Machine Learning HW2

ML TAs
mlta-2023-spring@googlegroups.com

Objectives

- Solve a classification problem with deep neural networks (DNNs).
- Understand recursive neural networks (RNNs).

Outline

- Task Introduction
- Dataset & Data Format
- Submission & Grading

Task Introduction

Task Introduction

Task: Multiclass Classification

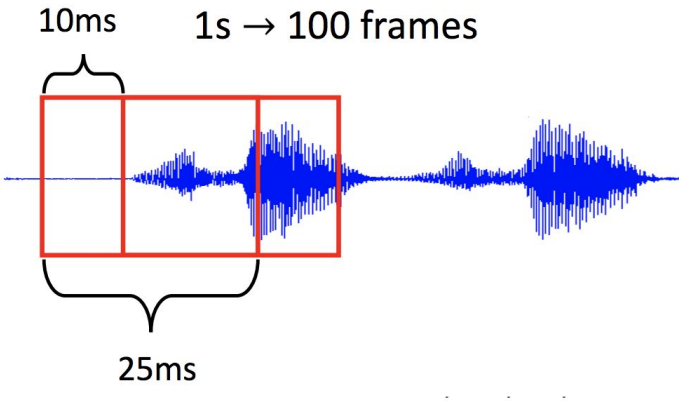
Framewise phoneme prediction from speech.

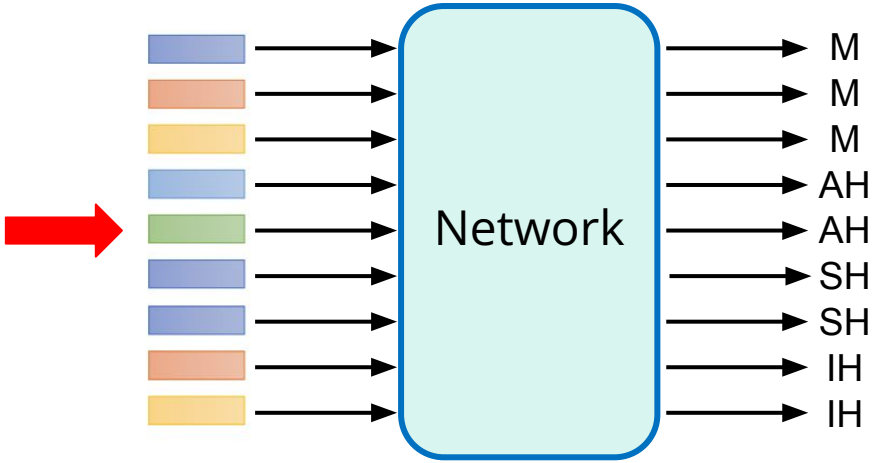
What is a phoneme?

Wiki: A unit of sound that can distinguish one word from another in a particular language.

- Machine Learning → M AH SH IH N L ER N IH NG

Task Introduction

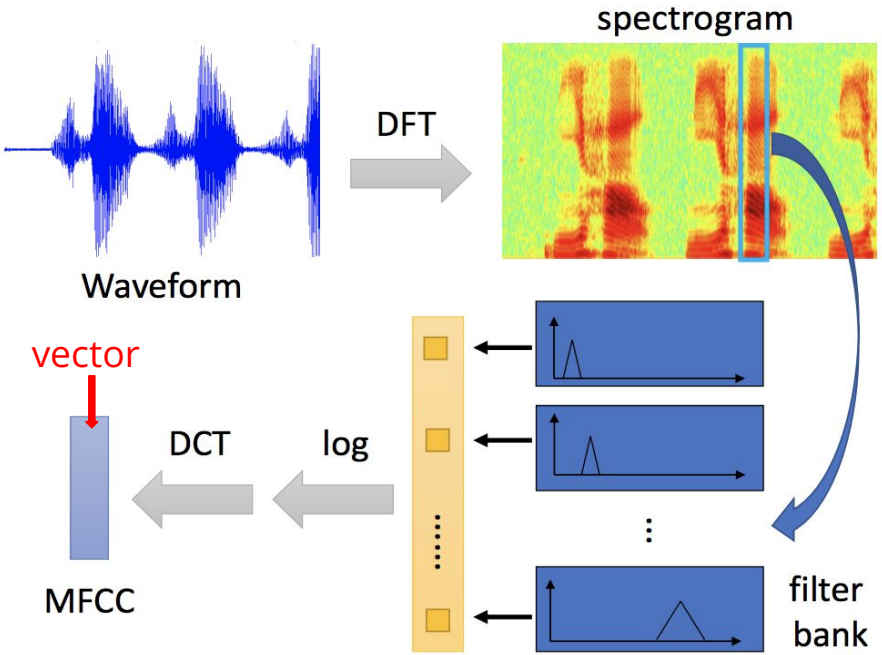




frame 39-dim MFCC

Data Preprocessing

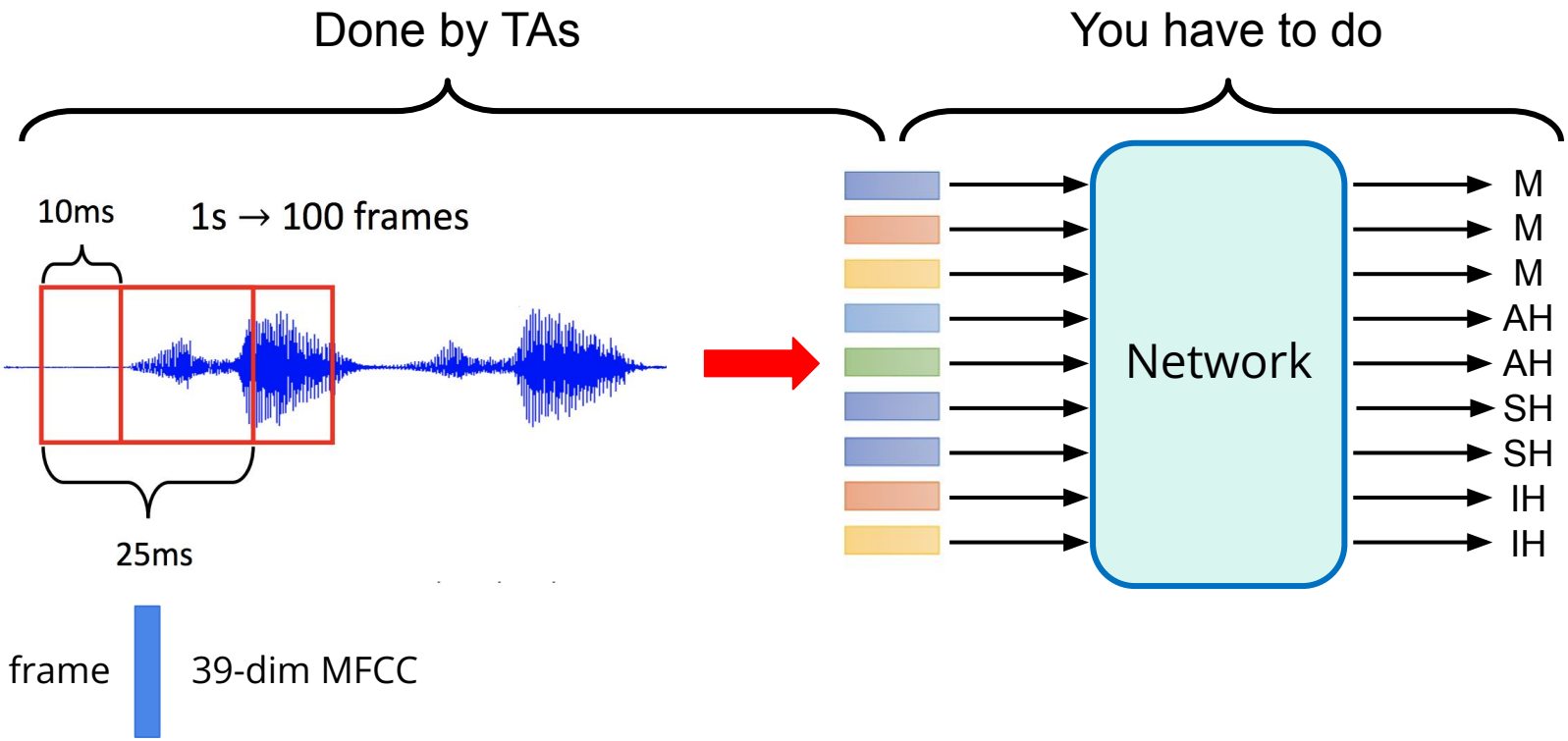
Acoustic Features - MFCCs (Mel Frequency Cepstral Coefficients)



For more details,
please refer to Prof. Lin-Shan Lee's
[Introduction to Digital Speech Processing] Chap.7

Image ref.
Prof. Hung-Yi Lee
[2020Spring DLHLP] Speech Recognition

Task Introduction



Dataset & Data Format

Dataset

LibriSpeech (subset of train-clean-100)

- Training: 3429 preprocessed audio features w/ labels (total 2116794 frames)
- Testing: 857 preprocessed audio features w/o labels (total 527364 frames)
- Label: 41 classes, each class represents a phoneme

Using additional data is prohibited. Your final grade will be multiplied by 0.9!

Data Format

- Data Format (The TAs have already extracted the features)

  libriphone/
  - train_split.txt (train metadata)
  - train_labels.txt (train labels)
  - test_split.txt (test metadata)
  - feat/
    - train/
    - test/

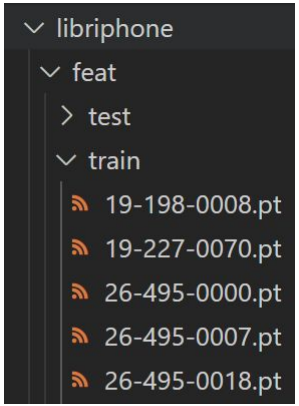
train_split.txt

```
1   4830-25898-0031
2   839-130898-0062
3   198-126831-0045
4   730-359-0022
5   1502-122619-0091
6   1246-124548-0045
7   5808-48608-0026
8   5049-25947-0011
9   1183-128659-0003
10  40-121026-0007
```

train_labels.txt

```
1   4830-25898-0031 0 0 0 0 0 0 0 0 0 0
2   839-130898-0062 0 0 0 0 0 0 0 0 0 0
3   198-126831-0045 0 0 0 0 0 0 0 0 0 0
4   730-359-0022 0 0 0 0 0 0 0 0 0 0 0
5   1502-122619-0091 0 0 0 0 0 0 0 0 0 0
6   1246-124548-0045 0 0 0 0 0 0 0 0 0 0
7   5808-48608-0026 0 0 0 0 0 0 0 0 0 0
8   5049-25947-0011 0 0 0 0 0 0 0 0 0 0
9   1183-128659-0003 0 0 0 0 0 0 0 0 0 0
10  40-121026-0007 0 0 0 0 0 0 0 0 0 0
```

features: 39-dim MFCC w/ CMVN
{filename}.pt for each utterance(audio)



Data Format

- Each .pt file is extracted from one original wav file
- Use torch.load() to read in .pt files as torch tensors
- Each tensor has a shape of (T, 39)

39 dims

T frames

```
tensor([[-0.9555, -0.9062,  0.9451,  ..., -1.4516, -1.5912, -1.3270],
        [-0.9434, -0.9633,  0.7211,  ...,  0.1566, -0.0150, -0.1353],
        [-0.8907, -0.9749,  0.6556,  ...,  1.1867,  0.4603, -0.0459],
        ...,
        [-1.0778, -0.7979,  0.8335,  ...,  0.6452, -0.3527, -0.7415],
        [-1.1911, -1.0670,  0.6462,  ...,  0.3025, -0.6755, -0.9707],
        [-1.1044, -1.0259,  0.7016,  ..., -0.1956, -0.4646, -0.5964]])
```

Submission & Grading

Submission & Grading

- Leaderboard (4%): Kaggle
- Code submission (2%): NTU COOL
- Report submission (4%): Gradescope

Kaggle Baselines

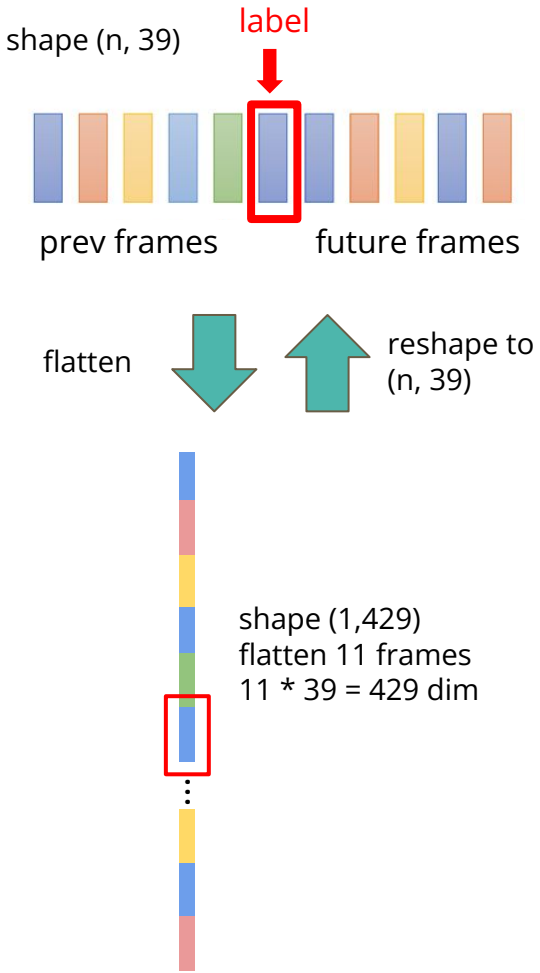
	Public Baseline	Hints	Training Time
(0.5%/0.5%) Simple	0.49798	sample code	~30 minutes
(0.5%/0.5%) Medium	0.66440	concat n frames, add layers	1~2 hours
(0.5%/0.5%) Strong	0.74944	batchnorm, dropout	3~4 hours
(0.5%/0.5%) Boss	0.83017	sequence-labeling (using RNN)	6~ hours

For boss baseline, you can refer to pervious course recording of RNN: video 1 and video 2.

Concat n frames

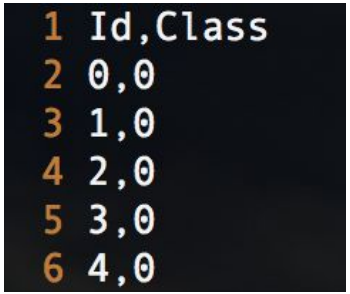
Since each frame only contains 25 ms of speech, a single frame is unlikely to represent a complete phoneme

- Usually, a phoneme will span several frames
- Concatenate the neighboring phonemes for training



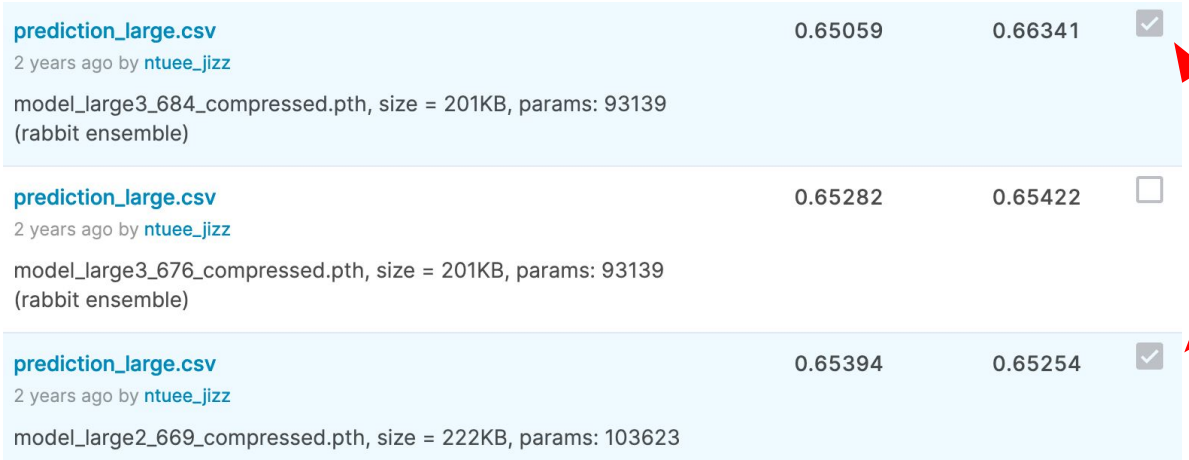
Kaggle Submission

- Displayed name: **<student ID>_<anything>**
  - e.g. b06901020_puipui
- You do NOT have to change your account name, just modify “team name” under the “team” tab
- Submission format: **.csv** file
- Evaluation metric: accuracy
- Submission deadline:
  - **2023/3/24 23:59 (UTC+8)**



Kaggle Submission

- You may submit up to **5** results each day (UTC).
- Up to **2** submissions will be considered for the private leaderboard.



remember to select **2** results for your final scores before the competition ends!

Code Submission (2%)

- Compress your code, then submit it to NTU COOL.

<student ID>_hw2.zip

e.g. b06901999_hw2.zip

- We can only see your last submission.
- Do not submit your model or dataset.
- If your code is not reasonable, your final grade will be multiplied by 0.9!
- Submission deadline:
    - **2023/3/24 23:59 (UTC+8)**

Report Questions

1. (2%) Implement 2 models with approximately the same number of parameters, (A) one narrower and deeper (e.g. hidden_layers=6, hidden_dim=1024) and (B) the other wider and shallower (e.g. hidden_layers=2, hidden_dim=1750). Report training/validation accuracies for both models.

2. (2%) Add dropout layers, and report training/validation accuracies with dropout rates equal to (A) 0.25/(B) 0.5/(C) 0.75 respectively.

Report Submission

- Submit with gradescope, no need to upload any files.
- We can only see your last submission.
- Submission deadline:
  - **2023/3/24 24:00 (UTC+8)**

Gradescope Submission

Active Assignments	Released	Due (CST)
HW2	Mar 03 at 2:00PM	Mar 25 at 12:00AM

Q2
2 Points

Add dropout layers, and report training/validation accuracies with dropout rates equal to (A) 0.25/(B) 0.5/(C) 0.75 respectively.

(A) Training accuracy:

answer

Validation accuracy:

answer

Regulations

(*) Academic Ethics Guidelines for Researchers by the Ministry of Science and Technology

- You should NOT plagiarize, if you use any other resource, you should cite it in the reference. (*)
- You should NOT modify your prediction files manually.
- Do NOT share codes or prediction files with any living creatures.
- Do NOT use any approaches to submit your results more than 5 times a day.
- Do NOT use additional data or pre-trained models.
- Your **assignment will not be graded** and your **final grade x 0.9** if you violate any of the above rules.
- Prof. Lee & TAs preserve the rights to change the rules & grades.

Deadline

- Kaggle (Leaderboard)
  - **2023/3/24 23:59 (UTC+8)**
- NTU COOL (Code submission)
  - **2023/3/24 23:59 (UTC+8)**
- Gradescope (Report submission)
  - **2023/3/24 24:00 (UTC+8)**

Link

- Course website
- NTU COOL
- Kaggle
- Gradescope
- Sample code (Colab)
- Sample code (Kaggle)

If you have any questions, you can ask us via...

- NTU COOL (recommended)
  - https://cool.ntu.edu.tw/courses/24108
- Email
  - mlta-2023-spring@googlegroups.com
  - The title should begin with “[hw2]”
- TA hour
  - Friday, 上課課餘時間
  - Friday, 19:00 ~ 21:00

Q & A

Gradescope enrollment

We will send the class invitation to your "NTU email".

Gradescope enrollment

If you have already registered a Gradescope account with your NTU email...

Gradescope <no-reply@gradescope.com>
寄給 曾亮軒

請謹慎處理這封郵件 寄件者並未驗證這封郵件，因此「Gmail」無法確認郵件是否確實由對方寄出。

回報為垃圾郵件　看起來沒有問題

曾亮軒 您好，

您 (b07502072@ntu.edu.tw) 已作為 學生 新增至 Gradescope 中的課程 ML 2023 (EE5184), Spring 2023。

檢視您的課程列表

忘記密碼？查看我們關於重設您帳號資訊密碼的指引。

不是正確的課程？請聯繫您的教師以取得進一步的協助。如有需要，您也可以 您即將退出此課程。

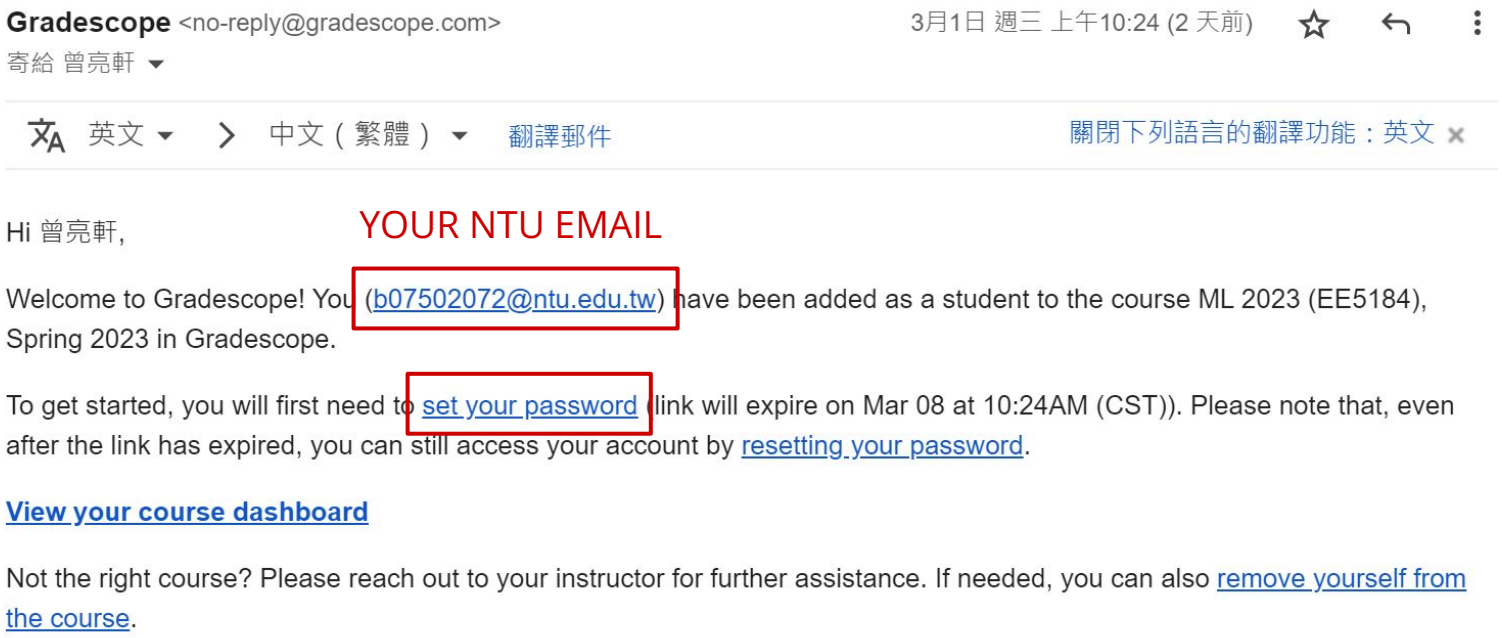
我們是來提供協助的。搜尋我們的說明中心 ，以取得您的任何 Gradescope 問題的答案，或傳送電子郵件至help@gradescope.com。

Gradescope 團隊

You can directly enter the course

Gradescope enrollment

Otherwise, please set your password to register an account.



Gradescope <no-reply@gradescope.com> 3月1日 週三 上午10:24 (2 天前)

寄給 曾亮軒

英文 > 中文（繁體） 翻譯郵件 關閉下列語言的翻譯功能：英文

Hi 曾亮軒,

YOUR NTU EMAIL

Welcome to Gradescope! You (b07502072@ntu.edu.tw) have been added as a student to the course ML 2023 (EE5184), Spring 2023 in Gradescope.

To get started, you will first need to set your password (link will expire on Mar 08 at 10:24AM (CST)). Please note that, even after the link has expired, you can still access your account by resetting your password.

View your course dashboard

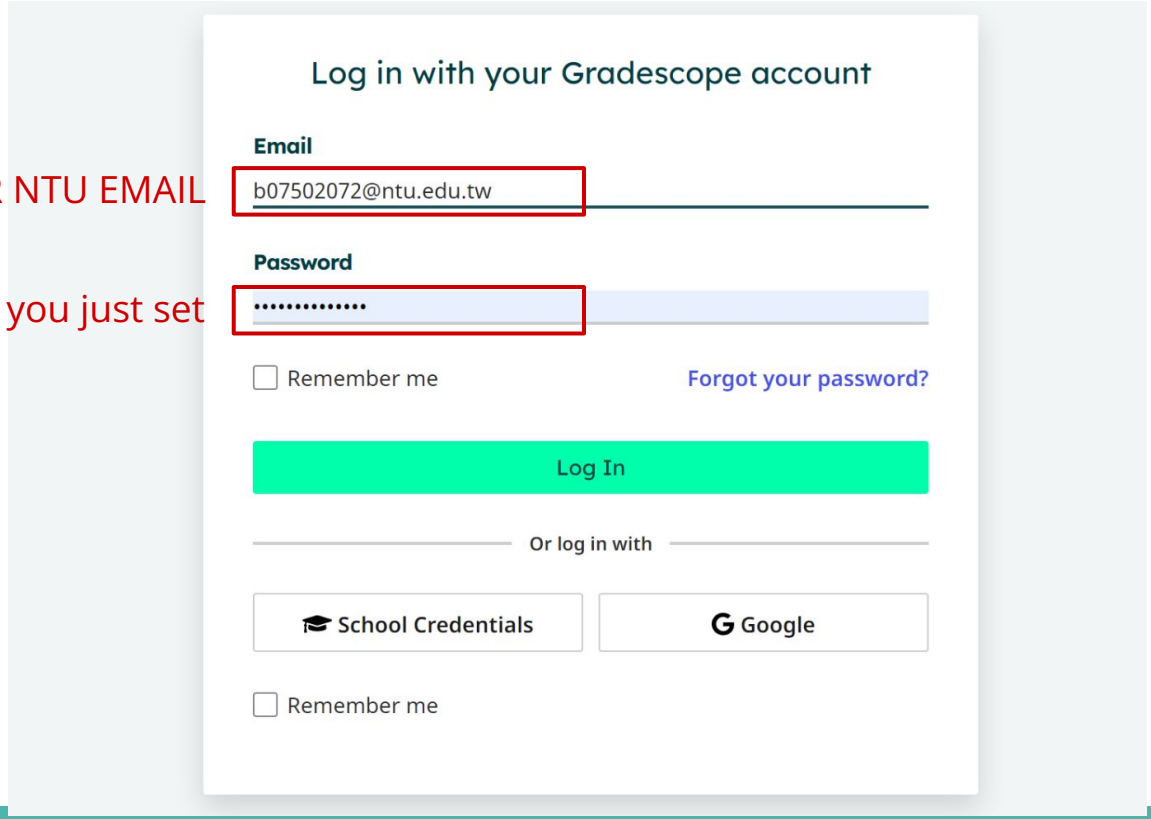
Not the right course? Please reach out to your instructor for further assistance. If needed, you can also remove yourself from the course.

Gradescope enrollment

Login

YOUR NTU EMAIL

The password you just set



Gradescope enrollment

If you haven't received the invitation **after 3/9 23:59**, please email to us (mlta-2023-spring@googlegroups.com, the title should begin with [Gradescope]).

Human evaluation

填問卷抽現金

- 我們是語音實驗室的學生，要請大家幫忙填問卷
- 一份問卷有 20 個音檔 (大概 3~5分鐘)，請大家**戴耳機**聽音檔，然後評量每個音檔的品質。詳細任務說明請見問卷內描述
- 請大家按照學號選擇一份問卷，多填的我們會直接把你的回答刪除
- 不用現場填沒關係，可以回去再填
- 學號沒有填就不能抽獎
- 最後會抽出 10 個人，每個人給300元
- 時間到本週日(03/05)

學號末碼	問卷連結
0	https://surveyjs.io/published?id=93a278ef-207f-4883-bbac-419e49d28644
1	https://surveyjs.io/published?id=c543089f-beda-4edf-a440-ea616fc66b9f
2	https://surveyjs.io/published?id=75ec34b4-78d8-4c68-aafc-5dc16c5c5b61
3	https://surveyjs.io/published?id=01cb45a2-cb6d-47f2-a437-01a0f4836863
4	https://surveyjs.io/published?id=36ca96c2-74c8-4c7c-b36c-d15a72301133
5	https://surveyjs.io/published?id=04620688-7c56-4ac7-b8c4-3c5247b2926f
6	https://surveyjs.io/published?id=45bd20c4-c77d-4a7e-8a93-b7f9cb96c6d8
7	https://surveyjs.io/published?id=17da5652-0513-46b1-bf57-626506327c25
8	https://surveyjs.io/published?id=27553609-eb22-474e-a7b3-3cd9f25d2f21
9	https://surveyjs.io/published?id=71a7bfe1-c849-46e1-9d48-a93c52f51aa0