CSC 2518

Topics:

- Robustness / Learned Representations.
- Reductive Architecture / Under-resourced Language.
- Intelligibility / Enhancement / Adaptation
- RL
- Time-domain / Phase Representation.

Automatic Speech Recognition.

Dynamic Time Warping (DTW)

Hidden Markov Model (HMM)

Large Vocabulary continuous dictation., Neural acoustic models outperformed by Gaussian mixtures.

Discriminative Training, Weighted finite-state transducers (WFSTs)

Neural architectures

Seide et al. 2011 : 30% RER on Switchboard.

Aim of ASR.

$$\hat{W} = \arg\max_W P(W|A) = \arg\max_W P(A|W)\,P(W).$$

W: most likely sentence.

A: speech audio.

Acoustic model: $P(A|W)$

Language model: $P(W)$

Evaluation : $F(X, w) = P(w|X)$.

Inference

$w = \arg$

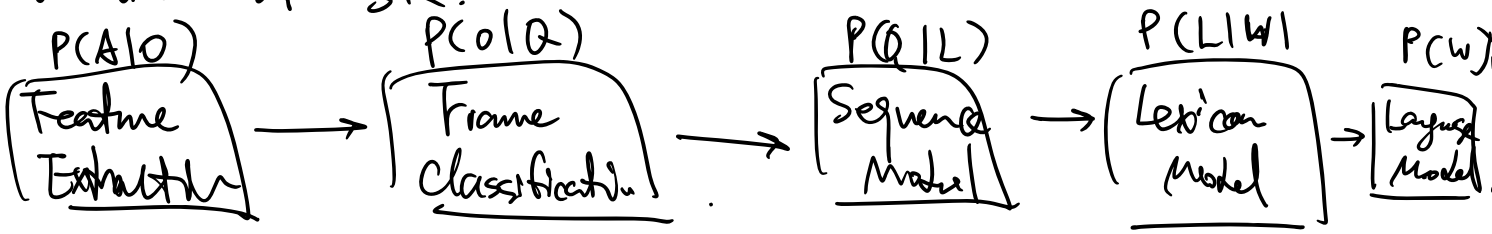
Language Modeling.

- Guide the search algorithm.
- Disambiguate between phrase which are acoustically similar

N-gram.

RNN.

Architecture of SR.



A: Audio

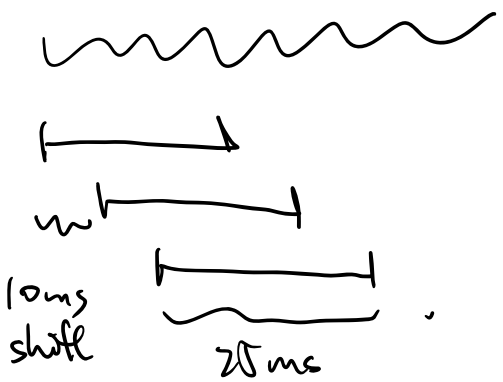
O: Feature Frames

Q: Sequence states

L: Phonemes

W: Words.

Feature Extraction.



Phoneme
basic unit.