Recent development of the HMM-based speech synthesis system (HTS)

Heiga Zen, Keiichiro Oura,
Takashi Nose, Junichi Yamagishi,
Shinji Sako, Tomoki Toda,
Takashi Masuko, Alan W. Black,
and Keiichi Tokuda

Outline

HMM-based speech synthesis

- Corpus-based speech synthesis
- Overview of HMM-based speech synthesis

HTS features in the past releases

- HTS version 1.0 ~ 1.0.1
- HTS version 2.0 ~ 2.0.1
- HTS version 2.1

Future release plans

Outline

HMM-based speech synthesis

- Corpus-based speech synthesis
- Overview of HMM-based speech synthesis

HTS features in the past releases

- HTS version 1.0 ~ 1.0.1
- HTS version 2.0 ~ 2.0.1
- HTS version 2.1

Future release plans

Corpus-based speech synthesis

Unit selection synthesis

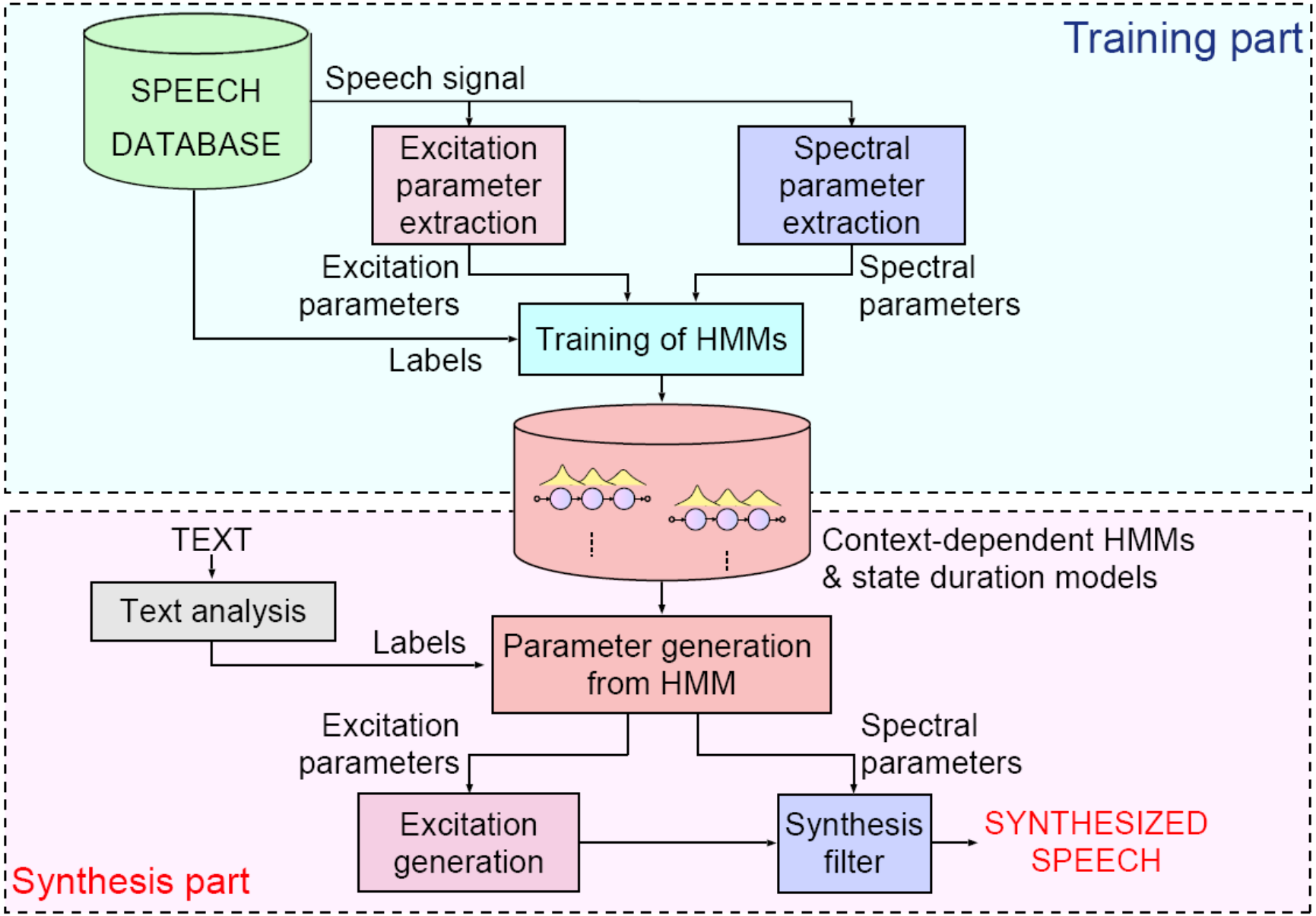
- Selects appropriate units from a speech database
- High quality (but sometimes discontinuous)
- Difficult to change its voice characteristics

HMM-based synthesis

- Generates speech parameters from statistical models
- Vocoded (but smooth & stable)
- Easy to change its voice characteristics

In last years, HMM-based approach is getting popular

Overview of HMM-based speech synthesis



HMM-based speech synthesis system

HTS: A toolkit for HMM-based speech synthesis

- Website: http://hts.sp.nitech.ac.jp/
- Provides a research platform for HMM-based synthesis
- Released as a patch code for HTK
- Open source, a new & simplified BSD-style license
- Over 5,000 downloads
- Used in various organizations
  e.g.) MSRA, iFlyTek, ATR, Cambridge, Edinburgh,
  CMU, KTH, IDIAP, DFKI, Bonn, & others…

Outline

HMM-based speech synthesis

- Corpus-based speech synthesis
- Overview of HMM-based speech synthesis

HTS features in the past releases

- HTS version 1.0 ~ 1.0.1
- HTS version 2.0 ~ 2.0.1
- HTS version 2.1

Future release plans

Features in the past releases ~ ‘02

HTS version 1.0 (Dec. 2002)

- Tree-based clustering based on the MDL criterion
- Stream-dependent tree-based clustering
- Multi-space probability distributions (MSD) for F0
- State-duration modeling & clustering
- Speech parameter generation algorithm
- Demo scripts using the CMU Communicator database

⇒ Speaker-dependent HTS voices can be constructed

Features in the past releases ~ ‘03

HTS version 1.1 (May 2003)

- Small run-time synthesis engine
- Demo scripts using the CSTR TIMIT database
- HTS voices for the Festival speech synthesis system

HTS version 1.1.1 (Dec. 2003)

- Variance flooring for MSD-HMMs
- Post-filtering
- HTS voices for the Galatea toolkit
- Demo scripts using the CMU ARCTIC database
- Demo scripts using the Nitech Japanese database

Features in the past releases ~ ‘06

HTS version 2.0 (Dec. 2006)

- Based on HTK version 3.4
- A number of new features
  - Speaker adaptation of multi-stream MSD-HMMs
  - Speaker adaptive training
  - EM-based speech parameter generation algorithm
  - Various model structures (e.g., full cov.)
  - Others…
- Bug fixes
- Speaker adaptation & adaptive training demo scripts

Features in the past releases ~ ‘07

HTS version 2.0.1 (Oct. 2007)

- Band structure for linear transforms
- Speaker interpolation
- Stream-dependent variance flooring scales
- Demo scripts support LSP-type spectral parameters

hts_engine API version 0.90 (Oct. 2007)

- A small stand-alone run-time synthesis engine
- A new & simplified BSD license

Features in the past releases ~ ‘08

HTS version 2.1 (Jul. 2008)

- Hidden semi-Markov model (HSMM) [Zen; ‘07]
- The speech parameter generation algorithm considering global variance (GV) [Toda; ‘07]
- Advanced adaptation technique (CSMAPLR) [Nakano; ‘06]
- Demo scripts using the STRAIGHT analysis/synthesis

hts_engine API version 1.00 (Jul. 2008)

- Stable version

Flite+hts_engine version 0.90 (Sep. 2008)

- An English TTS for embedded devices

Outline

HMM-based speech synthesis

- Corpus-based speech synthesis
- Overview of HMM-based speech synthesis

HTS features in the past releases

- HTS version 1.0 ~ 1.0.1
- HTS version 2.0 ~ 2.0.1
- HTS version 2.1

Future release plans

Internal versions

Additional features provided in the internal versions

- Variational Bayes [Nankaku; ‘03]
- Trajectory HMMs [Zen; ‘07]
- Minimum generation error training [Wu; ‘06]
- Shared tree construction [Yamagishi; ‘01]
- Eigenvoice [Shichiri; ’02]
- Multiple linear regression HMMs [Nose; ‘06]

Future release plans (Dec. 2009)

HTS version 2.1.1

- Based on HTK version 3.4.1
- Bug fixes & minor changes
- Context-dependent GV (?)
- Demo scripts for singing voice synthesis (?)

hts_engine API version 1.02

- Context-dependent GV (?)

Flite+hts_engine version 1.00

- Stable version

OpenJTalk version 1.00

- Japanese TTS using hts_engine API (?)

Open discussion

What functions should we add to HTS-2.1.1?

- Context-dependent GV
- Demo scripts for singing voice synthesis
- Post-filter for LSP parameter
- Mixed excitation
- State index question in context clustering
- HSMM based forced alignment tool using WFST
- Eigenvoice
- DAEM algorithm
- Phone duration modeling
- Other functions…

Thank you!!