

# UNCERTAINTY AND THE MEDICAL INTERVIEW

## TOWARDS SELF-ASSESSMENT IN MACHINE LEARNING MODELS

Jakob Drachmann Havtorn

# OVERVIEW Thesis



## CHAPTER 1-3 INTRODUCTION, RESEARCH QUESTIONS, AND BACKGROUND

CHAPTER 4 HIERARCHICAL VAES KNOW WHAT THEY DON'T KNOW

CHAPTER 5 MODEL-AGNOSTIC OUT-OF-DISTRIBUTION DETECTION  
USING COMBINED STATISTICAL TESTS

CHAPTER 6 A BRIEF OVERVIEW OF UNSUPERVISED SPEECH  
REPRESENTATION LEARNING

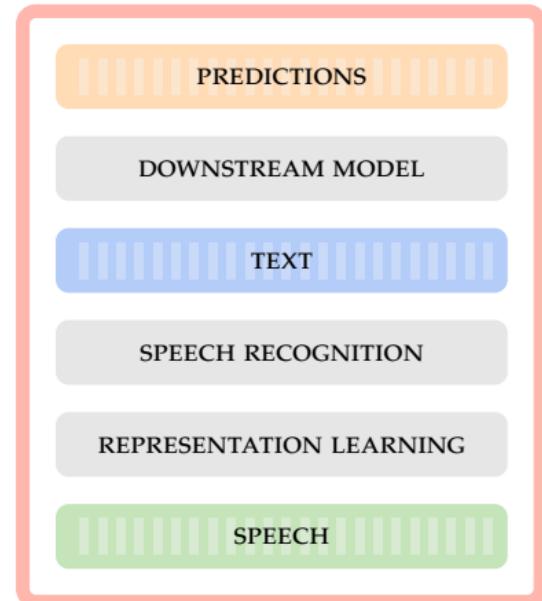
CHAPTER 7 BENCHMARKING LATENT VARIABLE MODELS FOR SPEECH

CHAPTER 8 AUTOMATED MEDICAL CODING ON MIMIC-III AND  
MIMIC-IV: A CRITICAL REVIEW AND REPLICABILITY STUDY

CHAPTER 9 A RETROSPECTIVE STUDY ON MACHINE LEARNING-  
ASSISTED STROKE RECOGNITION FOR MEDICAL HELPLINE CALLS

CHAPTER 10 DISCUSSION AND CONCLUSION

## UNCERTAINTY



# OVERVIEW Thesis



## CHAPTER 1-3 INTRODUCTION, RESEARCH QUESTIONS, AND BACKGROUND

**CHAPTER 4 HIERARCHICAL VAES KNOW WHAT THEY DON'T KNOW**

**CHAPTER 5 MODEL-AGNOSTIC OUT-OF-DISTRIBUTION DETECTION  
USING COMBINED STATISTICAL TESTS**

**CHAPTER 6 A BRIEF OVERVIEW OF UNSUPERVISED SPEECH  
REPRESENTATION LEARNING**

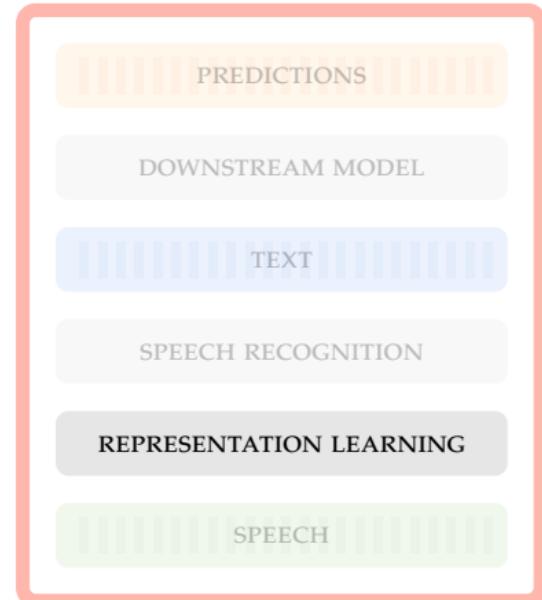
**CHAPTER 7 BENCHMARKING LATENT VARIABLE MODELS FOR SPEECH**

**CHAPTER 8 AUTOMATED MEDICAL CODING ON MIMIC-III AND  
MIMIC-IV: A CRITICAL REVIEW AND REPLICABILITY STUDY**

**CHAPTER 9 A RETROSPECTIVE STUDY ON MACHINE LEARNING-  
ASSISTED STROKE RECOGNITION FOR MEDICAL HELPLINE CALLS**

**CHAPTER 10 DISCUSSION AND CONCLUSION**

## UNCERTAINTY



# OVERVIEW Thesis



CHAPTER 1-3 INTRODUCTION, RESEARCH QUESTIONS, AND BACKGROUND

CHAPTER 4 HIERARCHICAL VAES KNOW WHAT THEY DON'T KNOW

CHAPTER 5 MODEL-AGNOSTIC OUT-OF-DISTRIBUTION DETECTION  
USING COMBINED STATISTICAL TESTS

CHAPTER 6 A BRIEF OVERVIEW OF UNSUPERVISED SPEECH  
REPRESENTATION LEARNING

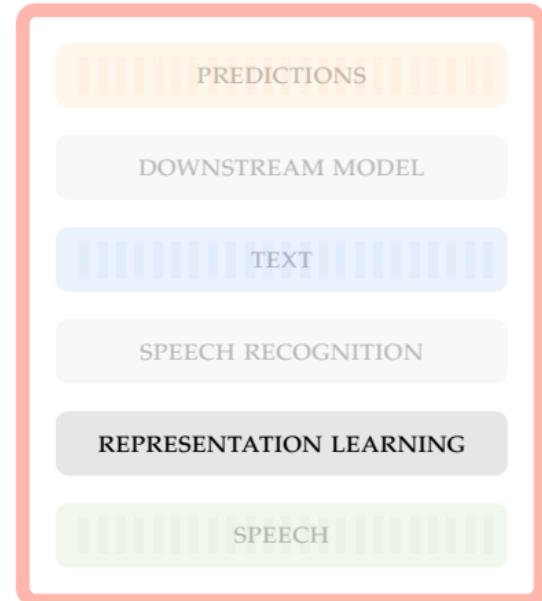
CHAPTER 7 BENCHMARKING LATENT VARIABLE MODELS FOR SPEECH

CHAPTER 8 AUTOMATED MEDICAL CODING ON MIMIC-III AND  
MIMIC-IV: A CRITICAL REVIEW AND REPLICABILITY STUDY

CHAPTER 9 A RETROSPECTIVE STUDY ON MACHINE LEARNING-  
ASSISTED STROKE RECOGNITION FOR MEDICAL HELPLINE CALLS

CHAPTER 10 DISCUSSION AND CONCLUSION

## UNCERTAINTY



# OVERVIEW Thesis



## CHAPTER 1-3 INTRODUCTION, RESEARCH QUESTIONS, AND BACKGROUND

**CHAPTER 4 HIERARCHICAL VAES KNOW WHAT THEY DON'T KNOW**

**CHAPTER 5 MODEL-AGNOSTIC OUT-OF-DISTRIBUTION DETECTION  
USING COMBINED STATISTICAL TESTS**

**CHAPTER 6 A BRIEF OVERVIEW OF UNSUPERVISED SPEECH  
REPRESENTATION LEARNING**

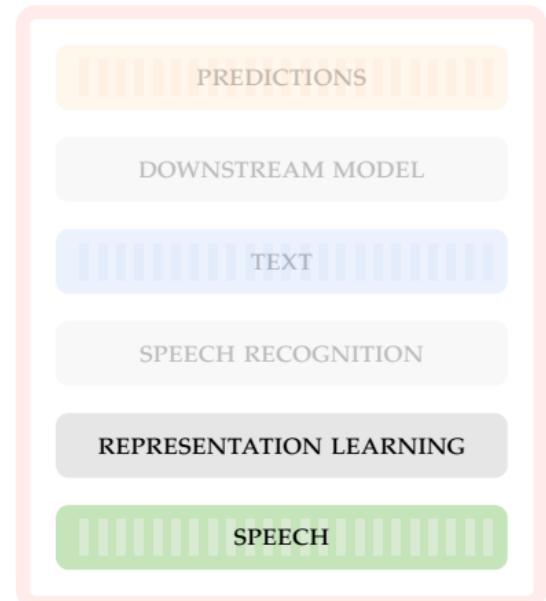
**CHAPTER 7 BENCHMARKING LATENT VARIABLE MODELS FOR SPEECH**

**CHAPTER 8 AUTOMATED MEDICAL CODING ON MIMIC-III AND  
MIMIC-IV: A CRITICAL REVIEW AND REPLICABILITY STUDY**

**CHAPTER 9 A RETROSPECTIVE STUDY ON MACHINE LEARNING-  
ASSISTED STROKE RECOGNITION FOR MEDICAL HELPLINE CALLS**

**CHAPTER 10 DISCUSSION AND CONCLUSION**

## UNCERTAINTY



# OVERVIEW Thesis

## CHAPTER 1-3 INTRODUCTION, RESEARCH QUESTIONS, AND BACKGROUND

**CHAPTER 4 HIERARCHICAL VAES KNOW WHAT THEY DON'T KNOW**

**CHAPTER 5 MODEL-AGNOSTIC OUT-OF-DISTRIBUTION DETECTION  
USING COMBINED STATISTICAL TESTS**

**CHAPTER 6 A BRIEF OVERVIEW OF UNSUPERVISED SPEECH  
REPRESENTATION LEARNING**

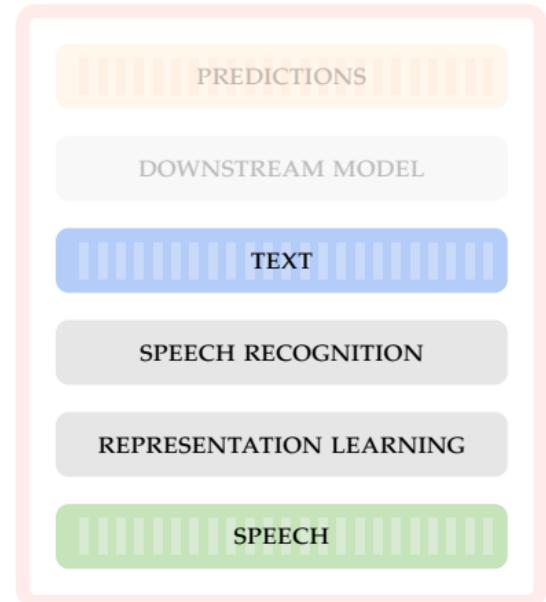
**CHAPTER 7 BENCHMARKING LATENT VARIABLE MODELS FOR SPEECH**

**CHAPTER 8 AUTOMATED MEDICAL CODING ON MIMIC-III AND  
MIMIC-IV: A CRITICAL REVIEW AND REPLICABILITY STUDY**

**CHAPTER 9 A RETROSPECTIVE STUDY ON MACHINE LEARNING-  
ASSISTED STROKE RECOGNITION FOR MEDICAL HELPLINE CALLS**

**CHAPTER 10 DISCUSSION AND CONCLUSION**

## UNCERTAINTY



# OVERVIEW Thesis



## CHAPTER 1-3 INTRODUCTION, RESEARCH QUESTIONS, AND BACKGROUND

CHAPTER 4 HIERARCHICAL VAES KNOW WHAT THEY DON'T KNOW

CHAPTER 5 MODEL-AGNOSTIC OUT-OF-DISTRIBUTION DETECTION  
USING COMBINED STATISTICAL TESTS

CHAPTER 6 A BRIEF OVERVIEW OF UNSUPERVISED SPEECH  
REPRESENTATION LEARNING

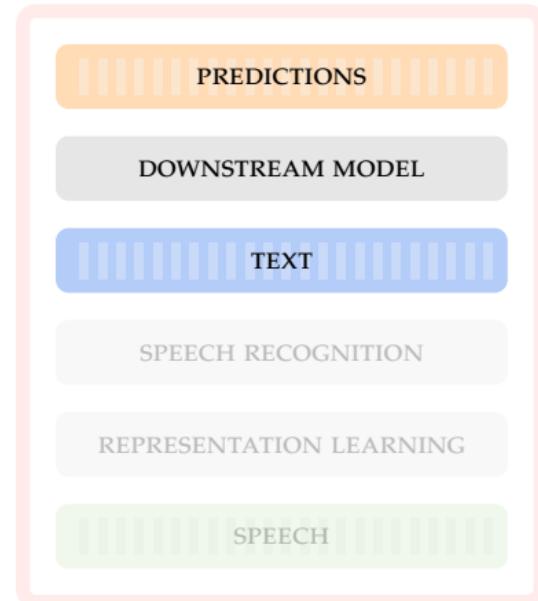
CHAPTER 7 BENCHMARKING LATENT VARIABLE MODELS FOR SPEECH

CHAPTER 8 AUTOMATED MEDICAL CODING ON MIMIC-III AND  
MIMIC-IV: A CRITICAL REVIEW AND REPLICABILITY STUDY

CHAPTER 9 A RETROSPECTIVE STUDY ON MACHINE LEARNING-  
ASSISTED STROKE RECOGNITION FOR MEDICAL HELPLINE CALLS

CHAPTER 10 DISCUSSION AND CONCLUSION

## UNCERTAINTY



# OVERVIEW Thesis

CHAPTER 1-3 INTRODUCTION, RESEARCH QUESTIONS, AND BACKGROUND

CHAPTER 4 HIERARCHICAL VAES KNOW WHAT THEY DON'T KNOW

CHAPTER 5 MODEL-AGNOSTIC OUT-OF-DISTRIBUTION DETECTION  
USING COMBINED STATISTICAL TESTS

CHAPTER 6 A BRIEF OVERVIEW OF UNSUPERVISED SPEECH  
REPRESENTATION LEARNING

CHAPTER 7 BENCHMARKING LATENT VARIABLE MODELS FOR SPEECH

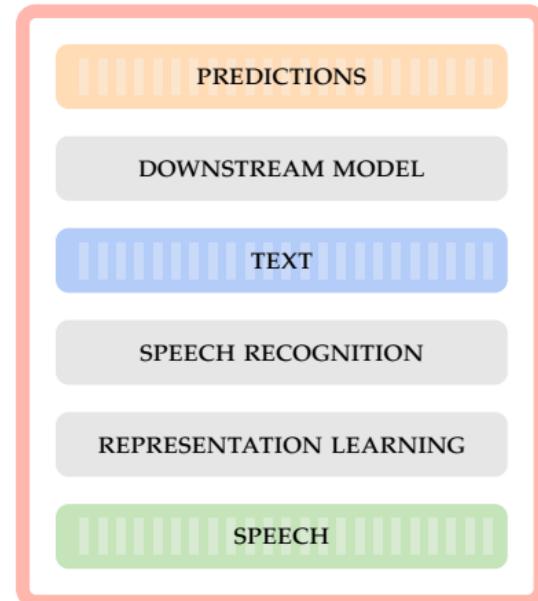
CHAPTER 8 AUTOMATED MEDICAL CODING ON MIMIC-III AND  
MIMIC-IV: A CRITICAL REVIEW AND REPLICABILITY STUDY

CHAPTER 9 A RETROSPECTIVE STUDY ON MACHINE LEARNING-  
ASSISTED STROKE RECOGNITION FOR MEDICAL HELPLINE CALLS

CHAPTER 10 DISCUSSION AND CONCLUSION



## UNCERTAINTY



# OVERVIEW Presentation



## CHAPTER 1-3 INTRODUCTION, RESEARCH QUESTIONS, AND BACKGROUND

CHAPTER 4 HIERARCHICAL VAES KNOW WHAT THEY DON'T KNOW

CHAPTER 5 MODEL-AGNOSTIC OUT-OF-DISTRIBUTION DETECTION  
USING COMBINED STATISTICAL TESTS

CHAPTER 6 A BRIEF OVERVIEW OF UNSUPERVISED SPEECH  
REPRESENTATION LEARNING

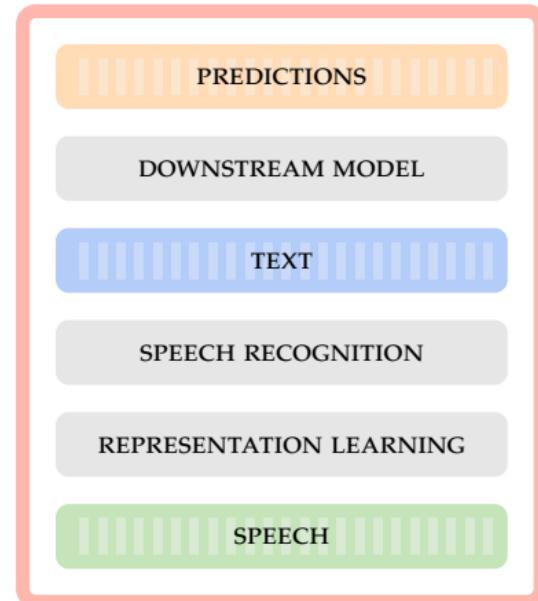
CHAPTER 7 BENCHMARKING LATENT VARIABLE MODELS FOR SPEECH

CHAPTER 8 AUTOMATED MEDICAL CODING ON MIMIC-III AND  
MIMIC-IV: A CRITICAL REVIEW AND REPLICABILITY STUDY

CHAPTER 9 A RETROSPECTIVE STUDY ON MACHINE LEARNING-  
ASSISTED STROKE RECOGNITION FOR MEDICAL HELPLINE CALLS

CHAPTER 10 DISCUSSION AND CONCLUSION

## UNCERTAINTY



# OVERVIEW Presentation



## CHAPTER 1-3 INTRODUCTION, RESEARCH QUESTIONS, AND BACKGROUND

CHAPTER 4 HIERARCHICAL VAES KNOW WHAT THEY DON'T KNOW

CHAPTER 5 MODEL-AGNOSTIC OUT-OF-DISTRIBUTION DETECTION  
USING COMBINED STATISTICAL TESTS

CHAPTER 6 A BRIEF OVERVIEW OF UNSUPERVISED SPEECH  
REPRESENTATION LEARNING

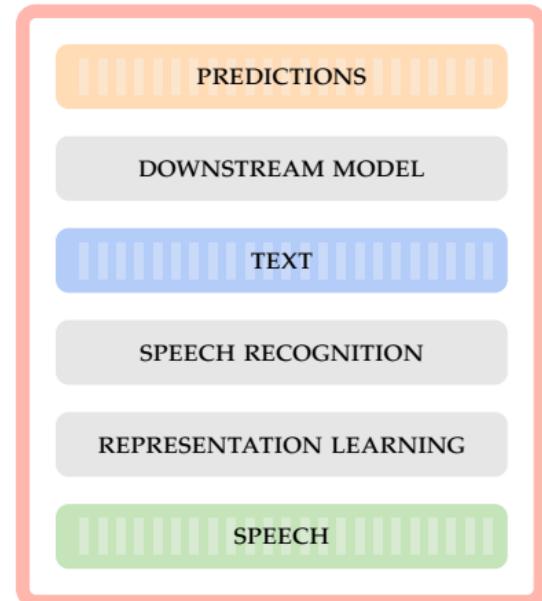
CHAPTER 7 BENCHMARKING LATENT VARIABLE MODELS FOR SPEECH

CHAPTER 8 AUTOMATED MEDICAL CODING ON MIMIC-III AND  
MIMIC-IV: A CRITICAL REVIEW AND REPLICABILITY STUDY

CHAPTER 9 A RETROSPECTIVE STUDY ON MACHINE LEARNING-  
ASSISTED STROKE RECOGNITION FOR MEDICAL HELPLINE CALLS

CHAPTER 10 DISCUSSION AND CONCLUSION

## UNCERTAINTY



# OVERVIEW Presentation

CHAPTER 1-3 INTRODUCTION, RESEARCH QUESTIONS, AND BACKGROUND

---

CHAPTER 4 HIERARCHICAL VAES KNOW WHAT THEY DON'T KNOW

CHAPTER 6 A BRIEF OVERVIEW OF UNSUPERVISED SPEECH  
REPRESENTATION LEARNING

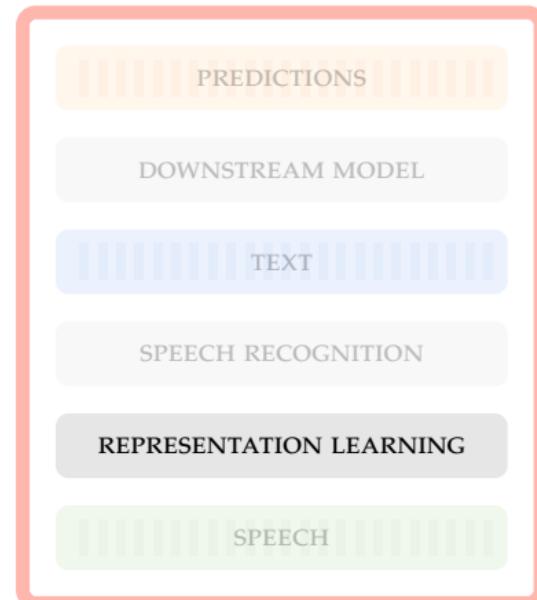
CHAPTER 9 A RETROSPECTIVE STUDY ON MACHINE LEARNING-  
ASSISTED STROKE RECOGNITION FOR MEDICAL HELPLINE CALLS

---

CHAPTER 10 DISCUSSION AND CONCLUSION



## UNCERTAINTY



# OVERVIEW Presentation

CHAPTER 1-3 INTRODUCTION, RESEARCH QUESTIONS, AND BACKGROUND

---

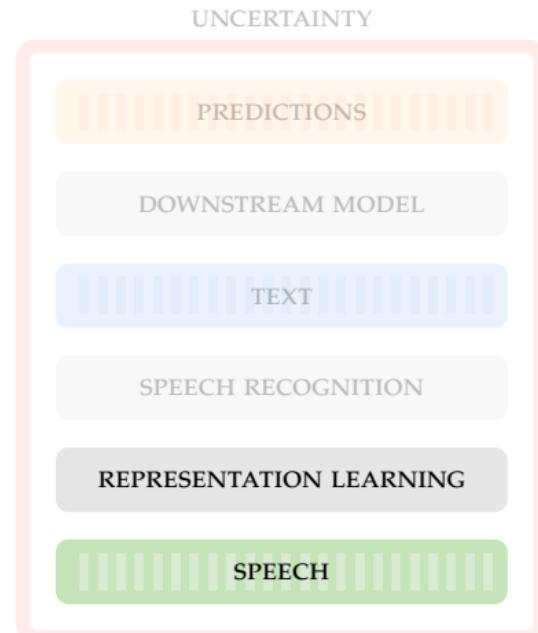
CHAPTER 4 HIERARCHICAL VAES KNOW WHAT THEY DON'T KNOW

CHAPTER 6 **A BRIEF OVERVIEW OF UNSUPERVISED SPEECH  
REPRESENTATION LEARNING**

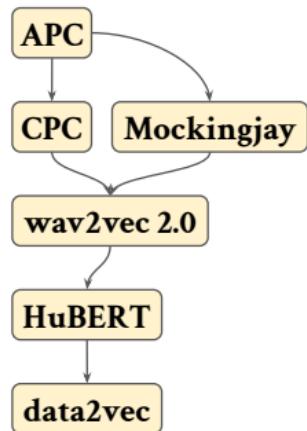
CHAPTER 9 A RETROSPECTIVE STUDY ON MACHINE LEARNING-  
ASSISTED STROKE RECOGNITION FOR MEDICAL HELPLINE CALLS

---

CHAPTER 10 DISCUSSION AND CONCLUSION

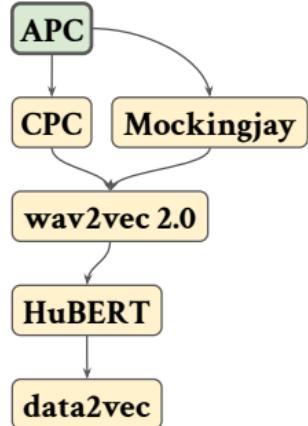


## Development of SSL for speech

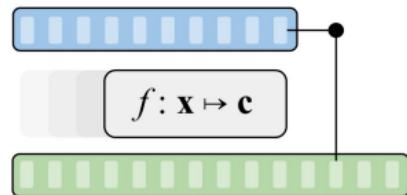


MODEL DESCRIPTION GOES HERE

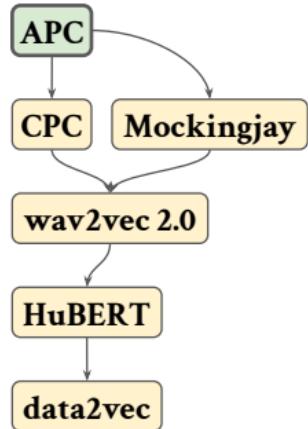
# Autoregressive Predictive Coding (APC)



- **Task:** Predict future inputs.
- **Input/target:** Log-mel spectrogram.
- **Architecture:** RNN/Transformer decoder.
- **Slow features:** Predict k steps ahead.

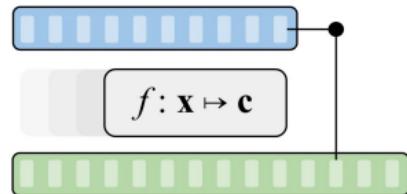


# Autoregressive Predictive Coding (APC)

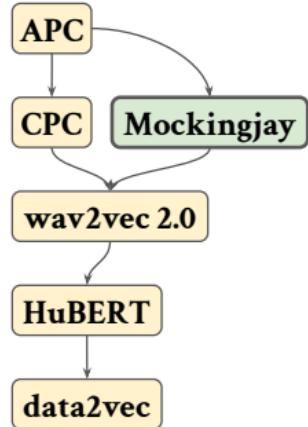


- Challenges:

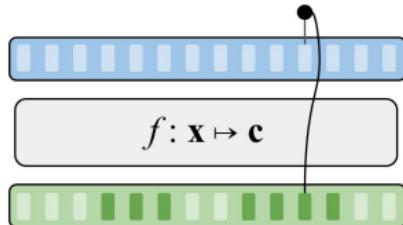
- Encodes only past inputs  $\times$
- Uses the input as target  $\times$



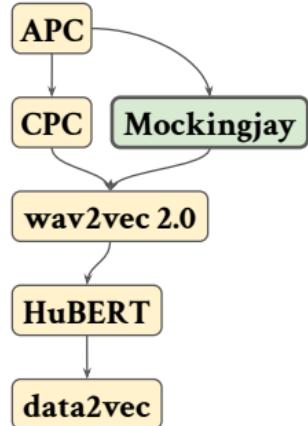
# Mockingjay



- **Task:** Reconstruct masked inputs.
- **Architecture:** Transformer encoder.
- **Masking:**
  - X% at random. (Mockingjay)
  - X% + N consecutive (wav2vec 2.0)
  - SpecAugment (Masked RNN)

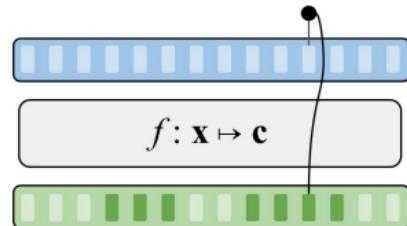


# Mockingjay

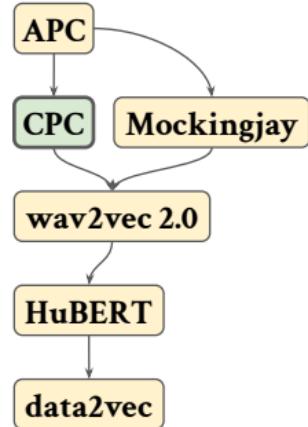


- Challenges:

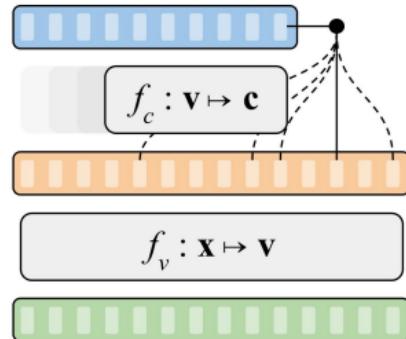
- Encodes the entire input ✓
- Uses the input as target ✗



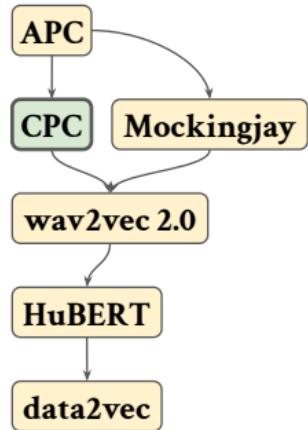
# Contrastive Predictive Coding (CPC)



- **Contrastive models:** Distinguish target samples from negative samples.
- **Learned target:** Discard details.
- **Sampling negatives:**
  - Sample sequence?
  - Same speaker?

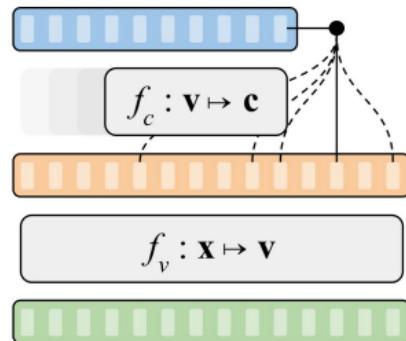


# Contrastive Predictive Coding (CPC)

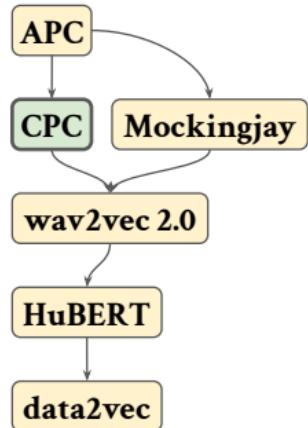


- Challenges:

- Only encodes past inputs  $\times$
- Uses a learned target  $\checkmark$

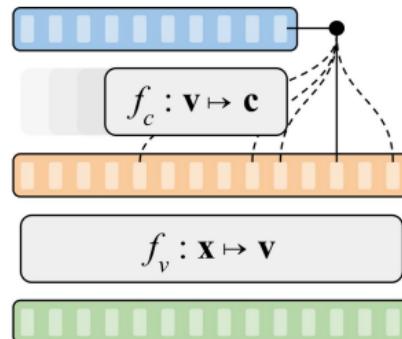


# Contrastive Predictive Coding (CPC)



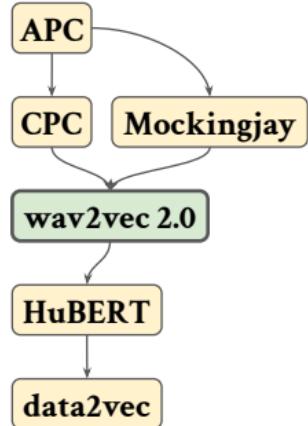
- Challenges:

- Only encodes past inputs ✗
- Uses a learned target ✓
- Sampling negatives ✗

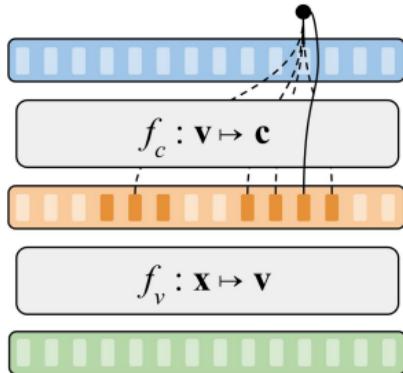


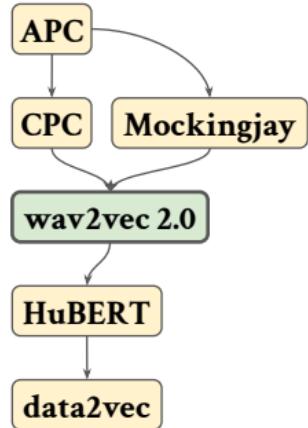
# A BRIEF OVERVIEW OF UNSUPERVISED SPEECH REPRESENTATION LEARNING

## wav2vec 2.0



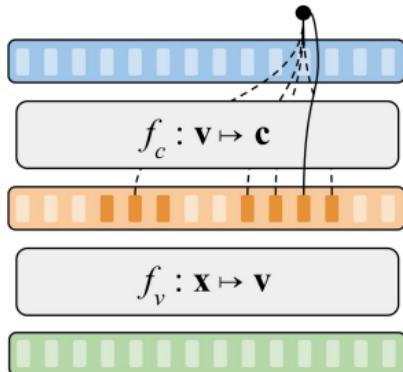
- Masking + contrastive learning.
- **Quantisation:** Better negative samples.
- **Results:**
  - 960 hours: **2.0%** WER.
  - 10 minutes: **4.8%** WER.



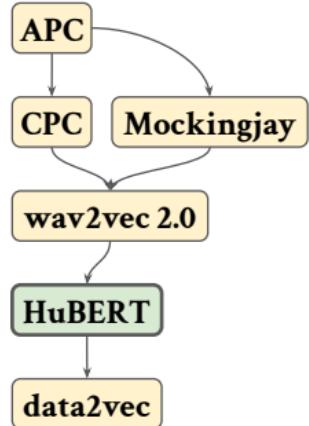


- Challenges:

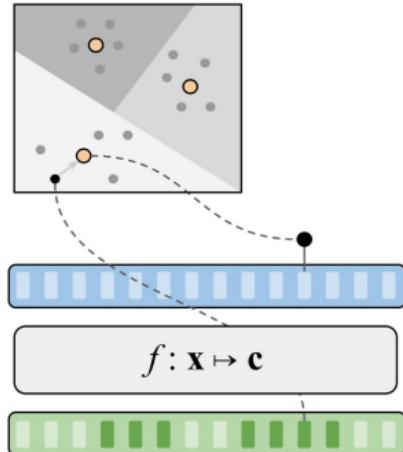
- Encodes the entire input ✓
- Uses a learned target ✓
- Sampling negatives ✗



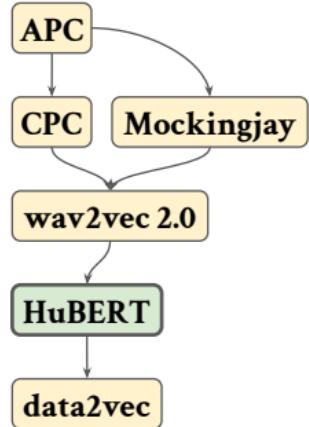
A BRIEF OVERVIEW OF UNSUPERVISED SPEECH REPRESENTATION LEARNING  
**Hidden-unit BERT (HuBERT)**



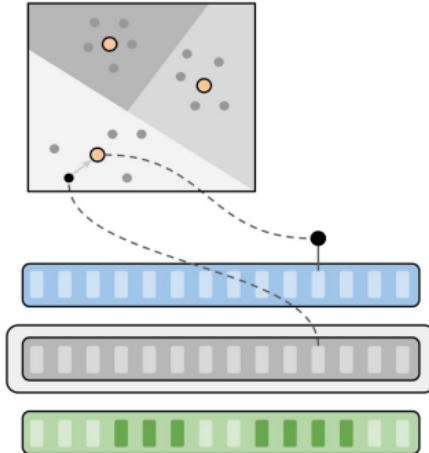
- Target: K-means teacher.
- Training: Simple cross-entropy loss.
- 1st iteration: K-means on inputs.



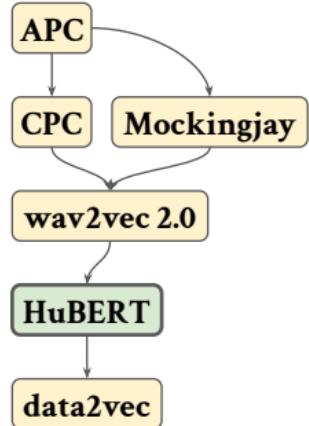
A BRIEF OVERVIEW OF UNSUPERVISED SPEECH REPRESENTATION LEARNING  
**Hidden-unit BERT (HuBERT)**



- **Target:** K-means teacher.
- **Training:** Simple cross-entropy loss.
- **1st iteration:** K-means on inputs.
- **2nd iteration:** K-means on hidden layers.

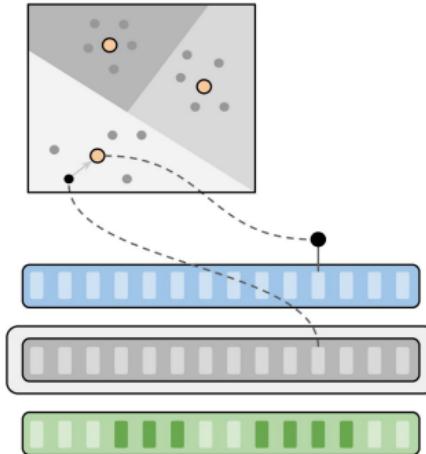


A BRIEF OVERVIEW OF UNSUPERVISED SPEECH REPRESENTATION LEARNING  
**Hidden-unit BERT (HuBERT)**

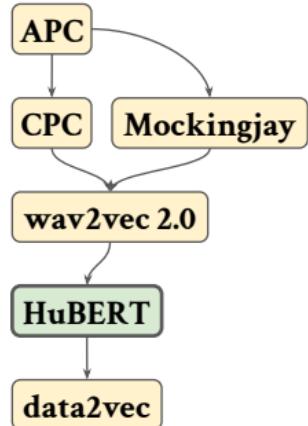


- Challenges:

- Encodes the entire input ✓
- Uses a learned target ✓
- No need for negative samples ✓

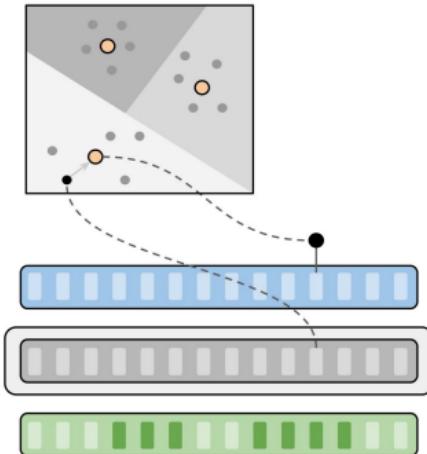


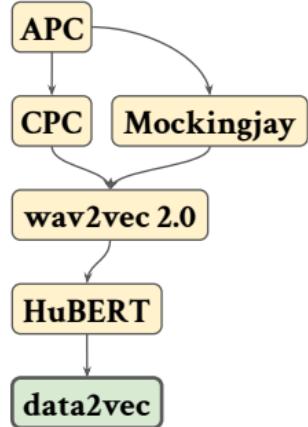
A BRIEF OVERVIEW OF UNSUPERVISED SPEECH REPRESENTATION LEARNING  
**Hidden-unit BERT (HuBERT)**



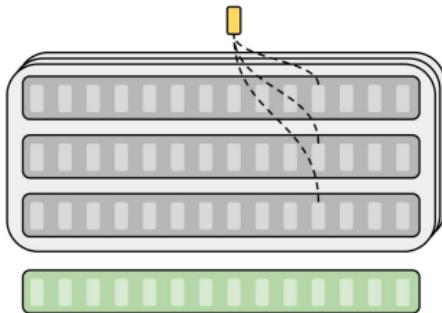
- Challenges:

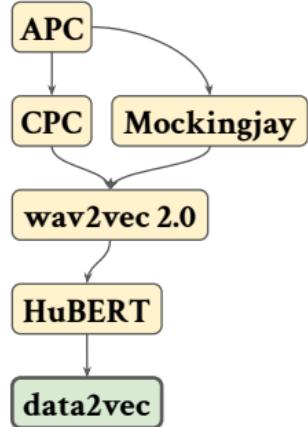
- Encodes the entire input ✓
- Uses a learned target ✓
- No need for negative samples ✓
- Targets updated infrequently ✗
- Quantized targets ✗



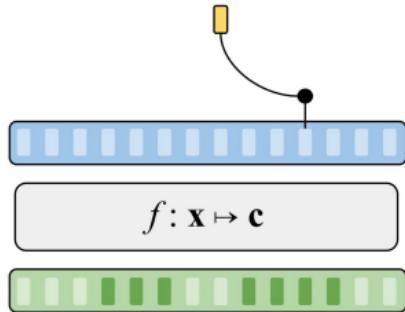


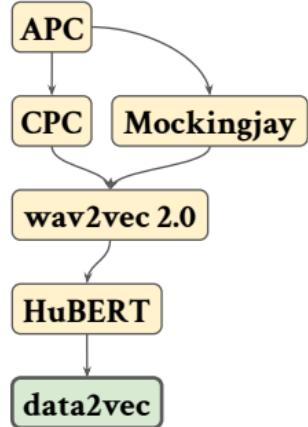
- Uses a teacher-student framework.
- Teacher:
  - EMA of student (online) ✓
  - Target is average of top K layers ✓





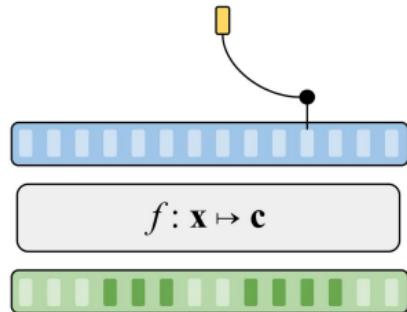
- Uses a teacher-student framework.
- Teacher:
  - EMA of student (online) ✓
  - Target is average of top K layers ✓
- Student training: Smooth  $\ell_1$  loss.





- Challenges:

- Encodes the entire input ✓
- Uses a learned target ✓
- No need for negative samples ✓
- Targets updated continuously ✓
- Continuous-valued targets ✓



## Conclusions

- **Main conclusions:**
  - The most popular self-supervised speech models can be compactly described by a few core design choices.
  - Many of these design choices are mirrored in earlier work on speech embedding models.
- **Open questions and limitations:**
  - Which design choices benefit which downstream tasks?
  - It is difficult to compare methods as model size and evaluation procedures differ widely between papers.

Thank you for your attention

## Bibliography I