

# Comparison of JFA, FDIC and FNAP

Mengke HU

ASPITRG Group, ECE Department  
Drexel University

*mengke.hu@gmail.com*

January 10, 2013

# Outline

## 1 Review Baseline system and session variability problem

- Baseline System for Speaker verification
- Speaker Verif. in the sense of Session Variability

## 2 Model Domain: Joint Factor Analysis for both inter-session and inter-speaker

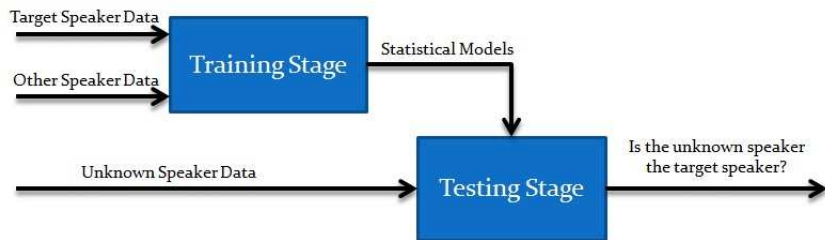
- JFA model review
- 3 submodels in JFA
- 2 methods of training in JFA

## 3 Feature Domain: FDIC and f-NAP for diminishing inter-session var.

- Inter-session Compensation in feature domain (FDIC)
- Nuisance Attribute Projection in feature domain (f-NAP)

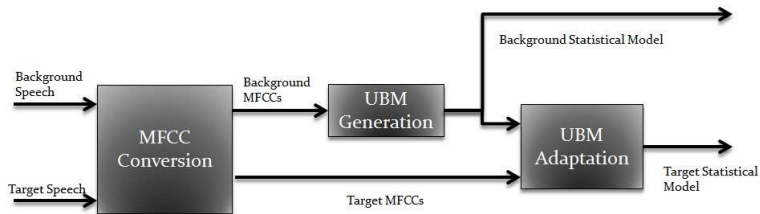
# Baseline System

## Overview



# Baseline System

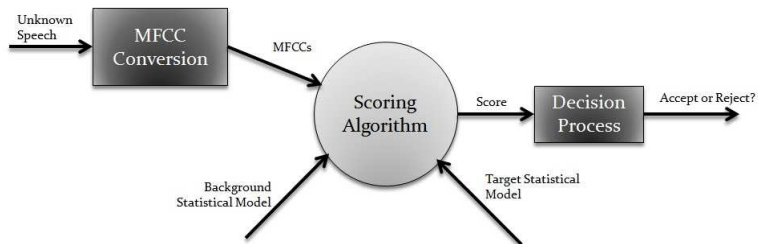
## Training



- UBM generation on ancillary **training data set**
- UBM adaptation for target model on enrollment set in the **evaluation data set**

# Baseline System

## Testing



- This step is on testing set in the **evaluation data set**

# Outline

## 1 Review Baseline system and session variability problem

- Baseline System for Speaker verification
- Speaker Verif. in the sense of Session Variability

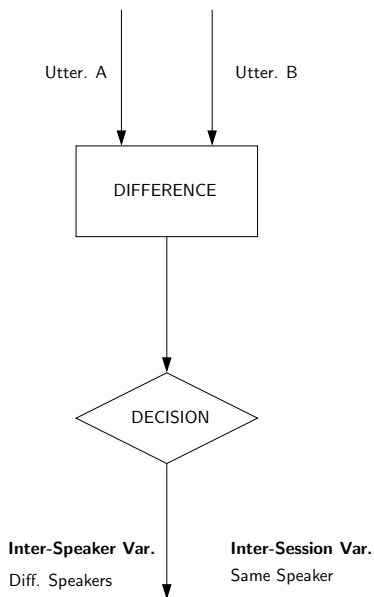
## 2 Model Domain: Joint Factor Analysis for both inter-session and inter-speaker

- JFA model review
- 3 submodels in JFA
- 2 methods of training in JFA

## 3 Feature Domain: FDIC and f-NAP for diminishing inter-session var.

- Inter-session Compensation in feature domain (FDIC)
- Nuisance Attribute Projection in feature domain (f-NAP)

# Speaker Verif. in the sense of Session Variability



# Inter Session Var. and Nuisance Compensation

## 1 Inter Session Var.

- ▶ Channel effects: different recording sessions
- ▶ intra-speaker var. : health, emotional state, etc.

## 2 Nuisance Compensation

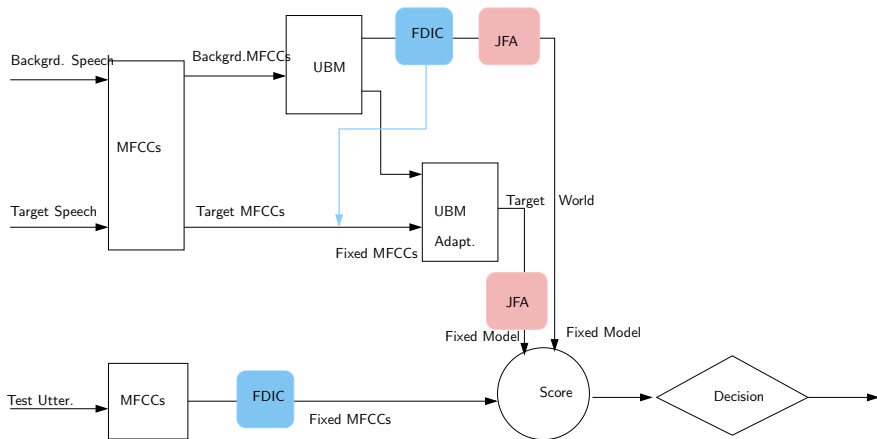
- ▶ Model Domain technique: fix Gaussian mixture models (means and covar. matrices)  
Eigen-channel in JFA
- ▶ Feature Domain techniques: fix features  
FDIC, f-NAP



## Summary: session var. in speaker verification

- 1 Diminish inter-session var. (nuisance for speaker verification)
- 2 Characterize inter-speaker var. (useful for speaker verification)

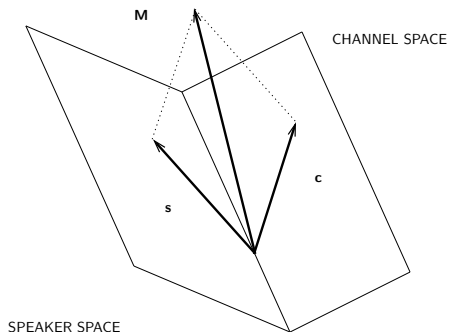
# JFA vs FDIC overview



# Outline

- 1 Review Baseline system and session variability problem
  - Baseline System for Speaker verification
  - Speaker Verif. in the sense of Session Variability
- 2 Model Domain: Joint Factor Analysis for both inter-session and inter-speaker
  - **JFA model review**
  - 3 submodels in JFA
  - 2 methods of training in JFA
- 3 Feature Domain: FDIC and f-NAP for diminishing inter-session var.
  - Inter-session Compensation in feature domain (FDIC)
  - Nuisance Attribute Projection in feature domain (f-NAP)

## JFA model review



$$\mathbf{M} = \mathbf{s} + \mathbf{c}$$

- $\mathbf{s}$ : speaker-dependent for inter-speaker var.
- $\mathbf{c}$ : channel-dependent for inter-session var.
- $\mathbf{M}$ : utterance-dependent (speaker-and-channel-dependent)

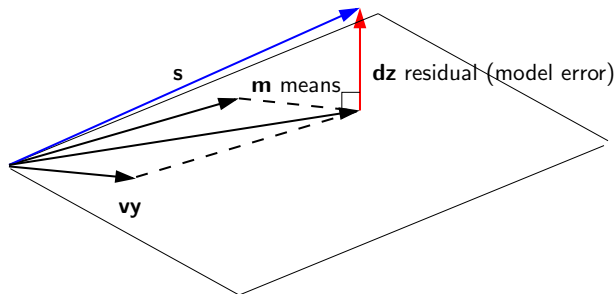
# JFA model review

General form:

$$\mathbf{M} = \mathbf{s} + \mathbf{c}$$

Speaker Space:

$$\mathbf{s} = \mathbf{m} + \mathbf{v}_y + \mathbf{d}_z$$



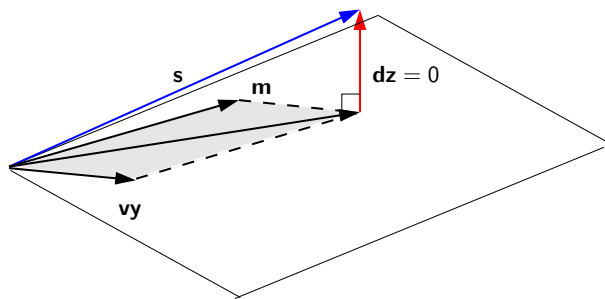
Channel Space:

$$\mathbf{c} = \mathbf{u}\mathbf{x}$$

# Outline

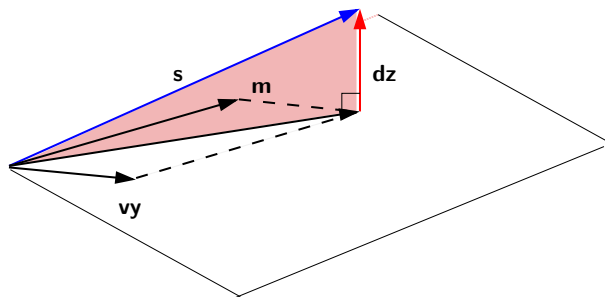
- 1 Review Baseline system and session variability problem
  - Baseline System for Speaker verification
  - Speaker Verif. in the sense of Session Variability
- 2 Model Domain: Joint Factor Analysis for both inter-session and inter-speaker
  - JFA model review
  - **3 submodels in JFA**
  - 2 methods of training in JFA
- 3 Feature Domain: FDIC and f-NAP for diminishing inter-session var.
  - Inter-session Compensation in feature domain (FDIC)
  - Nuisance Attribute Projection in feature domain (f-NAP)

# Eigenvoice Model on Speaker Space



$$s' = m + vy$$

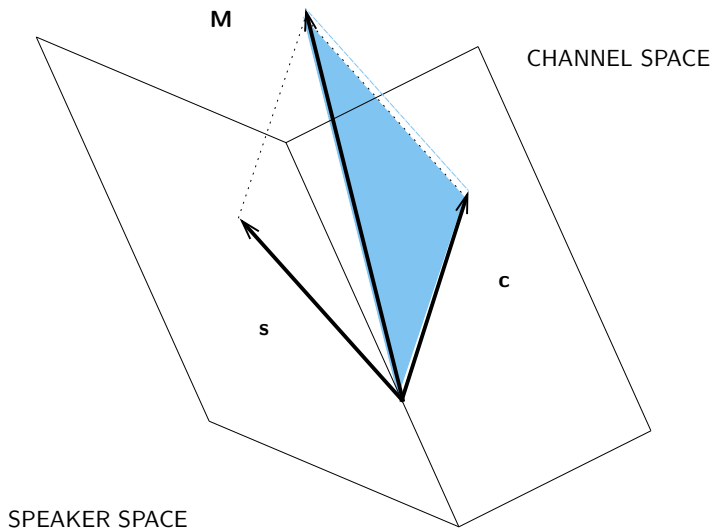
# Diagonal Model



$$s'' = m + dz$$



# Eigen-channel Model

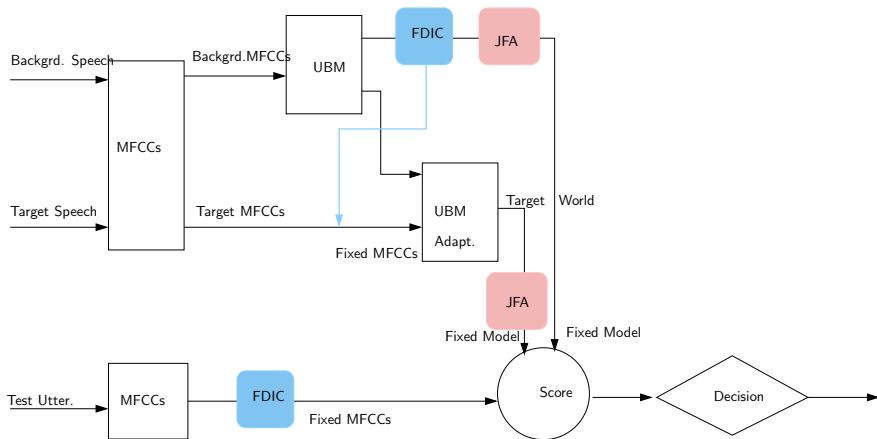


$$\mathbf{M} = \mathbf{s} + \mathbf{u}\mathbf{x}$$

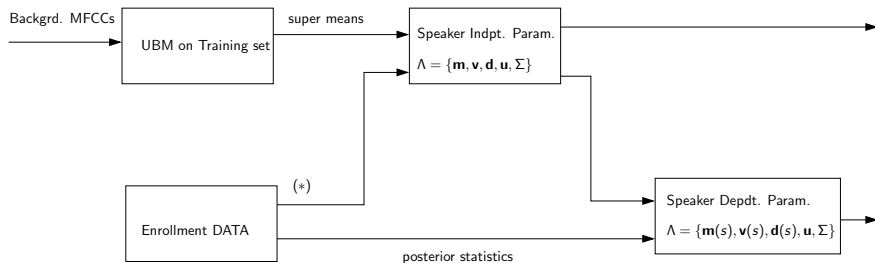
# Outline

- 1 Review Baseline system and session variability problem
  - Baseline System for Speaker verification
  - Speaker Verif. in the sense of Session Variability
- 2 Model Domain: Joint Factor Analysis for both inter-session and inter-speaker**
  - JFA model review
  - 3 submodels in JFA
  - 2 methods of training in JFA**
- 3 Feature Domain: FDIC and f-NAP for diminishing inter-session var.
  - Inter-session Compensation in feature domain (FDIC)
  - Nuisance Attribute Projection in feature domain (f-NAP)

# JFA vs FDIC overview



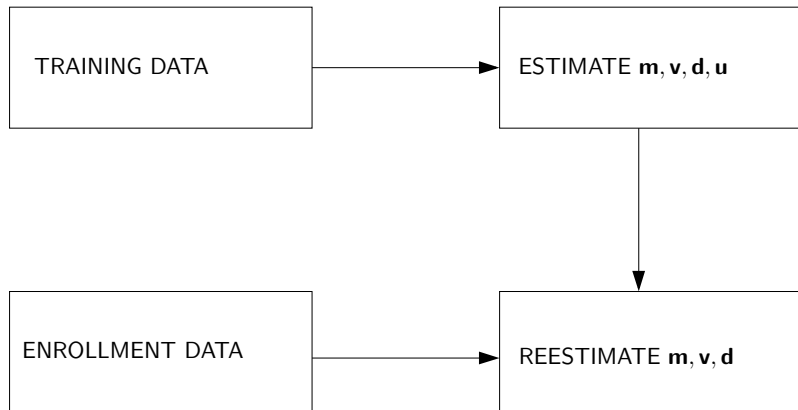
# JFA training stage



Remark: (\*) is used to eliminate the dis-match between Training and Evaluation set. Here minimum divergent estimation is applied.

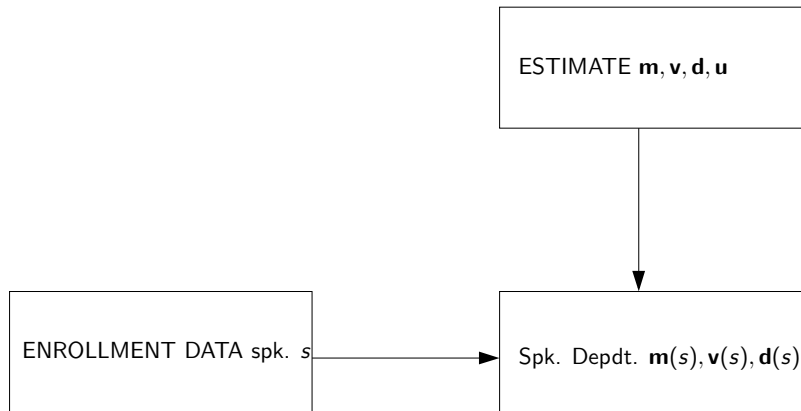
# JFA Joint Estimation

## Speaker Independent Parameters



# JFA Joint Estimation

## Speaker Dependent Parameters

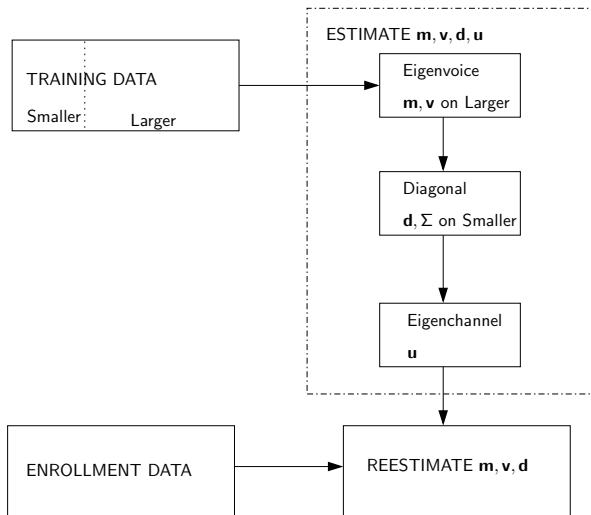


# Drawbacks of JFA Joint estimation

- 1 **d** plays very little role for speaker verification
- 2 **d** cannot be neglected since it is useful to catch differences between different speakers.
- 3 How to exaggerate the influence imposed by **d**?

# JFA Decoupling Estimation

Speaker Independent

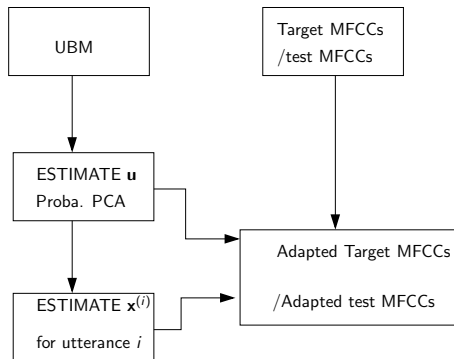




# Outline

- 1 Review Baseline system and session variability problem
  - Baseline System for Speaker verification
  - Speaker Verif. in the sense of Session Variability
- 2 Model Domain: Joint Factor Analysis for both inter-session and inter-speaker
  - JFA model review
  - 3 submodels in JFA
  - 2 methods of training in JFA
- 3 Feature Domain: FDIC and f-NAP for diminishing inter-session var.
  - Inter-session Compensation in feature domain (FDIC)
  - Nuisance Attribute Projection in feature domain (f-NAP)

$$\mathbf{M} = \mathbf{s} + \mathbf{u}\mathbf{x}$$



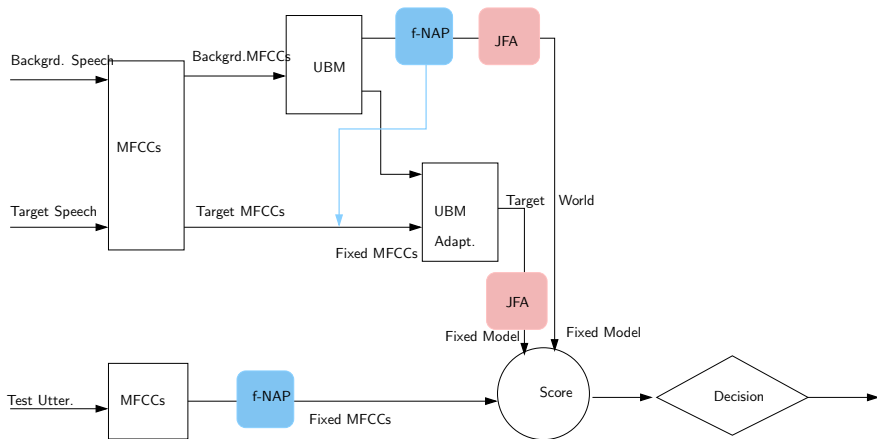
# FDIC vs JFA

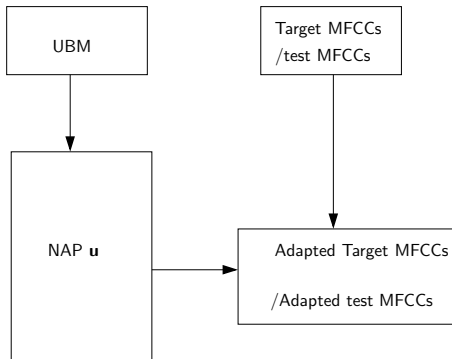
- 1 FDIC uses eigenchannel model in JFA model
- 2 FDIC adopt PPCA to estimate matrix  $\mathbf{u}$
- 3 FDIC calculates the channel factor  $\mathbf{x}$  for each utterance component-wise, while JFA only cares about the posterior mean  $\mathbb{E}[\mathbf{x}(s)]$

# Outline

- 1 Review Baseline system and session variability problem
  - Baseline System for Speaker verification
  - Speaker Verif. in the sense of Session Variability
- 2 Model Domain: Joint Factor Analysis for both inter-session and inter-speaker
  - JFA model review
  - 3 submodels in JFA
  - 2 methods of training in JFA
- 3 Feature Domain: FDIC and f-NAP for diminishing inter-session var.
  - Inter-session Compensation in feature domain (FDIC)
  - Nuisance Attribute Projection in feature domain (f-NAP)

# JFA vs FDIC overview





## FDIC vs f-NAP

- FDIC and f-NAP are the same in feature adaptation step
- The only difference is in the model adaptation step ( $\mathbf{u}$ ).
- f-NAP estimate the nuisance attribute needed to be removed, while FDIC estimate the  $\mathbf{u}$  directly

# NAP for GMM-SVM speaker verification

- 1 SVM is a 2-class classifier constructed from sums of a kernel function  $K(., .)$
- 2 Use KL divergence of the GMMs of 2 utterance to construct kernel
- 3 Estimate the nuisance attribute, which is the direction being removed
- 4 Use PCA to obtain the matrix  $\mathbf{u}$