# "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
- I Evaluation: 16 pages
- intelligibility, comprehension
- phoneme recognition, words recognition,
- comprehension
- □ Implementation: 5 pages:

TTL and software

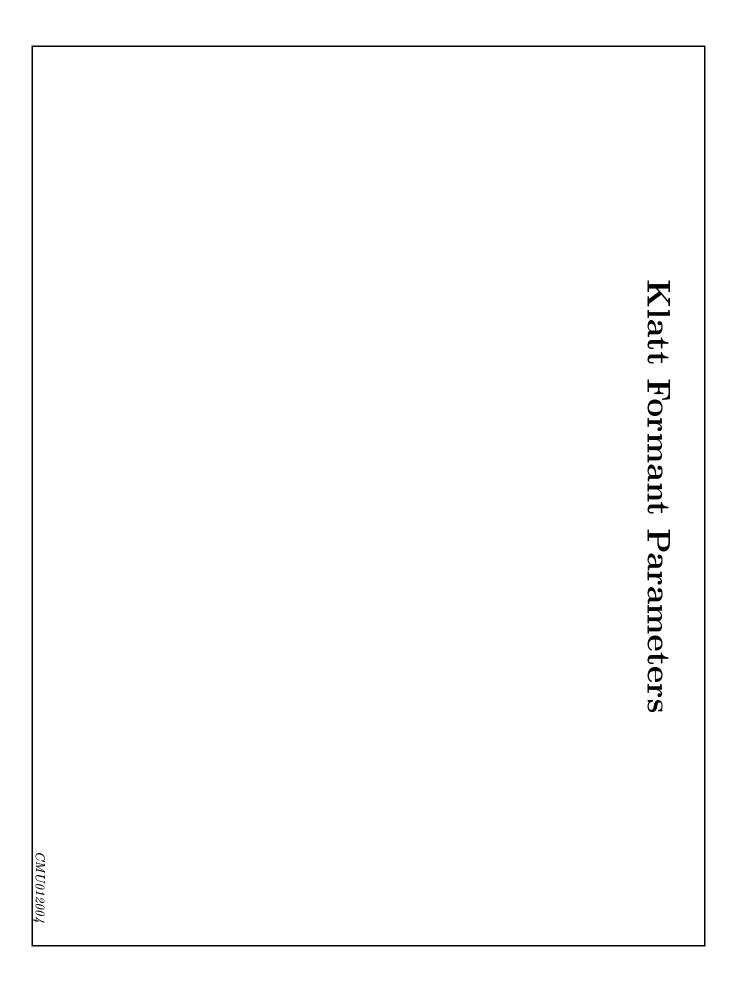
 $MU01200_{2}$ 

### Analysis

- ☐ Finite list of all Text Processing issues:
- abreviations 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
- I Phonetic component:
- post prosody so can care about timing
- generate formant values
- smooth between them



## Param vector every 5ms

- amplitude of voicing in dB
- AF amplitude of frication in dB
- AH amplitude of aspiration in dB
- AVS amplitude of aspiration in dB
- amplitude of sinusoidal voicing in dB
- F1first formant frequency in Hz
- F2second formant frequency in Hz
- F3third formant frequency in Hz
- F4fourth formant frequency in Hz
- 10 FNZ nasal zero frequency in Hz
- В1 first formant bandwidth in Hz
- B2 second formant bandwidth in Hz
- В3 third formant bandwidth in Hz
- A2 second parallel formant amplitude in Hz
- A3third parallel formant amplitude in Hz
- , A5 fourth parallel formant amplitude in Hz 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

- $\square$  cascade vs parallel generation
- □ Sounds sources:
- periodic/white noices (voiced/unvoices)
- ☐ Hirose: prosody synthesis
- □ Tokuda: HMM Generation
- Articulatory/Speech Production:
- Honda, Bateson etc
- $\square$  Lots of early work in synthesis:
- Fujimura, Bell Labs

#### Each require generators □ Others □ Nasal parameters □ Glottal resonator □ Formants (Hz, amplitude, bandwidth) □ Voicing/Frication/Aspiration The Klatt Formant Synthesizer Params

### $\square$ Are these the right parameters: $\square$ By expert crafting □ Analysis/Synthesis: - try all until you find them can we derive this for a large data set How do you find the params?

## Intelligibility and Comprehension

- □ Modified Diagnostic Rhyme Test (MRT)
- bad back bin bass bat bath ...
- 1 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