

專題研究

WEEK 4 - LIVE DEMO

Prof. Lin-shan Lee

TA. Ju-Chieh, Chou (r06922020@ntu.edu.tw)

Outline

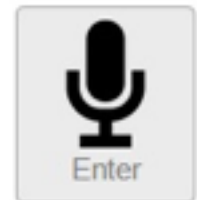
- **Review**
- **Live Demo introduction**
- **Restriction**
- **TODO**
- **To think**
- **Warning**
- **FAQ**

Review: achieving ASR system

- Automatic Speech Recognition System
- Input wave -> output text



DO YOU WANNA BE A SNOWMAN



Review

- **Week 1: feature extraction**
 - ▣ **compute-mfcc-feat**
 - ▣ **add-delta**
 - ▣ **compute-cmvn-stats**
 - ▣ **apply-cmvn**
 - ▣ **File format: scp,ark**

Review

- **Week 2: training acoustic model**
 - **monophone**
 - **clustering tree**
 - **triphone**
 - **Models: final.mdl, tree**

Review

- **Week 3: decoding and training lm**
 - **SRILM(ngram-count/ kn-smoothing)**
 - **Kaldi – WFST decoding**
 - **HTK – Viterbi decoding**
 - **Vulcan(kaldi format -> HTK format)**
 - **Models: final.mmf tiedlist**

Live Demo

- **Now we integrated them into a real-world ASR system for you.**
- **You could upload your own models.**
- **Now give a shot! Experience your own ASR in a “real” way.**

Live Demo

- <https://140.112.21.35:54285/>
- **Demo**
- **Remember: use http**s**.**
- **Ignore the warnings.**

Restriction

- **MFCC with dim 39 only.**
- **Fixed phone set. (Chinese phones)**
- **LM must be one of unigram/bigram/trigram model.**

To Do

- Sign up Live Demo with your account.
 - Please inform TA of your account name for activation.
 - FB Group: 數位語音專題
- Test with basic model embedded in the system.



- Upload your model
 - LM/LEX/TREE/MDL
 - For better performance, you may re-train your models.
- Test with your own models.

To Think

- **Compare the basic models with your own models, what is the main difference?**
- **Do you know of what kind your training data are?**
 - **train.text/dev.text/test.text**
- **How about manually tagging your own lexicon and train your own language model?**
- **Guess about the training data of the basic models.**
 - **Exemplify your description.**

Language Model : Training Text (2/2)

□ `cut -d ' ' -f 1 --complement $train_text > ./exp/lm/LM_train.text`

即 日 起 簡 章 備 索
每 梯 次 名 額 限 八 十 名
繪 畫 理 念 概 說
梧 棲 鎮 立 圖 書 館
彰 縣 文 化 中 心
南 縣 文 化 中 心
高 市 文 化 中 心
屏 東 中 正 藝 術 館
澎 縣 文 化 中 心
宜 縣 文 化 中 心
供 學 術 機 構 作 進 一 步 研 究
也 都 不 得 不 更 加 謹 慎 小 心
深 怕 下 一 著 錯 棋

Language Model : ngram-count (3/3)

□ Lexicon

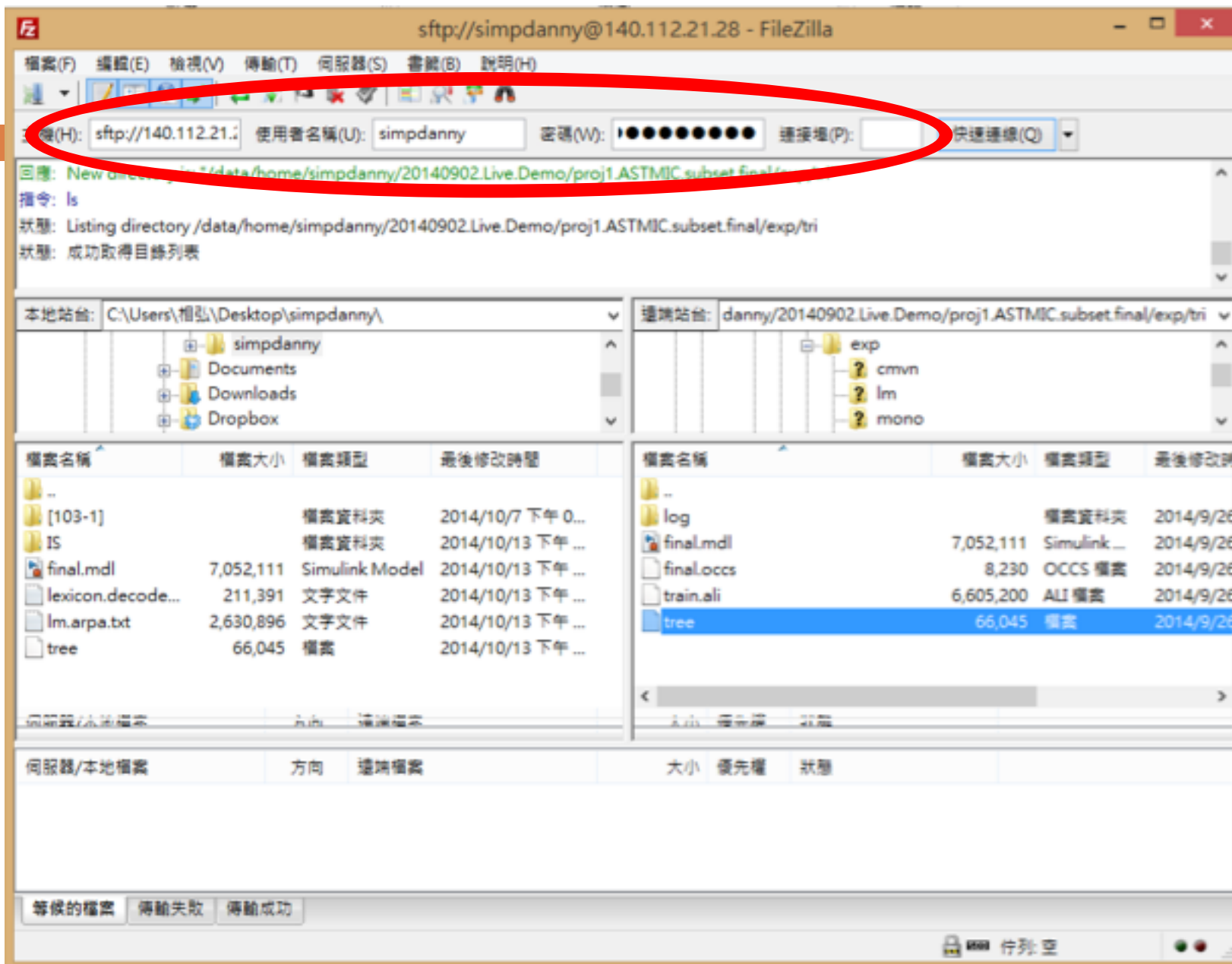
▣ lexicon=material/lexicon.train.txt

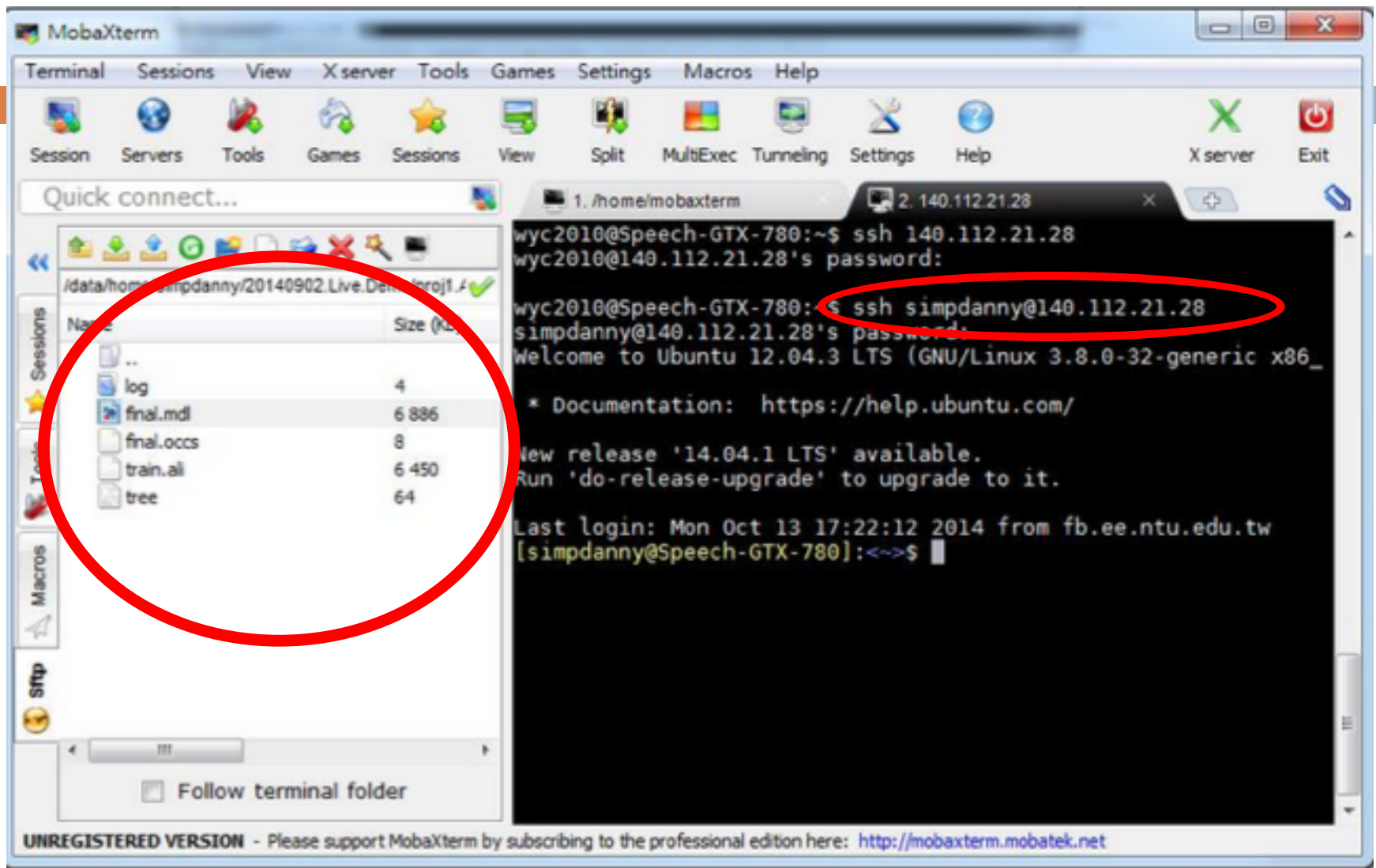
排	CH_p	CH_a	CH_i
牌	CH_p	CH_a	CH_i
派	CH_p	CH_a	CH_i
培	CH_p	CH_e	CH_i
培	CH_p	CH_o	CH_u
陪	CH_p	CH_e	CH_i
裴	CH_p	CH_e	CH_i
賠	CH_p	CH_e	CH_i
佩	CH_p	CH_e	CH_i
珮	CH_p	CH_e	CH_i
配	CH_p	CH_e	CH_i
拋	CH_p	CH_a	CH_u
跑	CH_p	CH_a	CH_u

與黨	CH_#	CH_y	CH_d	CH_a	CH_N#
與國	CH_#	CH_y	CH_g	CH_u	CH_o
與共	CH_#	CH_y	CH_g	CH_o	CH_N#
與決	CH_#	CH_y	CH_dzi	CH_y	CH_E
與其	CH_#	CH_y	CH_tsi	CH_i	
與會	CH_#	CH_y	CH_h	CH_u	CH_e CH_i
語法	CH_#	CH_y	CH_f	CH_a	
語調	CH_#	CH_y	CH_d	CH_i	CH_a CH_u
語彙	CH_#	CH_y	CH_h	CH_u	CH_e CH_i
語氣	CH_#	CH_y	CH_tsi	CH_i	
語詞	CH_#	CH_y	CH_ts	CH_U	
語言	CH_#	CH_y	CH_#	CH_i	CH_E CH_n#
語音	CH_#	CH_y	CH_#	CH_i	CH_n#
語文	CH_#	CH_y	CH_#	CH_u	CH_@ CH_n#
玉米	CH_#	CH_y	CH_m	CH_i	

FAQ

- **Q: How to download the models in the workstation?**
- **A:**
 - FileZilla
 - MobaXterm
 - **“sftp” or “scp” command in your linux OS.**





Warning

- ❑ **Do not press the “Record” button before you upload your model or choose the basic model, otherwise the server will shut down.**

FAQ

- **Q: Why I always got server error?**
- **A: Make sure you got models uploaded. Is the timestamp field empty?**

Last upload date:

• TREE:	2014-10-13 17:51
• MDL:	2014-10-13 17:51
• LM:	2014-10-13 17:51
• LEX:	2014-10-13 17:51

TREE: exp/tri/tree

MDL: exp/tri/final.mdl

LM: exp/lm/lm.arpa.txt

**Lexicon: material/lexicon.decode.txt
(or material/lexicon.train.txt)**

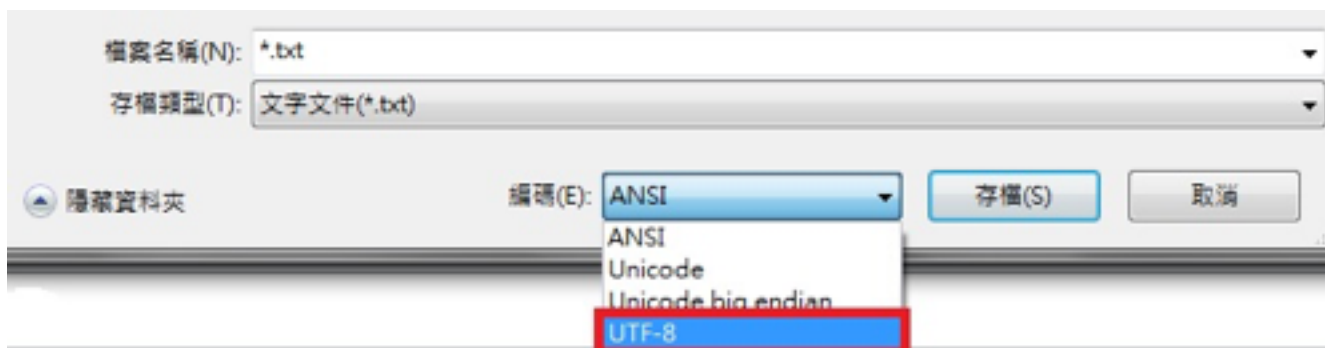
Warning: If using **material/lexicon.train.txt**, remember to add the following two lines to the top of the file.

<s> sil

</s> sil

FAQ

- **Q: In corpus mode, why I always got error or 0 accuracy?**
- **A:**
 - Make sure your corpus is written under UTF-8 encoding.
 - In notepad, the default is ANSI.
 - In vim, the default is UTF-8.



FAQ

- **Q: In corpus mode, why I got negative accuracy?**
- **A:**
 - **Accuracy is actually calculated by $(\text{length} - \text{error}) / \text{length}$.**
 - **請參考數位語音處理概論 ch8 page 11**

Q&A

- **A short tutorial video:**
 - <https://youtu.be/9x5e2W9KNfc>
- **This system is just online.**
 - Any bug is expected.
- **If you got any question, contact TA through FB group or email instantly.**
 - We need you feedback about UI/function.
 - Feel free saying about anything.
 - Email: r06922020@ntu.edu.tw