Technical Report on Arabic CAPT Databases

Building an Arabic CAPT system requires an Arabic Speech database that contains diverse speech and pronunciation errors. At the time, no database is available for Arabic L2 speakers with emphasis on pronunciation errors, and enough speakers with detailed phoneme annotations. Hence in this report, I describe the two Arabic CAPT databases that we developed in the Speech Processing Group, Arabic-CATP-1 and Arabic-CAPT-2.

1. Arabic-CAPT-1

We recorded a new speech dataset, Arabic-CAPT-1, having in mind the quality of the text selection and the coverage of most errors that L2 Arabic speakers might make. A team of experts from Arabic Language Institute (ALI-T) teams, proposed the text for recording and had many meetings and discussions with the team from College of Computer and Information Sciences (CCIS-T), until a set of 25 long length sentences and some 61 very special minimal pairs of words were finally selected. Details of text selection, database specifications, the recording system, registration of the speakers, speech recording are presented in below. Details of speech labeling and analysis will also be presented. Note that the detailed description of this database is included in our submitted work named " Computer-Aided Pronunciation Training (CAPT) System for Non-native Learners of the Arabic Language".

1.1. Selection of text for recording of the speech corpus

A main issue in CAPT systems is to select the optimal words and sentences that can cover the majority of errors in learning the pronunciation of L2. The selected texts must contain very specific phonemes that are difficult to pronounce by the speakers, and useful in improving the pronunciation. The text should have the following characteristics:

- Varied text containing rich diversity of phones and di-phones.
- Optimal number of sentences/words that can be pronounced by L2 Arabic learners in a minimal amount of time.

ALI-T team is well qualified for this task, as they have years of expertise in teaching the Arabic language for Non-Arabs, and conducted many research studies to enforce this expertise. Based on this expertise they were able to advice for a methodology to construct the text most suitable for the project, which took a considerable time and efforts. The proposed methodology is as below.

Methodology for the CAPT text selection

The selection of the CAPT sentences were subject to many constraints as follows:

a) Sounds
   - Many repetitions of the same sound or phoneme are preferable.
   - Appearance of the sound at the start, middle and end of the word.

b) Words
   - Common: Common words are preferred to specific words.
   - Diversity: words used in diverse Arabic countries are preferred.
   - Affinity: Usual and daily words are preferred.
   - Inclusion: Words used in many domains are preferred to words used in specific domains.
   - Importance: words needed by the learner are preferred.
   - Purity: Original Arabic words are preferred to Arabized Arabic words.

c) Sentences of the text
   - Sentences must be meaningful.
   - Must have Arabic cultural aspect
   - Must be valuable.
   - Short sentences are preferred, to avoid boringness.
   - Must be consistent and clear.
   - Must be in accordance to and respect to the Kingdom's beliefs and constants.

The selection of the CAPT text passed by two main steps. The first step, contributed by ALI experts, consisted of applying the above methodology and suggesting sentences and specific words that contain phonemes that learners of Arabic have problems in pronouncing correctly, in addition to simple phonemes that can be found in other languages, such as `m`, `n`, …etc. The second step, conducted by the project team, was refining the text to the mobile app and testing it. This second step passed by 4 stages, the first three stages were completed before starting the audio recording.

All the selected CAPT-texts were tested and adjusted over five main criteria:

- Reasonable time to read all the content.
- Complexity of the content, phoneme positions and length of words.
- Richness of the phonemic content in every sentence.
- Dual phones words must contain minimal pair diversity in phoneme pronunciations.
- It is well known that the more the sentences become long, the speaker starts damping its voice and were prone to more reading latency and stuttering, not in accordance to what the CAPT system aims to correct.

Both CCIS and ALI teams checked all the sentences and agreed to select an optimal number of 16 sentences with a minimal number of 21 words and a maximum number of words of 42 words. In addition, they selected a set of 61 minimal dual phones pairs, to target the phoneme dualities that can lead to pronunciation errors, these dual words differ by some phonemes but have a similar structure. Arabic experts from the ALI-T team stressed on the fact that these short and long dual phonetic words are very important in assessing and evaluating Non Arabs pronunciations.

Samples of two meaningful sentences and eight minimal dual phones pairs are presented in Table 1 below.

Table 1, Sample sentences and dual phonetic words

Sentence 1	أَفْضَلُ النَّاْسِ عِنْدَ اللهِ هُمْ أَصْحَابُ الأَعْمَالِ الفَاضِلَةِ، وَالضَمَائِرِ المُضِيْئَةِ، الَّذِينَ يَرْكُضُونَ إِلَى الخَيْرِ رَكْضاً، يُعِينُونَ الضَّعِيْفَ وَالْمَرِيضَ، وَيُنَاهِضُونَ الضَّلَالَ وَالإِضْرارَ أَمَلًا فِي مَرْضَاةِ اللَّهِ.
Sentence 2	أَفْطَرْتُ بِالأَمْسِ عِنْدَ طَارِقٍ عَلَى طَعَامٍ طَيِّبٍ، وَعِنْدَمَا خَرَجْتُ رَأَيْتُ طُيُورًا فَوْقَ مَبْنَى المَطْبَعَةِ تَطُوفُ بِهَا، ثُمَّ تَحطُّ عَلَيْها لِتَلْتَقِطَ الْحَبَّ، وبَعْضَ الأَطْعِمَةِ، فَطَابَتْ نَفْسِي بِرُؤْيَتِهَا.
Pair of minimal dual phone words 1	خَـيْر/ غَـيْر ** غَيْر/ خَيْل ** خَائِبٌ/ غَائِبٌ
Pair of minimal dual phone	كُلْ/ قُلْ ** رَقَدَ/ رَكَدَ ** قَالَ/ كَالَ ** كَتَمَ/ خَتَمَ ** مَكْتُومٌ/ مَخْتُوم

words 2	

The initial recording setup was designed to record the CAPT students in the Arabic Language Institute at King Saud University, in a controlled live session, face to face. If during the recording sessions, the speaker feels tired or bored, a short pause can be made, and the recording can continue after the pause session. Unfortunately, the COVID-19 restrictions imposed that students cannot come to the university, and we had to move to online recording, through a newly developed mobile app, that will be described later in the section. This online recording had some benefits and some drawbacks, as illustrated in Table 2.

Table 2, Benefits and drawbacks of the online recording solution

Benefits	Drawbacks
Any screen can be easily recorded again in a new time if any error is detected without the need to for the student to come back to the recording room (once the admin allows re-recording)	Speaker recordings had to be checked after each speaker completion. This induced EXTRA COSTS, for the listeners and Q/A checking stage.
Number of students at ALI was much lower than the number at time of submission of the proposal. Online recording allowed us to record students from inside/outside of Riyadh.	Recording solution must be compatible with diverse screens and phones, which increased the time for the testing and tuning of the App.
Diversity of the students, as students are not from one location or university.	Decrease of the total number of recorded phonemes. (see Figure 1)
Long sentences have been shortened, so the speakers could complete the recording in shorter time.	Additional explanations and discussions were necessary to make the students understand the installation and the use of the recording solution.

Due to the mobile recording constraints, the CAPT selected sentences have been additionally shortened, in order to fit into the screens of the students. Selecting the text in the app pages went into four versions. Statistics of the four versions of the texts are shown in Table 2. A detailed comparison, of the first three versions of the CAPT-Texts for each selection phase, is illustrated in Appendix A. Appendix B lists the text for version 4 (Mobile version) after further simplification and shortening so the read texts can be audio-recorded easily in the Mobile App.

Table 3, Statistics of the CAPT-Text selections

Statistics	V1	V2	V3	V4 (mobile)
Number of sentences	29	28	16	25
Total number of words (sentences)	1100	767	474	463
Maximum number of words / per sentence	72	41	16	27
Minimum number of words / per sentence	20	20	21	11
Pairs of Minimal Dual words.	113	113	113	61
Number of Screens : sentences	-	-	-	25
Number of screens: words	-	-	-	16

The numbers of phonemes, within the sentences and words of all the text selections versions, are presented in Figure 1. The phoneme distribution remained almost the same although we reduced the total number of phonemes to half from V1 to V3. V4 Mobile version is a reduced version of the V3, to fit in the screens of the Mobile, where sentences were shortened keeping the same meaning of the content, and 61 important dual pairs were kept for recording.

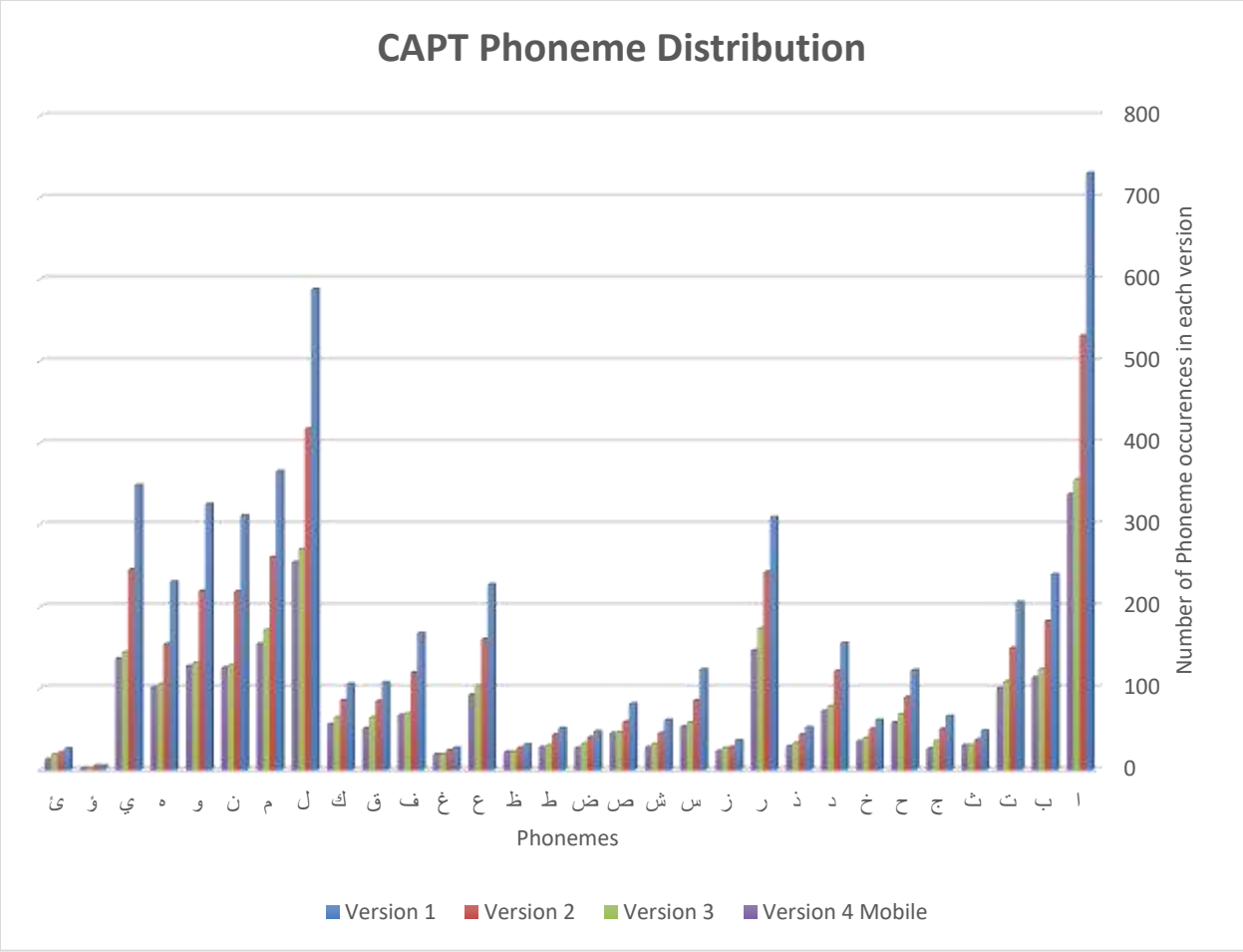


Figure 1, Phoneme distribution in the sentences and dual pairs of the various text selections

A comparative table of the phonemes of each version is detailed in Table 4.

Table 4, Comparative between the text selections of the various versions

Phoneme	Version 1	Version 2	Version 3	Version 4 Mobile
ا	730	531	355	337
ب	239	182	123	113
ت	205	149	108	100
ث	48	37	30	30
ج	66	50	35	26
ح	122	89	68	58
خ	61	50	39	35
د	155	121	78	72
ذ	52	43	33	29
ر	309	242	173	146
ز	36	28	27	23
س	123	85	58	53
ش	61	45	32	28

ص	81	59	46	45
ض	47	40	32	27
ط	51	43	30	28
ظ	31	27	22	22
ع	227	160	103	92
غ	27	24	19	19
ف	167	119	69	67
ق	107	84	64	51
ك	105	85	64	56
ل	588	417	270	254
م	365	260	171	154
ن	311	218	128	125
و	325	219	131	127
ه	230	154	105	101
ي	348	245	144	136
ؤ	5	5	2	2
ئ	26	21	19	13
Total Phonemes	**5248**	**3832**	**2578**	**2369**

1.2. Database Specifications

The recording step started by recruiting some sample speakers from the ALI institute, from the fourth level, in order to test the recording time and the quality of the reading. From the initial speakers' samples when recording version 3 of the text, we noticed that the duration of the recording varied between 40 and 45 minutes. The time was still long and we had to reduce the 16 long sentences (in version 3) to 25 short sentences (in version 4) with a maximum of 24 words and a minimum of 11 words per sentence. This was also a good consideration to fulfill the display constraint in reducing the displayed text on the phone screen, as mobile screens do not allow very crowded text and buttons in a convivial application. Sample screens from the Tahadath Mobile App are shown in Table 6.

Once the mobile app was developed and sent to diverse students at different geographical locations, we noticed that Non-Arabs had huge problems in reading texts without diacritics; we then updated the texts with a full diacritization. Screen shots of all the texts in the mobile app are presented in Appendix C.

The sampling rate of 8 KHz was decided upon two considerations:

1. Speech recorded at high sampling rates is in general reduced to 16 kHz or 8 kHz, for easiness of manipulation, and simplicity of use in training large models.

2. Recording from the microphone of the mobile phone allows only 8 kHz (sample metadata of the recordings are shown in Table 5.

Table 5, Sample Metadata of the CAPT recordings

Input #0, **'mp3'**: □ filetype
com.android.version: 10
Duration: 00:00:06.74, start: 0.000000, bitrate: 20 kb/s
Stream #0:0(eng): Audio: amr_nb (samr / 0x726D6173), **8000 Hz**, mono, flt, 12 kb/s (default)

1.3. Establishment of the recording system

As already mentioned, we opted to develop a mobile app instead of a computerized application. We had two options, either use unity to develop the app or android using Java. The development of the recording app was subject to the known software lifecycle, presented in Figure 2.

Figure 2, Lifecycle of the Tahadath Mobile app

We have already developed, similar applications while recording speech datasets for a previous KACST funded project [1]. The requirements step was deeply discussed between the members of the team, and the first orientation was the use of a computerized application, but due to the medical restrictions against gathering of students in the institutes, due to COVID-19, we opted to use an app.

The first tentative app was developed by a specialized developer in unity, and has been tested for diverse criteria of screen sizes, colors, etc… We noticed, on the long run, that unity did not support Arabic writing in a very smooth manner, and the app developer had to load and deal with images in the application instead of writing Arabic texts directly in the app. A screenshot of the initial application is presented in Figure 3.

Figure 3, Unity APP screenshot

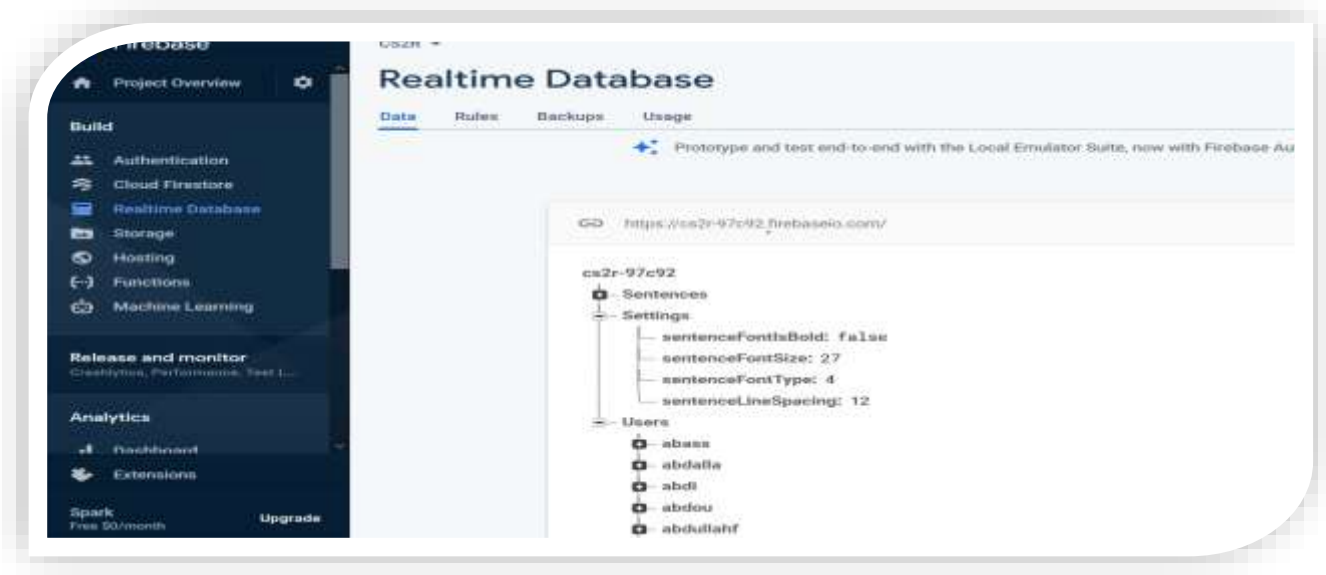
Unfortunately, with the numerous changes of the texts and fonts, we had to move to the development of another application in Java Android. The Java application felt more convivial to Arabic texts and font variations. The Java developer made diverse versions, as per the team requests (design-develop-test). The latest version is the 1.0.10 (10th version), in addition to a second similar application that was also developed for Arabs, as we wanted to split at the database level, Arabs from Non-Arabs recording, for a better management and checking. Some sample screens from the Android App Tahadath are presented in Table 6, a complete listing is also detailed in Appendix C.

Table 6, Sample screens from the Tahadath Mobile App

The backbone of the Android application is the Google firebase, and the application was subject to very strict access, as speakers are invited by their WhatsApp number and the access to

the application is subject to a fixed name and password, generated by our database manager. Access can be done only when a recording flag is enabled, once the speaker reads all the lists and approves its recordings, the recording flag is disabled, and no more access to the database is allowed to that speaker, unless it is reactivated by the admin for checking purposes or recording repetitions. A screenshot of the Google Firebase management platform is shown in Figure 4.

Figure 4, cs2r firebase real-time database



We can notice from Figure 4, that the control of the display font and size can be easily controlled from a centralized part of the app, in addition to the possibilities to change the display sizes at the mobile level.

Additional tests have been also made by the database team, in order to test the app in different mobiles and different versions of android. Some problems appeared in fonts and positions and were fixed as per the maintenance step of the software lifetime cycle.

1.4. Additional Improvements to the app

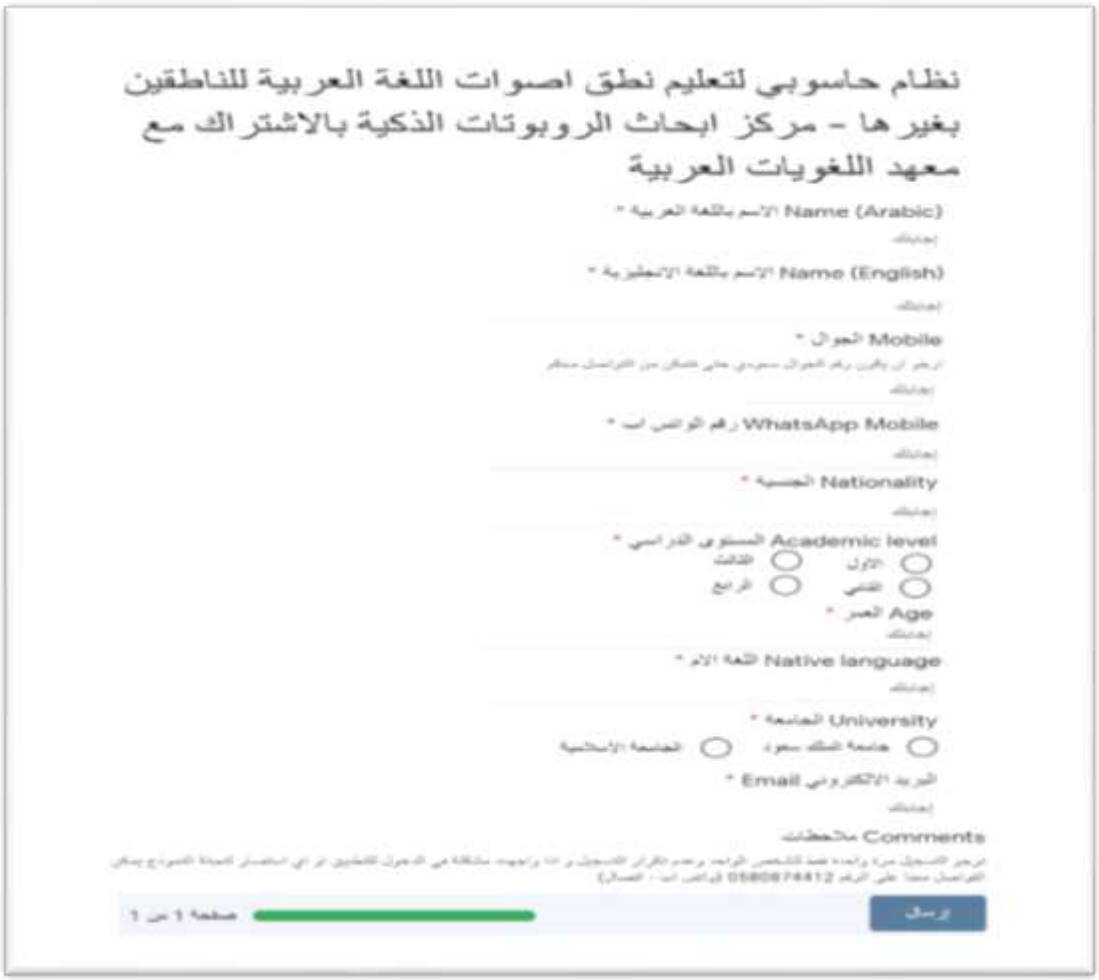
The app that has been developed is a one-way communication, i.e., the speaker records then the recordings are checked. This lead to many problems in terms of quality and recording durations, see Table 2, for benefits and drawbacks of the distant recording. When a speaker records his voice, we had to wait until he finishes his recordings to start checking because different students may be recording at the same time. Hence it was not possible to control every speaker in real time, because each speaker can record at his pace when he feels himself ready.

1.5. Speaker Registration

In the project proposal, we intended to record 200 Non-Arabs and 100 Arabs in the whole project. In order to manage such huge number of students, we followed a sample work methodology, where we start by a small number then increase to the target quota.

To ease the enrollment of the students who were mostly at distant locations, a google form, as shown in *Figure 5*, has been established and sent to the volunteers directly or to a coordinator from each institute who will send to students that he recruit at his institute. Each student needed to fill all the required fields and send it back. Once the forms are collected, the students are contacted by our team for further explanation of the recording steps or to answer any question.

Figure 5, Screen shot of the Google form sent to the students for the CAPT enrollment



In the following part, we will present some statistics of the enrolled students. These statistics include number of students, country of origin, language spoken, level of education, etc.

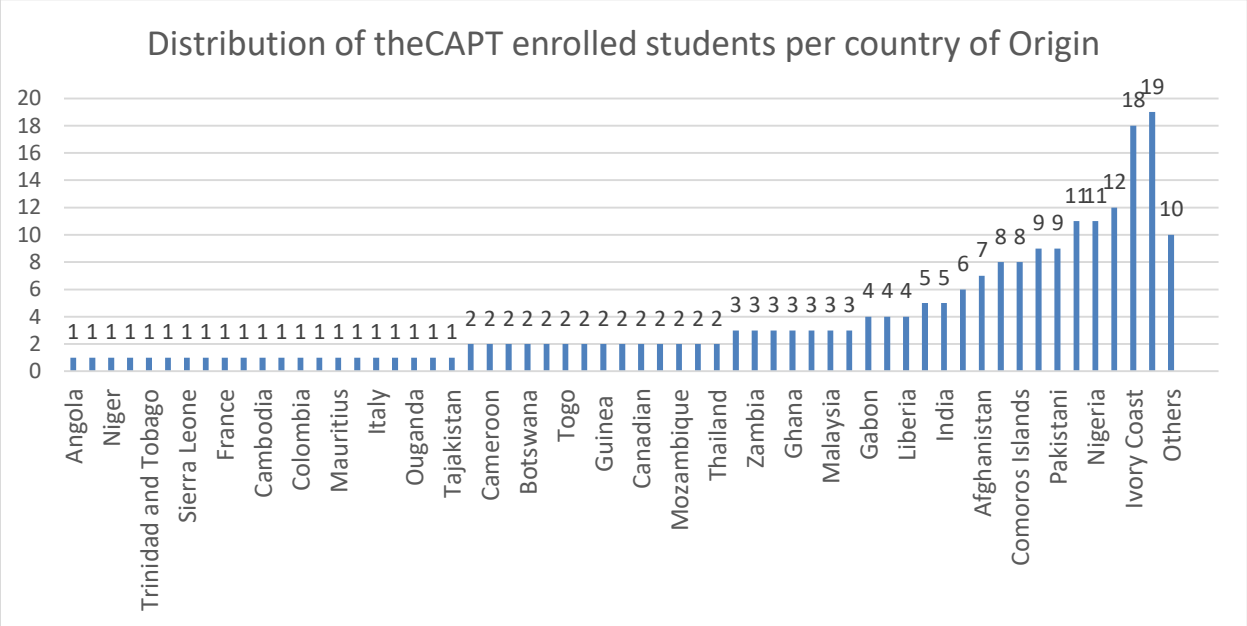
Most of the 371 collected google forms were from the Level 3 and Level 2. After many checking and controls for the validity of the pronunciation of the speakers, and testing their aptitude to pronounce Arabic even with errors, but with a minimal fluency, only 220 students had valid recordings, as shown in Table 7.

Table 7, Number of students that enrolled via the Google Form App

Academic level	Google Form Enrolled Students	Completed Valid Recordings
LEVEL 1	30	13
LEVEL 2	107	82
LEVEL 3	143	77
LEVEL 4	90	48
	370	**220**

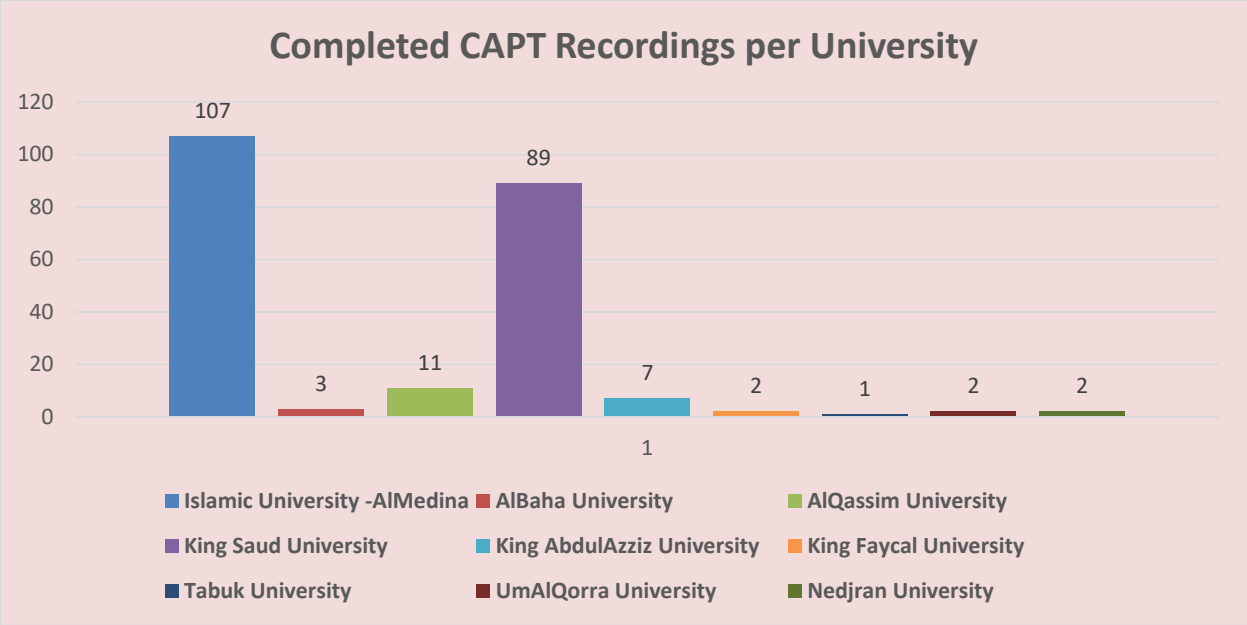
In Figure 6, we present the distribution of the nationalities (more than 59 countries) of the speakers that participated to the recording of the session 1.

Figure 6, Statistics of the Nationalities of the CAPT speech recording at session 1



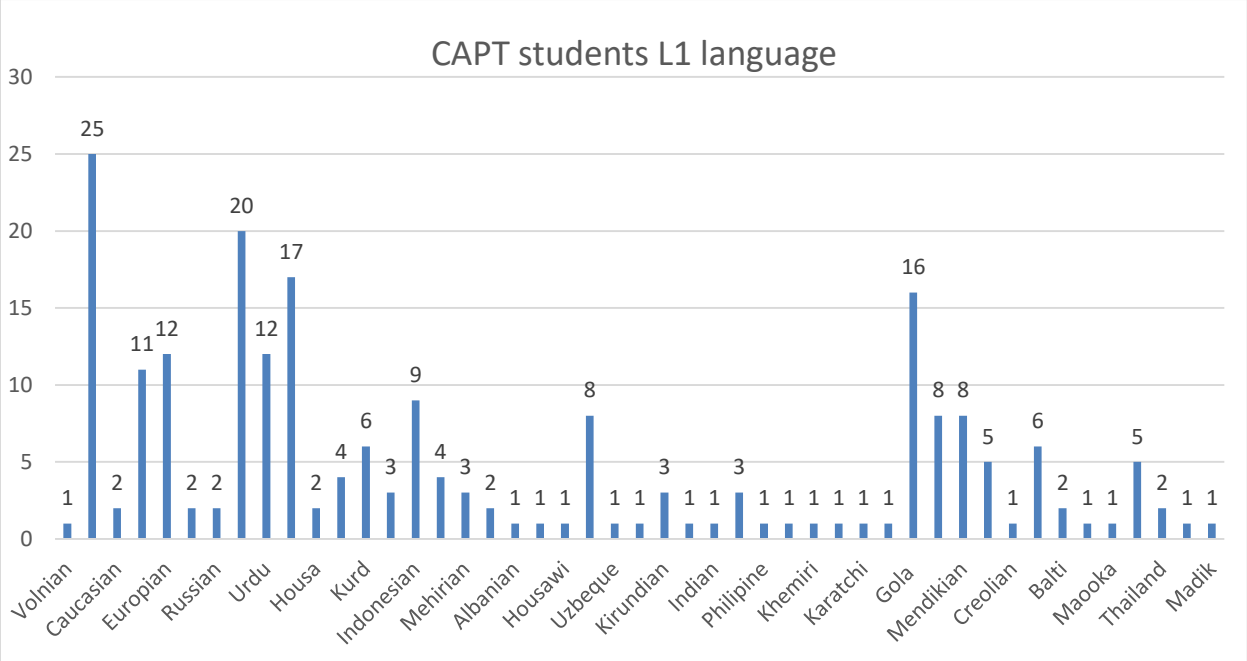
Statistics of the number of completed recordings per university are presented in Figure 7.

Figure 7, CAPT Completed Recordings per University



In Figure 8, we present the number of recorded Non-Arab students from 47 different L1.

Figure 8, Statistics of the L1 of the CAPT speech recording for the session 1 – Non-Arabs



1.6. Database Recording (Session 1)

The enrolled students at session 1, were contacted via WhatsApp or by phone in order to officially make them understand the scope of the recording procedure. Each student was

provided with a multimedia video tutorial, as a demo of the whole recording made by our database manager, in addition to a manual in Arabic and English, explaining all the steps of the use of the app. Each student received all the items listed in Table 8. A copy of the manual in both Arabic and English sent to every enrolled speaker is appended in Appendix F.

Table 8, Credentials received from /sent to the enrolled speakers (students)

	Item	Destination
Student ID :	Ahmed-05555555555	Received within the Google form
Username :	ahmed1	Sent to the student
Password :	123	Sent to the student
Android Application :	Apk format (through WhatsApp)	Sent to the student
Manual :	Tahadath-Manual.pdf	Sent to the student
Video :	App-Demo-Tutorial.avi	Sent to the student
Use of the recorded speech :	Consent screen in the app.	Within the Mobile App

We tried to be as clear as possible, in order to avoid any inconvenience in the use of the app.

1.6.1. Recording Constraints

- Due to the coronavirus, the decrease in the number of students at the ALI institute forced us to turn to the Islamic University of Al-Madinah, as they have more than 1500 students at their premises from more than 117 nationalities, and this helped us a lot in selecting the quality /quantity required by the project.
- The reason for selecting most of the students from outside of Riyadh, is that ALI student dropped from 300 students at the time of writing the proposal to 70 students at recording time and half of them were not physically present in Riyadh.
- The response from students at Islamic University of Al-Madinah was good at the beginning then stalled, so we recruited students from other universities in KSA

1.6.2. Additional Remarks

- A consent text was written in Arabic in the app, the student needed to approve by clicking a check box, before staring the speech recording session.

- The recording of each speaker was accepted, when it is has been double-checked, and passed the quality control criteria defined by the team.
- The student received an honorarium against his participation to the CAPT recordings.
- Many students from the level 1 could not read the texts completely, and were discarded from the recordings.

1.6.3. Recording Arab Speakers

Recording of Arab speakers started after recording of Non-Arabs. The project team tried hard to recruit Arab volunteers by personal invitation and by sending the request to participate in many WhatsApp groups. The response is very slow but the team is trying hard. The number of those who registered in the system database is 58 and among them 32 recorded their speech. The team is still trying hard to recruit. Vacation may be a major reason for the slow response.

2. Arabic-CAPT-2

2.1. Selecting a new text for the recording of Arabic-CAPT-2

We cooperated with a linguistic scholar who is also an experienced instructor of Arabic as a second language to propose a new methodology to choose the text to complement the methodology that we used to select the text of session 1. Below is the new methodology in Arabic.

ما زالت دراسة نطق الأصوات العربية تفتقر إلى قاعدة بيانات مناسبة (**Data Base**) والتي يمكن أن يعتمد عليها في إنشاء برامج حاسوبية تتميز بدقة عالية لتعليم العربية لغير الناطقين بها أو التعرف على أصواتها. ونظرا لصعوبة دراسة نطق جميع الألفاظ العربية في فترة زمنية محدودة، فقد تصل الألفاظ التي تمثل الظواهر الصوتية المختلفة لنطق الصوت الواحد إلى المئات، وهو أمر غير ممكن في ضوء الإمكانات والحدود الزمنية للمشروع الواحد؛ نشأت الحاجة الملحة إلى اختيار عينة مناسبة من الألفاظ بحيث تشمل الظواهر الصوتية المختلفة قدر الإمكان. ومن هنا، قمنا بعمل هذا النموذج ليحقق هذه الغاية، أي اختيار ألفاظ تشمل الظواهر المختلفة لنطق الأصوات العربية لغير الناطقين بها مع مراعاة المدة الزمنية التي تستغرقها قراءة الكلمات بحيث لا تزيد عن عشرين دقيقة تقريبا؛ حيث قمنا باتباع الأسس العلمية والخطوات المنهجية –المبينة أدناه– التي تضمن لنا تحقيق هذا الهدف، وذلك بعد الرجوع إلى أهم الدراسات الصوتية إلى جانب برامج وكتب تعليم اللغة العربية لغير الناطقين بها، فضلا عن خبرة الباحث ومعرفته الطويلة في الظواهر الصوتية العربية ومشكلاتها.

منهج الاختيار (**Methodology**):

لقد ركز هذا النموذج بشكل أكبر على الأصوات التي لا تتوفر في اللغات الأخرى، كالأصوات الحلقية وصوت الضاد والظاء والطاء، إلى جانب الأصوات اللغوية الأخرى التي تكمن صعوبتها للمتعلم غير العربي في كونها تجاور أصواتا أو توجد في مقاطع صوتية غير معتادة لدى المتعلم للغة العربية. وقد اتبعنا المعايير العلمية والخطوات المنهجية (سنذكر خطوة واحدة منها لانها مازالت في طور النشر) :

1- التجاور الصوتي (Juxtaposition): لا شك أن المجاورة الصوتية إحدى أهم العوامل المؤثرة في نطق الأصوات، فالتأثير المتبادل بين الأصوات المتجاورة ظاهرة لغوية مهمة في دراسة الأصوات، وتحدٍّ دائما ما يواجه متعلم اللغة الأجنبي. ومن مظاهر هذا التأثير، الإبدال، أي تغيير صوت بالتعويض عنه بصوت آخر. ولا نقصد بالإبدال هنا التغير الأصلي لنطق الصوت، كنطق التاء دالا في (ازدهر)، وإنما نقصد به تبديل المتعلم للغة العربية صوتا ما بصوت آخر مقارب تسمح به لغته الأصلية، كنطق القاف كاف أو خاء. وقد يكون الصوت موجودا في لغة المتعلم غير العربي لكنه لا يأتي في لغته مجاورا لصوت آخر، مثل لفظ (جِصّ)، الصوتان موجودان في اللغة الإنجليزية، لكن لا يوجد مثل هذا التجاور الصوتي فيها، ومن هنا تنشأ إشكالية النطق.

2.2. Completing the recording of Arabic-CAPT-2 (male and female speakers).

We re-contacted the non-native Arabic speakers of Arabic-CAPT-1 to record the new selected text. This time we provided a training sample hosted on a website to allow the speakers to listen to the recordings of an Arab expert, before recording the 42 screens of the mobile app. The average recording duration ranged between 15 to 20 min per speaker. 230 non-native speakers have been recorded in this session. In addition, we followed up by recording 80 non-native females from different nationalities.

2.2.1. Samples of text selection of session 2

Below is an example of the selected text for the Arabic phoneme (/ق/), which consists of isolated phonemes, words, and minimal pairs. The same procedure has been done for the remaining Arabic phonemes.

أَقْ، أَقْ، إِقْ، قَا، قُو، قِي

قَتادَة قُنْبُرَةْ قِصَّةْ بُرْقُوقْ مُبَرْقَعْ اِقْتَصِدْ تَقَطَّرَ اِسْتَقْرَأْ مِقْياس تَقَوقُعْ غاسِقْ يُشْفِقْ

قَسَا	كَسَا
يَقِيْل	يَكِيْل
تَقْدِيْر	تَكْدِير
شَقَّ	شَكَّ

مُشْرِقِ	مُشْرِكْ

2.2.2. Statistics from the recording of session 2

Figures 3-6 show the statistics of session 2 recording in terms of nationality, education level, mother language, and university.

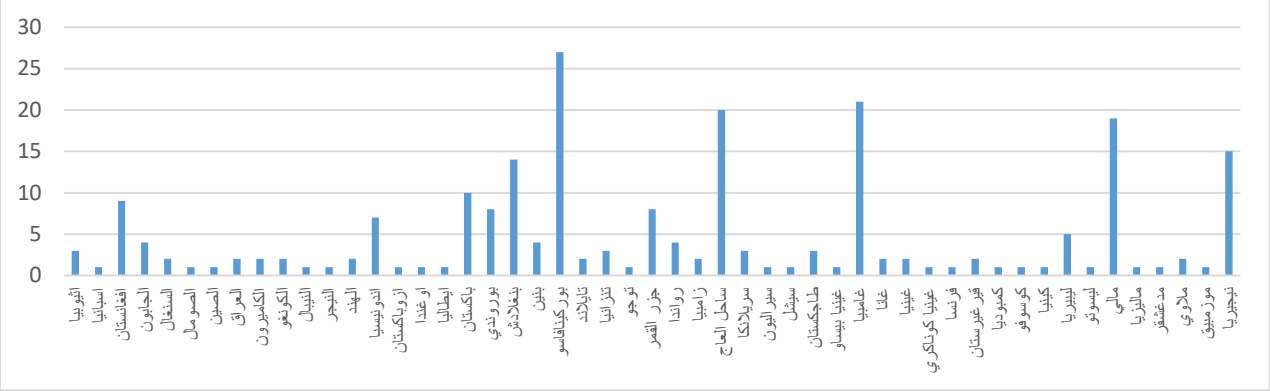


Figure 9: Nationalities distribution of the speakers of session2.

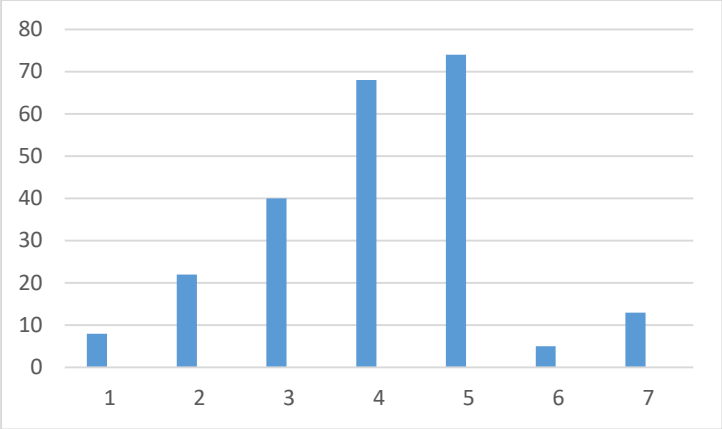


Figure 10:Level distribution of the speakers of session2.

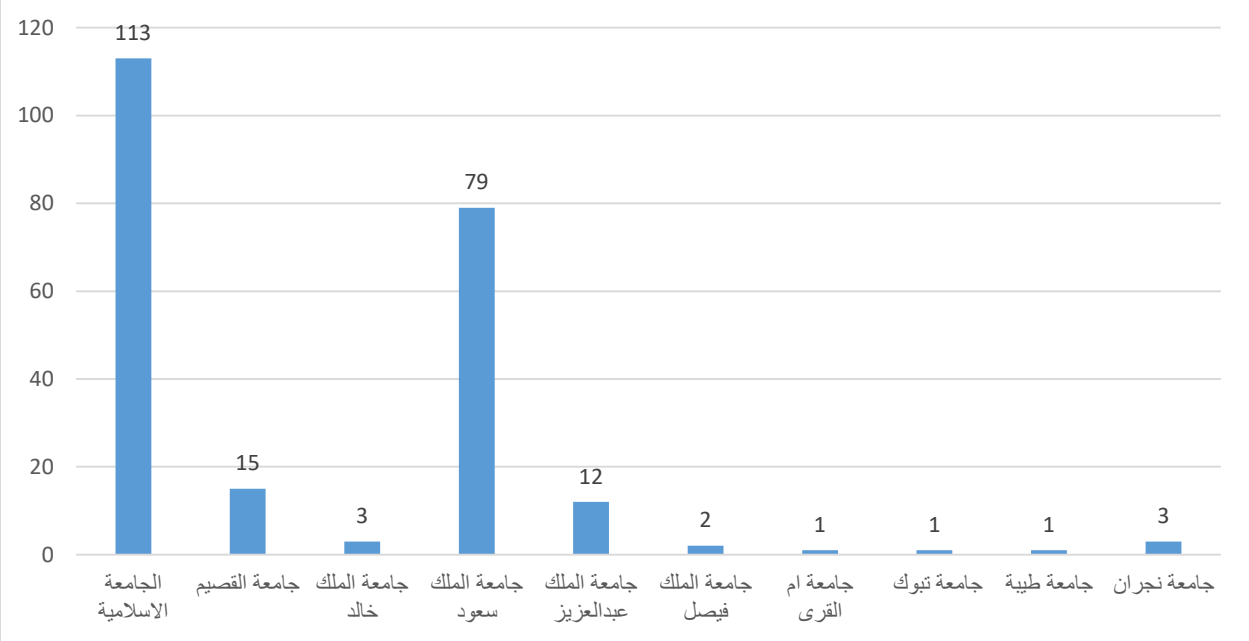


Figure 11: Universities distribution of the speakers of session2.

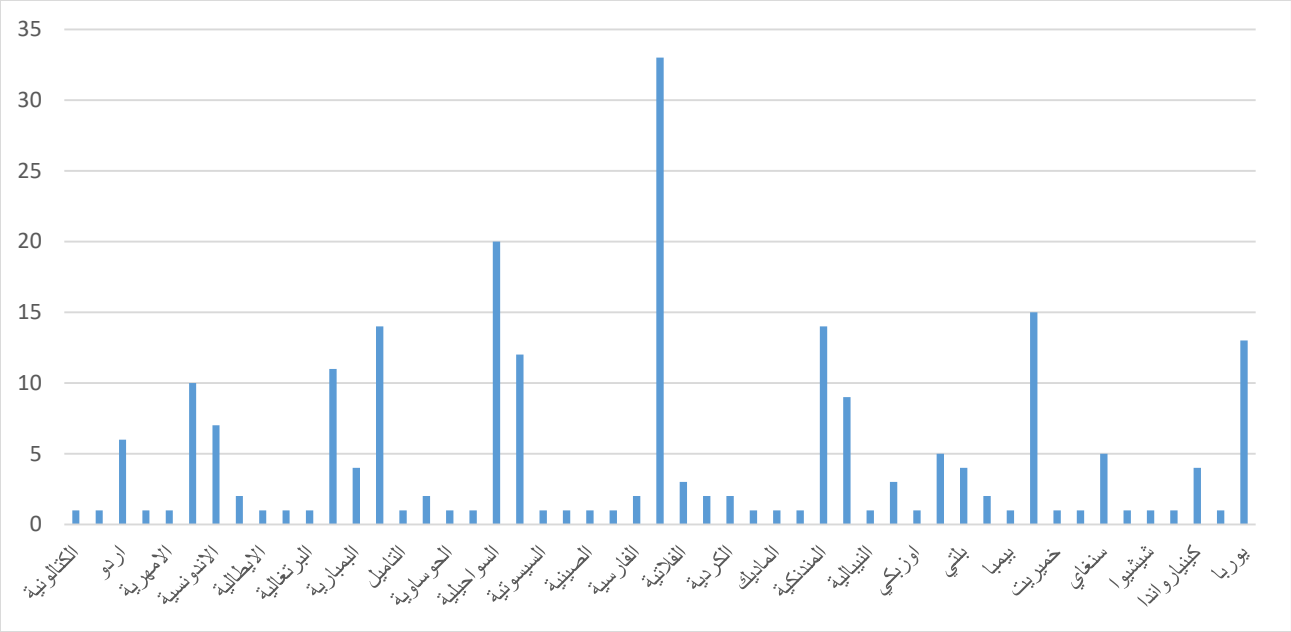


Figure 12: L1-Language distribution of the speakers of session2.