7. 音声の認識：高度な音響モデル

7.1 実際の音響モデル
7.2 識別的学習
7.3 深層学習

7.1 実際の音響モデル

- 混合分布の学習
  - 各音素の特徴ベクトルは、一つの正規分布で近似できるほど単純ではない
    例）男女差、方言、…
  - 複雑な確率密度関数を複数の正規分布の重み付き和で表現　→　混合分布

$$\phi = \sum_{i=1}^{N} w_i \phi_i$$

Φ_i：i 番目の正規分布
w_i：i 番目の正規分布の重み
N：混合数

  - 重みはEMアルゴリズムで学習

7.1 実際の音響モデル

- 話者適応
  - 不特定話者用音響モデルのパラメータを、少数の特定話者データを用いて調整
  - MLLR (Maximum Likelihood Linear Regression) 法
    - 学習済みHMMにおいて、平均ベクトルを以下の式で変換

$$\mu' = \boldsymbol{A}\mu + b$$

    - 特定話者データの尤度が最大となるような行列$\boldsymbol{A}$と定数項bを推定

7.2 識別的学習

- 学習データの尤度計算

$$p(\boldsymbol{W}|\boldsymbol{X}) = \frac{P(\boldsymbol{X}|\boldsymbol{W})P(\boldsymbol{W})}{P(\boldsymbol{X})} = \frac{P(\boldsymbol{X}|\boldsymbol{W})P(\boldsymbol{W})}{\sum_{\boldsymbol{W}} P(\boldsymbol{X}|\boldsymbol{W})P(\boldsymbol{W})}$$

  - 生成モデル：$P(\boldsymbol{X}|\boldsymbol{W})$ が大きくなるようにパラメータを求めた
  - 識別モデルの考え方：$\Sigma_{\boldsymbol{W}} P(\boldsymbol{X}|\boldsymbol{W})P(\boldsymbol{W})$ を小さくすればよい
    → 正解以外の単語列に対して $P(\boldsymbol{X}|\boldsymbol{W})$ が小さくなるように学習

- 相互情報量最大化基準

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \log P(\boldsymbol{W}|\boldsymbol{X})$$

$$= \arg\max_{\boldsymbol{\theta}} \sum_{r} \log \frac{P(\boldsymbol{W}_r, \boldsymbol{X}_r; \boldsymbol{\theta})}{\sum_{\tilde{\boldsymbol{W}}} P(\tilde{\boldsymbol{W}}, \boldsymbol{X}_r; \boldsymbol{\theta})}$$

$\tilde{\boldsymbol{W}}$：対立仮説
r：学習データのインデックス

7.3 深層学習

- DNN-HMM法
  - HMMの各状態で特徴ベクトルを出力する確率 $b_i(\boldsymbol{x})$ を $p(\boldsymbol{x}|s_i)$ と書き換え
  - ベイズの定理

$$p(\boldsymbol{x}|s_i) = \frac{P(s_i|\boldsymbol{x})}{P(s_i)} p(\boldsymbol{x})$$

DNNで計算

学習データから最尤推定

定数

  - $\boldsymbol{x}$ はMFCCではなく、メルフィルタバンクの出力（またはもとの音声信号）で特徴抽出もDNNで学習

7.3 深層学習

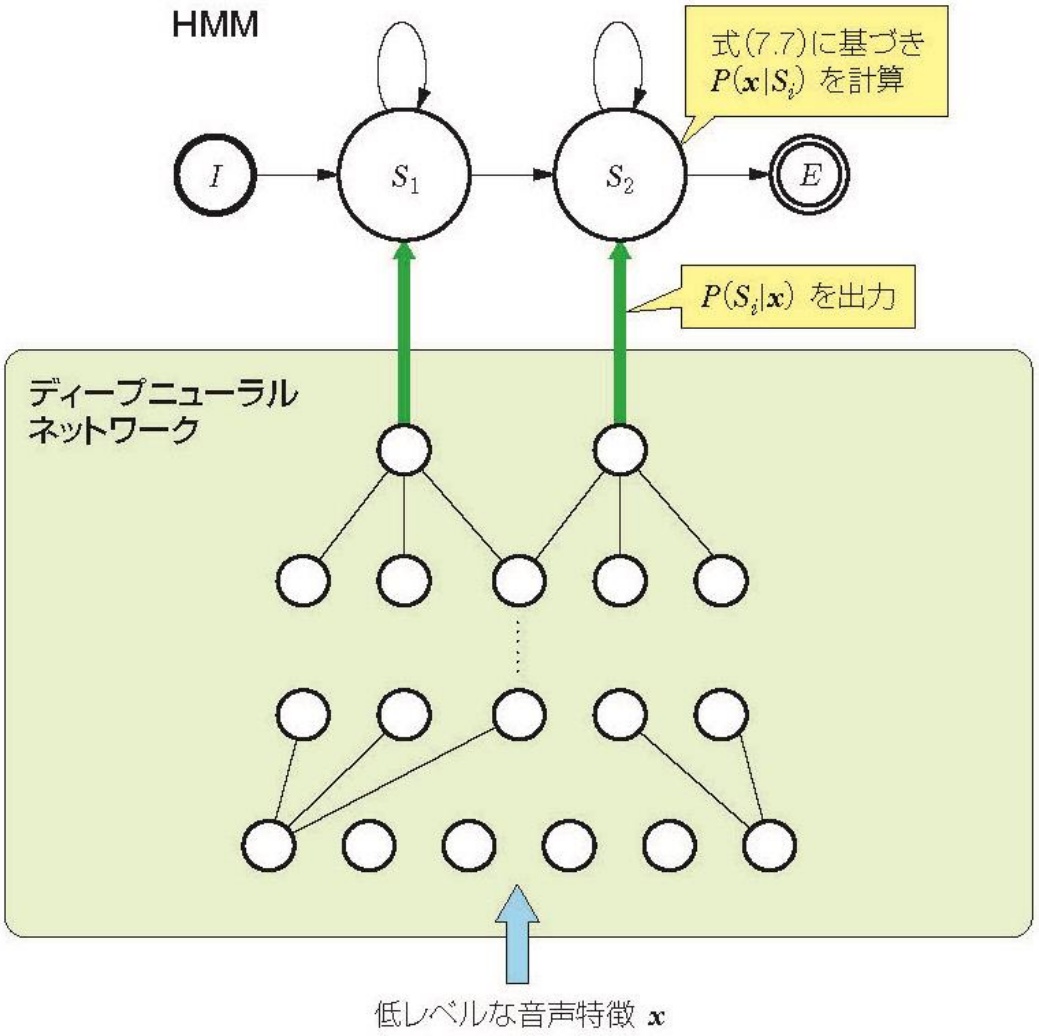

図 7.4 深層学習による音声認識