# Automatic Lecture Annotation

Nada Nasser
*Media Engineering and Technology*
*German University in Cairo*
Cairo, Egypt
nada.nasser@guc.edu.eg

Jailan Salah
*Media Engineering and Technology*
*German University in Cairo*
Cairo, Egypt
jailan.salah@gmail.com

Nada Sharaf
*Media Engineering and Technology*
*German University in Cairo*
Cairo, Egypt
nada.hamed@guc.edu.eg

Slim Abdennadher
*Media Engineering and Technology*
*German University in Cairo*
Cairo, Egypt
slim.abdennadher@guc.edu.eg

***Abstract*—This Research to Practice Full Paper presents an application (AutoNotes) that improves the students' learning outcomes by using natural language processing techniques (NLP) such as speech recognition and keywords extraction. AutoNotes is an android application that converts the speech of the instructor to text in real time during the lecture and then it synchronizes this text to the original lecture slides to produce new slides annotated with the explanation of the instructor. These new lecture slides can be displayed in a classroom. Moreover, it can be accessed by the students on their mobile phones. AutoNotes also provides a feature for the students who like to write their notes with their own way. A focus group was conducted to gather target audience opinions. The results have shown that students liked the idea of AutoNotes. They saw it beneficial for them especially if the original lecture slides are not detailed. They also like the idea of putting the notes with its corresponding slide. To evaluate AutoNotes, a between group design setup of two identical lectures with different groups of participants was conducted. In the first lecture, the lecturer used the original lecture slides. However, in the second one the annotated lecture slides was used. The participants were given several tests to compare their overall learning experiences. The results of the tests reveal that the participants that were exposed to the annotated lecture slides had a higher engagement level and learning gain than the other group exposed to the original lecture slides. Additionally, they had a workload level less than the other group. For the system usability test, the average usability scale that AutoNotes got is 78 which can be considered above average. Finally, the results of the accuracy testing reveal that the accuracy of the speech recognition was 83.2%. The precision and recall metrics were used to evaluate the synchronization of the speech with the slides. The average precision score was 75.6%. However, the average recall score was 51.1%.**

***Index Terms*—Speech Recognition, Classroom, Learning, Lecture, Slides, Usability, Accuracy**

## I. Introduction

Students can be either auditory learners who learn better through their sense of hearing so that they remember and understand information when explained out loud [1]. Another type of learners is visual learners who tend to learn through their sense of sight and this student is usually found at the front of the classroom soaking up whatever the instructor writes on the board during the lecture [1]. Thus, the instantaneous display of the speech of the instructor as text in the classroom will let the students not only hear the explanation of the instructor but also read it in the same time. This will thus give the chance to all of the students to better learn and understand the lecture content [1].

Most of the instructors design their lecture slides only to guide them during their lectures. The content of these slides is typically organized in brief points that outline the main ideas, but do not provide detailed explanations. Thus, the students could not depend on these lecture slides as their studying material because they are not enough for studying. For this reason, taking notes is an essential part of every student's life to retain the explanation and the information said by the instructor during the lectures to use these notes to study. However, students face several problems in taking notes. It is very difficult for most of them to concentrate on the explanation of the instructor and take notes at the same time. It is also difficult for them to write notes as fast as their instructors speak specially for students with disabilities such as hearing impaired students. So, having an application that eliminates these difficulties and provides them with annotated lecture slides will facilitate their learning process.

The aim of this project is to design and implement an android application that improves the students' learning outcomes. This is done by providing the students with both auditory and visual learning and eliminating all the difficulties that the students face to take notes. This also facilitates the study process for the students by providing them a good and complete study material.

This paper is organized as follows: Section II introduces the prior works, Section III tackles all the features included in AutoNotes. Then, Section IV describes the system architecture. Section V explains the experimental design approach used to test AutoNotes. Section VI includes the detailed analysis of the results of all the tests held during the experiments. Finally Section VII is the conclusion of the paper which includes an overall summary and states the recommended future work.

## II. Related Work

The real time captioning systems are the systems that use real time speech recognition technology to translate the speech of the instructors to text and display this text for the students in the classrooms. Additionally, this text also can be used as an alternative to traditional classroom note-taking [2] [3] [4].

However, the non real time synchronization systems are the systems that use non real time speech recognition technology to synchronize an audio soundtrack or a speaker video with the electronic slides from a presentation [5] [6] [7] [4].

The systems that can do both are very rare. There are few systems that can convert the speech of the instructor to text in real time and after the lecture it synchronizes this text with the lecture slides.

There is a study [4] that combines and compares these two categories. It introduces two systems. The first one is a real-time captioning system that was implemented as a client-server application for instant viewing during class on a projection screen or directly to the laptop personal computers of the students. ViaScribe (version 2.3.7) [8] was chosen for real-time captioning because it had a proven track record for reliable captioning and had a client-server platform for streaming live transcription to the laptop PCs of the students during lectures. The second system is a post-lecture synchronization system that employed a digital audio recording of the lecture to provide transcripts. Then, these transcripts are synchronized with the audio recording and class PowerPoint slides for students to view online or download after class. In this system, IBM Hosted Transcription Service (HTS) was chosen to be used because of its higher word recognition accuracy rates. IBM HTS is a speaker-independent speech recognition software developed by IBM Research that automatically transcribes a variety of standard audio or video file formats through a cloud service [4]. Students in both cases access the multimedia files combining audio, transcripts, and PowerPoint slide images by using a website called Synote [4]. Both systems were evaluated in four phases during both social science and life science lecture courses at a public university to explore diversity in lecture material and teaching styles [4]. Students stated that they benefited the most from having multimedia class notes [4]. They have been able to pay more attention to the instructor instead of focusing on recording complete class notes [4]. Moreover, they had the ability to review the lecture material multiple times which facilitates their study process. Additionally, they liked the idea of the synchronization of the lecture audio with the transcripts and the slides because it provides them with a better learning experience [4].

A study was performed to implement a system that automatically synchronize an audio soundtrack with electronic slides from a presentation [5]. This system was mainly implemented to record the university and college lectures. The new slides generated are made available online by lecturers for private study by those attending the lectures. They are also utilized for distance learning by students physically remote to the institution. This system focuses on synchronization of digital audio recordings and slide presentations using linguistic content to provide searching of the actual information delivered in the lecture. It demonstrates the application of a combination of techniques from information retrieval to segment and align slides with the spoken contents of the audio presentation. Additionally, due to the fact that lectures often do not exactly follow the planned structure, the system was able to detect the slides within a presentation that are not covered in the audio of the lecture. Then, it processes these parts to automatically create for them new slides in the presentation [5].

There is also a system addressing the task of automatically synchronizing presentation slides with the speaker video, through a global adaptive curve fitting algorithm [7]. The results showed that the accuracy of the system is sufficient that the manual checking and correction is many times faster than the existing completely manual approach.

Another systems have different objective. A method to correct automatically generated speech transcripts of talks and lecture videos using text from accompanying presentation slides was presented [6]. The approach solves the problem of dealing with technical terms which are often outside the vocabulary of speech recognizer. It align the transcript to the slide word sequence so that it can improve the organization of closed captioning for hearing impaired users, and improve automatic highlighting or magnification for visually impaired users. It shows that it is feasible to correct speech transcripts using slide words, despite the limited overlap of words between the two sources. it had achieved this using an alignment approach that can absorb the many spurious words. The method is effective because many of the correctable words are both important and difficult for speech recognizer. Furthermore, the alignment of these two disparate data streams is beneficial for improving access to educational video, especially for users with disabilities [6].

There are some common issues detected in the evaluation of most of the non real time synchronization systems [5] [7]. These issues are related to the task of automatically synchronizing presentation slides with the transcription. First, the summary or concluding slides which feature a high density of keywords can incorrectly match with any section of the transcription. Additionally, there are some slides match very poorly with any part of the transcript,even though they were covered in the presentation. This was most often because the slide contains little text, for example if the slide contains a figure, or if the lecturer happens to use different vocabulary to describe the topic covered by the slide. This issue is difficult to address because a lecturer may discuss at length a slide showing a graph or image, but there is no information to match to the transcription. So the system is not always correct and the output should ideally be checked by a human operator [5] [7] [4].

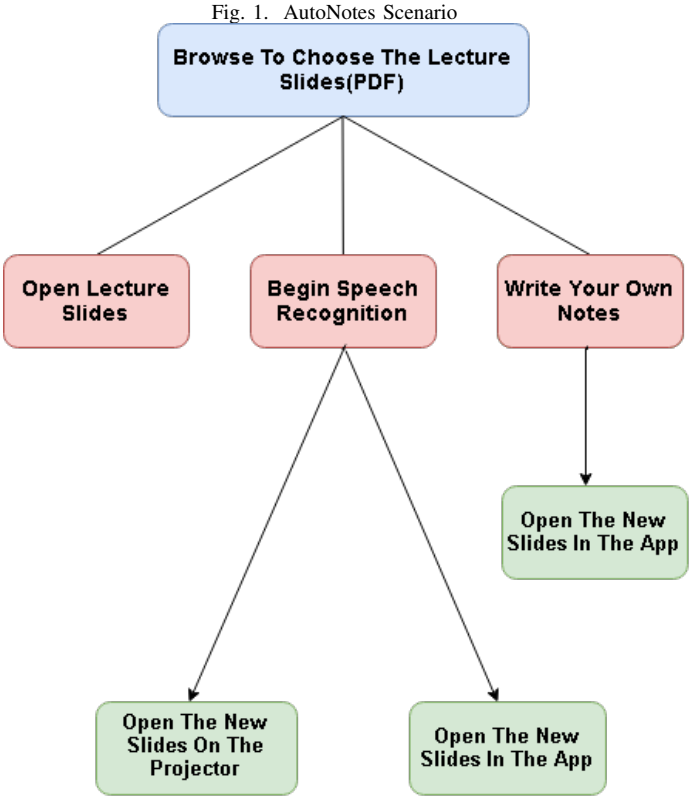
AutoNotes is different from the similar existing systems. As mentioned above, some of these systems perform real time captioning so that the speech of the instructor is translated to text and displayed on a screen in the classrooms. Other systems perform automatic synchronisation of the speech transcript of

the instructors with the lecture slides. This synchronisation process is done after the lectures. However, in AutoNotes the speech of the instructor is translated to text then it is synchronised with the original lecture slides to produce the new annotated lecture slides that can be displayed in real time on a screen in the classrooms or accessed by the students' mobile phones during the lectures or at home. Additionally, to the best of our knowledge, there is no similar systems developed as mobile applications. AutoNotes was developed as an android mobile application to be less costly. The instructors and the students use only their mobile phones without the need of any additional tools to be available in the classrooms.

## III. AutoNotes Features

AutoNotes combines several features to be able to facilitate the note taking and the study processes for the students. Figure 1 shows the scenario of AutoNotes. First, the users must

Fig. 1. AutoNotes Scenario




browse and choose the lecture file to be processed as shown in figure 2

Then, the instructors have to choose between open the original lecture slides, begin speech recognition or write their own notes. If a user chose *Begin Speech Recognition*, the real time captioning of the instructor's speech will begin as shown in figure 3.

In addition to the synchronization of the speech transcript with the lecture slides, the new annotated lecture slides can be displayed in classroom on a projector or it can be accessed from the user's mobile phone. If a user chose *Write Your Own Notes*, the user can write notes to any slide in the lecture slides as shown in figure 4.

Every note the user writes will be added to its corresponding slide and the new annotated lecture slides will be saved on the user's mobile phone.

Fig. 2. AutoNotes Features

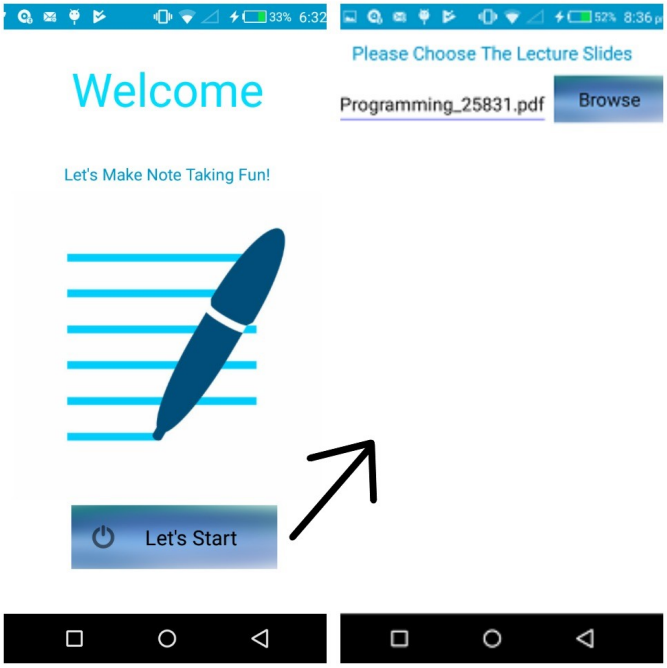



Fig. 3. Speech Recognition

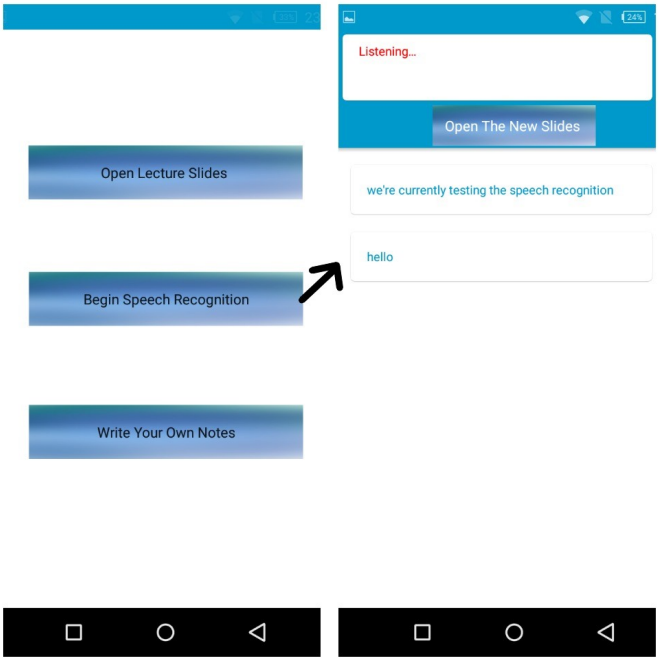
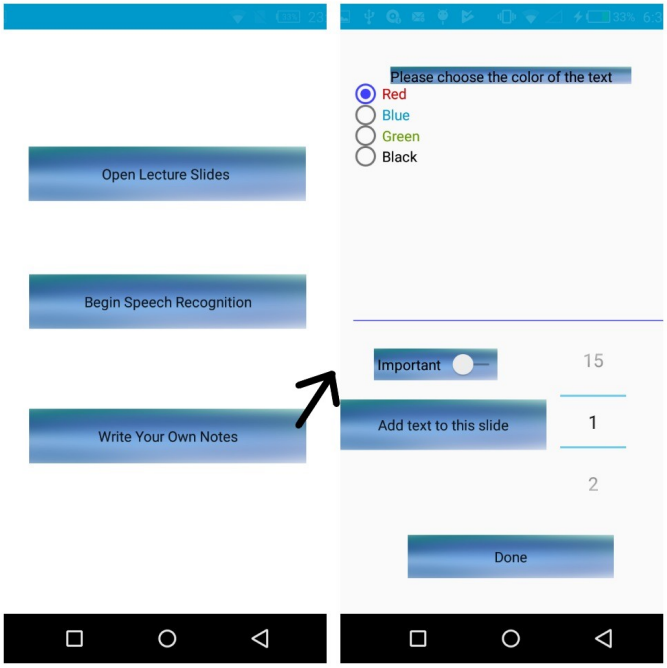


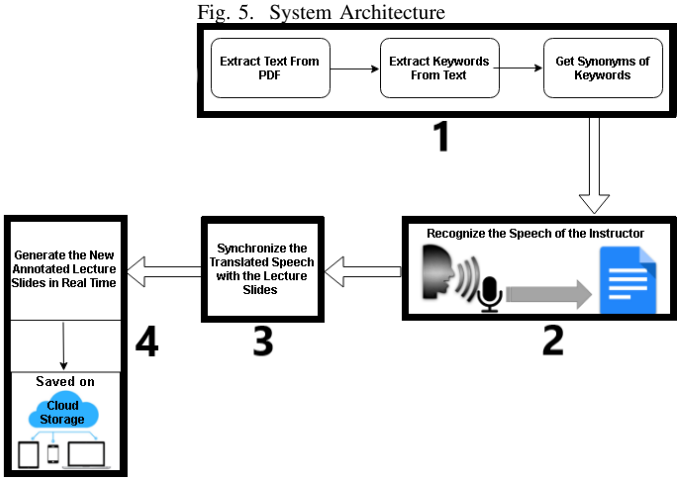
Fig. 4. AutoNotes Features




## IV. System Architecture

Figure shows 5 the four main modules that AutoNotes consists of. This section discusses all these modules and how they communicate with each other.

Fig. 5. System Architecture




### A. Text Processing

This is the module responsible of extracting text from the lecture PDF. The text extraction was implemented using itext PDF library which is a Java library that allows to create PDF, read PDF and manipulate them. Additionally, iText can parse PDFs to extract the content of a page. Thus, the text was extracted from each page in the lecture PDF separately.

After the text extraction, the text extracted is used to extract keywords from it. The Watson keywords API which is named as *Natural Language Understanding* API which is developed for advanced text analytic was found to be very useful to do this task. This API can extract from text emotions, sentiments and keywords. However, it was used only to extract the keywords from the text of each slide of the PDF. It also gives the option to the users to choose the maximum number of keywords required. So, 20 keywords are extracted from each slide. However, sometimes it returns less than that if a slide has no enough text to extract 20 keywords from. The text extracted from each slide in the lecture was used to call the Watson API. So that, each response of the API requests contains the keywords of each slide separately.

As most of the lecturers do not explain the lecture with the same words that exist in the lecture slides. Therefore, synchronizing the speech transcript of the lecturer with the lecture slides using only the keywords extracted from each slide is not sufficient and will lead to a low accuracy. For this reason, the Big Huge Thesaurus API was used which is a very simple API for retrieving the synonyms for any word. Thus, in this module the synonyms of all the keywords extracted from the lecture slides are retrieved.

### B. Speech Recognition

The real time captioning feature was implemented using the speech recognition technology. After a lot of searching and trials of several speech recognition APIs, it was found that Google Cloud Speech-to-Text API is the best to use for several reasons. First, it is one of the most accurate speech recognition APIs. Additionally, it can translate the speech to text in real time as it returns results while the user is still speaking. It is also a speaker independent speech recognition which is designed to recognize the voice of anyone without the need of any training. In this module, the speech of the instructor is streamed to chunks and each chunk is converted to text in real time. To do that, several android components were used. A voice recorder class was implemented as the first step in the implementation of the speech recognition. To implement it, we have used the Media-Recorder API which is provided by the Android multimedia framework. Additionally, a service called *The Speech Service* was implemented which is mainly responsible of Google Cloud Speech-to-Text API request and response. It overrides the regular methods of any android service. Another android activity was used which represents the speech recognition screen. It has an instance from the voice recorder class and also an instance from the speech service. To display the translated text as a list on this screen by using a recycler view and an array adapter.

### C. Synchronization of Transcript with Lecture Slides

This is a very important module which takes all the outputs of the previous mentioned modules as its inputs. This module is mainly for implementing a matching algorithm which takes all the keywords and their synonyms got from all the lecture slides to match between these lecture slides and the speech transcript of the instructor.

As mentioned in Speech Recognition Section, each chunk of the speech of the instructor is recognized and converted to text in real time. Thus, each chunk is taken as an input in this module. Then, each chunk is matched against each slide to find the slide with the highest matching score. This matching task is done by checking the occurrences of any of the keywords or their synonyms from a specific slide in the transcript chunk. Each keyword or synonym from a specific slide found in this chunk increases the matching score of this slide.

So that the chunk of text is added automatically to the slide with the highest matching score. This matching algorithm is almost synchronized with the time and the normal flow of any lecture by using the following constraints:

- Sometimes the first chunk of the speech transcript does not match with any of the lecture slides which means that all the matching scores are zeros. Thus, it is automatically added to the first lecture slide as most of the instructors begin the explanation from the first lecture slide.
- There could be also some chunks that match with two or more slides with the same matching score. This case is handled by choosing the slide that is closer to the slide that was matched with the previous chunk.
- Moreover, there could be some chunks that do not match with any slide of the lecture slides. In this case, these chunks are added to the slide that was matched with the previous chunk.

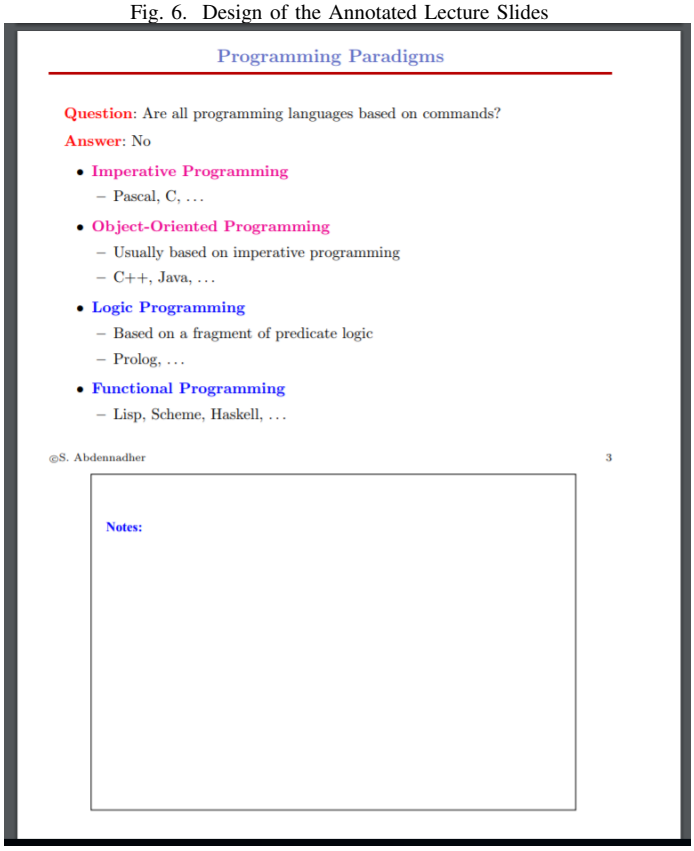
By adding these constraints to the matching algorithm, it is almost synchronized with the time. Thus, each chunk of the speech transcript should found its corresponding lecture slide. However, the accuracy of this matching algorithm depends on the content of the lecture slides. If a slide contains a high density of keywords, it can incorrectly match with a lot of wrong chunks of the speech transcript. Additionally, if a slide contains little text or contains only a figure, it will not match with any of the chunks.

### D. Generate Annotated Lecture Slides

The itext library was used in this module to create the new PDF of the annotated lecture slides. Due to the fact that every instructor uses different template, fonts and colors in their lecture slides. The students are also familiar with the way of every instructor in designing the lecture slides. The new slides are designed to be the same as the original lecture slides with the same design. Only a rectangle that contains the notes is added at the bottom of each slide as shown in Figure 6. Every note is added to its corresponding slide in real time. So that the new annotated lecture slides are also generated in real time.

The new PDF of the annotated lecture slides is saved automatically in a cloud to let the lecturer be able to display it on the projector of the classroom. The Firebase Cloud Storage [9] was chosen to save the new lectures on it because it is powerful, simple and cost-effective object storage service. It is mainly built for application developers who need to store and serve user-generated content. Additionally, the new PDF is also saved on the users' mobile phone to give them the ability to open and view it on their mobile phones.

Fig. 6. Design of the Annotated Lecture Slides




## V. Experimental Design

### A. Focus Group

A focus group was conducted after the implementation of a prototype to gather the target audience opinions and attitudes about AutoNotes. It was conducted with 10 participants who are college students from diversity of majors. A comfortable venue and recording method were chosen. The focus group was held in a private area where participants were feeling relaxed and comfortable. The seats were organized into a circle to make all participants feel more equal and comfortable in participating. Thirteen questions were designed in a way to encourage participants to open up and talk about their opinions in depth so we have used a lot of open-ended questions.

### B. Between-Group Design

To evaluate AutoNotes, we have conducted a between group design setup of two identical lectures with different groups of participants. In the first lecture, the lecturer used the original lecture slides. However, in the second one the annotated lecture slides that is generated in real time by AutoNotes was used. As this experiment needs two groups of participants to be conducted. 22 participants took part in this experiment so that each group contains 11 participants. The participants were all college students from different ages, gender and majors. They consisted of 7 males and 15 females. Their average age was 21 years old. They were from Engineering (Media Engineering and Technology, Mechatronics and Information Engineering

and Technology), Pharmacy and management faculties. Most of these participants had no knowledge about the topic that would be explained in the lecture to be able to test their knowledge gain from the lecture.

A test on the content that was explained during the lecture was given to the participants before and after the lecture. It is a 10 MCQ questions test that contains some questions on the information that are already written in the original slides and the other questions are on the information said by the lecturer during the lecture (See Appendix ). The idea behind this test is to assess the participants' knowledge gain by comparing the results of those pre and post tests. Then, by comparing the participants' knowledge gain of the two groups we can conclude how the annotated lecture slides generated by AutoNotes has affected the participants' knowledge gain and how it was useful for them. For this reason, The pre and post tests were identical to ensure that they have the same level of difficulty.

As one of our objectives is improving student engagement in lectures. The Engagement Test was used to compare the students' engagement with the lecture content in both lectures with the two different groups of participants. This test measures the overall experience that the participants have gone through by attending one of the two lectures with the different lecture slides [10].

Then, the Official NASA Task Load Index (NASA-TLX) test [11] was used to compare between the participants' workload of the two groups to conclude how the annotated lecture slides helped the participants in the lecture and how it affected their workload.The NASA-TLX is a subjective workload assessment tool that calculates perceived workload in order to assess a task.

Finally, the System Usability Scale (SUS) [12] which is a reliable tool for measuring the ease of use of any system was given to the participants that used AutoNotes.

### C. Accuracy Testing

To evaluate the accuracy of AutoNotes, it was important to test it in the normal classroom environment without any preparation. The aim of this evaluation was to test the speech recognition accuracy in the classroom environment. This is done to the fact that recognizing clearly spoken well formed sentences in a quiet environment is not the same as recognizing spontaneous speech in a classroom with background noise. The dynamism present in this environment which generates false starts, dis-fluencies, hesitations, ungrammatical constructs reduces the accuracy of the speech recognition. Additionally, this evaluation was also to test the accuracy of the speech transcripts synchronization with the presentation slides.

AutoNotes was used by an instructor during a research colloquium talk conducted in the university. The instructor was presenting using 36 presentation slides which are enough to test the accuracy of the synchronization and to observe all its technical issues. Unfortunately, the presentation slides and the speech of the instructor was containing a lot of Arabic parts. Thus, due to the fact that AutoNotes recognizes only English words these parts were incorrectly recognized which decreased the accuracy of the speech recognition as well as the synchronization of the speech with the slides. Additionally, The speed of the speech of the instructor also affected the accuracy of the speech recognition which then affected the accuracy of the synchronization of the speech with the slides.

It was observed that there are several factors that can affect the accuracy of the speech recognition which are:

- The speed of the speech of the instructor
- The English pronunciation of the instructor
- The amount of code switching in the speech of the instructor
- The distance between the microphone and the instructor's mouth
- The loudness of background noises
- The quality of the microphone

There are also several factors that can affect the accuracy of the synchronization of the speech with the slides which are:

- The number of slides that contain little text
- The number of slides which feature high density of keywords
- The speech recognition accuracy

TABLE I
INDEPENDENT T-TEST RESULTS OF THE LEARNING GAIN

| Group | N | Mean | SD | SE |
|---|---|---|---|---|
| 1 | 11 | 4.364 | 2.111 | 0.636 |
| 2 | 11 | 6.364 | 1.286 | 0.388 |

## VI. RESULTS

In this section, The results of all the tests that have been held during the experiments are reported which are: The learning gain test, engagement test, task load index test and finally the system usability scale test. The result of the accuracy testing is also analyzed.

### A. Learning Gain Test Results

The aim of this test is to measure the difference between the learning outcome of the participants of the group that was exposed to the original lecture slides and the other group that was exposed to the annotated lecture slides. The difference between the results shows that the learning gain of the participants that had the chance of read and listen to the explanation at the same time using the annotated lecture slides was higher than the learning gain of the participants that only listened to the explanation. The independent t-test also confirmed that the learning gain of group two which is the group exposed to the annotated lecture slides (M=6.364, SD=1.286) was significantly higher than the gain of the first group (M=4.364, SD=2.111) (t=2.684, p=0.014) as shown in Tables I and II.

TABLE II
INDEPENDENT T-TEST RESULTS OF THE LEARNING GAIN

| t | df | p |
|---|---|---|
| -2.684 | 20.00 | 0.014 |

TABLE III
INDEPENDENT T-TEST RESULTS OF THE ENGAGEMENT LEVEL

| Group | N | Mean | SD | SE |
|---|---|---|---|---|
| 1 | 11 | 26.64 | 3.501 | 1.055 |
| 2 | 11 | 29.73 | 3.165 | 0.954 |

### B. Engagement Level Test Results

As the aim of this test is to measure the overall engagement and involvement level of the participants of both groups exposed to the original and the annotated lecture slides. By comparing the results of the two groups, we have determined the group that achieved the highest engagement level.

Tables III and IV show the independent t-test results of the two groups which indicate that the engagement level of the participants of the group that was exposed to the annotated lecture slides (M=29.73, SD=3.165) was significantly higher than the engagement level of the other group (M=26.64, SD=3.501) (t=2.172, p=0.042).

### C. Task Load Index Test Results

The aim of this test was to compare the workload done by the participants of the group exposed to the original lecture slides and the other group that was exposed to the annotated lecture slides.

The results of the independent t-test between the two groups are shown in tables V and VI. These results show that the participants' workload of group two which is exposed to the annotated slides (M=3.924, SD=0.835) is significantly less than the workload of the first group (M=4.679, SD=0.711) (t=2.282, p=0.034).

### D. System Usability Scale Results

The System Usability Scale measures the usability of a system and the satisfaction of their users. The average SUS score is 68. A SUS score above a 68 would be considered above average and anything below 68 is below average.

This questionnaire was taken only by the participants of the group that was exposed to the annotated slides and tested the application. Figure 7 shows the SUS score which is out of 100 given by each participant in the group. The average SUS score that the system got is 78/100.

TABLE IV
INDEPENDENT T-TEST RESULTS OF THE ENGAGEMENT LEVEL

| t | df | p |
|---|---|---|
| -2.172 | 20.00 | 0.042 |

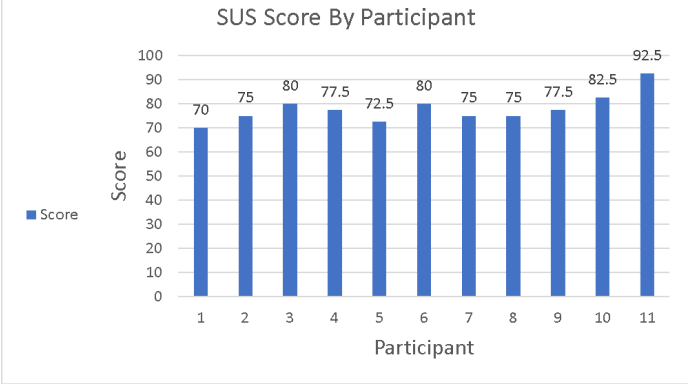
TABLE V
INDEPENDENT T-TEST RESULTS OF THE WORKLOAD

| Group | N | Mean | SD | SE |
|---|---|---|---|---|
| 1 | 11 | 4.679 | 0.711 | 0.214 |
| 2 | 11 | 3.924 | 0.835 | 0.252 |

TABLE VI
INDEPENDENT T-TEST RESULTS OF THE WORKLOAD

| t | df | p |
|---|---|---|
| 2.282 | 20.00 | 0.034 |

Fig. 7. The SUS Scores Given to AutoNotes by the 11 Participants




### E. Accuracy Testing

The aim of this accuracy testing is to evaluate the accuracy of the speech recognition in the normal classroom environment as well as evaluating the accuracy of the synchronization process of the speech of the instructor with the presentation slides. The speech recognition accuracy was simply calculated by getting an accuracy score for the notes of each slide. Then, the average of all the scores was retrieved. Finally, the average accuracy percentage was calculated which is 83.2 %.

Precision and recall metrics were used to analyze the result of the synchronization process of the speech of the instructor with the presentation slides. The precision and recall evaluation metrics are used in machine learning and information retrieval. The precision is the fraction of relevant instances among the retrieved instances. However, the recall is the fraction of relevant instances that have been retrieved over the total amount of relevant instances. High precision means that an algorithm returned more relevant results than irrelevant ones. High recall means that an algorithm returned most of the relevant results [13] [14].

The recall is the number of true positives divided by the number of true positives plus the number of false negatives. True positives are data points classified as positive that actually are positive (correctly classified), and false negatives are data points identified as negative that actually are positive (incorrectly identified). The precision is defined as the number of

true positives divided by the number of true positives plus the number of false positives. False positives are cases incorrectly labels as positive that are actually negative [13] [14].

A precision score was calculated for each slide in the presentation by dividing the number of the correctly matched sentences with this slide over the total number of sentences that matched with this slide. However, the recall score was calculated for each slide by dividing the number of the correctly matched sentences with this slide over the number of sentences that must be matched with this slide. The average precision score got for all the lecture slides was significantly higher than the average recall score. The average precision score was 75.6 %. However, the average recall score got was 51.1 %. This was due to the fact that AutoNotes can only recognize English words. However, the instructor used a lot of Arabic sentences in her speech in the presentation. Thus, a lot of parts in the speech was not recognized correctly which decreases the average recall score.

## VII. Conclusion and Future Work

### A. Conclusion

AutoNotes is an android application that can convert the speech of the instructor to text. Then, it synchronizes this text to the lecture slides to generate new annotated lecture slides. All of that is done in real time and the new lecture slides is saved on a cloud and on the mobile phone used.

The results of the focus group conducted show that all of the students like the idea of AutoNotes and they see it very beneficial for them specially if the original lecture slides are not detailed. They like the idea of putting the notes with its corresponding slide. They like also the fact that it writes all the information said by the instructor as notes so it saves their time spent in listening to the lectures recordings and also it saves their effort done to write notes during lectures.

The results of the tests taken by the participants in the experiment done reveal that the participants that used AutoNotes and were exposed to the annotated lecture slides had a higher engagement level and learning gain than the other group exposed to the original lecture slides. Additionally, they had a workload level less than the other group. The average usability scale that AutoNotes got is 78 which can be considered above average.

Finally, the results of the accuracy testing reveal that the accuracy of the speech recognition was 83.2%. The precision and recall metrics were used to evaluate the synchronization of the speech with the slides. The average precision score was 75.6%. However, the average recall score was 51.1%.

### B. Future Work

Wide variety of ideas can be added to AutoNotes. One of the very useful ideas is to use multilingual speech recognition software which can recognize speech that contains words from more than one language because a lot of instructors while speaking in English during the lectures they occasionally insert some Arabic words. Moreover, another idea can be implemented such as use of speaker dependent speech recognition software which works by learning the unique characteristics of a single person's voice. New users must first train the software by recording certain speech with their voice and based on that speech the computer can analyze how the person talks. The aim of implementing this idea is reaching the optimal speech recognition accuracy. Images recognition can also be used to recognize the figures that exist in the lecture slides. This will increase the accuracy of the synchronization of the speech with the lecture slides. Additionally, based on the feedbacks collected from the experiments done. There are some extra features that can be added to AutoNotes such as summarizing notes or highlighting important notes.

## References


[1] Sanjanaashree P, Anand Kumar M and Soman K.P, "Language learning for visual and auditory learners using scratch toolkit," 2014 International Conference on Computer Communication and Informatics, Coimbatore, 2014, pp. 1-5.

[2] Bain, Keith & Basson, Sara & Wald, Mike. (2002). Speech recognition in university classrooms. 192-196. 10.1145/638249.638284.

[3] Stinson, Michael & Elliot, Lisa & Kelly, Ronald & Liu, Yufang. (2008). Deaf and Hard-of-Hearing Students' Memory of Lectures With Speech-to-Text and Interpreting/Note Taking Services. Journal of Special Education - J SPEC EDUC. 43. 52-64. 10.1177/0022466907313453.

[4] Ranchal, Rohit & Doughty, Teresa & Guo, Yiren & Bain, Keith & Martin, Heather & Robinson, Joseph & Duerstock, Bradley. (2013). Using Speech Recognition for Real-Time Captioning and Lecture Transcription in the Classroom. Learning Technologies, IEEE Transactions on. 6. 299-311. 10.1109/TLT.2013.21.

[5] Jones, Gareth & Edens, Richard. (2002). Automated Alignment and Annotation of Audio-Visual Presentations. 2458. 276-291. 10.1007/3-540-45747-X21.

[6] R. Swaminathan, M. E. Thompson, S. Fong, A. Efrat, A. Amir and K. Barnard, "Improving and Aligning Speech with Presentation Slides," 2010 20th International Conference on Pattern Recognition, Istanbul, 2010, pp. 3280-3283.

[7] Yu Chen and Wei Jyh Heng, "Automatic synchronization of speech transcript and slides in presentation," Proceedings of the 2003 International Symposium on Circuits and Systems, 2003. ISCAS '03., Bangkok, 2003, pp. II-II.

[8] IBM, "IBM ViaScribe," http://www-03.ibm.com/able/ accessibility_services/ViaScribe-accessible.pdf, 2011.

[9] https://firebase.google.com/docs/storage.

[10] Pearce, J.M., Ainley, M., Howard, S.: The ebb and flow of online learning. Comput. Hum. Behav. 21(5), 745–771 (2005).

[11] https://humansystems.arc.nasa.gov/groups/tlx/.

[12] https://usabilitygeek.com/how-to-use-the-system-usability-scale-sus-to-evaluate-the-usability-of-your-website/.

[13] https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c.

[14] https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall.

[15] https://www.periscopedata.com/blog/precision-recall-and-roc-curves-for-pregnancy-tests.

## APPENDIX

### A. Learning Gain Test

1. What is the most popular Cryptocurrency?
(a) Ethereum (b) Bitcoin

2. The main feature of Ethereum over Bitcoin is:
(a) Public Transactions
(b) Consensus
(c) Smart Contracts
(d) Mining

3. Smart Contracts Content is:
(a) Public (b) Private

4. The most challenging part in the project was:
(a) Using Smart Contracts
(b) Implementing Ring Signatures
(c) Achieving Consensus

5. To provide the insurance service users some privacy features:
(a) Smart Contracts were used
(b) Ethereum was chosen
(c) Ring Signatures were implemented

6. What is Blockchains Network model?
(a) Centralized Network (b) Decentralized Network

7. A blockchain transaction is considered valid:
(a) Once users reach consensus on the data
(b) The transaction is received correctly
(c) The transaction is broadcasted to the network

8. Merging IoT with blockchains introduces new:
(a) Energy Applications
(b) Administrative Applications
(c) Financial Applications

9. All of the following is a feature of the insurance application except:
(a) Anonymous
(b) Decentralized
(c) Adaptive
(d) Automated

10. Calculating the secret signature is a role of:
(a) User Wallet
(b) IoT Node
(c) Smart Contract