TTS Evaluation

Julia Hirschberg
CS 4706
TTS Evaluation

- Intelligibility Tests
- Mean Opinion Scores
- Preference Tests
Intelligibility Tests

• Diagnostic Rhyme Test (DRT)
  – Listening test
  – Listeners choose between two words differing by a single phonetic feature (voicing, nasality, sustenation, sibilation)
  – DRT: 96 rhyming pairs
    • Dense/tense, bond/pond, ...
      – Subject hears dense, chooses either dense or tense
      – % of correct answers is intelligibility score
  – Problem: Only tests single word synthesis
• Modified DRT:
  – 300 words, 50 sets of 6 words (*went*, *sent*, *bent*, *tent*, *dent*, *rent*)
  – Embedded in carrier phrases:
    • Now we will say *dense* again

• Mean Opinion Score
  – Have listeners rate output on a scale from 1 (bad) to 5 (excellent)

• Preference tests:
  – Reading addresses out loud, reading news text, using two different systems or systems against human voice
  – Do a **preference test** (prefer A, prefer B)