セグメントモデルによる音声認識

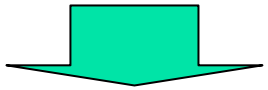
NTTコミュニケーション科学基礎研究所

南　泰浩

セグメントモデルとは？

HMMの欠点

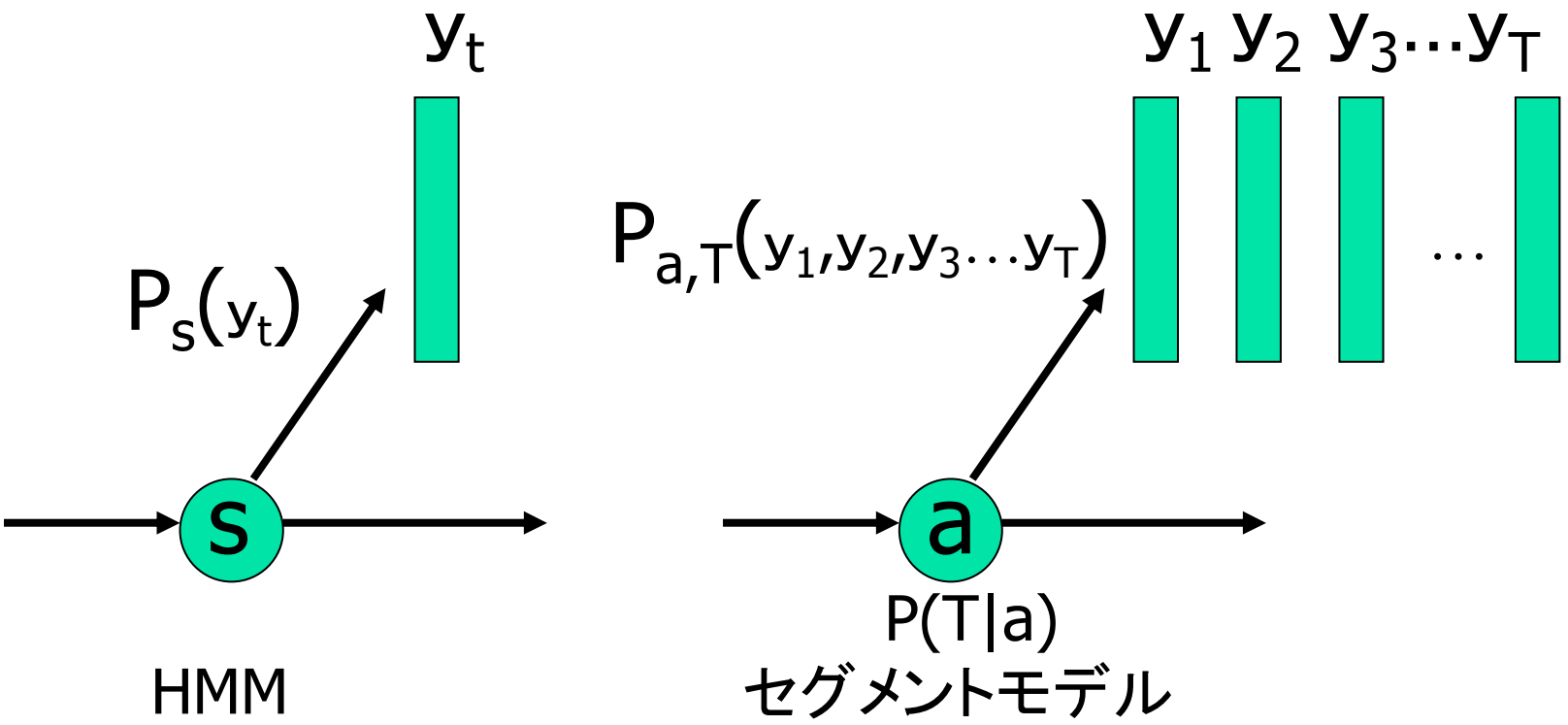
- 継続時間モデルが導入されていない
- 状態内の観測系列の時間依存性を反映できない

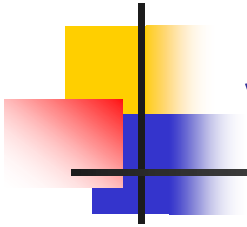


改良

セグメントモデル

HMMとセグメントモデルの違い

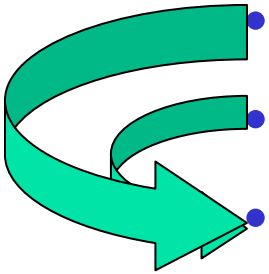




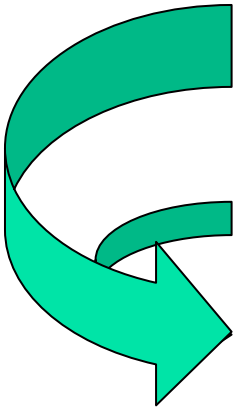
セグメントモデルの分類

- 継続時間長制御モデル
- 条件付ガウスモデル
- 動的システムモデル
- グラフィカルモデル
- （ベイジアンネットワーク）
- 制約付平均トラジェクトリモデル
- パラメトリック、ノンパラメトリック
- 非線形モデル
- 生成モデル

セグメントモデルの分類



- 継続時間長制御モデル
- 条件付分布（ガウス）モデル
- 動的システムモデル
- グラフィカルモデル
  （ベイジアンネットワーク）



- 制約付平均トラジェクトリモデル
  パラメトリック、ノンパラメトリック
- 非線形モデル
- 生成モデル

条件付分布(ガウス)モデル

- 出力確率を以下のように近似

$P_{a,T}(y_1,y_2,y_3\cdots y_T)$

$= \prod P(y_t|y_{t-1},a)$

(Wellekens, 高橋)

- より複雑な条件付確率

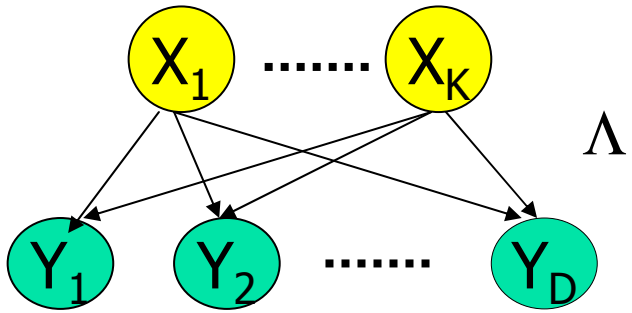
$P(y_t|y_{t-3},y_{t-2},y_{t-1},a)$ (中川)

グラフィカルモデル

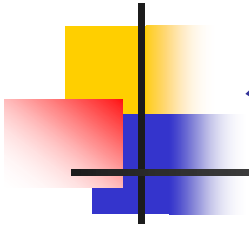
グラフィカルモデルとは？

依存関係をグラフを用いる統計モデルで表現したもの

例　因子分析、確率的主成分分析
　　独立成分分析



以下Ghahramani に基づくグラフィカルモデルの分類



グラフィカルモデル

無向グラフィカルモデル

マルコフネットワーク

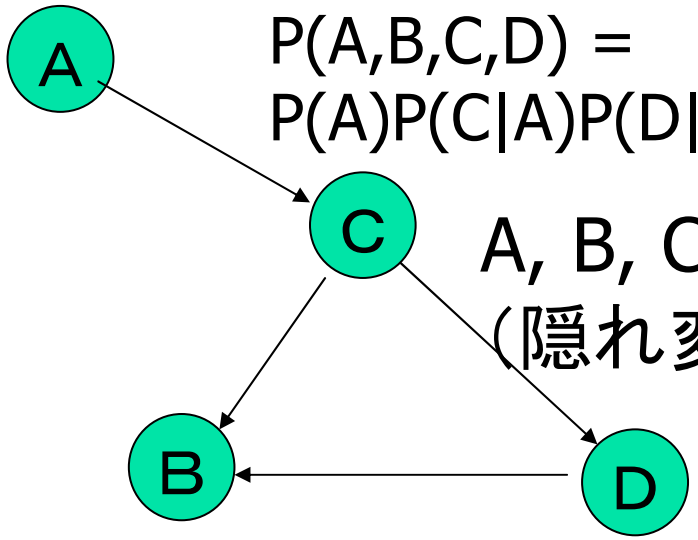
例　ボルツマンマシン

マルコフランダムフィールド

グラフィカルモデル

- 有向グラフィカルモデル、非巡回

⇒　ベイジアンネットワーク



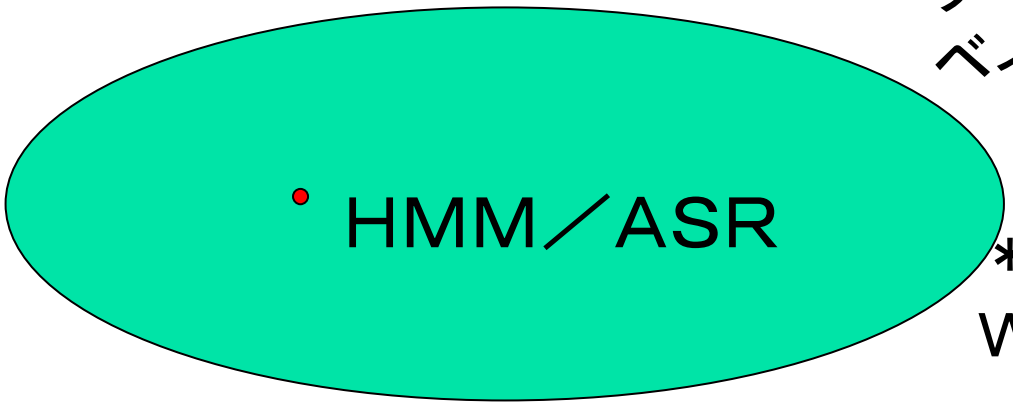
P(A,B,C,D) =
P(A)P(C|A)P(D|C)P(B|C,D)

A, B, C, Dは離散、連続確率変数
（隠れ変数の場合もある。）

グラフィカルモデル

- ダイナミックベイジアンネットワーク（DBN）
  ベイジアンネットワークを時系列変数
  を扱えるように拡張

HMM、線形動的システム
を含む大きなモデル



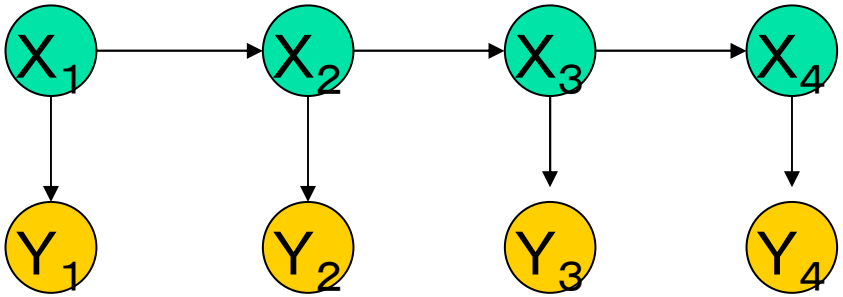
ダイナミック
ベイジアンネット

＊CLSP
WS2001より

グラフィカルモデル

- 例　HMMとの対応

（隠れ変数が離散値）



$P(x_1,x_2 \ldots x_T,y_1,y_2 \ldots y_T)=P(x_1)P(y_1|x_1)\ \Pi\ P(x_t|x_{t-1})P(y_t|x_t)$

$P(y_t|x_t)$: 出力確率（密度）

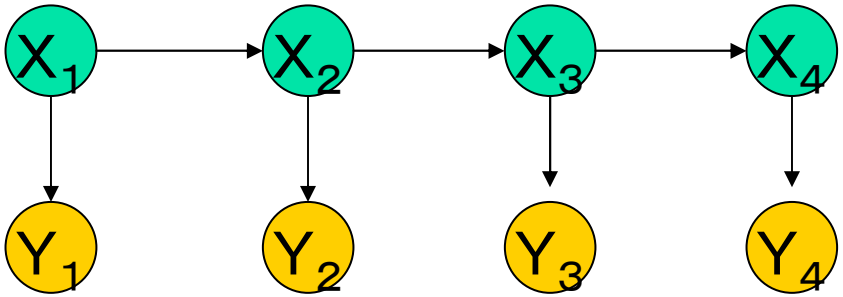
$P(x_t|x_{t-1})$:　遷移確率

X_t: 状態確率変数（離散値）

グラフィカルモデル

例　線形ガウス状態モデルとの対応

（隠れ変数が連続値）



$P(x_1,x_2 \ldots x_T,y_1,y_2 \ldots y_T)=P(x_1)P(y_1|x_1)\ \Pi\ P(x_t|x_{t-1})P(y_t|x_t)$

$P(y_t|x_t)$:　$\leftarrow y_t = Cx_t+v_t$

$P(x_t|x_{t-1})$: $\leftarrow x_t = Ax_{t-1}+w_t$

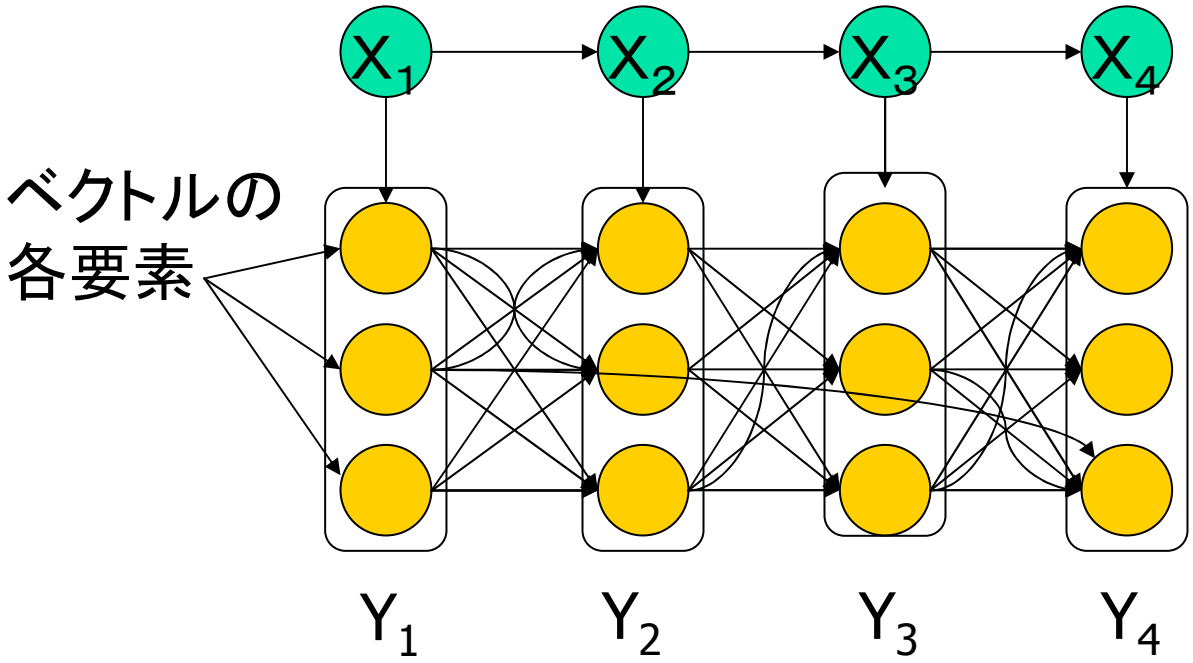
X_t: 状態確率変数（連続値）

w_t, v_t: 無相関、0平均ガウスノイズベクトル

グラフィカルモデル

- ベイジアンネットによる音声認識

  条件付分布(ガウス)モデルの拡張(Zweig)

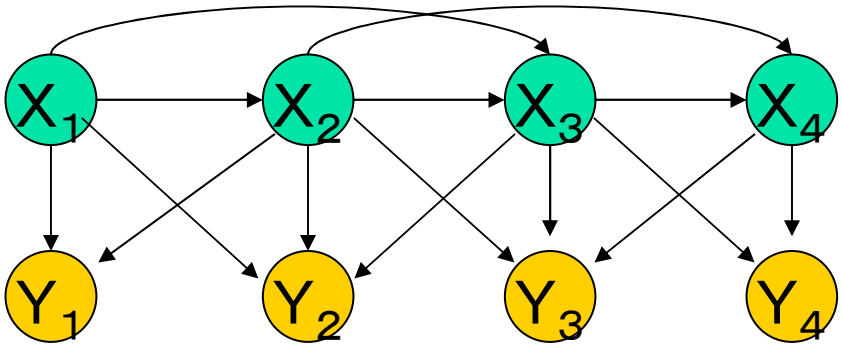


$P(Y_t|Y_{t-1})$を表現

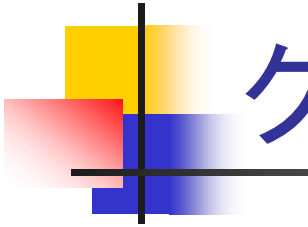
より複雑条件付モデル

グラフィカルモデル

- ベイジアンネットによる音声認識
  様々な変数の依存関係を記述できる

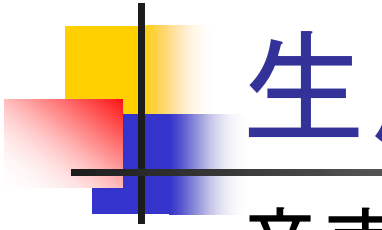


Deviren



グラフィカルモデル

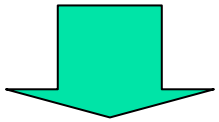
- ベイジアンネットによる音声認識
- モデル構造（依存関係）の決定も可能
  - 尤度＋MDL
  - 識別的構造学習法（Zweig）
- ベイジアンネットの難しさ
  - 理論　　実際の音声認識
- →GMTK：ツールキット(Bilmes)
  - パラメータ数　⇔　データ量



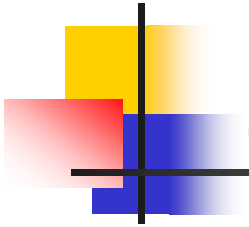
生成モデル

音声生成系を考慮して音声認識をモデル化

- 音声の連続性の拘束条件を導入したい。
- セグメント間の影響をモデル化したい



調音結合のモデル化



生成モデル

生成モデルの例

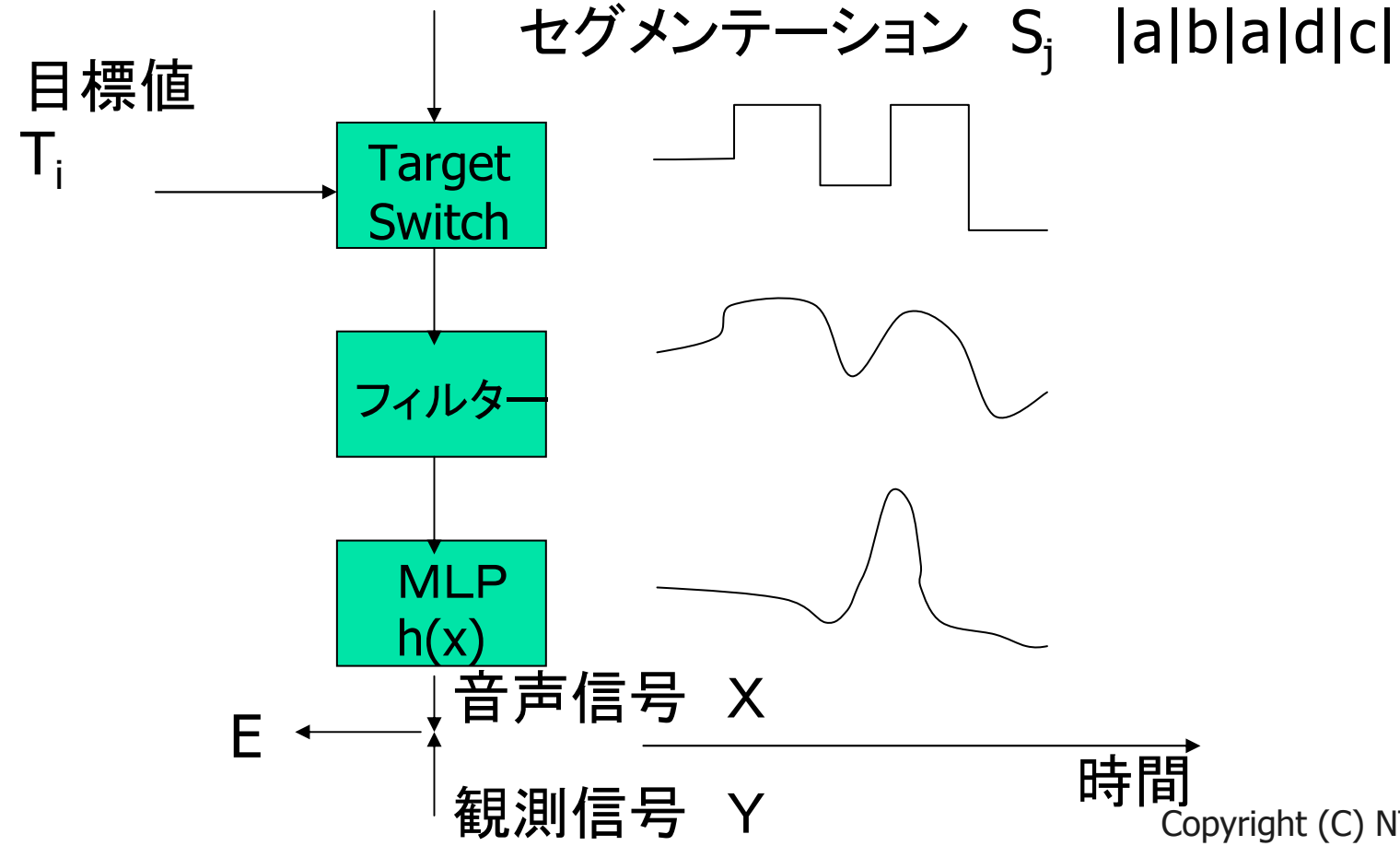
- 線形動的システムと非線形処理を組み合わせたもの

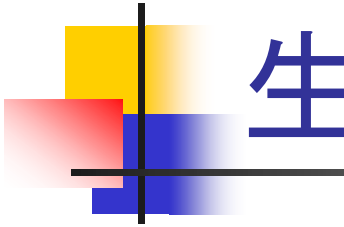
  Hidden Dynamic Model (Richards)

  Vocal Tract Resonance dynamic model (Deng)

生成モデル

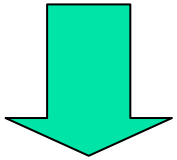
Hidden Dynamic Model (Richards, Bridle)





生成モデル

Center for Language and Speech Processing (CLSP) Summer Work Shop 98の結果

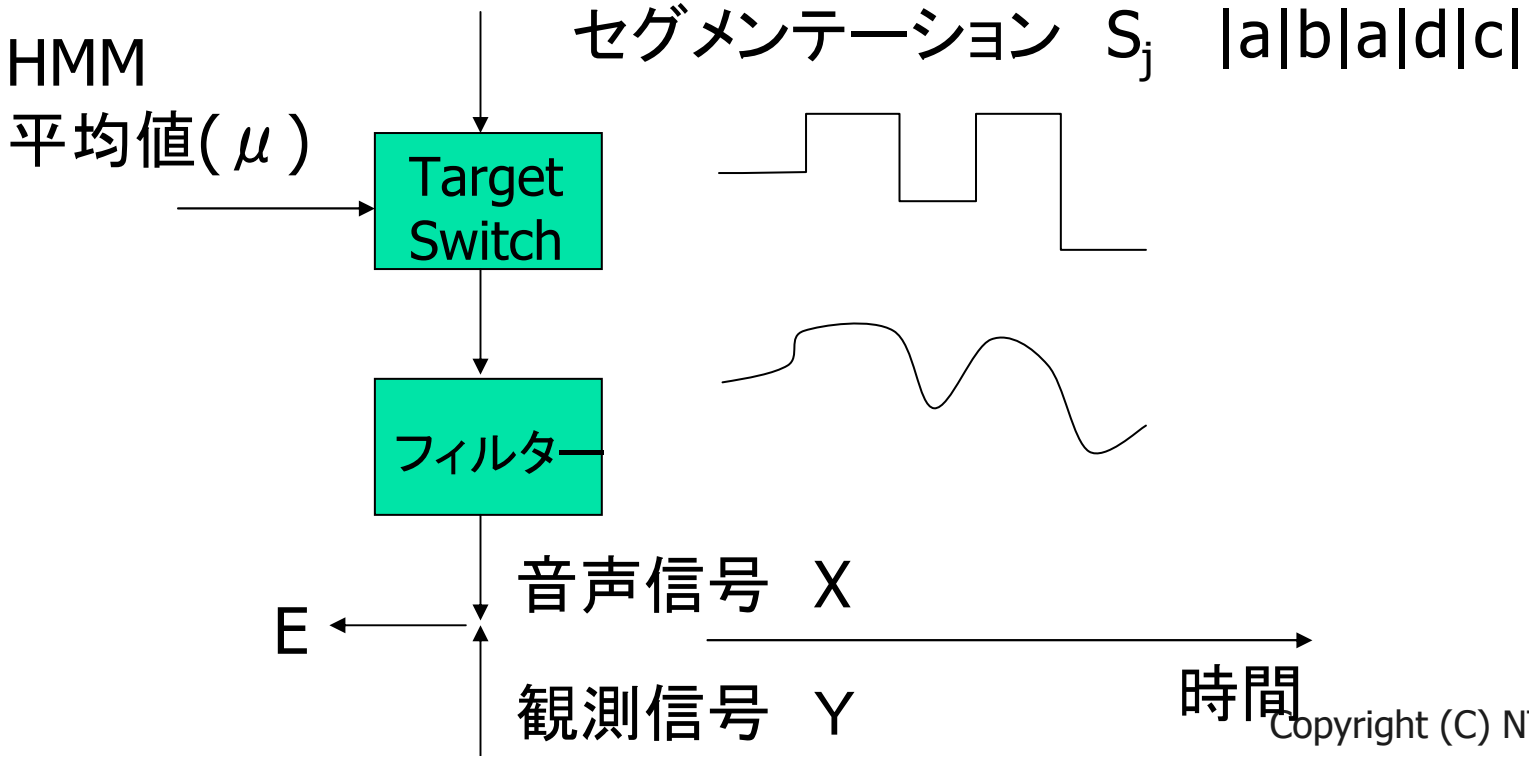


HMMに比べ優位ではない

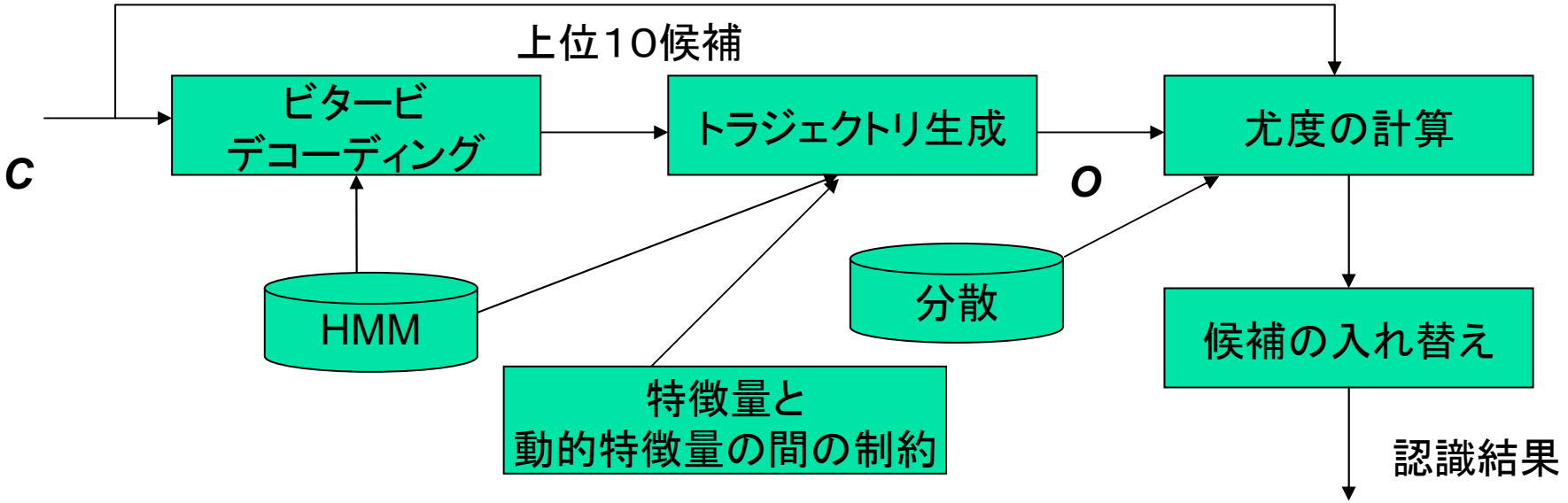
少なくとも音声特徴量上での連続性拘束は必要

生成モデル

HMMからのトラジェクトリ合成を使う(南)



HMMを用いて制約付平均トラジェクトリを生成する手法（南）



認識実験による評価

- 学習データ：音響学会の503音韻バランス文
- 各状態の正規分布数：3
- HMMタイプ：環境依存型HMM
- 評価データ：100都市発声（男女各35人）

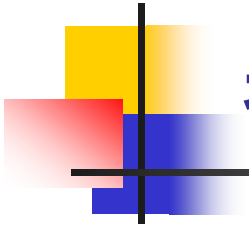
100都市認識での誤認識率

提案手法	HMM	エラー削減率
1.8%	2.2%	18.2%

連続性に関する拘束条件は重要

セグメントモデル

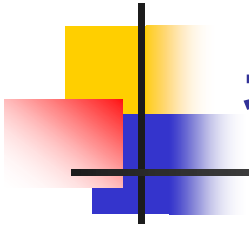
- グラフィカルモデル
  HMM、線形動的システムを含む記述力の高いモデル
  - 成功例（古山、Zweig、Deviren）
  - 学習データの量　⇔　パラメータの数
  - 数学モデル→実際の音声認識
- 生成モデル
  音声生成系を考慮して音声認識をモデル化
  - 成功例少
    - NTTの研究　→　パラメータの連続性は重要



参考文献

セグメントモデルの分類のために参照とした文献

- “From HMM’s to Segment Models: A Unified View of Stochastic Modeling for Speech Recognition”, M. Ostendorf et al., IEEE Trans. SAP, 1996.
- 音声認識においてHMMとトライグラムを超えるもの
  （中川　人工知能学会誌　2002年1月）.
- 音声認識研究の動向
  （中川　電子情報通信学会　D－II　2000年2月）.



参考文献

- 条件付ガウスモデル

（様々な文献があるが、以下のものだけあげておく）

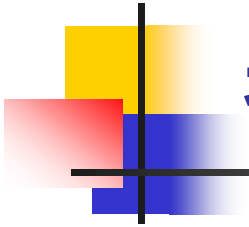
Wellekens et al, “Explicit correlation in hidden Markov model for speech recognition”, ICASSP 1987.（条件付ガウスモデルを提案した初期の文献）

高橋他、“フレーム間相関を利用した音韻HMMによる音声認識、信学論、1994.（データの少なさを平滑化）

中川他、“セグメント統計量を用いた隠れマルコフモデルによる音声認識”、信学論、1996.

（時間的に長い条件を導入）

より詳しく調べたい方は、Ostendorfの文献を参照してください。



参考文献

- グラフィカルモデル（音声認識音声認識関連）

Zweig et al. “Structurally discriminative graphical models for automatic speech recognition – results from the 2001 Johns Hopkins summer workshop”, ICASSP 2002.2.

Zweig et al. “Probabilistic modeling with Bayesian networks for automatic speech recognition”, Australian Journal of Intelligent Information Processing, 1999.

Deviren et al. “Structural learning of dynamic Bayesian networks in speech recognition”, Eurospeech 2001.

Murphy “Dynamic Bayesian networks: representation, inference and learning”, UC, Berkeley Dr. thesis, 2002.

Murphyのホームページ

http://www.ai.mit.edu/~murphyk/

Bilmes, “Buried Markov models for speech recognition”, ICASSP 1999.

Bilmes et al. “The graphical models toolkit: an open source software system for speech and time-series processing”, ICASSP 2002.

Ozgur Cetin et al. “The 2001 GMTK-based SPINE ASR system”, ICSLP 2002.

Bilmes の発表論文のページ

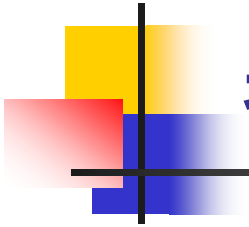
http://ssli.ee.washington.edu/people/bilmes/pubs-frame.html

グラフィカルモデルツールキットのページ

http://ssli.ee.washington.edu/~bilmes/gmtk/

WS2001のページ（CLSP）

http://www.clsp.jhu.edu/ws2001/groups/gmsr/



参考文献

- グラフィカルモデルの一般的な説明
  統計数理研究所公開講座のページ
  http://juban.ism.ac.jp/seminar.html
  Zoubin Ghahramani のレクチャページ
  http://www-2.cs.cmu.edu/~zoubin/SALD/
  “Learning dynamic Bayesian networks”
  “Statistical approaches to learning and discovery”
  グラフィカルモデル　朝倉書店
- ベイジアンネットの一般的な説明
  ベイジアンネットセミナー
  http://www.aist.go.jp/ETL/~motomura/bn2001/
  http://www.aist.go.jp/ETL/~motomura/bn2002/paper.html
  ベイジアンネットの概要
  http://www.etl.go.jp/~motomura/DS/

参考文献

- 調音器官モデル

Deng et al., “A statistical coarticulatory model for the hidden vocal-tract-resonance dynamics”, EUROSPEECH 1999.

Deng et al., “Spontaneous speech recognition using a statistical model of VTR-dynamics”, WS98 Slide.

http://www.clsp.jhu.edu/ws98/projects/dynamic/presentations/final/ldeng/sld001.htm

R. Togneri et al., “An EKF-based algorithm for learning statistical hidden dynamic model parameters for phonetic recognition”, ICASSP 2001.

Richards et al., “The HDM: a segmental hidden dynamic model of coarticulation”, ICASSP 1999.

Bridal et al., “An investigation of segmental hidden dynamic models of speech coarticulation for automatic speech recognition”, WS98 final report.

Picone et all., “Initial evaluation of hidden dynamic models on conversational speech”, ICASSP, 1999.