

**“From Text to speech: the MITalk system”**

Jonathan Allen, M. Sharon Hunnicutt and Dennis Klatt  
with Robert C. Armstrong and David B. Pisoni

# Overview

- Overview: 12 pages
  - history of early TTS systems
- Analysis: 59 pages
  - Text pre-processing, Mophonological analysis
  - phrase-level parser, morphophonemics
  - letter-to-sound and lexical stress
- Synthesis: 79 pages
  - overview, phonology, prosody
  - (durations, F0), phonetics
  - klatt formant synthesizer
- Evaluation: 16 pages
  - intelligibility, comprehension
  - phoneme recognition, words recognition,
  - comprehension
- Implementation: 5 pages:

~~TTS and software~~

# Analysis

- Finite list of all Text Processing issues:
  - abbreviations etc.
- There was no large lexicon:
  - context dependent LTS rules
  - exception list
  - morphological decomposition
  - stress assignment
  - morphophonemics (post lexical rules)

# Synthesis

- Phonological component:
  - phones, stress, metrical boundaries
  - post lexical rules, co-articulation effects
- Prosodic component:
  - 11 Klatt Duration rules
  - (what we not train with CART)
  - F0 Contour Generator (O'Shaughnessy))
  - Sentence, Clause, Word contours with modification
- Phonetic component:
  - post prosody so can care about timing
  - generate formant values
  - smooth between them

# **Klatt Formant Parameters**

Param vector every 5ms

- 1 AV amplitude of voicing in dB
- 2 AF amplitude of frication in dB
- 3 AH amplitude of aspiration in dB
- 4 AVS amplitude of aspiration in dB
- 5 F0 amplitude of sinusoidal voicing in dB
- 6 F1 first formant frequency in Hz
- 7 F2 second formant frequency in Hz
- 8 F3 third formant frequency in Hz
- 9 F4 fourth formant frequency in Hz
- 10 FNZ nasal zero frequency in Hz
- 11 B1 first formant bandwidth in Hz
- 12 B2 second formant bandwidth in Hz
- 13 B3 third formant bandwidth in Hz
- 14 A2 second parallel formant amplitude in Hz
- 15 A3 third parallel formant amplitude in Hz
- 16 A4 fourth parallel formant amplitude in Hz
- 17 A5 fifth parallel formant amplitude in Hz
- 18 A6 sixth parallel formant amplitude in Hz
- 19 AB bypass path amplitude in db

# The Klatt Formant Synthesizer

- cascade vs parallel generation
- Sounds sources:
  - periodic/white noises (voiced/unvoiced)
- Hirose: prosody synthesis
- Tokuda: HMM Generation
- Articulatory/Speech Production:
  - Honda, Bateson etc
- Lots of early work in synthesis:
  - Fujimura, Bell Labs

## The Klatt Formant Synthesizer Params

- Voicing/Frication/Aspiration
- Formants (Hz, amplitude, bandwidth)
- Glottal resonator
- Nasal parameters
- Others

Each require generators



## How do you find the params?

- By expert crafting
- Analysis/Synthesis:
  - try all until you find them
- Are these the right parameters:
  - can we derive this for a large data set

# Intelligibility and Comprehension

- Modified Diagnostic Rhyme Test (MRT)
  - bad back bin bass bat bath ...
- Harvard Psychoacoustic Sentences:
  - The birch canoe slid on the smooth planks
  - 92.7% word correct. 21 listeners/500 utts total
- Haskins anomalous Sentences
  - The wrong shot led the farm
  - 78.7% current 23 listeners
- Comprehension:
  - paragraph of text
  - answer comprehension questions

Note not MOS, but measurement of understandability