DT2112
Introductiuon to the Automatic Speech Recognition Exercise

Giampiero Salvi

KTH/CSC/TMH giampi@kth.se

VT 2016

Objective

Get familiarity with practical problems in Automatic Speech Recognition

You will not need to:

- implement ASR algorithms
- perform complex calculations

. . . but if interested all the source code is available

The Task

Build and test a "four digit" speech recogniser:
Example: "ett tre åtta fyra"

Speaking mode: continuous speech

Speaking style: read speech

Speakers: speaker dependent

Vocabulary: small (10 words)

Robustness: low noise, close microphone

The Task

1. define a pronunciation dictionary
2. record training and test data
3. train monophone, single-mixture phone models
4. evaluate the recognition performance on the test data and on-line
5. answer the questions in the exercise protocol

Work in groups

- you have to test cross-speaker recognition
- form groups of maximum 3 people
- notify the group composition (KTH Social)
- borrow a USB headset for the recordings

The Environment

- the exercise is performed on Linux (Ubuntu)
- preferably use the CSC Ubuntu machines
- familiarity with Unix shell helps
- ... but there are detailed instructions and we are willing to help

You can try to perform the exercise on your own computers, but we can not help with the practical problems you may encounter

The Tools

you will get:

- a document with detailed instructions and the protocol to fill
- a package with all the relevant scripts and files
- the Hidden Markov Model Toolkit (HTK) for ASR related algorithms

To Pass

you have to:

- perform the steps described in the lab
- answer the questions described in the protocol
- hand in the answers to me electronically or on paper.
- send me an archive with the all the relevant files

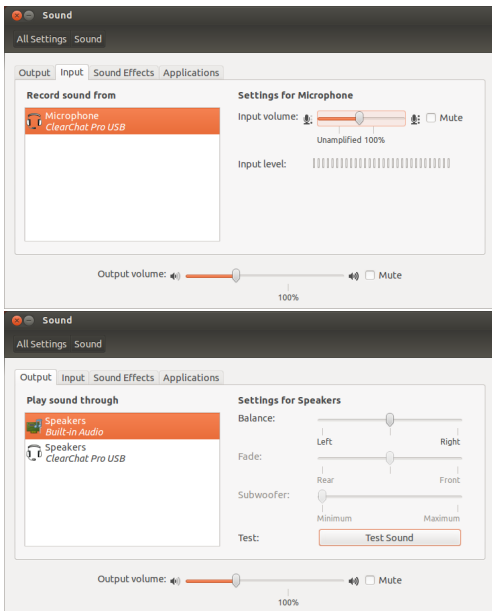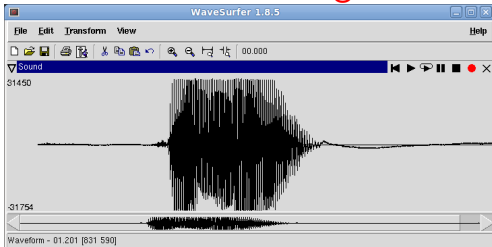# Practical Issues: Broken Audio

Follow instructions at
`https://www.kth.se/en/csc/it-support-csc/`
`environment/linux/precise/`
`ubuntu-12-04-precise-pangolin-1.318854`

Practical Issues: Audio Settings

Practical Issues: Recording Levels

Bad recording!

Good recording: