HMM-based very low bit rate speech coder
- Encoder: HMM-based phoneme recognition
- Decoder: HMM-based speech synthesis
- Efficient F0 quantization
- MSD-VQ (Multi-Space Probability Distribution VQ)
- Japanese system is proposed in 100 bit/s
  [Hoshiya et al., ’04]

Lower bit rate keeping speech quality
- VQ using phoneme decision tree
- Phonetic decision tree based CB selection
- Constructed this system for US English phoneme

HMM-Based Speech Coding System
- Feature of this system
  - Can to transmit only encoded phonemes, F0 sequences and state durations
  - Decoded voice is obtained by the HMM-based speech synthesis

Input Speech
- F0 Extraction
- Mel-Cepstral Analysis
- Segmentation
- F0
- MSD-VQ
- VQ

Encoder
- Parameter Generation
- Encoding
- Phoneme Sequence
- Phone Recognition
- State Duration
- Phonetetic HMM

Synthetic Speech
- MLSA Filter
- Huffman Coding
- Decision Trees

Decoding
- Parameter Generation
- Encoding
- Encoding
- Encoding

Constrained by the demonstration
- [http://kt-lab.ics.nitech.ac.jp/alex/demo.html]

Quantizing segmented F0 seq. for each phoneme HMM
- HMM
- Observation Vector $o$:
  - $l_0 = (o_1, o_2, ..., o_T)$
  - $S_0 = (S_1, S_2, S_3)$

VQ Using Phonetic Decision Tree
- Combination of phoneme decision tree that considers phonological information and VQ
  - Conventional [Hoshiya et al., ’04]
  - VQ using phoneme decision tree
    - A codebook is trained at each leaf node
    - Each codebook has a fixed size
  - Proposed
    - Construct decision trees with different sizes
    - Shrink large tree to small one
    - Each leaf node has codewords
    - Apply Huffman coding at each leaf node

Experimental Conditions
- Database
  - CMU ARCTIC English corpus speaker BDL (1131 sentences)
- Training data: 1091
- Test data: 40

Analysis conditions
- Sampling frequency: 16 kHz
- Frame shift: 5 ms
- Analysis conditions: STRAIGHT

Result
- Bit rates
  - Can transmit at 0.1 kbit/s → 0.2 kbit/s
    - The transmission bit rate of the cellular phone is about 10 kbit/s
- Decoding speech
  - A part of the synthetic speech is heard by the demonstration
    - [http://kt-lab.ics.nitech.ac.jp/alex/demo.html]

Conclusion
- Conclusion
  - Examined the new quantization technique based on the phonetic decision tree
  - Constructed HMM based very low bit rate speech coder for US English
- Future works
  - Improvement of recognition accuracy
  - Need to lower the bit rate of phoneme