



Speech Processing 15-492/18-492

Speech Synthesis
Building Voices

Building a Voice

- ◆ *Designing the Prompts*
- ◆ *Recording the Prompts*
- ◆ *Labeling the Utterances*
- ◆ *Finding parameters (F0, MCEP)*
- ◆ *Building the synthesis voice*
- ◆ *Tuning and Testing*

Software Requirements

◆ *Festival Speech Synthesizer*

- *Free software language independent synthesizer*
- *Multiplatform: Windows, Linux, OSX*
- *Used for research and commercial synthesis*

◆ *Festvox*

- *Voice building tools*
- *Scripts, instructions, example databases*
- *Used for over 40 different languages*

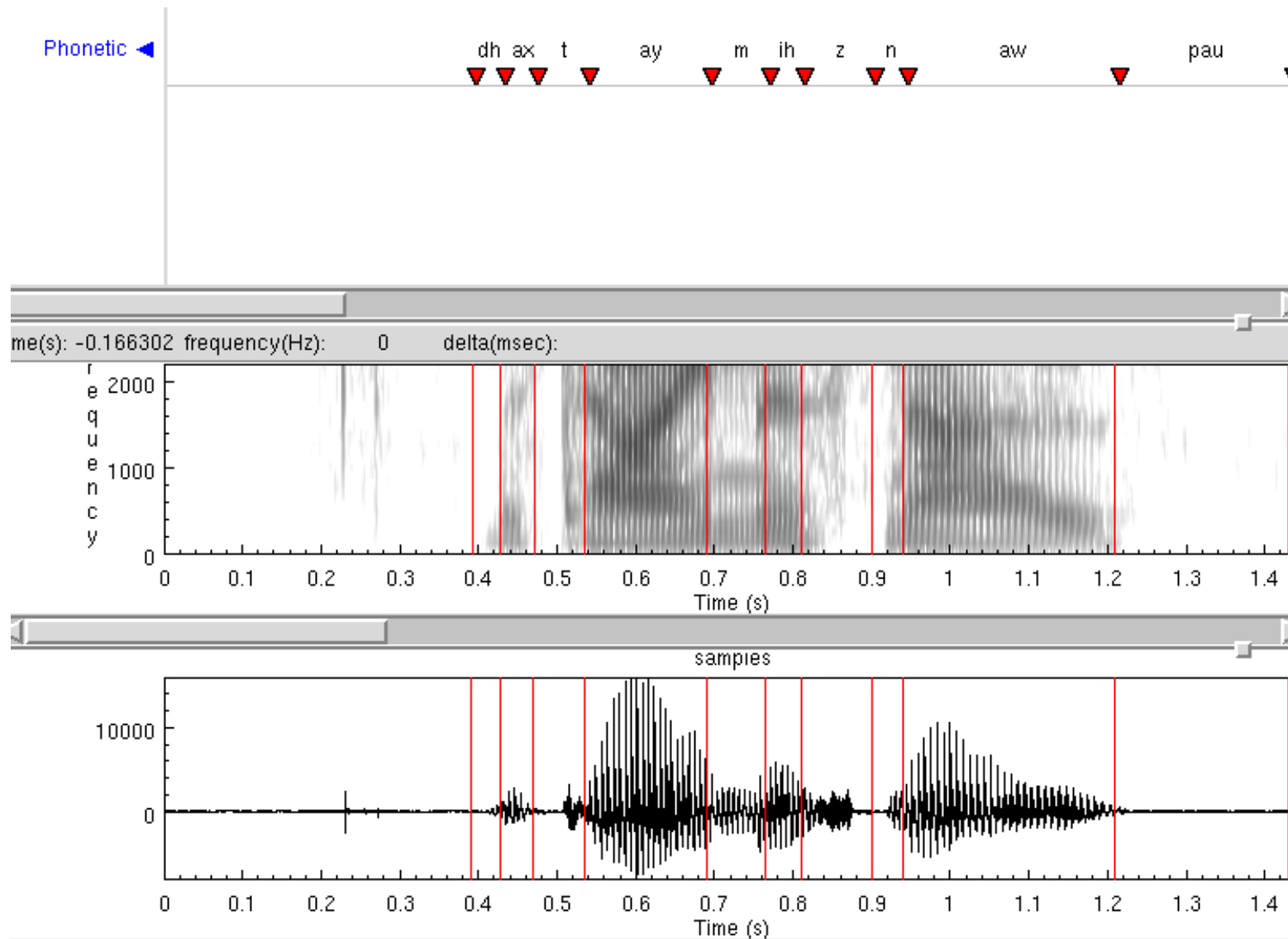
Festival Speech Synthesis

- ◆ *After Installation*
- ◆ *festival -tts stuff.txt*
- ◆ *festival*
- ◆ *festival> (SayText "hello world")*

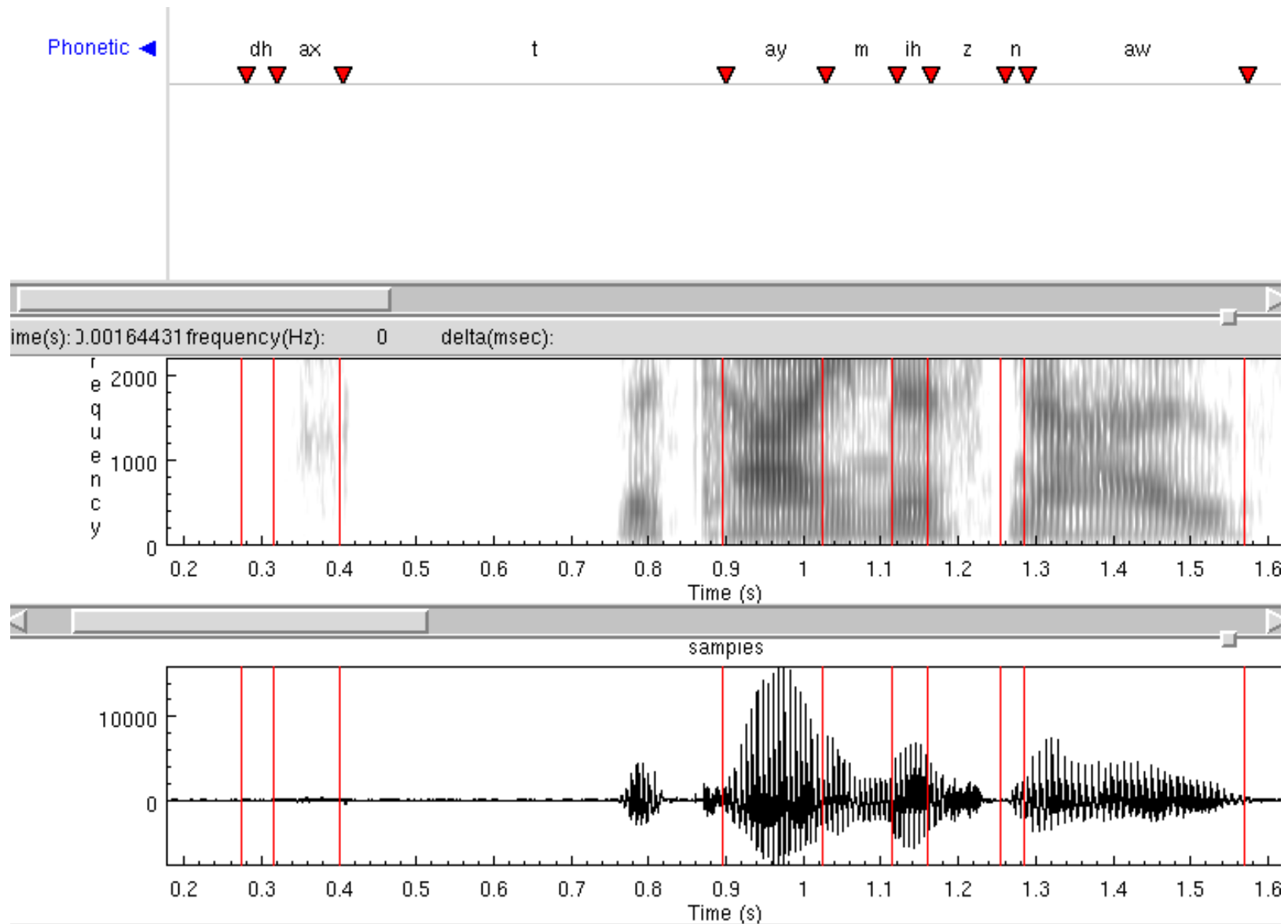
Building Synthetic Voices

- ◆ <http://festvox.org/bsv>
 - *Look at section on “Telling the Time”*

Automatic Labeling



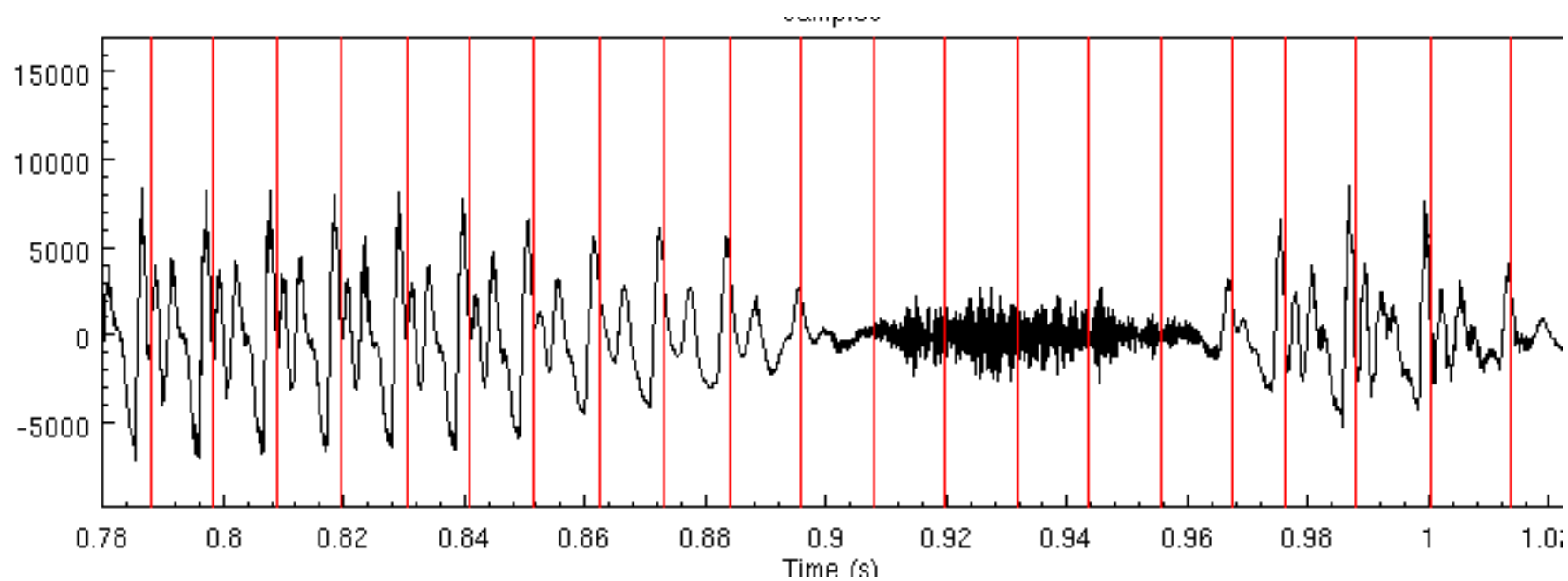
Automatic Labeling (bad)



Parameterization

- ◆ *Extract pitch marks from data*
 - *Find voices/unvoiced regions*
 - *Add “fake” pitch marks during unvoiced regions*
- ◆ *Extract MFCC pitch synchronously*
 - *Instead of a fixed frame advance (e.g. 5ms)*
 - *Extract it at each pitch mark*
 - *Try to capture the spectrum at the pitch period*

Pitchmarks



Building a LDOM synthesizer

- ◆ *Build cluster tree on each unit type*
 - *Not just on phones*
 - *Tag phones with word they come from*
 - *d_limited and d_domain are treated as different*

Tuning and Testing

- ◆ *Test it on some real data*
 - *Ensure number/symbol expansions are correct*
- ◆ *Prompts should probably be word expanded*
 - *Flight US187 -> flight u s one eight seven*
- ◆ *Remove bad prompts*
 - *Or fix labels*
- ◆ *Remember to keep access to the speaker*
 - *If you have to update the system, you need the same speaker available*

Summary

- ◆ *Building a voice*
 - *Databases design, recording, labeling*
 - *Parameter extraction and model building*
- ◆ *Limited domain synthesis*

