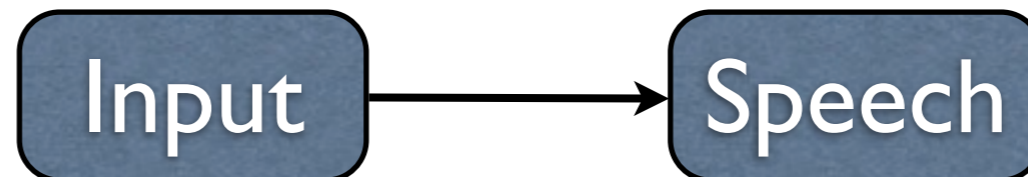


Recent advances in Speech Generation using Deep Learning Techniques

Hamid Mohammadi
Advanced Machine Learning Course
CSLU OHSU
09/28/2015

Speech Generation

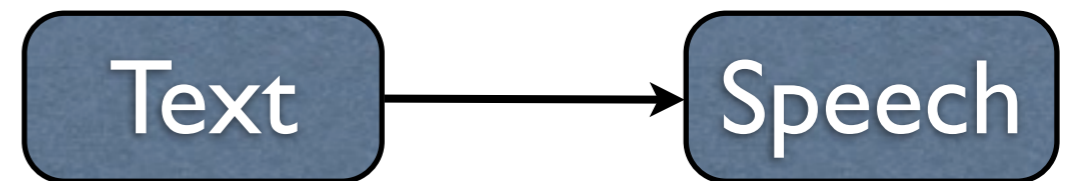
- Speech Generation (SG): **Regression** tasks in speech processing that are aimed to generate speech signals from various types of inputs.



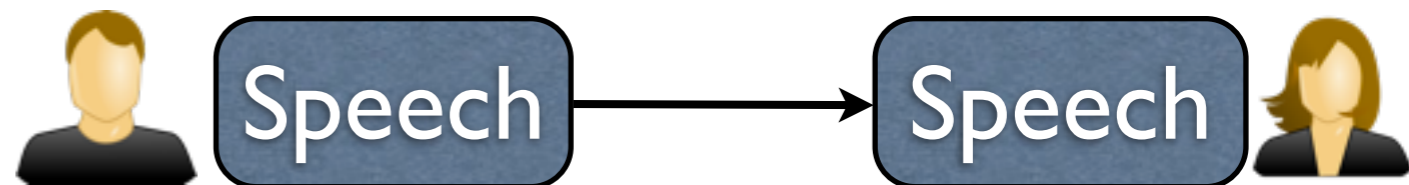
Speech Generation

- Some examples of SG:

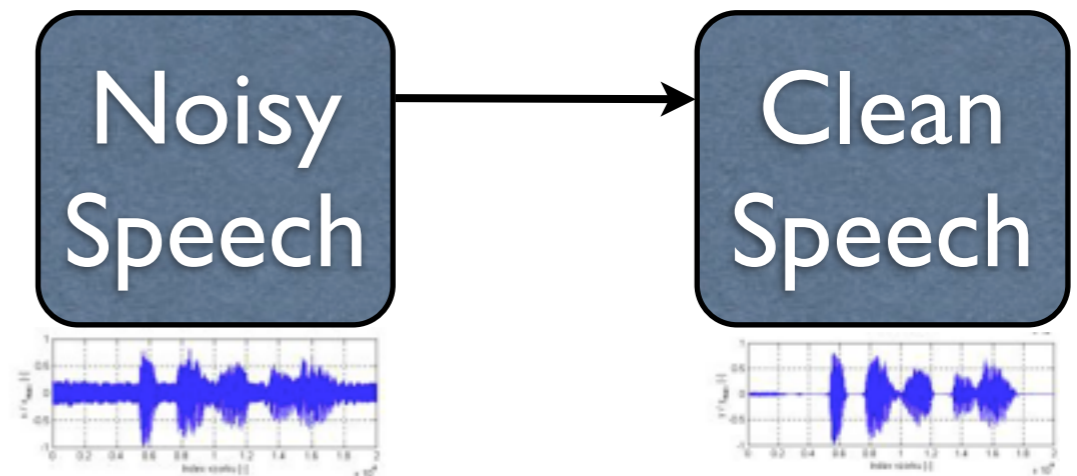
- Text-to-Speech Synthesis



- Voice Conversion



- Speech Enhancement



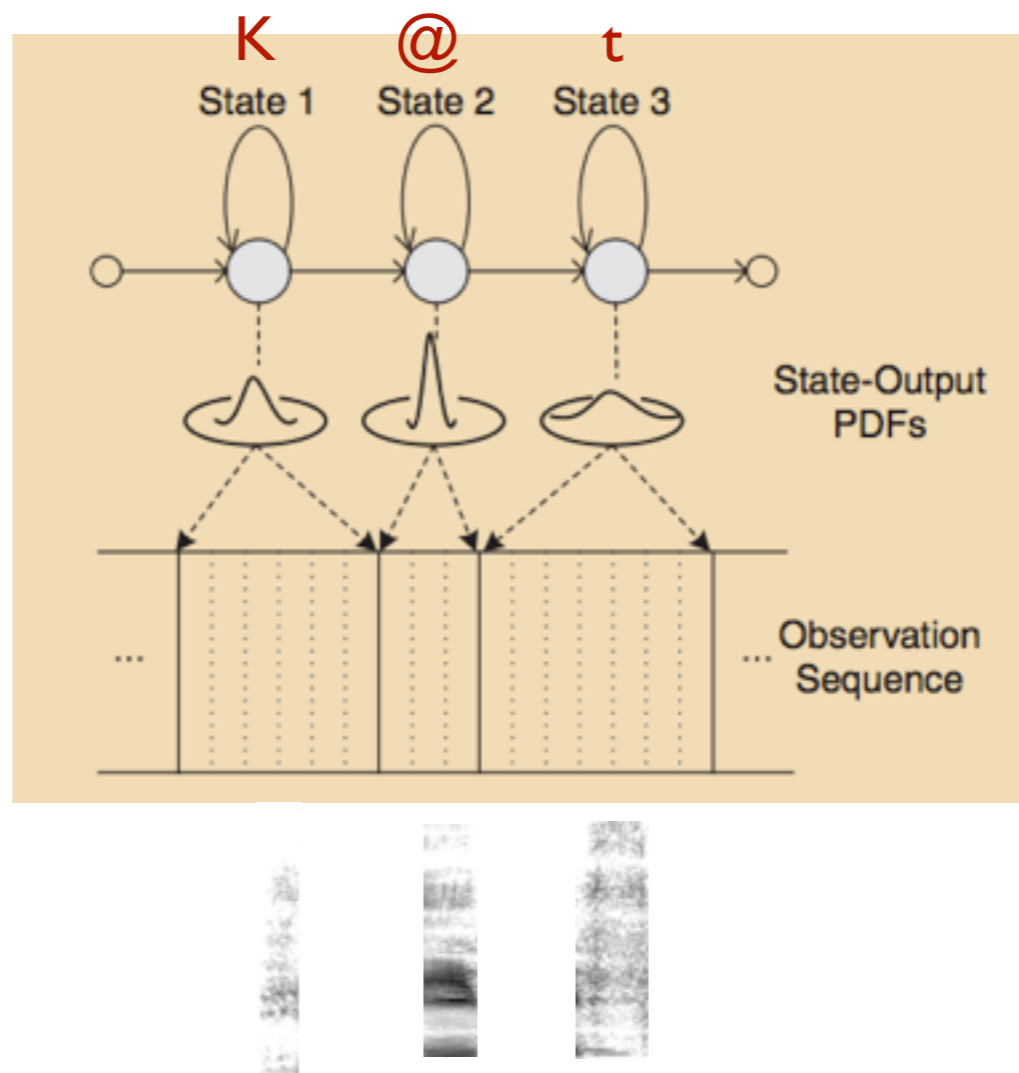
Speech Generation

- Statistical parametric speech generation (SPSG):
- generates speech wave by combining:
 - statistical acoustic models, and
 - vocoding techniques (e.g. Linear Predictive Coding)



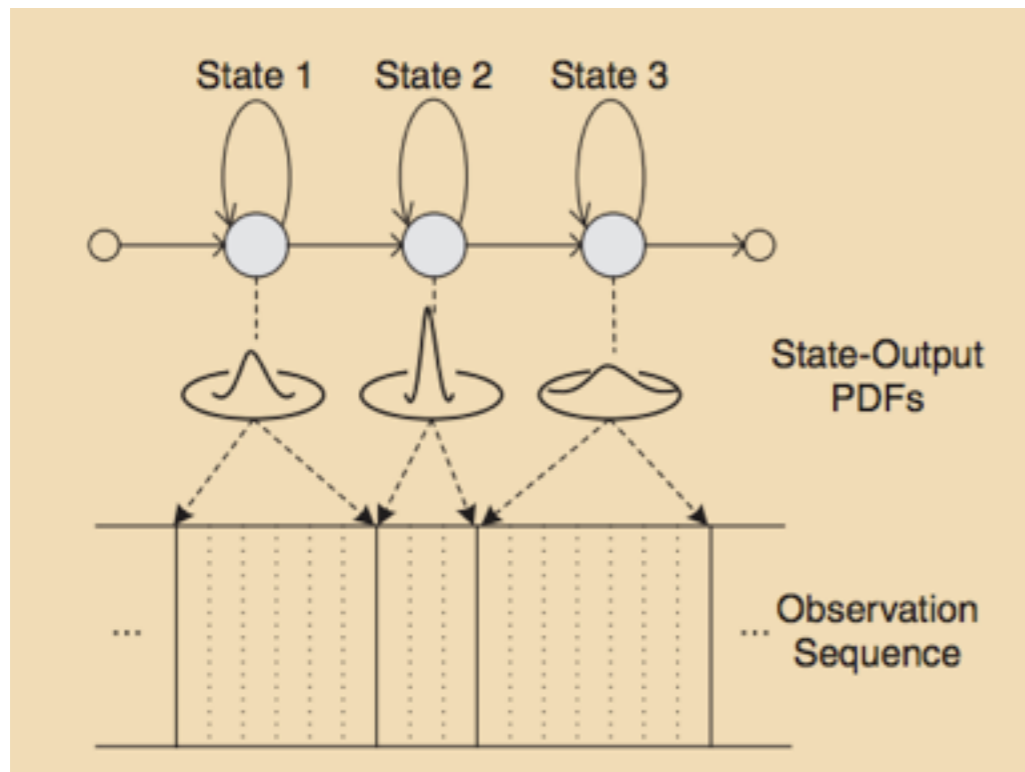
SPSG

- Models:
 - Hidden Markov Model (HMM)

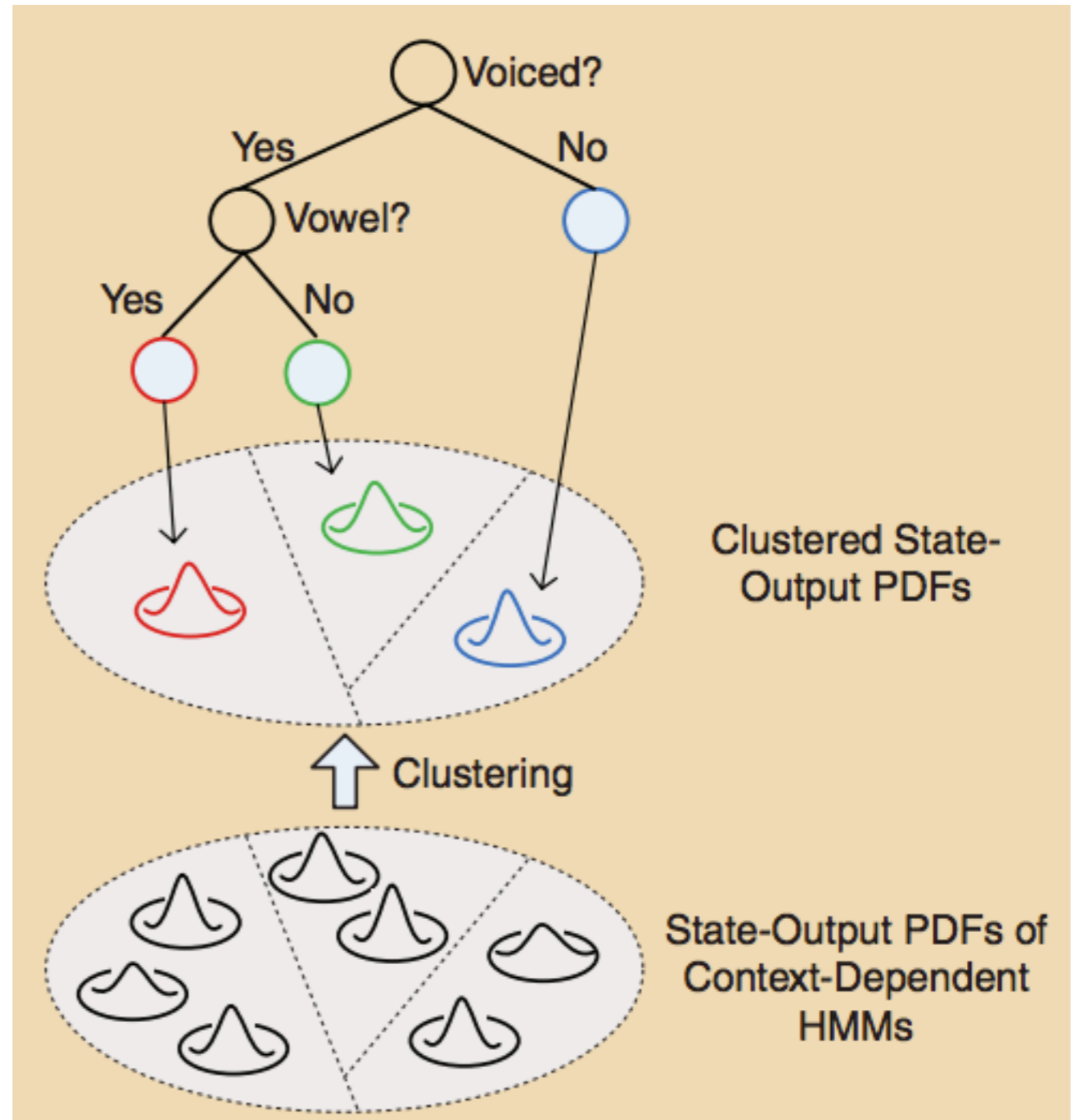


HMM

Text



Speech



Bottleneck Features

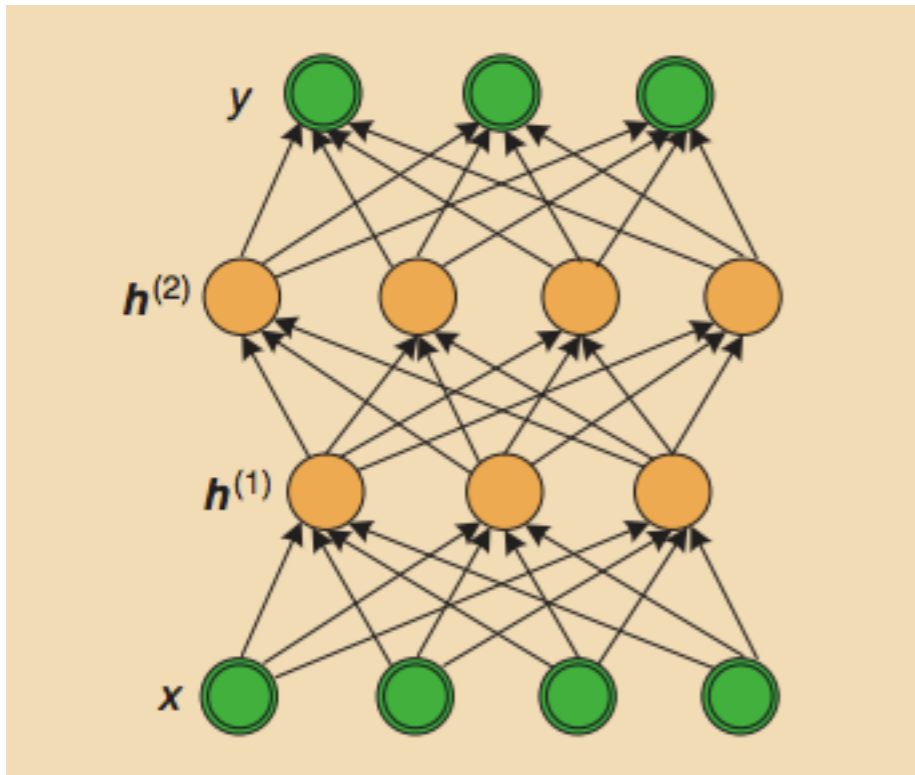
- Instead of using a speech parametrization such as LPCs or Mel-Cepstrum,
- Learn Features using Deep architectures such as
 - Denoising Autoencoders, or
 - Restricted Boltzmann Machine (RBMs)
- Use those features in HMM-based TTS

Deep Learning Networks

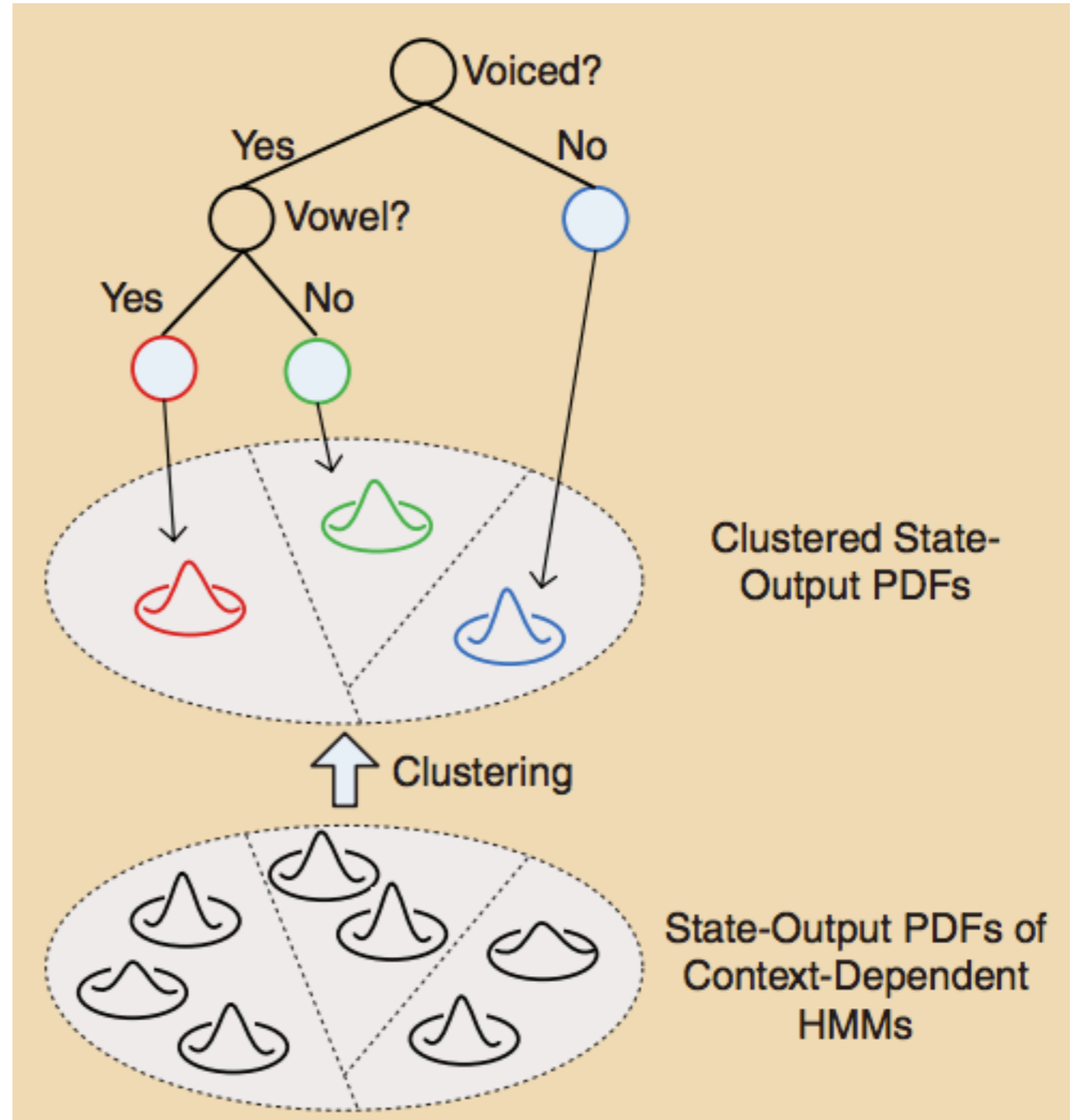
- Deep Neural Network (DNN)
- Recurrent Neural Network (RNN)
- Long Short Term Memory (LSTM)
architecture
-

DNN

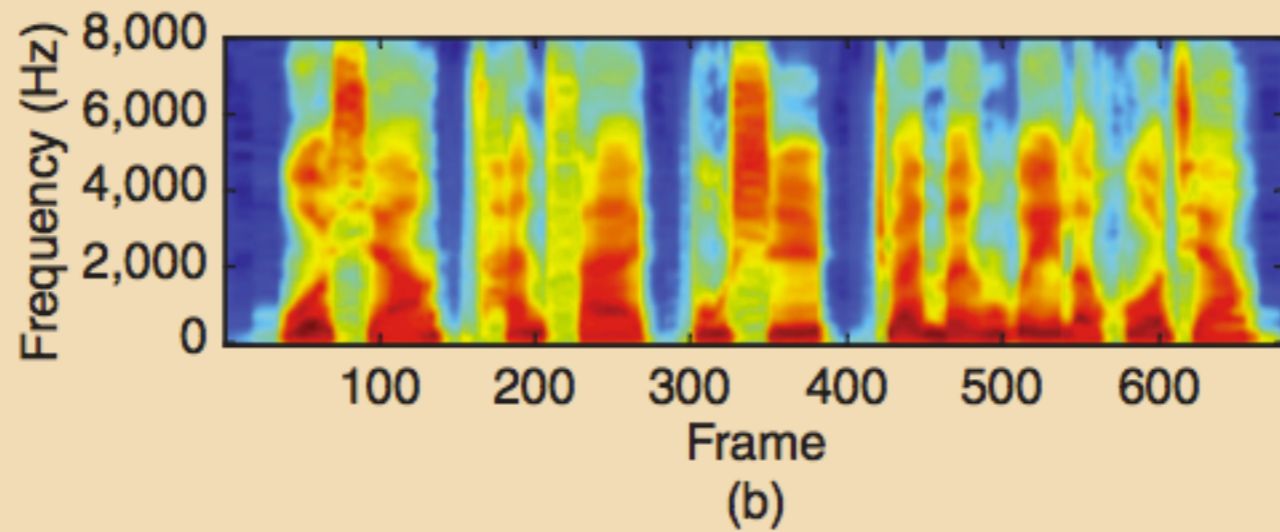
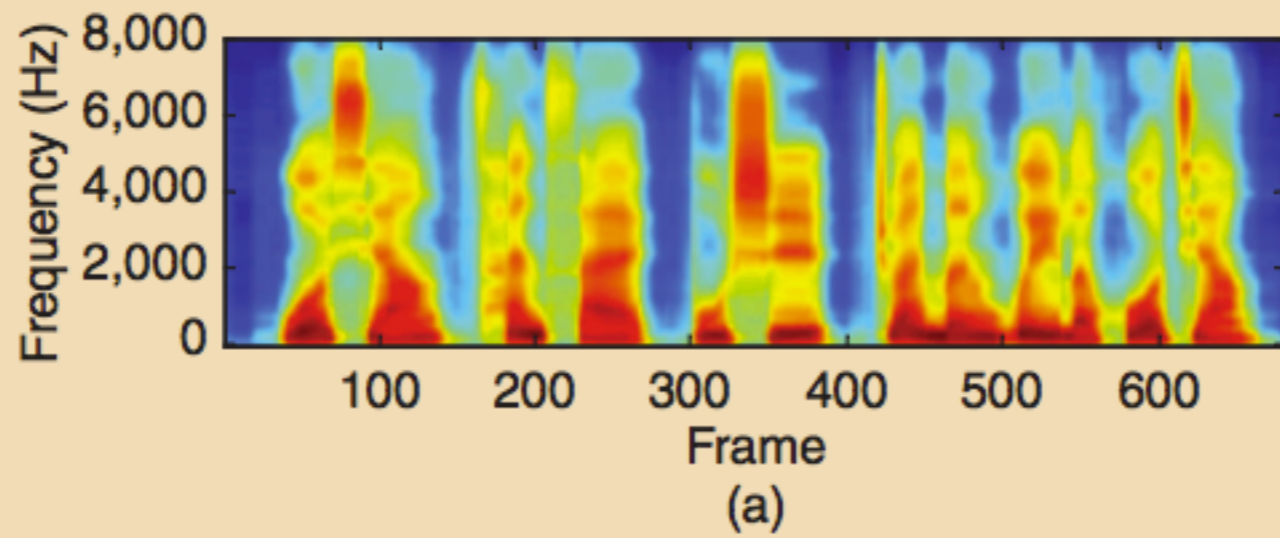
Text



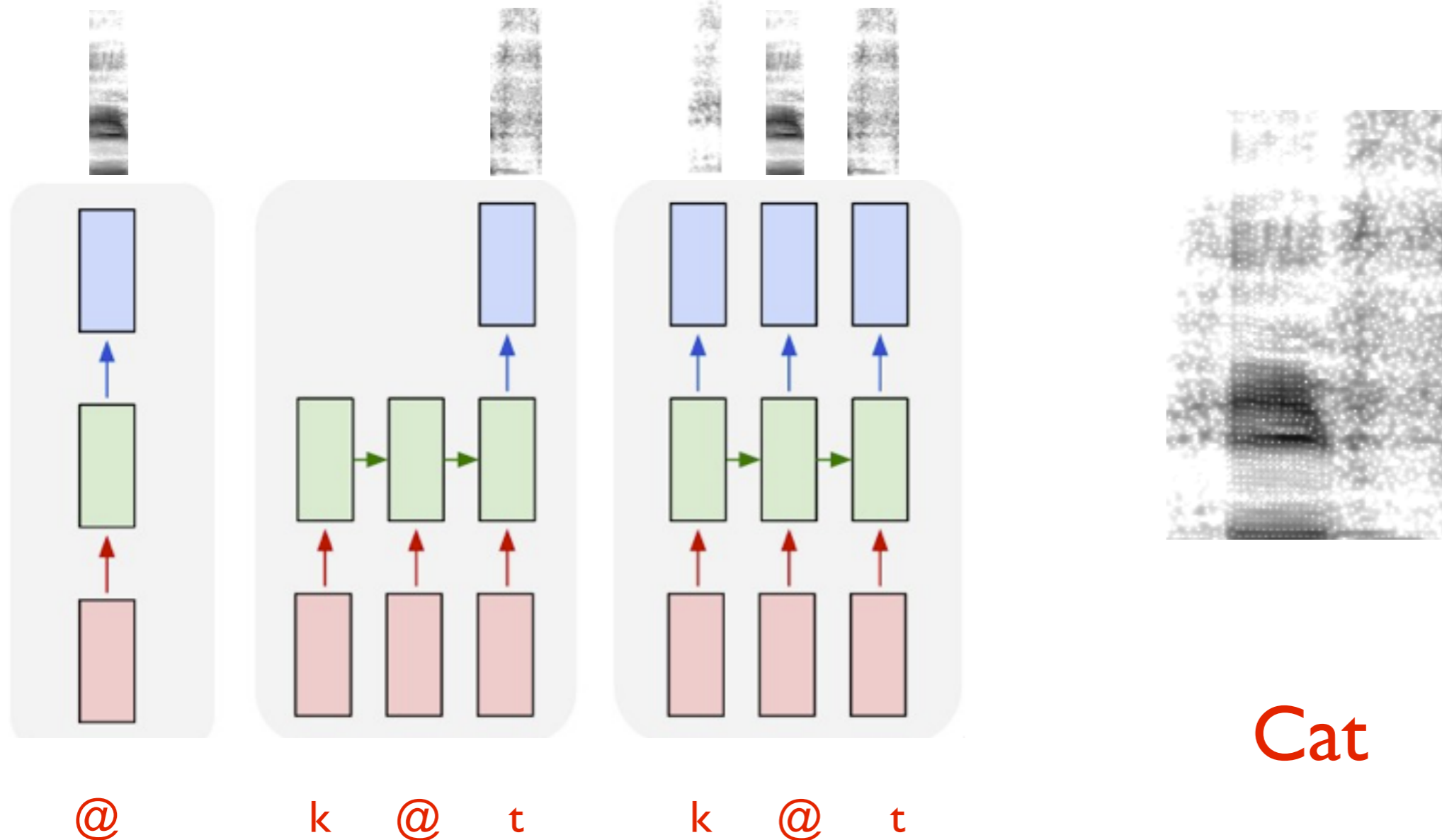
Speech



DNN

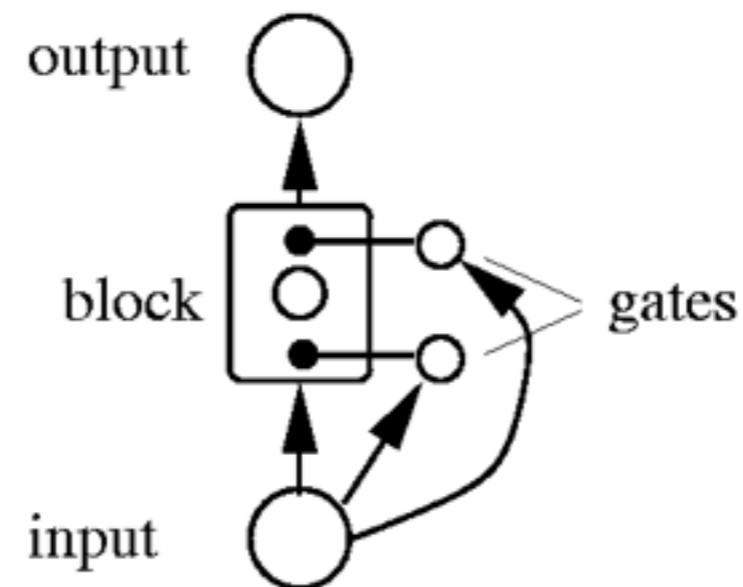


Recurrent NNs



Long-Short Term Memory

- The LSTM: A RNN architecture is designed to model temporal sequences and their long-term dependencies.
- It has some memory-blocks



Recurrent NNs

- Microsoft Research RNN samples
- <http://research.microsoft.com/en-us/projects/dnntts/>