

Four areas of synthesis

1. Architecture
2. Text processing
3. Linguistic/prosodic processing
4. Waveform synthesis

Architecture

- Utterance structure:
 - Relations, items and features
- List, trees, lattices of items
- Uniform access method:
 - pathnames:
 - n. R.:SYLStructure.parent.stress

Text processing

From files of characters to words

- Tokenization
- Token type identification:
 - numbers (years, dates etc)
 - words, abbreviations, letter sequences
 - homograph disambiguation
- Token to word(s):
 - numbers, money, expansion etc.

Linguistic processing

From words to phones, duration and F0

- Lexical Lookup:
 - LTS for out of vocabulary
- Prosody:
 - Intonation: accents and F0
 - Duration
- Post-lexical rules:
 - word context dependent pronunciations

Waveform synthesis

From phones, duration and F0 to waveforms

- Synthesizer types:
 - Articulatory
 - Formant
 - Concatenative

- Concatenative synthesis:
 - diphones
 - unit selection
 - limited domain (word/phrase concatenation)

Synthesis

- From **utterance** (text, tokens etc.)
 - apply **modules**
 - creating **relations** and **items**

Text to speech

- Tokenize raw file
- Chunk tokens into utterances
- Apply tts functions (synthesize and play)

Text to speech

From utterances to waveforms

- Tokens to words
- POS identification
- Phrasing
- Words to syllables and phones
- Accent assignment
- Post lexical rules
- Segmental durations
- F0 contour generation
- waveform generation

Tailoring the synthesizer

simple mode specific

- email, addresses, task specific
- lexical forms
- token to word rules
- prosodics:
 - global duration
 - F0 range

Tailoring the text

More help from the application

- Specific markup:
 - Sable, XML markup
 - breaks,
 - voices,
 - token identification
- Lexical restrictions
- Phrasal restrictions

Optimizing both

Limited domain synthesis

- Add specific mode
- Build limited domain waveform synth:
 - *only* speaks in domain
 - most common phrases in database
 - slot/filler type applications

Spoken Output

- Natural:
 - human sounding, appropriate style
- Flexible:
 - Easily controllable
- Efficient:
 - low resources at run-time and build-time

Project presentations

- 13th May and 15th May
- 10:30-12:00 short presentations:
 - 15 minute talks
 - with demos
 - hand in short write-up by end of that day
 - Tell me if you other constraints