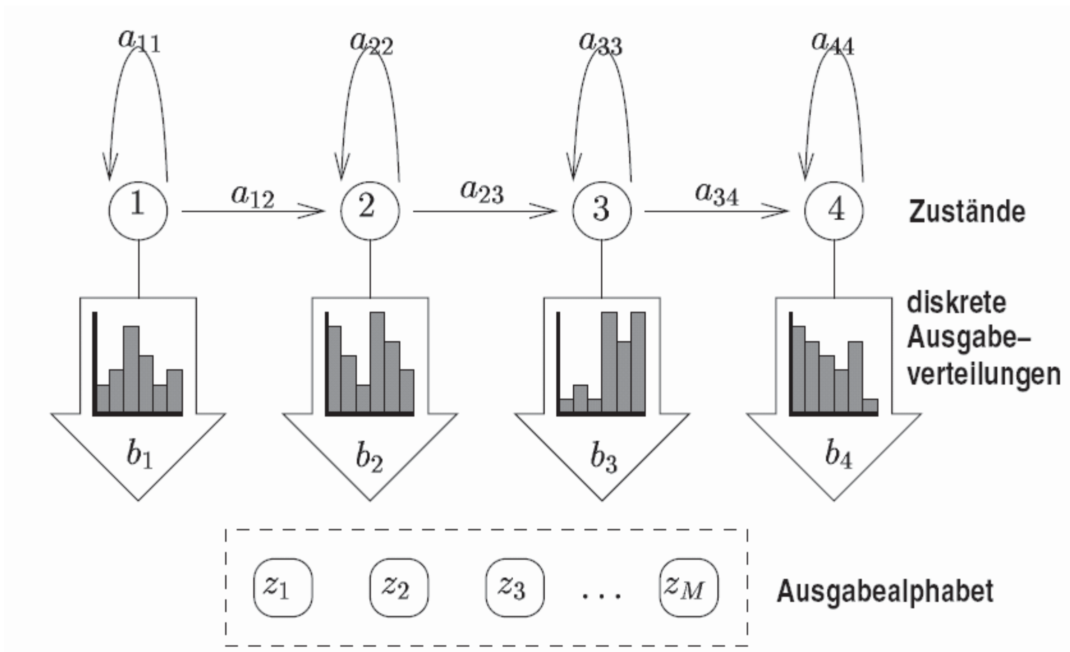
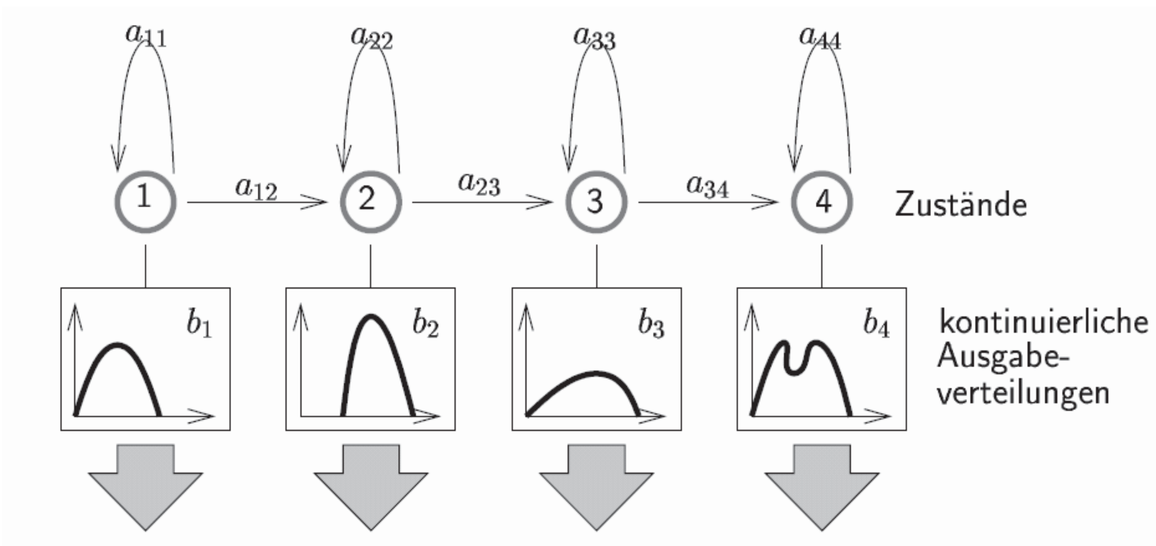
12. HMMs in Speech Recognition

Discrete Production Probabilities



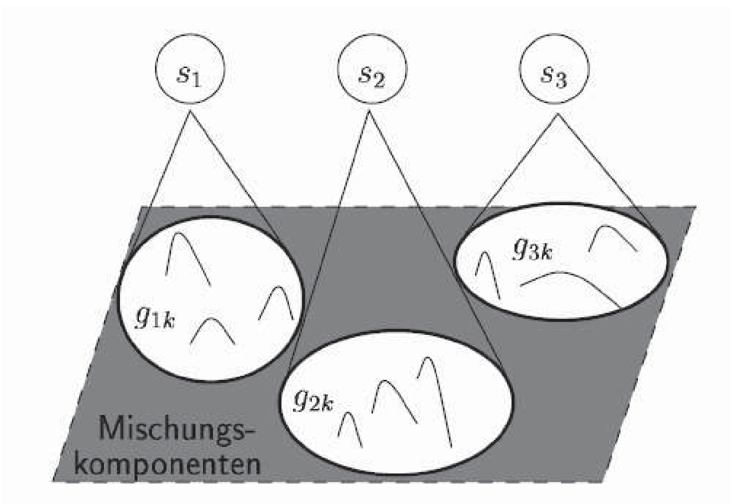
 From: Schukat-Talamazzini

Continuous Production Probabilities



 From: Schukat-Talamazzini

Continuous Production Probabilities



Each state has its own emission Probabilities (e.g. a GMM)

 From: Schukat-Talamazzini

Use GMMs as Production Probabilities

- Use EM-Algorithm to estimate
  - Transition probabilities a_{ij}
  - Emission probabilities $b_j(x_t)$
- "Continuous HMM"
- Two hidden properties:
  - State j
  - Id *k* of Gaussian

Update Parameters for GMM in HMM

Define auxiliary function $$\zeta_t(j,k) = \frac{P(\vec{x}, q_t = j, k_t = k \mid \theta)}{P(\vec{x} \mid \theta)}$$

$$= \frac{\sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij} p_{jk} b_{jk}(x_t) \beta_t(j)}{\sum_{i=1}^{N} \alpha_T(i)}$$

k: labels the Gaussian to be mixed

Mixture weights in state j: p_{jk}

Specific emission probability: $b_{jk}(x_t)$

Update Parameters for GMM in HMM

Update mixture weights:

$$\overline{p}_{j,k} = \frac{1}{\sum_{k'=1}^{M}\sum_{t=1}^{T}\zeta_t(j,k')}\sum_{t=1}^{T}\zeta_t(j,k)$$

Update means:

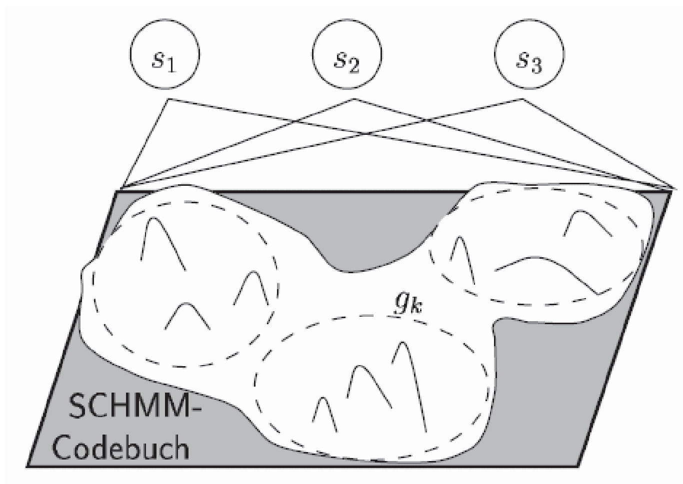
$$\overline{\mu}_{j,k} = \frac{1}{\sum_{t=1}^{T}\zeta_t(j,k)}\sum_{t=1}^{T}\zeta_t(j,k)x_t$$

Update covariances:

$$\overline{\Sigma}_{j,k} = \frac{1}{\sum_{t=1}^{T}\zeta_t(j,k)}\sum_{t=1}^{T}\zeta_t(j,k)(x_t-\mu_{j,k})(x_t-\mu_{j,k})^t$$

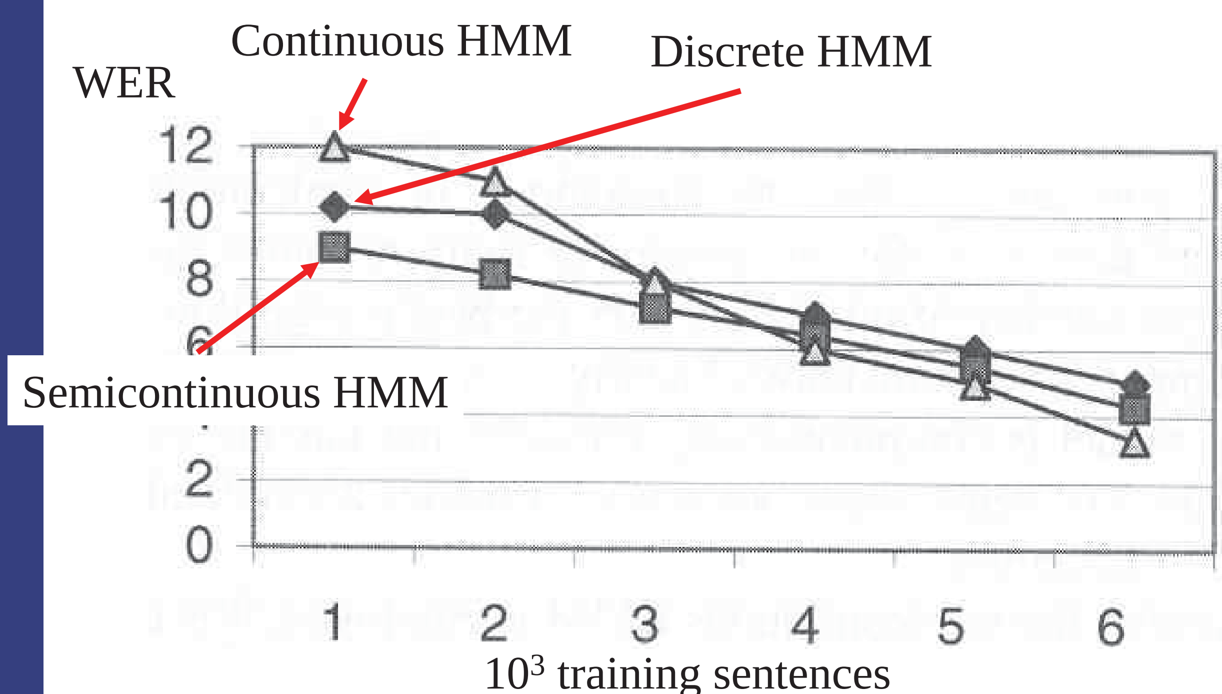
Semicontinuous HMMs

Idea: one set of Gaussians for all states
$\mapsto$mixture weights are state specific



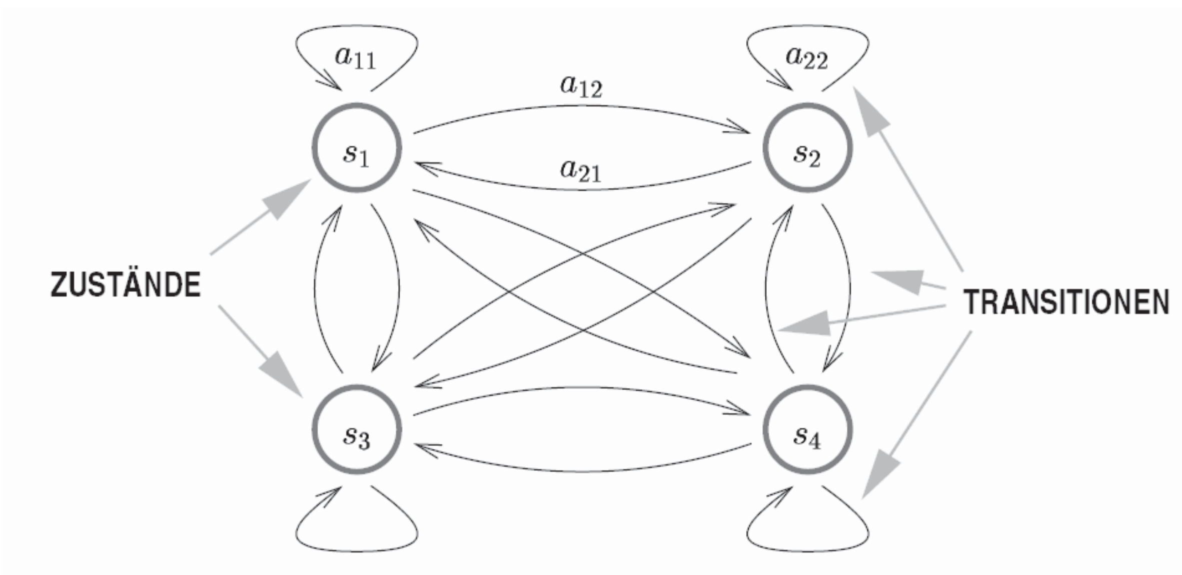
From: Schukat-Talamazzini

Comparing Discrete/Continuous HMMs



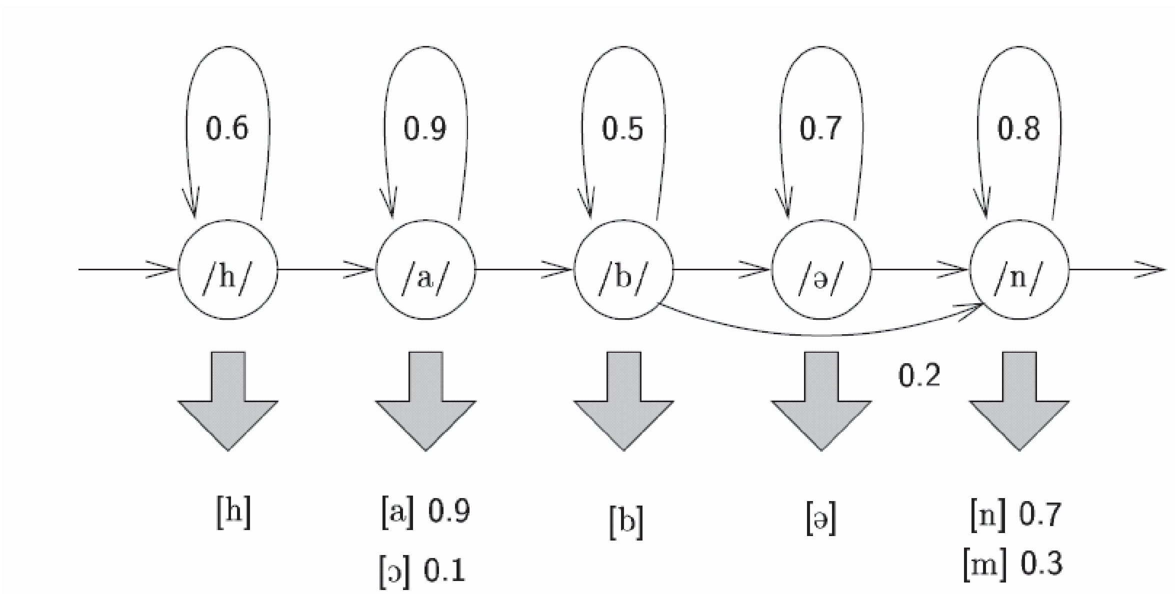
If you have enough training data use a continuous HMM

Constrain Transition Probabilities: Possible Model Topologies



 From: Schukat-Talamazzini

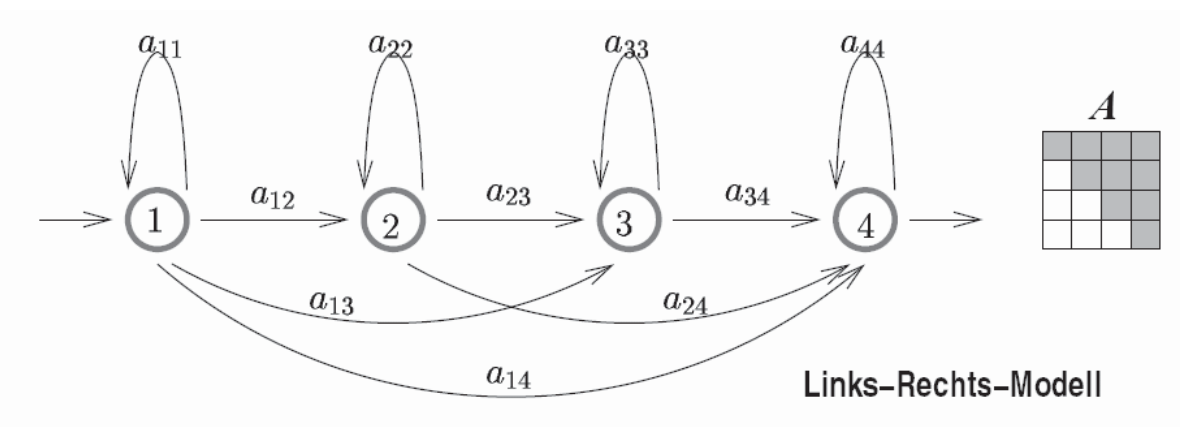
Structure of a Hidden Markov Model



Topology constructed on a case basis

 From: Schukat-Talamazzini

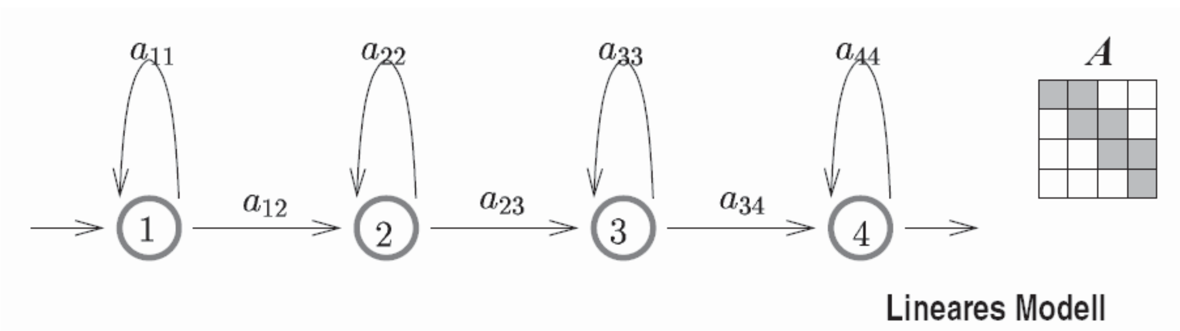
Standard Model Topologies: Left-to-Right Model



Idea: you never return in a phonem to a sound already articulated

 From: Schukat-Talamazzini

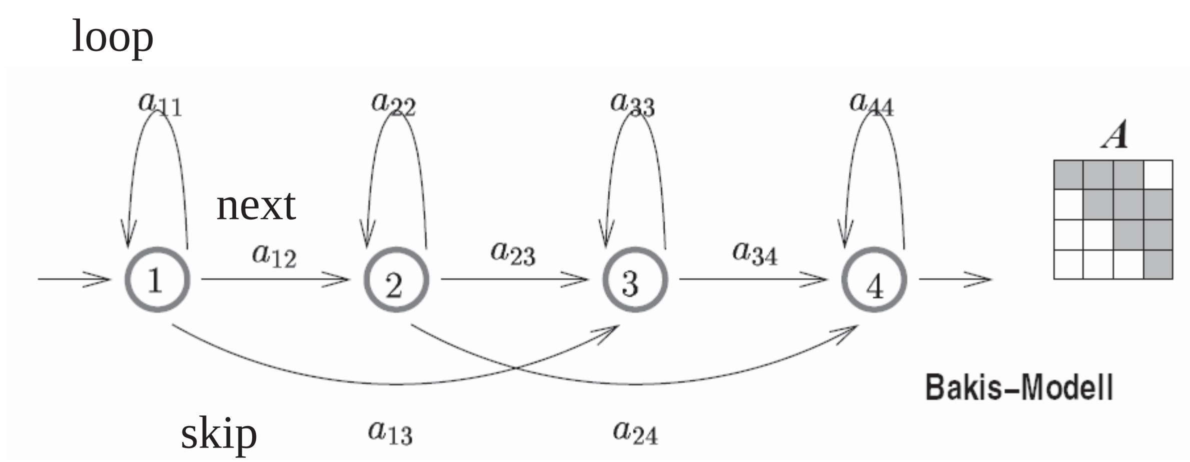
Standard Model Topologies: Linear Model



Idea: you may stretch a sound but you articulate all sounds that make up a phoneme

 From: Schukat-Talamazzini

Standard Model Topologies: Bakis Model



Idea: allow for a certain amount of sloppiness.
Bakis model is the one most widely used in ASR

 From: Schukat-Talamazzini

Summary

- Issues in speech recognition
  - Modelling emssions probabilities
  - Tying
  - Models for state transitions