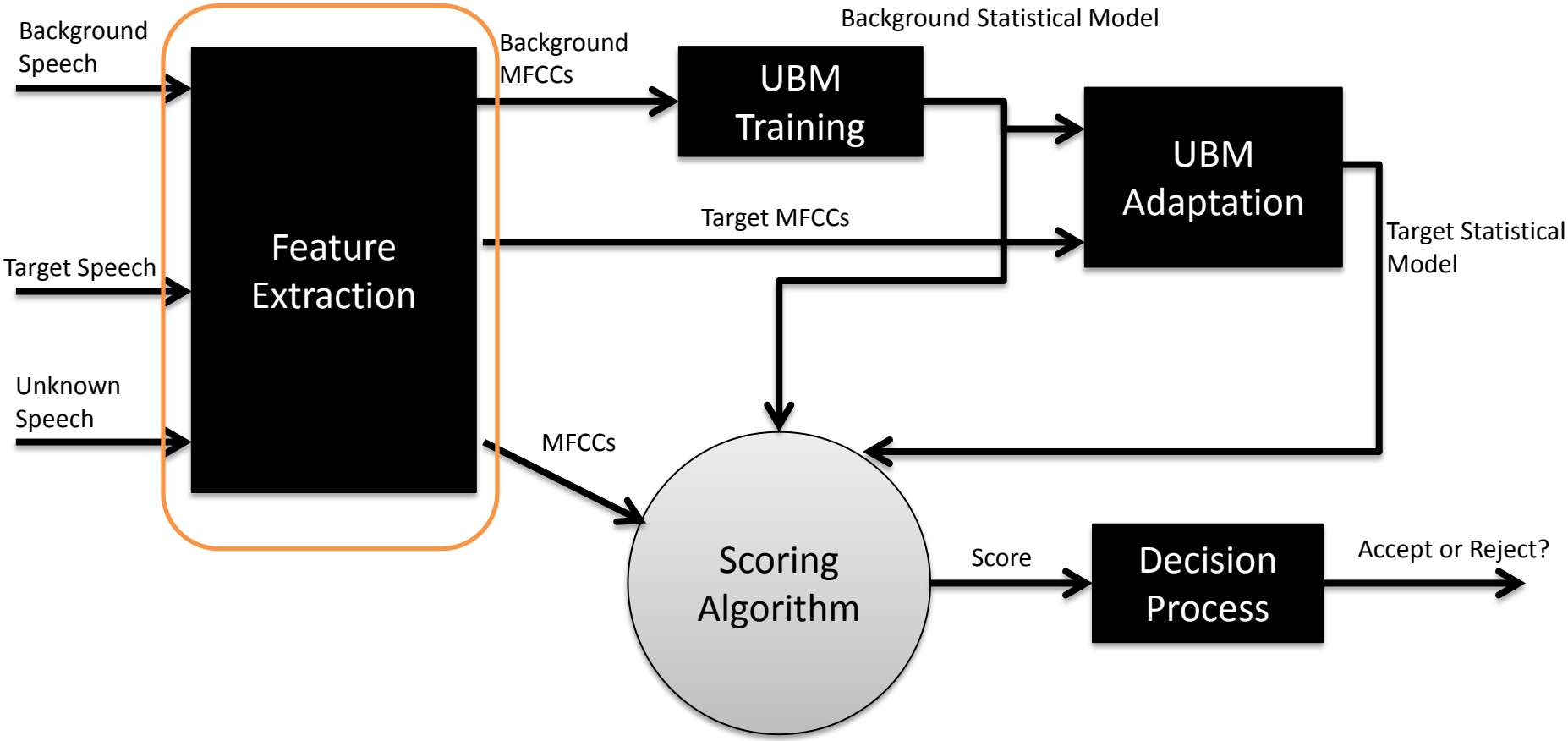
What to Do?

- Want to remove channel properties
- Can map channel dependent features to channel independent features
- Feature Mapping is one such way



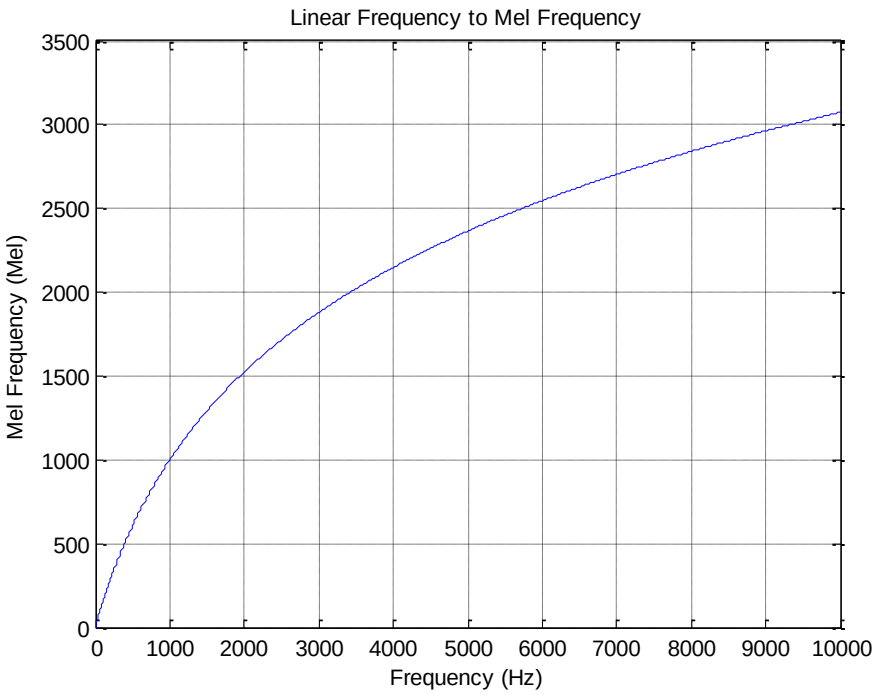
Outline

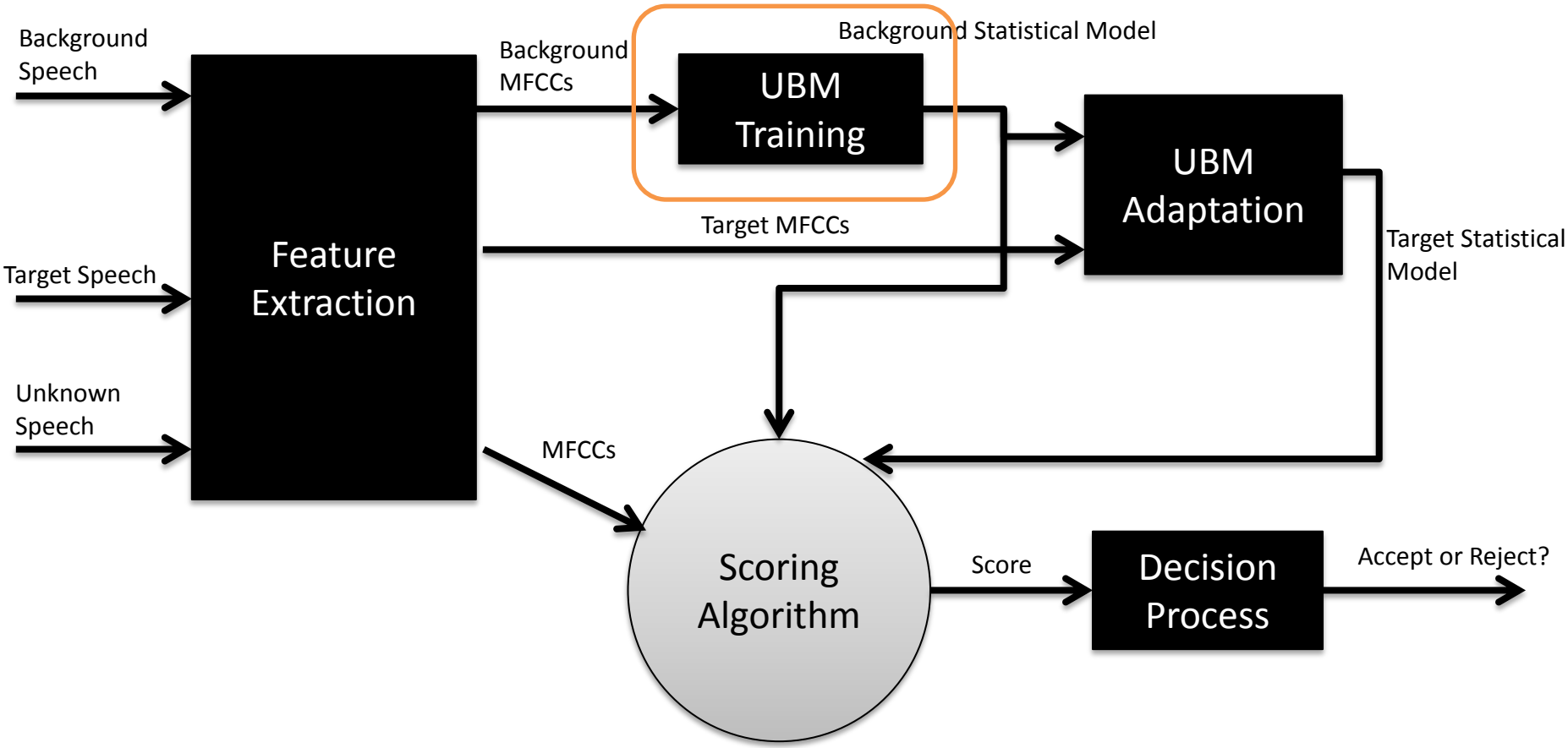
- Introduction
- **Overview of Speaker Recognition**
- Feature Mapping
- Conclusion


Background Speech
Target Speech
Unknown Speech
Feature Extraction
Background MFCCs
Background Statistical Model
UBM Training
UBM Adaptation
Target MFCCs
Target Statistical Model
MFCCs
Scoring Algorithm
Score
Decision Process
Accept or Reject?

Feature Extraction

- Ear response is
  - Logarithmic to Frequency Changes
  - Invariant to Phase Differences
  - Logarithmic to Amplitude
- Need Features that can map that



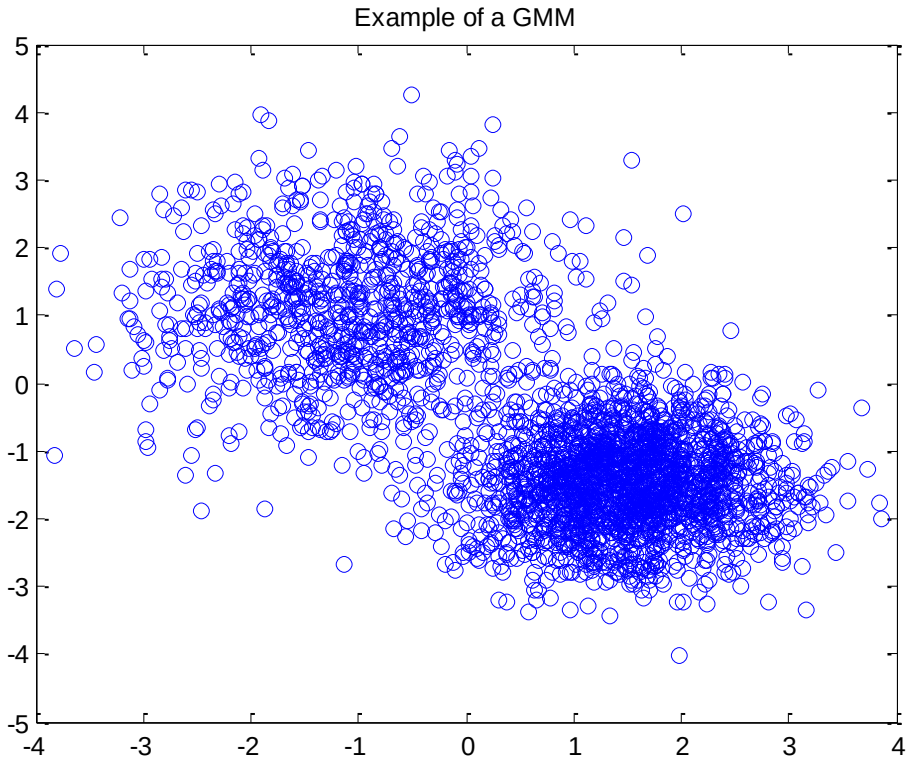

Background Speech
Target Speech
Unknown Speech
Feature Extraction
Background MFCCs
Target MFCCs
MFCCs
UBM Training
Background Statistical Model
UBM Adaptation
Target Statistical Model
Scoring Algorithm
Score
Decision Process
Accept or Reject?

UBM Training

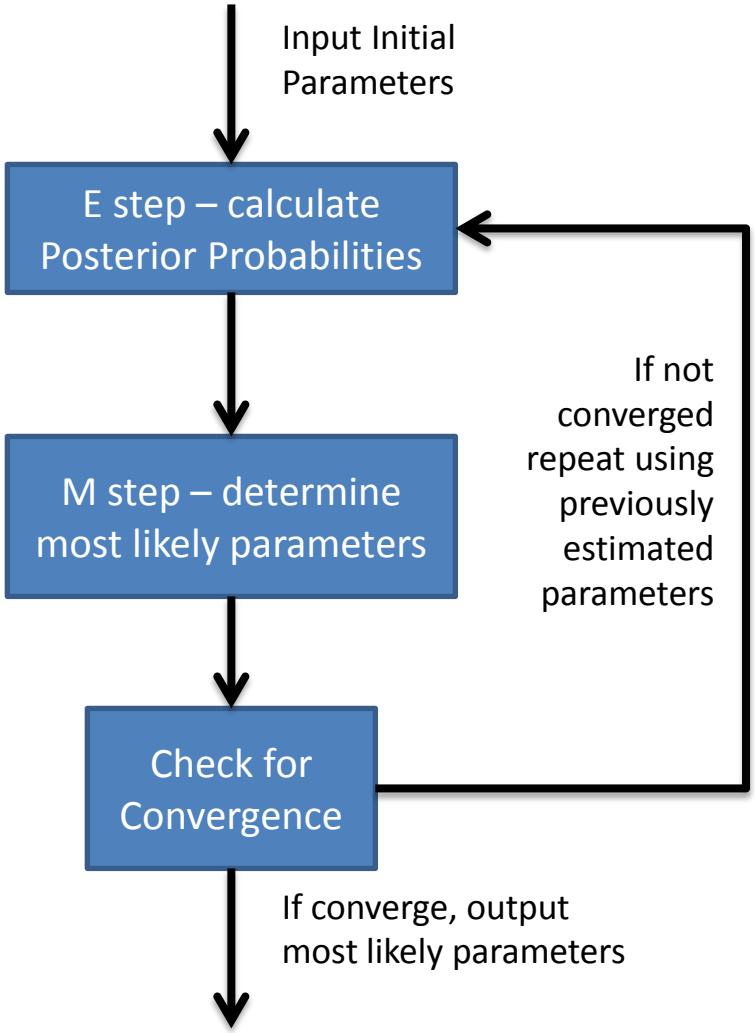
- UBM – Background Model and Null Hypothesis of the Neyman Pearson Detector
- Represents global independent properties of speech
- Set used to train must contain all expected variations in speech
- Feature vector distribution can be estimated by a GMM

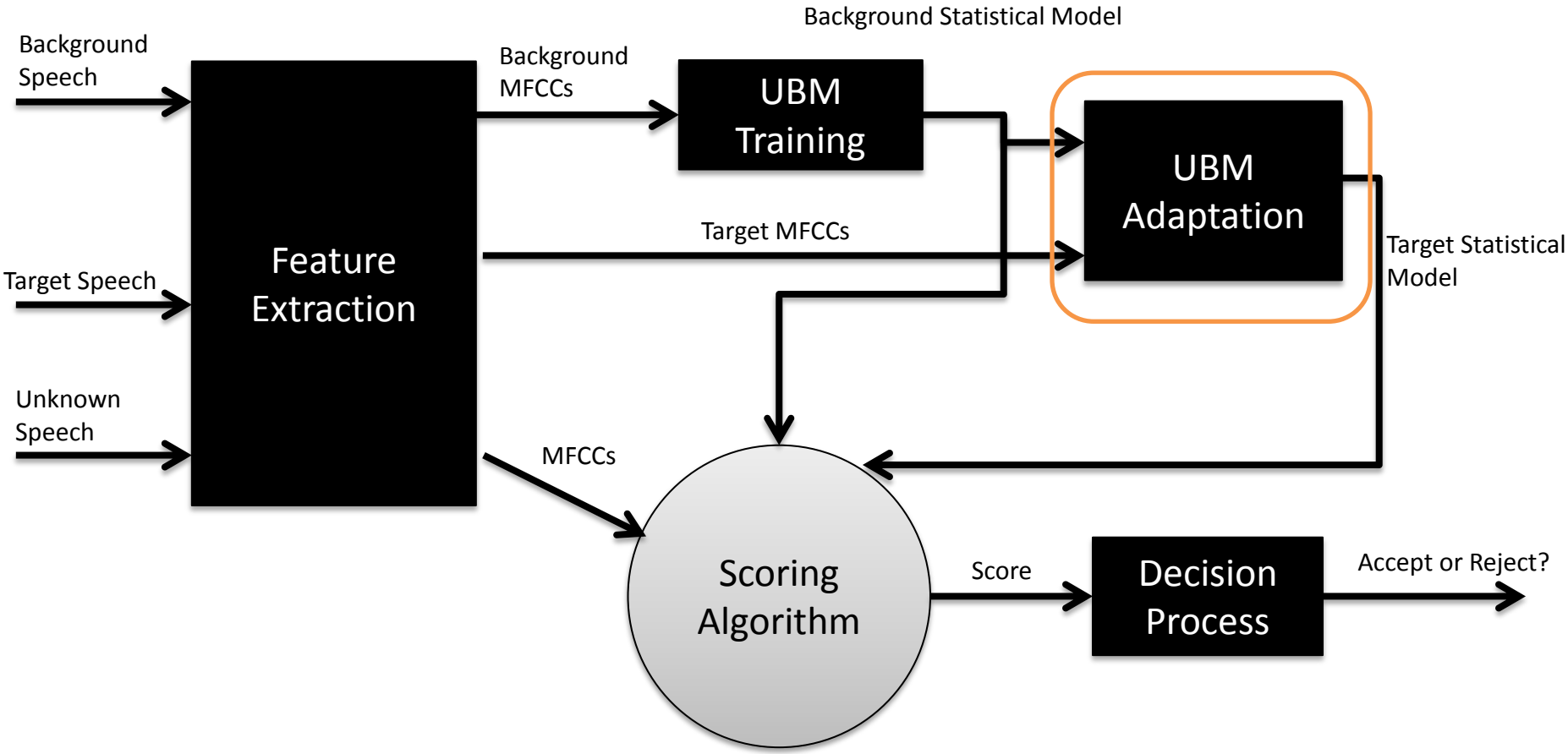
UBM Training – The GMM

- Gaussian Mixture Model
- Sum of multiple weighted Gaussians



UBM Training – EM Algorithm




Background Statistical Model
Background Speech
Background MFCCs
UBM Training
UBM Adaptation
Target Statistical Model
Target MFCCs
Target Speech
Feature Extraction
Unknown Speech
MFCCs
Scoring Algorithm
Score
Decision Process
Accept or Reject?

Model Adaptation

- One Speaker: 10 seconds – 80 minutes
- Too Little Data for GMM Training
- Adapt only the Mean Vectors

Model Adaptation – The Algorithm

- Need to first calculate the posterior probabilities for each mixture given a point

$$\gamma_i(\mathbf{x}_t) = \frac{w_i^{(j)} g(\mathbf{x}_t | \boldsymbol{\mu}_i^{(j)}, \Sigma_i^{(j)})}{\sum_{s=1}^{M} w_s^{(j)} g(\mathbf{x}_t | \boldsymbol{\mu}_s^{(j)}, \Sigma_s^{(j)})}$$

Model Adaptation – The Algorithm

- Calculate sufficient statistics for mean and variance

$$n_i = \sum_{t=1}^{T} \gamma_i(\mathbf{x}_t)$$

$$E_i(\mathbf{x}_t) = \frac{1}{n_i} \sum_{t=1}^{T} \gamma_i(\mathbf{x}_t)\, \mathbf{x}_t$$

$$E_i(\mathbf{x}_t^2) = \frac{1}{n_i} \sum_{t=1}^{T} \gamma_i(\mathbf{x}_t)\, \mathbf{x}_t^2$$

Model Adaptation – The Algorithm

- New estimate of the means, covariances, and weights

$$w_i' = \left[\frac{\alpha_i n_i}{T} + (1-\alpha_i) w_i\right] g$$

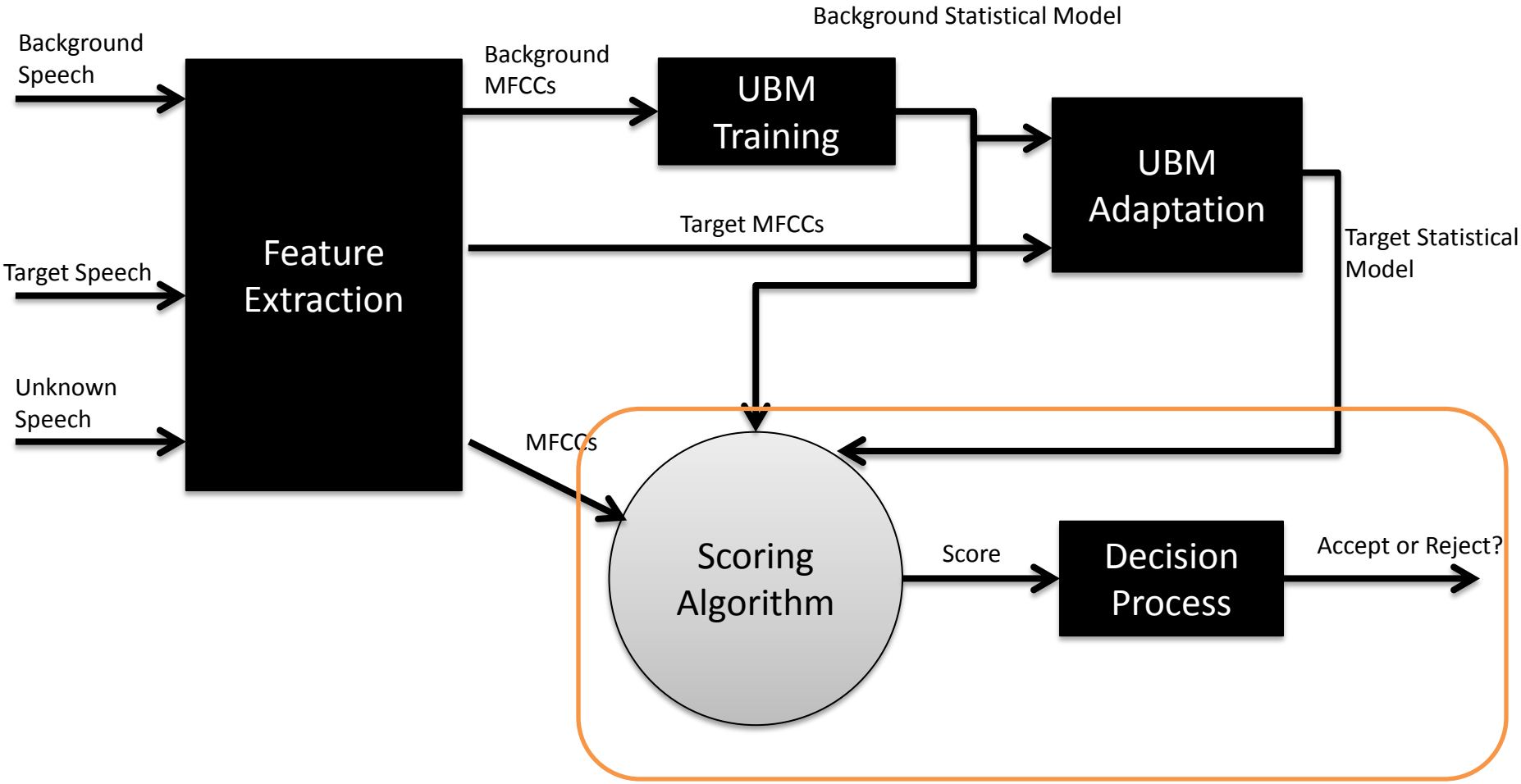
$$\mu_i' = \alpha_i E_i(\mathbf{x}) + (1-\alpha_i)\mu_i$$

$$\Sigma_i = \alpha_i E_i(\mathbf{x}^2) + (1-\alpha_i)\left(\Sigma_i + \mu_i^2\right) - \mu_i'^2$$

- Where g is a scale factor so the weights sum to 1 and α_i is

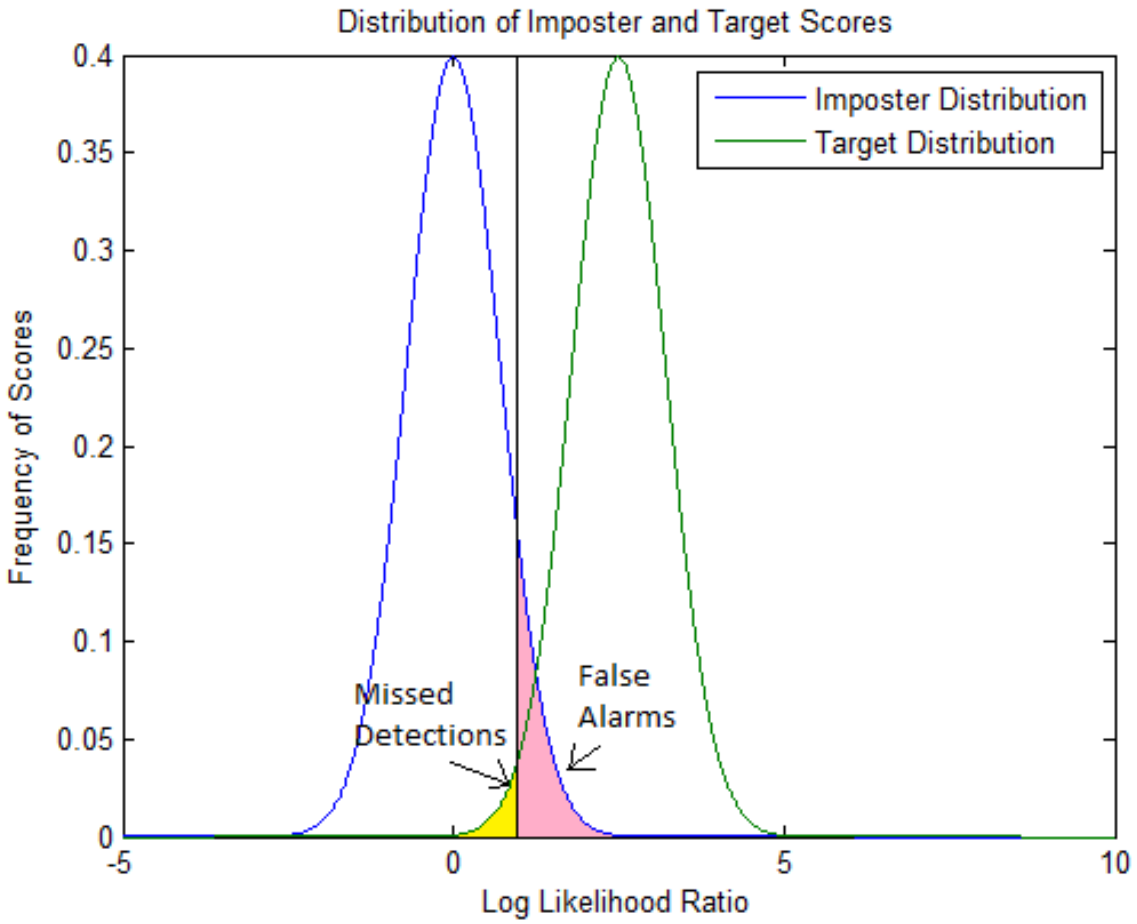
$$\alpha_i = \frac{n_i}{n_i + r}$$

- Where r is a fixed relevance parameter


Background Statistical Model
Background Speech
Target Speech
Unknown Speech
Feature Extraction
Background MFCCs
UBM Training
Target MFCCs
UBM Adaptation
Target Statistical Model
MFCCs
Scoring Algorithm
Score
Decision Process
Accept or Reject?

Scoring

- Log Likelihood test to Determine if someone is target or not

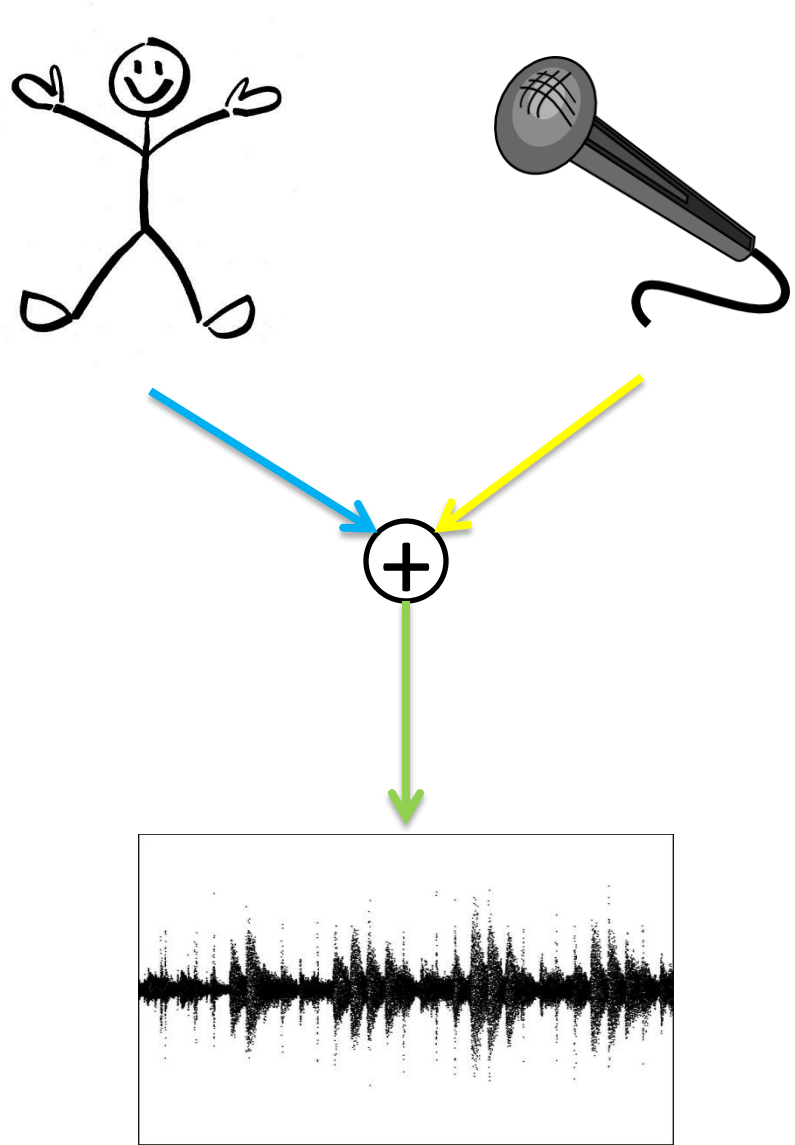


Outline

- Introduction
- Overview of Speaker Recognition
- **Feature Mapping**
- Conclusion

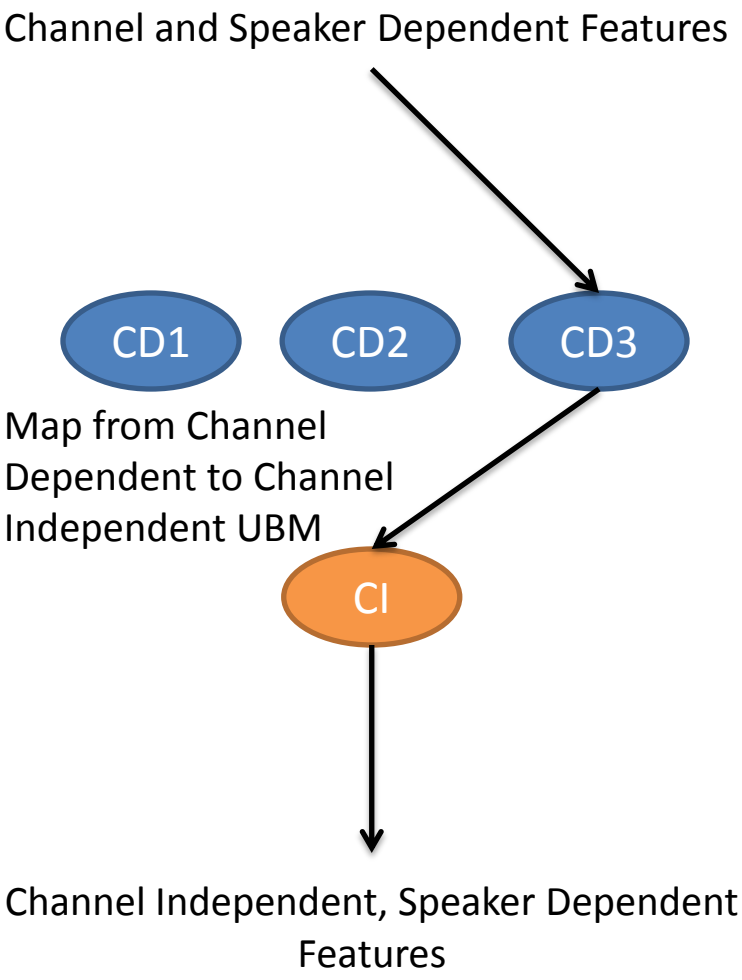
Feature Mapping

- Features = functions of speaker and channel
- Want to remove channel portion



Feature Mapping

- Maps the features from a speaker and channel dependent set to one dependent on just speaker
- Does this by mapping features from a channel dependent UBM to a channel independent UBM

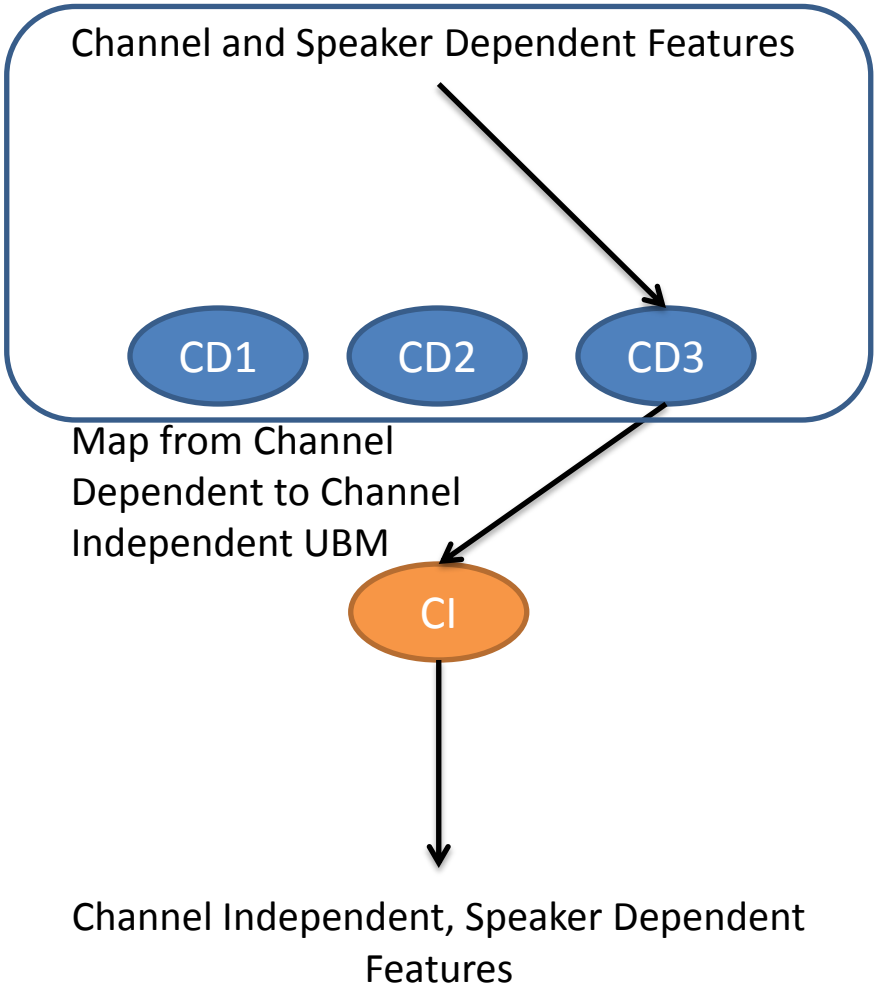


Channel Dependent UBM

- Formed by Model Adaptation
  - Adapt Means, Covariances, and Weights
- Trained using subset of data known to come from specific channel
- Must have one for all combinations of channel types

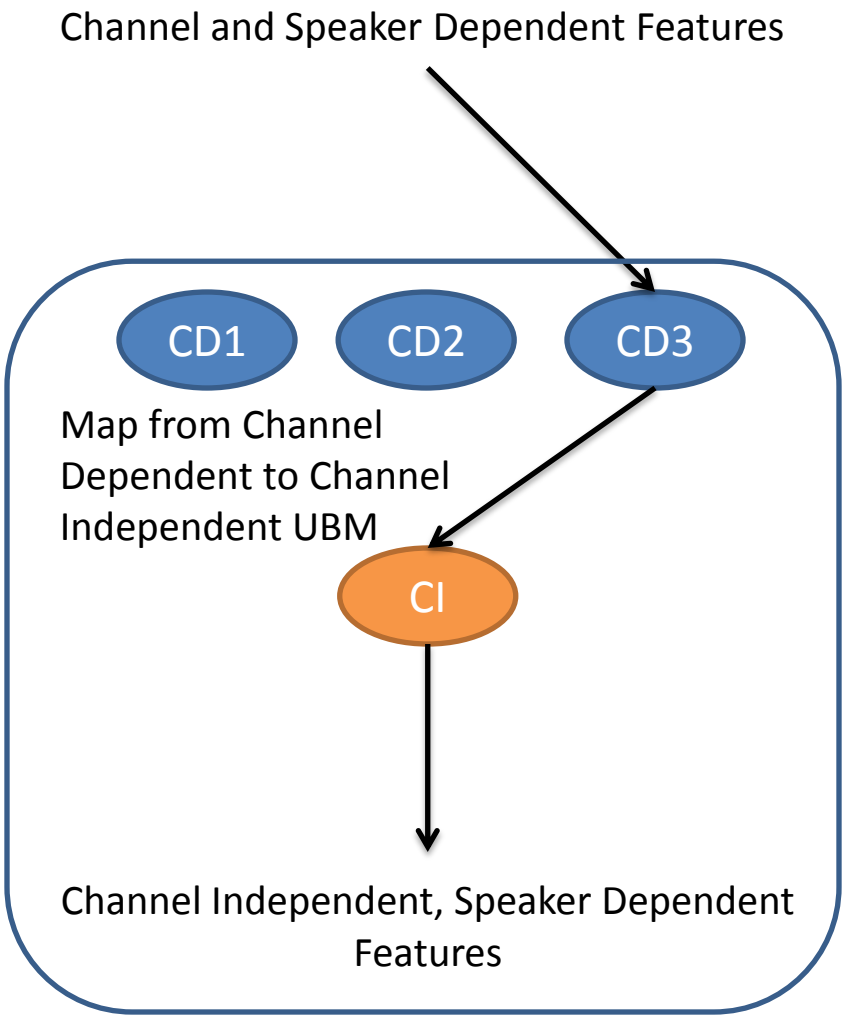
Mapping Features

- For each utterance
  - Determine most likely Channel Dependent model
- If summed channel speech is used, can do this for short intervals of time instead



Mapping Features

- For each feature vector
  - Find the most likely mixture component
    - $i = \underset{1 \le j \le M}{\operatorname{argmax}} w_i p_i(x)$
    - Where $p_i(x)$ is the probability of the feature under mixture i
  - Map channel dependent feature to channel independent feature
    - $y = \left(x - \mu_i^{CD}\right) \frac{\sigma_i^{CI}}{\sigma_i^{CD}} + \mu_i^{CI}$



After Feature Mapping

- New Features have been mapped to channel independent model
- From there, take channel independent features and perform standard speaker verification

Outline

- Introduction
- Overview of Speaker Recognition
- Feature Mapping
- **Conclusion**

Conclusion

- Features are functions of both speaker and channel
- Speaker Verification systems suffer when different channel effects are present.
- Feature Mapping works to reduce channel effect by mapping from a channel dependent model to channel independent model