Decision Tree based Simultaneous Clustering of Phonetic Contexts, Dimensions and State Positions for Acoustic Modeling
Heiga Zen, Keiichi Tokuda, Tadashi Kitamura (Nagoya Institute of Technology, Japan)

1. Introduction
Context-dependent HMMs (ex. triphone HMMs)
→ Too many free-parameters in system
Unseen model

Phonetic Decision Tree (P-DT) based state tying
→ P-DT reduces free-parameters in a system
→ P-DT can generate unseen models
→ All dimensions have the same sharing structure
→ Each state position has different decision tree

All dimensions have the same context-dependency?
→ All dimensions have the same state-position-dependency?

Phonetic, Dimensional & State Positional Decision Tree (PDS-DT) based clustering technique
→ PDS-DT is an extension of P-DT
→ Can construct different tying structure for each dimension
→ Evaluate PDS-DT in speech recognition experiment

2. Phonetic Decision Tree-based State Tying Technique
P-DT based on ML criterion [S. J. Young et al.; 1994]
P-DT based on MDL criterion [K. Shinoda et al.; 1997]

Split \( S \) into \( S_p \) and \( S_{\bar{p}} \) by question \( q \),
The difference of DL value, \( \Delta \), is given by
\[
\Delta_q = \frac{1}{2} \left[ \sum \frac{\Gamma(S_q)}{\Gamma(S_{\bar{q}})} \log \frac{\sum \frac{\Gamma(S_q)}{\Gamma(S_{\bar{q}})}}{\Gamma(S_{\bar{q}})} + \sum \frac{\Gamma(S_{\bar{q}})}{\Gamma(S_q)} \log \frac{\sum \frac{\Gamma(S_{\bar{q}})}{\Gamma(S_q)}}{\Gamma(S_q)} \right] - (s(S_q) \log \frac{S_q}{\bar{S}_q} + s(S_{\bar{q}}) \log \frac{S_{\bar{q}}}{S_q})
\]
\( s \) : Dimensionality of Feature Vector
\( \Sigma \) : Covariance Matrix of Each Cluster
\( \Gamma(\cdot) \) : State Occurrence Count for Each Cluster

3. Phonetic & State Positional Decision Tree
→ Introduce questions about state positions into P-DT
→ Can construct state-tying structure across state positions
→ Almost the same as P-DT based state tying

MDL-based dimensional-split [H. Zen, et al., 2002]
→ Phonetic & Dimensional Decision Tree (PD-DT)
→ Phonetic & Positional Decision Tree (PD-DT)
→ PD-DT construct proper context-dependent sharing structure for grouped dimensions

4. Phonetic & Dimensional Decision Tree

 Phonetics, Dimensional & State Positional Decision Tree (PDS-DT) based clustering technique
→ Introduce dimensional-split technique into PS-DT
→ Phonetic contexts, dimensions and state positions are clustered simultaneously
→ Each distribution has different dimensionality
→ Each distribution is composed of different dimensions
→ Unified technique for decision-tree based acoustic modeling based on MDL criterion

5. Phonetic, Dimensional & State-positional Decision Tree based Clustering

6. Constructed Parameter Sharing Structure

7. Experimental Conditions

8. Example of constructed PDS-DT

9. Constructed HMMs

10. Recognition Experimental Results

[Tables and diagrams with data and comparisons, not transcribed here]