Creating Voices for a Diphone Based Text to Speech System

Supervisors: A. Lobb
S. Bangay

RHODES UNIVERSITY
Where leaders learn

M. Hood—g01h0708@campus.ru.ac.za
http://www.cs.ru.ac.za/research/students/g01H0708/

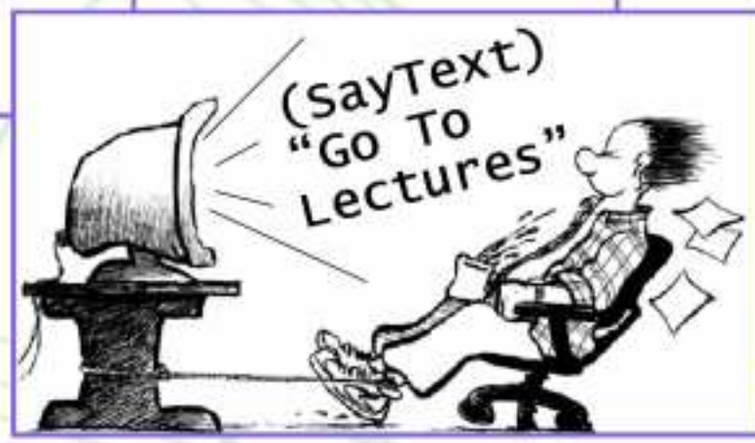
Festival:

- The Text to Speech system I use.
- Developed at Centre for Speech Technology Research at Edinburgh University.
- Open source and free to researchers.
- Allows multiple languages and voices.
- Has FestVox voice generation tools.

Recording:

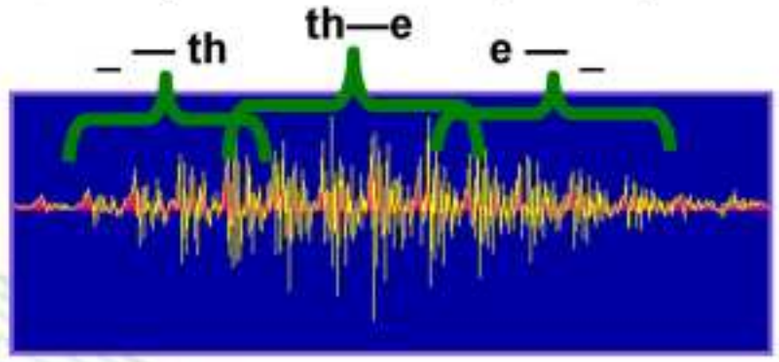
- First step in creating a voice.
- Record all the diphones that make up a language
- Repeat sounds after the computer.

```
[mly59@mly59 3 ru_us_matt_diphone after esu]$ bin/prompt_thea
 etc/usdiph.list 1000
1000 ( us_1000 "pau t aa s - ch aa t aa pau" ("s-ch") )
start recording for 2 seconds ...
... end recording
1001 ( us_1001 "pau t aa s - jh aa t aa pau" ("s-jh") )
start recording for 2 seconds ...
... end recording
```

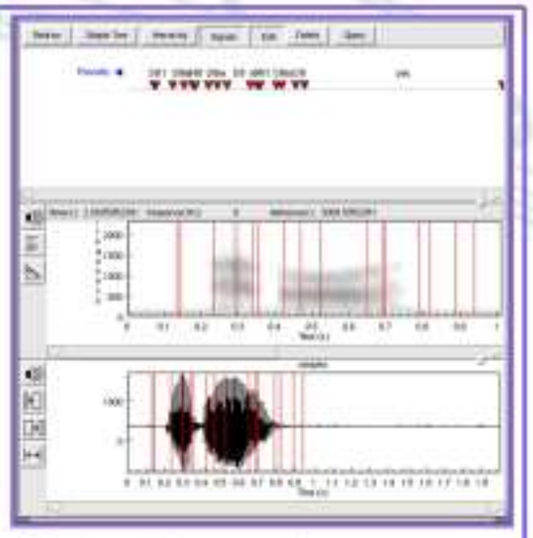


Diphones:

- Diphones are phonetic pairs
- Speech is made not by joining the phonetic sounds, but a pair at a time. E.g. to say "the"



Labelling:



- After recording, the phonetic sounds must be labelled.
- Auto labelling not very successful.
- Tedious, human intensive process.
- Labelling makes a huge difference to final voice quality.

Progress:

- Recorded 1396 diphone pairs in US-English.
- Auto labelled sound clips.
- Created voice: ru_us_matt_diphone.
- Continue work on labelling to improve the voice quality.
- Look into voice adaptation as an alternative to creating from scratch.