

# Multimodal Fusion using Dynamic Hybrid Models

Mohamed Amer, Behjat Siddiquie, Saad Khan,  
Ajay Divakaran and Harpreet Sawhney



**SRI International**

# Problem

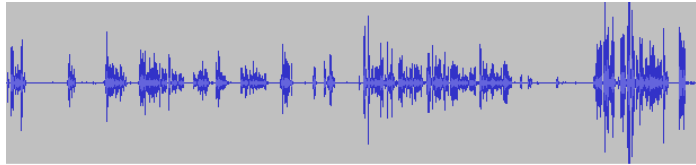
## Event Detection in Multimodal Sequences

- Human Behavioral Analysis

- Multimodal data

  - » Audio

  - » Video



Audio



Video

- Goal:

  - » Classify a set of multimodal events

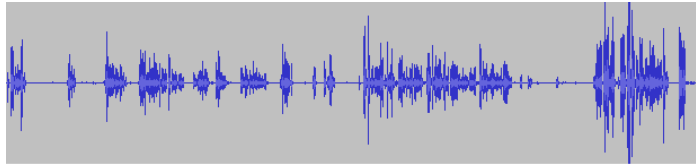
# Problem

## Event Detection in Multimodal Sequences

- Human Behavioral Analysis
- Multimodal data

- » Audio

- » Video



Audio



Video

- **Challenges:**

- » Temporal Data from Multiple Heterogeneous Modalities
- » Multiple Temporal Scales
- » Missing Data

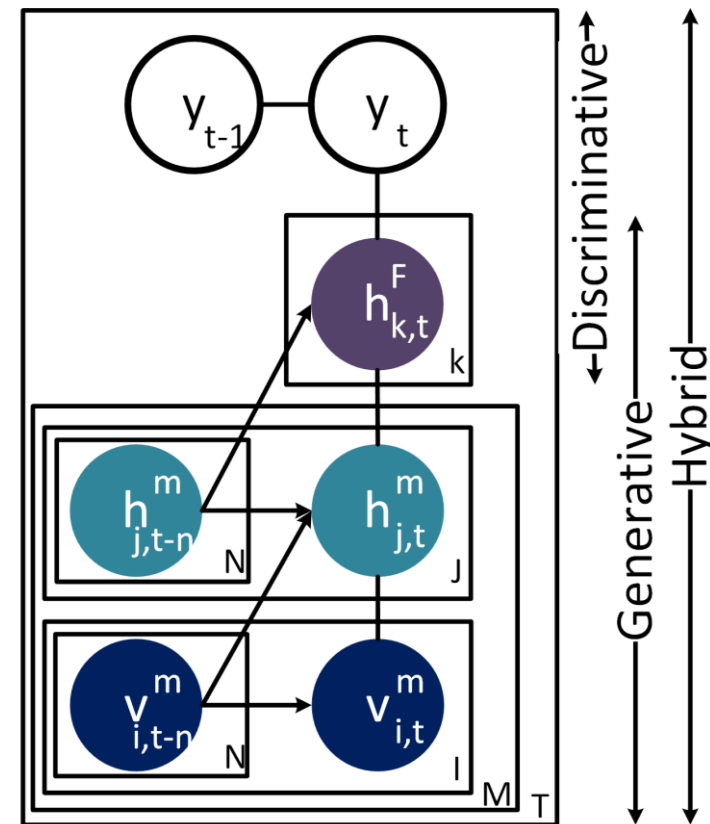
# Approach

We propose a Staged Dynamic Hybrid model

- **Generative Component:** Multimodal Conditional Restricted Boltzmann Machines (MMCRBMs)
- **Discriminative Component:** Conditional Random Field (CRF)

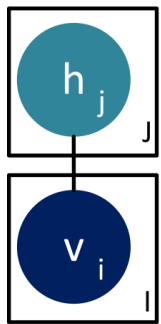
# Approach

- **(Hybrid)** Allows for exploiting the generative model's expressiveness and the discriminative model's classification power.
- **(Staged)** Allows for training each model separately, where the discriminative model trained on representations learned by the generative model.
- **(Dynamic)** Modeling the temporal content of time varying data is important.

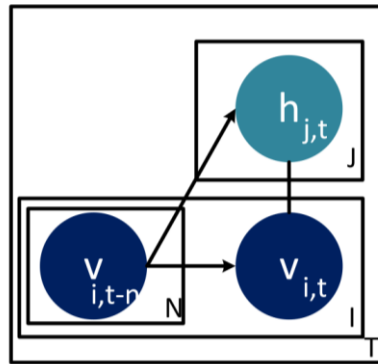


# Model: Generative Component

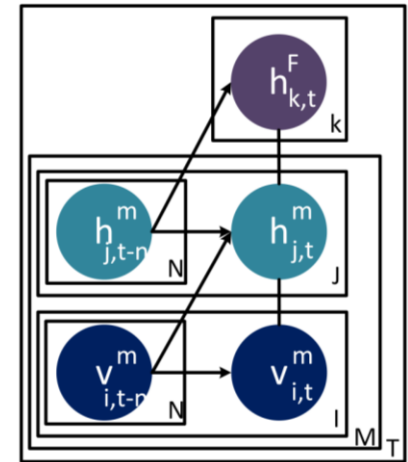
## Multimodal Conditional Restricted Boltzmann Machines (MMCRBMs)



Restricted Boltzmann Machine (RBM)



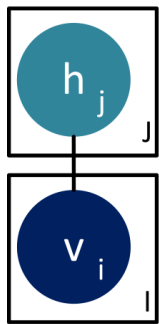
Conditional Restricted Boltzmann Machine (CRBM)



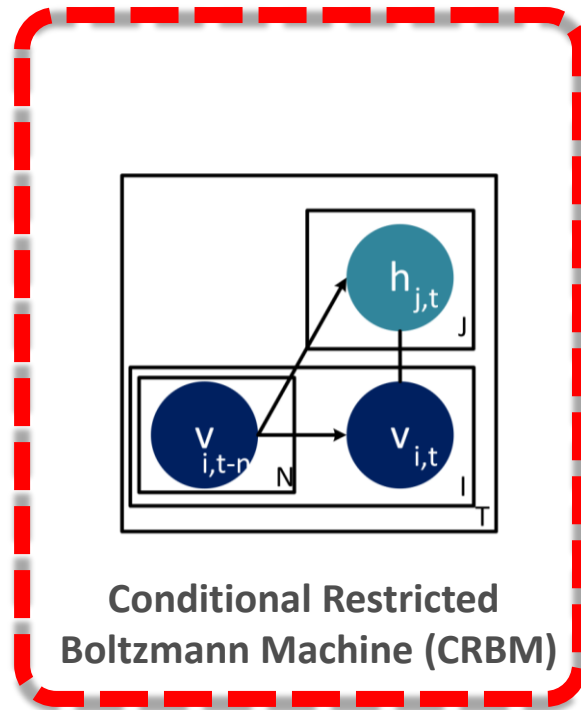
Multimodal Conditional Restricted Boltzmann Machine (MMCRBM)

# Model: Generative Component

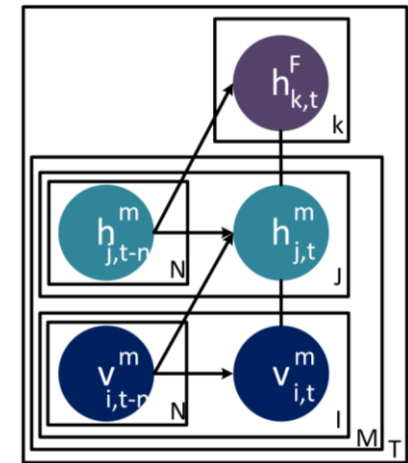
## Multimodal Conditional Restricted Boltzmann Machines (MMCRBMs)



Restricted Boltzmann Machine (RBM)



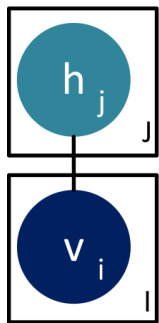
Conditional Restricted Boltzmann Machine (CRBM)



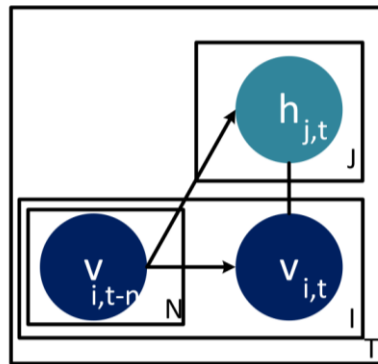
Multimodal Conditional Restricted Boltzmann Machine (MMCRBM)

# Model: Generative Component

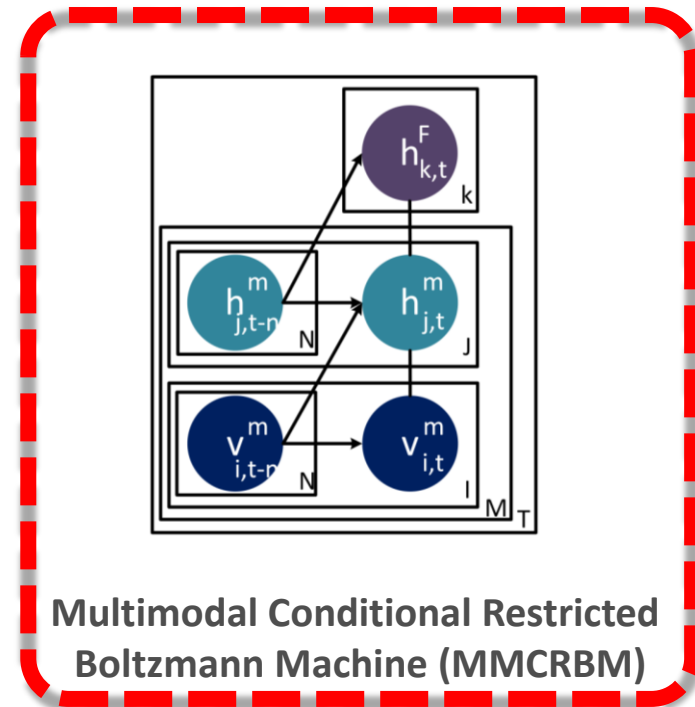
## Multimodal Conditional Restricted Boltzmann Machines (MMCRBMs)



Restricted Boltzmann  
Machine (RBM)



Conditional Restricted  
Boltzmann Machine (CRBM)

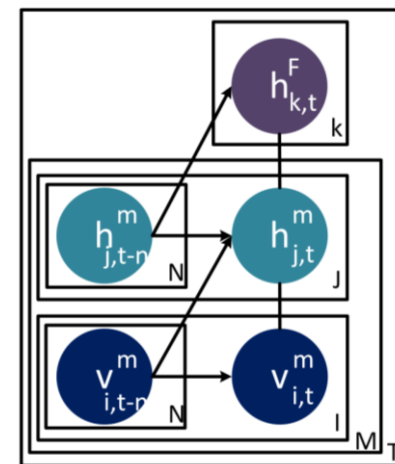
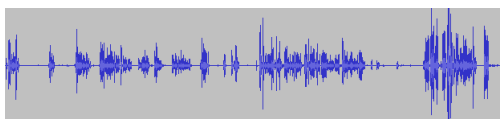
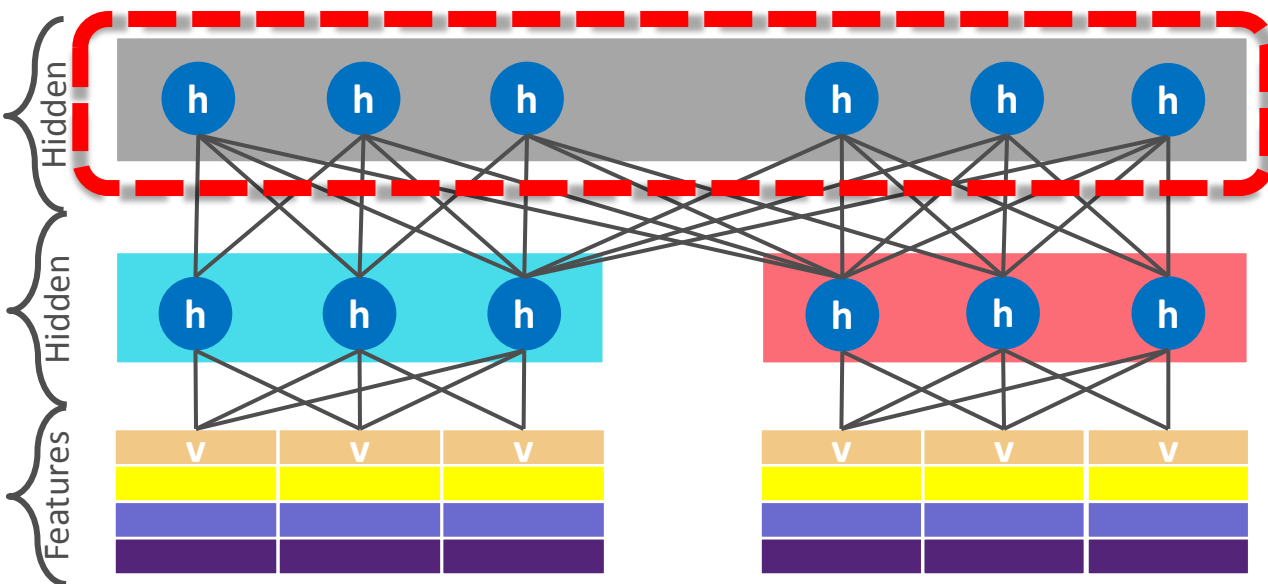


Multimodal Conditional Restricted  
Boltzmann Machine (MMCRBM)



# Model: Generative Component

## Multimodal Conditional Restricted Boltzmann Machines (MMCRBMs)

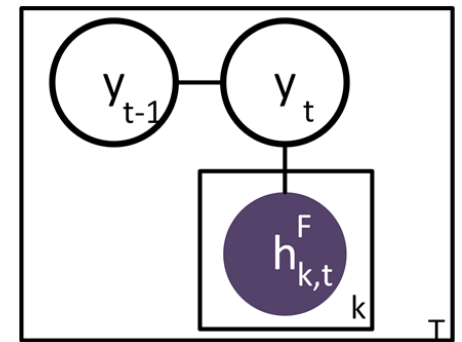


Multimodal Conditional Restricted Boltzmann Machine (MMCRBM)

# Model: Discriminative Component

## Conditional Random Field (CRFs)

- Models non-stationarities and long-term temporal dynamics

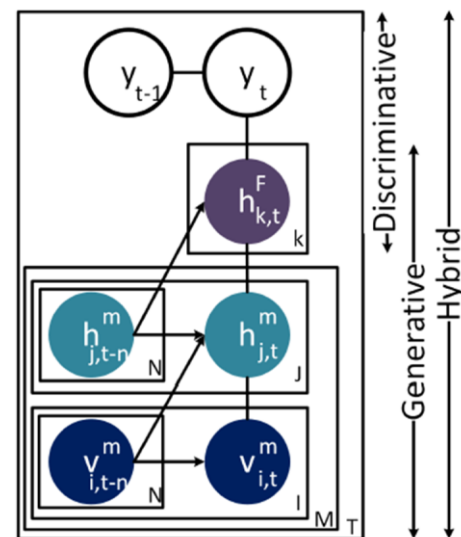


Conditional Random Field (CRF)

# Model: Hybrid Model

## Staged Model: Combination of CRFs and MMCRBM

$$\underbrace{p(\mathbf{y}_t, \mathbf{v}_t, \mathbf{h}_t | \mathbf{v}_{<t})}_{\text{Hybrid}} = \underbrace{p(\mathbf{y}_t | \mathbf{v}_t, \mathbf{h}_t)}_{\text{Discriminative}} \cdot \underbrace{p(\mathbf{v}_t, \mathbf{h}_t | \mathbf{v}_{<t})}_{\text{Generative}}$$



Hybrid Model (CRF-MMCRBM)

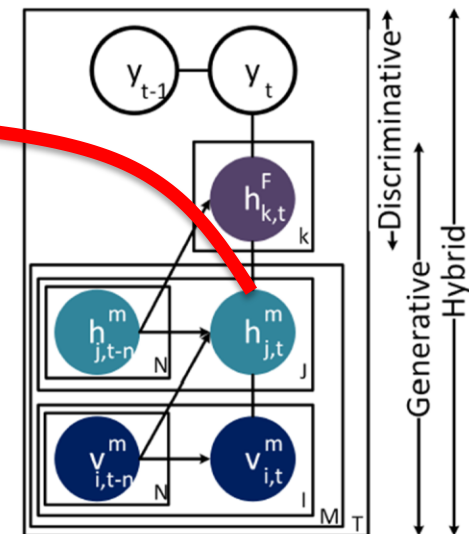
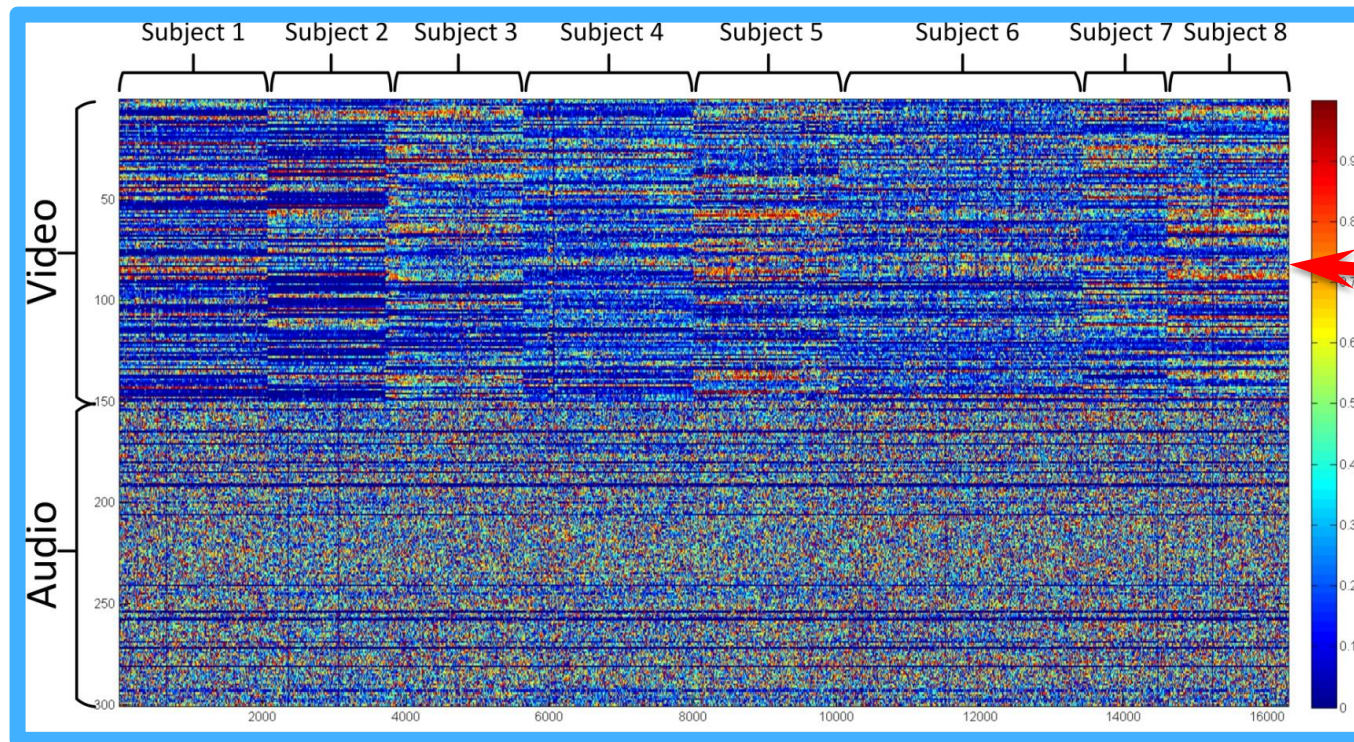
# Learning

## Staged Dynamic Hybrid model

- Multimodal Conditional Restricted Boltzmann Machines (MMCRBMs): **Contrastive Divergence**
- Conditional Random Field (CRF): **Maximum Likelihood Estimation**

# Inference

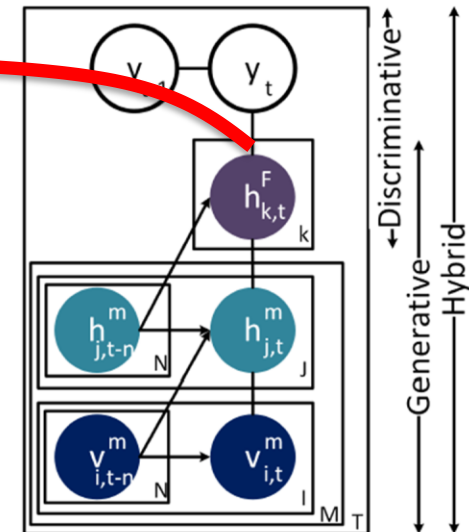
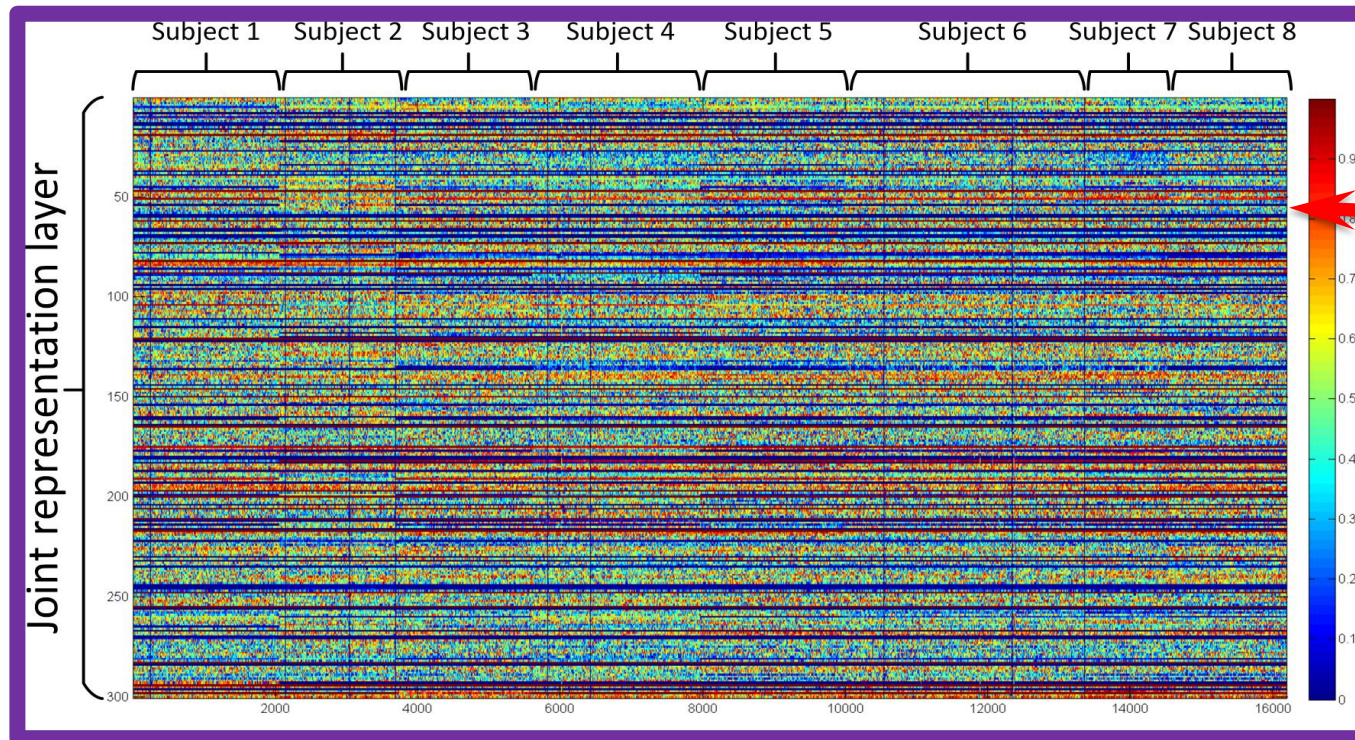
## Unimodal Representations





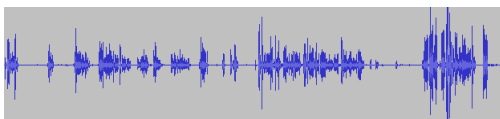
# Inference

## Fused Representation

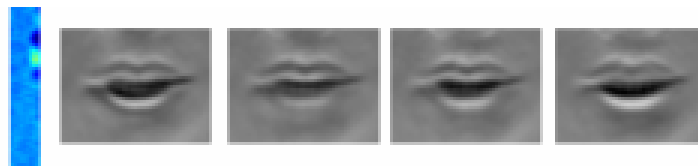


# Results

## AVEC



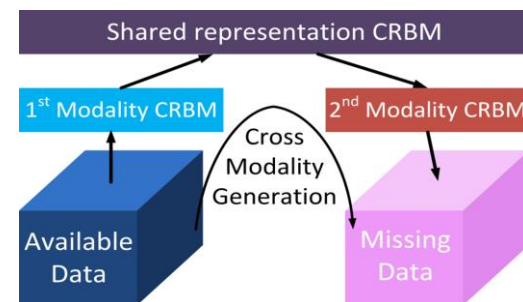
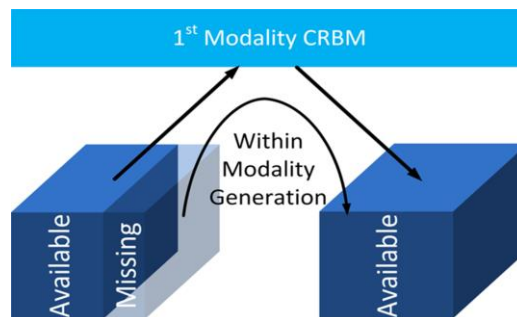
## AVLetters & CUAVE



Model/Dataset	AVEC-A	AVEC-V	AVEC-AV	AVLetters-A	AVLetters-V	AVLetters-AV	CUAVE-A	CUAVE-V	CUAVE-AV
SVM-RAW	64.8	62.4	67.4	55.8	56.2	58.5	61.5	58.4	65.0
CRF-RAW	<b>68.1</b>	<b>63.5</b>	<b>69.9</b>	<b>58.4</b>	<b>59.3</b>	<b>60.0</b>	<b>64.3</b>	<b>62.0</b>	<b>66.8</b>
SVM-RBM	61.8	63.9	67.8	58.4	62.1	62.9	65.1	61.8	65.4
CRF-RBM	<b>67.6</b>	<b>65.4</b>	<b>68.3</b>	<b>62.6</b>	<b>64.6</b>	<b>63.8</b>	<b>67.6</b>	<b>65.2</b>	<b>68.6</b>
SVM-CRBM	65.8	66.9	68.2	61.2	62.6	64.8	65.3	64.6	66.7
CRF-CRBM	<b>69.2</b>	<b>70.1</b>	<b>70.8</b>	<b>66.9</b>	<b>64.8</b>	<b>67.1</b>	<b>67.9</b>	<b>66.3</b>	<b>69.1</b>

Model/Dataset	AVEC-A	AVEC-V	AVLetters-A	AVLetters-V	CUAVE-A	CUAVE-V
SVM-RBM (0%)	61.8	63.9	58.4	62.1	65.1	61.8
SVM-CRBM (0%)	<b>65.8</b>	<b>66.9</b>	<b>61.2</b>	<b>62.6</b>	<b>65.3</b>	<b>64.6</b>
SVM-RBM (10%)	48.6	46.5	50.7	54.5	59.7	42.8
SVM-CRBM (10%)	<b>54.9</b>	<b>52.1</b>	<b>53.6</b>	<b>58.2</b>	<b>63.1</b>	<b>52.6</b>
SVM-RBM (30%)	35.5	31.2	39.2	32.1	36.1	31.9
SVM-CRBM (30%)	<b>42.7</b>	<b>40.2</b>	<b>45.8</b>	<b>41.6</b>	<b>43.7</b>	<b>41.2</b>

## Missing Data



# Conclusion

## Staged Hybrid Model

- Combines advantages of Gen./Disc. Models
- Generative Model
  - » Learns rich represent. modeling short-term dynamics
  - » Handles missing data
- Discriminative Model
  - » Models long-range dynamics
  - » Superior Classification accuracy

**Please visit our poster #21**