DSP-1 Applications

Workload 2021-2022

It is required to form groups of 3-5 students. Once the group is formed they must do all the deliverables together, i.e. no change for teams' members. The work load of the course consists of the following items:

I. The course Project (deadline: 9th of December !): (12 marks)

The purpose of the project is to help the speakers and the children to correctly pronounce the phonemes and some words of the Arabic language.

1. Data Collection:

- Each person should record the list of words in the end of this documents. Each person, should get a complete list of recordings for a:
  a. A male b. a female c. a child (less than 12 years)

(Hint: if the student is a male; he should record by himself and for a female outside the course students and a child. If the student is a female, she should record by herself then add recordings of a male speaker and a child. Not a single person to record more than once.

- Recordings must respect these specifications:
  1. Mono channel
  2. 16 bits
  3. 16KHz sampling frequency
  4. Quite room (low background noise)
  5. No clipping: means that the maximum amplitude of the recorded waveform not to exceed the maximum amplitude limits
  6. If there is a mistake in the recording process the speaker could repeat the correct utterance and the student can delete the wrong utterance using any audio editing tool like Cool Edit Pro.
  7. The recordings can be as follows:
     - Using a recording tool like "Cool Edit Pro" application

- Or mobile recordings (after making sure that the recording specifications are met).
- Or using WhatsApp recorded messages that can be collected by the student.
- We will add to the naming of the collected file one character to indicate the method of collection: C (computer), M (mobile recorders), W (WhatsApp).

8. Each utterance should be recorded as follows:
    - word1, salience for about 1 second then the second word,
    - then a salience of about 2 seconds
    - then the following two words separated by about a second and so on.
9. The name of any recording set should abide to the naming specifications as follows:
    - Group no.: Gxx, example G07 for students in Group 7
    - Student no.: Sx: example S3 for student no. 3
    - Male, Female or Child + age: Mxx, Fxx, Cxx, examples: M20, F40 or C09 for a male of 20 years old, a female for 40 years old or a child of 9 years old.
    - As per now each person has a complete set of recordings will have a naming like that: GxxSx(M/F/C)xx:
    - The recording method: C/M/W.
    - example: G05S2F20W means: the student is from Group 05, the student no. 2, she is a female of 20 years old and the recording method is the WhatsApp.
    - An **automatic tool** should be used to segment each person recordings to separated words. Each person must review his three recordings sets by word to be sure that each separated word recording is correct. Then an automatic assignment for the word recording naming should be as follows:
        - Word order: there are 62 recordings: PxxWx: example: P25W2 means Pair 25 word 2.
        - each Pair has two different words (except the last utterance).

10. The data should be classified into three groups:
    - Male speakers
    - Female Speakers
    - Children Speakers
11. From each group we should select the reference speakers. A reference speaker is the one who pronounce all characters and words correctly. This should be done using an expert (Dr. Mohsen Can help)
    - The naming should include the characters R (Reference) and T (Test).
    - The whole Naming should be like that:
      G10S5M21WP55W2R
      means: Group 10, Student 5 in group 10, Male Speaker (not necessary the gender of the student but the speaker), 21 the age of the speaker, through WhatsApp, Pair 55, Word 2, Reference Speaker.
12. After selecting the references, the remaining speakers will be taken as test speakers.
13. An **automatic tool** should be used to run on all the data to generate the MFCC coefficients for each utterance. Then both the raw data and the MFCC data will be available for each team.

- We will need a group to take care of the data:
    1. Each group must review its data and be sure that the recordings are per the order agreed upon.
    2. The data group should:
        - Use an automatic tool for segmenting the utterances to words.
        - Review the separated words.
        - Assign a naming to each word as shown above.
        - Run an MFCC on all the data

    This group must have good programmers, to adapt two engines:

    1. A segmentation tool
    2. MFCC tool

This group will be given a bonus 2 marks if it handles the data in a correct way. Their project will be about handling the whole data as mentioned above. They must also share in collect data by themselves.

2. These steps should be done to complete application:

1. To use one utterance (last one in the list) to select the nearest speaker from the reference list to the test speaker under testing:
    - Using the utterance no. 62, and the DTW to measure the distance between this test speaker and the nearest one from the reference list.
2. then screening all the words uttered by the test speakers, we like to be sure if each word is correctly uttered or not. That will be done by measuring the distance between word(ik) (where i is the word group no. and k either 1 or 2) versus words (i1) and (i2) form the reference. We will count a correct score if the word Test(i1) is closer to Reference(i1) and Test(i2) is closer to Reference(i2) and zero otherwise. It is to be noted that the utterance could be closer to the corresponding reference however the pronunciation is not correct, therefore we will put a limit to the distance between any two words (test vs. reference). We should select experimentally a threshold value for the distance to be accepted.
3. The results should be tabulated in a confusion matrix like that:

				Tot. Words	judgement				
Pair	Word	Word2	Word1	Males	other*	Word2	Word1	Correct	Wrong
1	1		ذَابَ	70	4	6	60	60	10
	2	ثابَ		70	5	56	9	56	14
2	1		ثائر	68	8	10	50	50	18
	2	سائر		68	9	55	4	55	13
3	1		ثمين	70	5	10	55	55	15
	2	سمين		69	4	56	9	56	13
4	1		ثناء	71	7	4	59	59	12
	2	سناء		70	11	50	9	50	20

Total				556				441	115
%								79%	21%

* if the distance exceeds the threshold for acceptance for any of the two words

				Tot. Words	judgement				
Pair	Word	Word2	Word1	**Females**	other*	Word2	Word1	Correct	Wrong
1	1		ذَابَ	65	4	6	55	55	10
	2	ثابَ		70	5	56	9	56	14
2	1		ثائر	68	8	10	50	50	18
	2	سائر		68	9	55	4	55	13
3	1		ثمين	70	5	10	55	55	15
	2	سمين		69	4	56	9	56	13
4	1		ثناء	71	7	4	59	59	12
	2	سناء		70	11	50	9	50	20

Total				551				436	115
%								79%	21%
				Tot. Words	judgement				
Pair	Word	Word2	Word1	**Children**	other*	Word2	Word1	Correct	Wrong
1	1		ذَابَ	75	4	6	65	65	10
	2	ثابَ		70	5	56	9	56	14
2	1		ثائر	68	8	10	50	50	18
	2	سائر		68	9	55	4	55	13
3	1		ثمين	70	5	10	55	55	15
	2	سمين		69	4	56	9	56	13
4	1		ثناء	71	7	4	59	59	12
	2	سناء		70	11	50	9	50	20

Total				561				446	115
%								80%	20%

Total for all data				1668				1323	345
								79%	21%

4. A graph should be drawn to some samples (5 for example) of the mismatched words. The x axis will be the time along the reference, and the distance along the y axis to show the regions of maximum

mismatches. A graph like the one drawn below. This graph should be drawn after using the DTW alignment.

Figure 1: to show the per frame distance to show the mismatched region(s)

5. Using all the test speaker pairs, a report should be presented to show the mistakes in pronunciation that this speaker has like: take care of your: ذ, ث, ق, ص...الخ

Mark distribution of the project:

- **3 marks** for correct and complete data collection and revised. This mark will be assigned to each student in the group. 1 full mark for a complete (all word pairs) and correct (keep the word pair order, the right word then the left word from the word pair list) recording of a speaker.
- **2 marks** for the correct use of the DTW
- **3 marks** for the application (User Interface). Basically dealing with a pre-recorded complete set.
    - If the optional (described below) part is implemented a bonus of 1-3 marks will be granted.

- **4 marks** for the report, presentation, and discussion
- **Extra 1 mark for any additional speaker specially children (after the 3 per student). This will be given for the student who fulfil this additional work.**

Deliverables:

1. Write a **paper** of about than 8-10 pages (please review the English carefully), must abide to the specifications of the given template. The names of each team must be in Arabic (4 names) on the first page.
2. Each group should prepare a **PPT presentation** and a demo for the project results and/or output. The code used should be delivered with the project material. Each student contribution should be shown in the last slide.
3. Each group should be prepared to present their work.
4. All project materials (project paper, demo, code and the PPT presentation) must be delivered digital form on the web.
5. All materials should be uploaded to a single place to be managed by the course for all the teams to benefit from and for evaluation.
6. The instructor will arrange for a time to review the presentations; however, each group must record a video of no more than 3 minutes to show their work. They will use their PPT with concentration on their added contributions.

This part is an optional part with bonus marks:

The application should allow the use of an external **online test speaker:**

a. Should allow any new speaker to record his/her sentence (الحمد لله)
b. The application will select the closest reference speaker
c. Then to allow this new speaker to record another word, the application will make the comparisons and show the results as follows:
    i. Correct or incorrect

ii. The mismatched region in the time and correspondingly the mismatched phonemes as shown in Figure 1 above.
iii. If the new speaker completed the whole set or some of it, a report should be generated to tell what are the list of words that are not correct.
iv. The bonus will be for implementing this function and will depend of the speed of the response (1-3 marks).

II. Assignment 1 for the speech processing (3 marks)
III. The image processing part will have assignments for (8 marks)
IV. The data processing and machine learning part workload (8 marks)
V. Attendance is important and absence can reduce the total mark.

No mid-term exam

The list of words to be uttered:

ملاحظات:

1. اخرج لسانك فى الثاء والذال والظاء
2. بين الفروق بين: ت/ط , س/ص, د/ض, ز/س, ك/ق, ذ/ظ...الخ
3. الجيم القاهرية هى الاصل
4. استخدم آخر جملة لاختيار المتحدث المرجعى

1.	ذَابَ / ثابَ
2.	ثَائِرْ / سَائِرْ
3.	ثَمِينْ / سَمِينْ
4.	ثَنَاءْ / سَنَاءْ
5.	نَذِيرْ / نَظِيرْ
6.	نَفَذَ / نَفَثَ
7.	ظَلِيلْ / ذَلِيلْ
8.	نَذَرَ / نَظَرَ
9.	مَحْظُورْ / مَحْذُورْ
10.	تِينْ / طِينْ
11.	تَابِعْ / طَابِعْ
12.	أَمَاتَ / أَمَاطَ
13.	طَابَ / تَابَ
14.	أَمْطَارْ / أَمْتَارْ
15.	دَلَّ / ضَلَّ
16.	دَرْب / ضَرْب
17.	عَدَّ / عَضَّ
18.	صَامِدْ / صَامِتْ
19.	دَلَالْ / ضَلَالْ
20.	فَائِدَةْ / فَائِضَةْ
21.	أَفَادَ / أَفَاضَ
22.	ضَنَّ / ظَنَّ
23.	مَضَارْ / مَدَارْ
24.	سُورَةْ / صُورَةْ
25.	عَسِيرْ / عَصِيرْ,

.26	لَـمَسَ / لَـمَزَ
.27	نَسَلَ / نَزَلَ
.28	مَسَحَ / مَسَخَ
.29	صَعَدَ / سَعَدَ
.30	صَرِيرْ/سَرِيرْ
.31	مَصِيرُهْ/ مَسِيرُهْ
.32	صَفَرْ / سَفَرْ
.33	ذَكِيْ / زَكِيْ
.34	ذَلَّ / زَلَّ
.35	زَفَرْ / ظَفَرْ
.36	غَمَزَ / غَمَسَ
.37	خَـيْر / غَـيْر
.38	غَيْر / خَيْل
.39	خَائِبْ / غَائِبْ
.40	قَلْبْ / كَلْبْ
.41	تَقْدِيرْ / تَكْدِيرْ
.42	قَالَ / كَالَ
.43	خَائِنْ / كَائِنْ
.44	خَانَ / كَانَ
.45	خَبِيرْ / كَبِيرْ
.46	خَامِلْ / كَامِلْ
.47	مَسَكَ / مَسَخَ
.48	وَعَدَ / وَأَدَ
.49	عَلَمْ / أَلَـمْ
.50	مُتَعَلِّمْ / مُتَأَلِّـمْ

.51	عِصَابَاتْ / إِصَابَاتْ
.52	حَمْزَةْ / هَمْزَةْ
.53	حَالَةْ / هَالَةْ
.54	أَشْبَاحْ / أَشْبَاهْ
.55	حَرَمْ / هَرَمْ
.56	حَامِلْ / خَامِلْ
.57	حَامِلْ / هَامِلْ
.58	هَزَمَ / حَزَمَ
.59	هَانَ / حَانَ
.60	هَوَى / حَوَى
.61	هُرُوبْ / حُرُوبْ
62.	الحمدُ لله