

Bao Nguyen

# Assisting Language Learning with Appla Application

Helsinki Metropolia University of Applied Sciences

Bachelor of Engineering

Information Technology

Thesis

25 April 2017

Author(s) Title	Bao Nguyen Assisting Language Learning with Appla Application
Number of Pages Date	45 pages 25 April 2017
Degree	Bachelor of Engineering
Degree Programme	Information Technology
Specialisation option	Software Engineering
Instructor(s)	Olli Alm, Senior Lecturer
<p>Appla is an application that utilizes speech recognition technologies to explore new approaches to language learning. The project aimed to build a mobile application that will assist the learning process of full-time working employees, who have little time to go to language courses, while preserving the context of a professional environment.</p> <p>The project used cutting-edged speech recognition services offered from Google to investigate the case of transcribing real-life conversations in real-time. As a result, two persons could have a talk with each other, while Appla records and simultaneously displays the transcripts of such a conversation. These texts could be reviewed, visualised and analysed to help the purposes of language learning.</p> <p>As a result, Appla project delivered successfully a working prototype to test and evaluate the original idea of assisting language learning using mobile learning. The project proposed a versatile solution that can further be used as a foundation for future research and development.</p>	
Keywords	Speech recognition, mobile-assisted language learning, nodejs, android, websocket, angularjs, mongodb, google cloud speech api

## Contents

1	Introduction	1
2	Mobile Assisted Language Learning (MALL)	2
2.1	Overview of MALL	2
2.2	Trends and Current Solutions to Language Learning	5
2.2.1	Commonly Investigated Topics in MALL Studies	5
2.2.2	Current Solutions to Language Learning	8
2.3	Appla's Proposal	12
2.4	Speech Recognition	14
3	Appla Overview	17
3.1	Overview of the Appla Application	17
3.2	User Cases	21
3.2.1	Recognizing Speeches	21
3.2.2	Authenticating User Account	25
3.2.3	Managing Rooms	28
3.3	Architecture Design and Diagram	30
3.4	Database Schema	33
3.4.1	Android Models	33
3.4.2	Models and Mongoose Schemas	34
4	Evaluation and Future Development	38
4.1	Results	38
4.1.1	Evaluation	38
4.1.2	Limitations	39
4.2	Future Development	40
5	Conclusion	42
	References	44

## 1 Introduction

The world has always been constantly on the move. In 2015, there was a drastic increase of 41 percent in the number of international immigrants in EU that reached 244 million, compared to year 2000 [12, 1]. According to the Statistics of the Finnish Immigration Service, Finland issued a total of more than 20.000 first residence permits in 2015 [1, 1]. A majority of immigrants coming to Finland is either for working or studying. With such a diversity in the job-hunting market, it has proven to be difficult to integrate successfully without overcoming a significant challenge: the language barrier. This is a huge problem that will eventually lead to a situation that the society cannot fully profit from the potential workforce and vice versa, when immigrants have to work at jobs they are overqualified for.

Language learning requires practice and training. The formal and traditional way of learning a foreign language is to go to offline courses offered by language schools. In addition, language learning also involves other activities like speaking to natives, watching movies, playing language-related games to help the students to improve one's skills. Among those activities, mobile-assisted language learning applications also have been increasingly developed to help users in the process (of learning a foreign language). These applications usually have pre-defined content and sometimes they are not related to students' interests and background.

It would be good to have an application that would assist and help learners to improve the language skills right in their professional working environment. While working, foreign employees often need to communicate with colleagues and customers in their non-native language. In fact, these verbal tasks can be quite challenging for those who are still learning the language. Meanwhile, the employees can also use these conversations as study materials to learn and reflect on their language skills. This is what Appla is trying to achieve.

The goal of the project is to investigate a new approach using available technologies to help learners in the process of language learning. I was assigned to develop a mobile application using speech recognition technology to aid the users with communication in the workplace and improve their language skills while working. I also built a web server that is responsible for handling data storage.

The technological point of view of the project was provided by me in collaboration with Senior Lecturer Olli Alm and Principal Lecturer Harri Airaksinen from Metropolia UAS. The team was interdisciplinary connecting University of Helsinki, Haaga-Helia UAS and Metropolia UAS, and is led by Dr. Johanna Komppa from the University of Helsinki.

## **2 Mobile Assisted Language Learning (MALL)**

### **2.1 Overview of MALL**

As an overview, this section will be explaining the definition of Mobile Assisted Language Learning (MALL) applications as well as discussing the advantages and disadvantages of such applications.

#### **Definition of MALL**

According to Stephen Krashen (1982), in his *Principles and Practice in Second Language Acquisition*, the acquisition of a language is a natural process, while learning a language is a conscious one [13]. Learning involves building up the knowledge system or architecture which over time and through practice becomes automatically accessible in receiving and generating. Vygotsky (1978), from a social cultural perspective, emphasized that language learning is based on social interaction with more proficient others [14, 35]. According to Scarino and Liddicoat (2009), the 'metacognition' characteristic, or awareness of how we learn, is essential to language learning. Students should have a good understanding of how they learn a language. They should constantly reflect on their study and develop self-awareness of themselves as learners. There is a strong association between learning and identity: students need to negotiate constantly who they are, and how they can be/ should be/ would like to be in the language and culture they are learning. [15, 31.]

The last few decades have witnessed drastic advances in technologies, which in turn opened new possibilities and alternative approaches to language learning that was before largely dependent on going to language classes. The solutions covered computer-assisted and e-learning, both in formal and informal fashion. Since 2000, mobile technologies have spread globally and resulted in a new evolution of technology-enhanced learning. With many innovations and improvements in the field such as pro-

cessing power, storage capacity, high-speed wireless connectivity, Bluetooth and 3G, mobile technologies and not limited to their core features (that were communication and entertainment) but have also showed other potentials in other fields. [2, 197-198.]

Mobile phones have grown into an unseparated and natural part in our daily lives. Children nowadays have already had access to multiple platforms of mobile technologies and become comfortable around such technologies from very early stages of life. The nature of how our brain works has been changing in many interesting ways, now that we carry very sophisticated devices on our hands that can handle a stream of information that we never experienced before in history. Such accelerating speed of adopting mobile technologies is currently changing the perspectives of learning in general. Mobile technologies can be referred to the next generation of e-learning. [3, 309.] According to a paper on research trends in Mobile Assisted Language Learning (MALL) from 2000 to 2012, the number of research publications regarding MALL for language teaching increased rapidly in 2008 and reached its peak in 2012. This confirms that interests have been spreading in this area.

Mobile learning (m-learning) as a term, in its essence, refers to teaching and learning supported by mobile technologies such as smartphones, media players and tablet computers, which are usually available anytime and anywhere. Possessing the characteristic such as rapid increasing of processing power and storage capacity, mobile technologies can be spontaneous, informal, contextual, portable ubiquitous, pervasive, and personal. As a result, these mobile devices gradually turn to be useful tools to help achieve educational goals. This has opened many doors for those who would not otherwise have the chances to participate in education. [2, 198.]

In general, Mobile Assisted Language Learning, which is a specialization within m-learning, has been noted for its differences to computer-assisted language learning because of the more personalized approach, coming from personal and portable smartphones and tablets. Their portability enables new ways of learning that maintain continuity as well as providing spontaneity of access and contextual interaction. [2, 198.] Owning such devices gives the learner the ability to control his or her own learning process and eventually take an active role based on their own objectives [3, 310].

## Advantages and Disadvantages of MALL

The main characteristics of mobile devices, according to Tayebeh Mosavi Miangah and Amin Nezarat, comprises two factors: portability and connectivity. Connectivity means that the mobile system designed for language learning must have the capability to be connected and communicated to access to the virtual learning material. [3, 310.] On the other hand, as for the character of portability, the learner should be able to take the learning material provided along with the mobile devices if he or she needed to. Klopfer and his colleagues have also defined some other properties of mobile devices such as social interactivity, context sensitivity and individuality [4].

Because of the portability of smartphones, learners are now given the ability to take their studies anywhere they could - on the bus, outside or at their part-time job, without being restricted to a physical classroom location or their personal computer at home (e-learning) [3, 310]. Such advantages of mobility eventually lead to many informal ways of language learning that do not involve pre-created materials from teachers and any other formal structured learning in classes or courses. While still important for students for building up a solid foundation and exploring in-depth knowledge, the formal structured materials given by teachers or textbooks are sometimes not relevant to the learners' interests.

However, if the students could contextually create their own materials, for example a personal dictionary based on their interests or their work environment, that would encourage the language learners to take a more active role in study. Furthermore, the cost of both time and money resources in mobile learning is considerably low – the production cost of mobile devices is much more affordable than before and switching on a device right on one's hand to access study materials is faster than going to class. As a result, this brings many more opportunities to students who cannot afford formal education and adults who cannot commit enough time for language studies because of work and family. Undoubtedly, mobile learning strongly emphasizes the role of learners rather than teachers - but without guidance and help from language professionals, MALL still has disadvantages compared to the traditional approach of language learning (to learn languages in classes).

According to Mianga and Nezarat (2012), many disadvantages of MALL lie on the limitation of the devices. For example, the small screens on smartphones proved to be occasionally difficult for students to read, especially for the older adults. Data connec-

tion is expensive and limited in some areas that can cause disruption in learning. Also, battery life is a problem as extensive usage of the devices is needed for such educational purposes. These problems are specifically mentioned in Stockwell's experiment where students had the tendency to switch to computers to finish the assigned tasks because it took too long to complete them on mobile devices. Consequently, the most common tasks that mobile learning excels in are flashcards and quizzes, which do not require a steep learning curve nor resources to get started. [5.] Such apps are beneficial in low-level thinking such as memorizing vocabulary; however, they still show weakness for other areas of higher level of thinking, for example: comparing/contrasting, summarizing or evaluating.

Furthermore, performance tracking in a big system that consists of multiple students learning in parallel can be an issue for teachers. However, as technologies develop, research and development around the globe are making important steps in the advancement of mobile technologies dedicated for language learning.

## **2.2 Trends and Current Solutions to Language Learning**

To better understand the reasons behind the Applia application, this section will be discussing about the most popular topics among MALL studies during 2000-2012 period and the current mobile solutions to learning Finnish language. Based on these studies, the proposal of Applia will be introduced.

### **2.2.1 Commonly Investigated Topics in MALL Studies**

The mobile applications designed for language learning have been going under significant changes over the years as the mobile technologies developed. According to Duman, Orhon and Gedik in their research paper about the trends in mobile assisted language learning from 2000 to 2012, there were no published studies during 2000-2003 period that met the criteria of the research data collection.

According to Godwin-Jones (2011) in his study abroad program, the useful features of smartphones used in language learning were: picture taking, text messaging and dual-language dictionary. However, the intended task was to write travel diaries - which later found to be problematic because of the T9 keyboard that was slow and error-prone to



write longer phrases efficiently. This limited input issue was very common at the time. [8, 2.]

In addition, Chinnery (2006) made a survey that turned into a report on projects which utilized mobile phones for vocabulary practice, quiz delivery, live tutoring and email lesson content delivery. He also featured file sharing, video playback and stylus text entry on PDAs which was popular at the time. Nonetheless, many projects faced many technical issues that limits the advancing of MALL: small and low-resolution screens caused troubles in text-reading or displaying pictures, audio quality (both input and output), error-prone text entry, storage and processor capacity. [9, 9-10.] On top of everything, Internet connectivity was slow and limited-access which poorly-designed web pages which eventually made communications troublesome - this was an essential factor that generally slowed down MALL progress. According to the study, there were only nine studies over the period of 2004-2007 [2, 201].

As a game changer in the smartphone market, the Apple iPhone in 2007 made a huge impact on many different development fields and eventually led the language learning researchers to re-investigate their possibilities. As a result, a notable increase of studies in 2008 (n=13) confirmed that the interests in the fields of MALL was increasing quickly [2, 201]. These new technologies - e.g. larger screens, touch-input, decent web navigation and fast Internet connectivity - have solved the issues arising from the earlier years in MALL and therefore enabled rapid development in the field.

The cost to manufacture smartphones were decreasing gradually and becoming much more affordable than before, which eventually enabled those who before cannot afford formal education can now own a smartphone and have unlimited access to language study materials that are available on the Internet. In 2012, the research trend peaked at 14 studies, convincing that learning new languages with smartphones is a new medium [2, 201]. There were a variety of topics investigated during the period of 2000 and 2012, as illustrated in figure 1 below.

Topic	2004	2005	2006	2007	2008	2009	2010	2011	2012	Total
Vocabulary	2	1	–	2	7	–	8	4	4	28
Grammar	–	–	–	–	1	–	–	–	–	1
Listening	–	–	–	–	2	1	–	3	2	8
Speaking/Pronunciation	–	–	–	–	2	2	1	1	–	6
Reading	1	–	–	1	1	–	1	1	–	5
Writing	–	–	–	–	–	–	–	–	1	1
Integrated skills	–	–	–	–	1	1	1	1	–	4
Dictionary use	–	–	–	1	1	1	–	1	–	4
Assessment-evaluation	–	–	–	–	–	–	–	1	1	2
Multimedia use/Design	–	–	–	–	1	1	2	2	–	6
Instructional design	–	–	–	–	–	–	–	1	1	2
Identity/Sense of community	–	–	–	–	1	1	1	–	–	3
Usability	–	1	–	3	5	1	4	1	1	16
Potential uses/Drawbacks	–	–	1	–	3	1	2	1	1	9
Interaction/Collaboration	1	–	–	–	–	2	1	–	–	4
Perception/Attitude	–	–	–	–	1	1	1	3	5	11
Academic achievement	–	–	–	–	–	1	–	–	1	2
<b>Total</b>	<b>4</b>	<b>2</b>	<b>1</b>	<b>2</b>	<b>26</b>	<b>13</b>	<b>22</b>	<b>20</b>	<b>16</b>	

Figure 1. Summary of commonly studied topics in MALL researches from 2004 to 2012 (Coped from Duman, Orhon and Gedik (2012) [2, 202])

It is important to note that figure 1 only shows the topics that were investigated in MALL studies, not the actual MALL application in practice. For example, “vocabulary” shows a total of 28 during the whole period meaning there are 28 studies that partially or fully investigated the vocabulary learning aspect of MALL apps. Usability shows 16 studies that investigated the usability of mobile applications for language learning. These studies were not actual MALL applications that teach grammar, vocabulary or usability.

As seen in figure 1, vocabulary remained the most popular topic to be investigated in MALL studies throughout the period. Understandably, vocabulary teaching in many forms (e.g. flashcards, dictionary) can be spontaneous, instant and utilizes the main advantages of MALL because smartphones are designed to be portable and on-the-go. Total 28 MALL studies that covered the topic of vocabulary learning indeed confirm a strength of MALL, which is a useful tool to assist the learners to gain knowledge in terminologies and words. Mobile system usability is also a common topic with total 16 studies. A sudden rise of 3 topics in 2007 shown that researchers have taken interest in this area quite early – just after the introduction Apple iPhone. Without a doubt, the studies of usability are going to be beneficial for other language learning applications to have a solid base foundation to be built on.

Writing and grammar skills are two of the least common topics investigated in MALL studies. Certainly, the limitations in technologies, e.g. slow input and small screens, as well as characters of mobile learning are not suitable of writing tasks – it is laborious to write long paragraphs on a mobile phone. In addition, assessment-evaluation was not investigated until 2011 and still unpopular in researches with only one research each year (that met the criteria of the paper) until 2012. It seemed that a feedback system can be problematic and difficult to be conducted in other to benefit language learning.

Taking a closer look at Figure 1, it is clear to point out that there were many language skills in MALL topics gradually became popular since 2008, as mobile apps were increasing rapidly in quantity with better multimedia features and capacity to catch up with the smartphone race between large technology companies. With such rich multimedia features, for example: enhanced audio and video quality, fast data connection, large storage and processor capacity, many language skills such as listening, speaking (pronunciation), multimedia use, and learners' perception and attitudes became popular topics for investigation in MALL studies. Technologies like voice/speech recognition began to mature and eventually became available to be adopted to smartphones, which in turn led researchers to investigate the opportunities to enhance learners' speaking and pronunciation skills. [2,202.]

Above is a brief introduction about the trends and difficulties while developing applications to assist language learning in general. As technologies evolve rapidly, it is equally interesting and challenging to predict what comes next that will shape the world that we know of. The next section will inspect the modern technologies that are available right now for language learning, particularly regarding Finnish language, to analyze their current issues.

### **2.2.2 Current Solutions to Language Learning**

The need to learn and understand Finnish has been increasing quickly as a way to integrate into everyday life as well as to land a job in Finland. Utilising the experience and expertise in the Finnish language of its project members, Appla focused on analyzing the current Finnish learning solutions to understand clearly their challenges and issues.

There are currently many available options and solutions to learn Finnish, for example: evening courses for those who works or study full-time, summer courses for students or intensive courses that held 4-5 days a week and so on. Many universities in Finland not only hold these courses for their international personnel and students but also open to those who do not attend at their universities – making it possible for everyone to access their language learning materials (with a fee). The courses are available at different levels and schedule so the learners are given the freedom to choose what works for them. In addition, studying in a language course at summer universities is also a popular choice among international students who want to advance their language skills when they have more free time. Summer universities are in tight contract and permanent collaboration with Finnish universities and their courses are usually short-term, not degree programmes [10].

While having its advantages in helping language learners to gain knowledge and expertise in depth with professional guidance, the courses usually cost money and time resources and require strong commitment from the students to progress further. Eventually, the courses are increasingly available via the Internet which leads to prices are lowered due to eliminating the facility cost. Many mobile applications for Finnish language learning are also introduced to assist the students to keep their knowledge sharp even on their busy hours.

Among them, WordDive and Finnish by Nemo are two popular options for mobile learning that teach vocabulary and grammar with a quite comprehensive list of learning topics for language learners. WordDive provides a mobile application and a website for learning. Finnish by Nemo has a feature that allow users to record their voices to imitate the native speaker's voice, hence practicing their pronunciation skills. These functionalities represent some similarities with Appla, hence chosen to be investigated in this study.

WordDive, developed by a growing Finnish start-up, is a choice that is mentioned on many media channels and Finnish courses websites which offer many study options depending on the students' needs. Their most popular course – “Comprehensive Finnish” – costs approximately 10 euros for a monthly subscription, which is considerably affordable compared to offline courses. At the first glance, the web application offered by WordDive is delivered with comprehensive lessons and good audio quality. The web app interface is shown in figure 2.

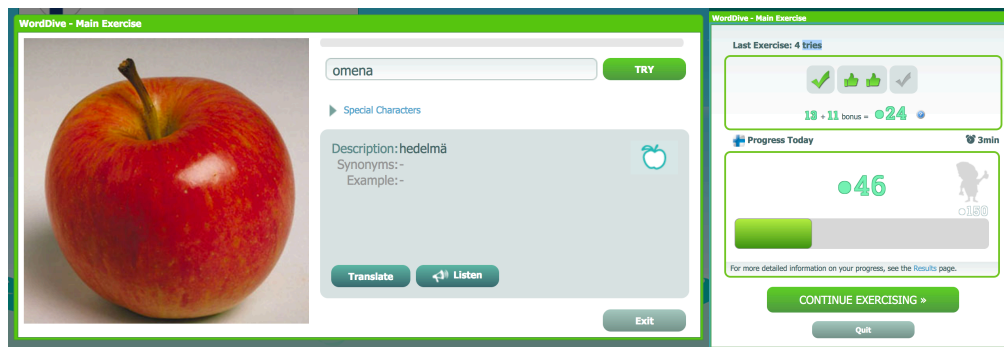


Figure 2. Exercises and progress tracking of WordDive (Screens captured from WordDive)

The first main exercises tested were quite straight forward. The user is shown an image and prompted to type the correct word of the object that is illustrated in the image. There is a Listen button for the user to listen to the pre-recorded audio snippet of the word he or she is supposed to fill in the input form. Correct answers will yield points which will be then added to the progress board for the user to track his or her own progress in the application. Not limited at web applications, WordDive also offers mobile apps that are available on Android and iOS platforms. The Android version is shown in figure 3.



Figure 3. WordDive exercise and progress tracking on Android (Screens captured from mobile app)

As seen from figure 3, the idea behind exercises and progress tracking remains constant through web and mobile applications. Knowing the strengths of MALL in developing vocabulary skills, as mentioned in the section about MALL trends, WordDive focus-

es on building vocabulary and pronunciation knowledge for the users through many built-in exercises and audio snippets. The short games focusing on various topics can be played instantaneously and spontaneously on the train or while waiting for the bus. The user can choose to practice grammar by writing the correct tenses according to the audio speech and English hints. It is also intended to help improve the listening skills of the user by pronouncing with a native Finnish accent. Indeed, WordDive is an interesting choice among current mobile applications that help sharpen different language skills.

Similarly, Finnish by Nemo is also available for download on Google Play and App Store. It is offered by a company that has many other educational applications that teach different languages – Finnish by Nemo is a modified version to adapt to Finnish language. In parts, the application also pursues the same theme as WordDive – focusing on phrases and pre-arranged exercises to work on speaking or listening skills. Figure 4 contains the screenshots of the application on Android.

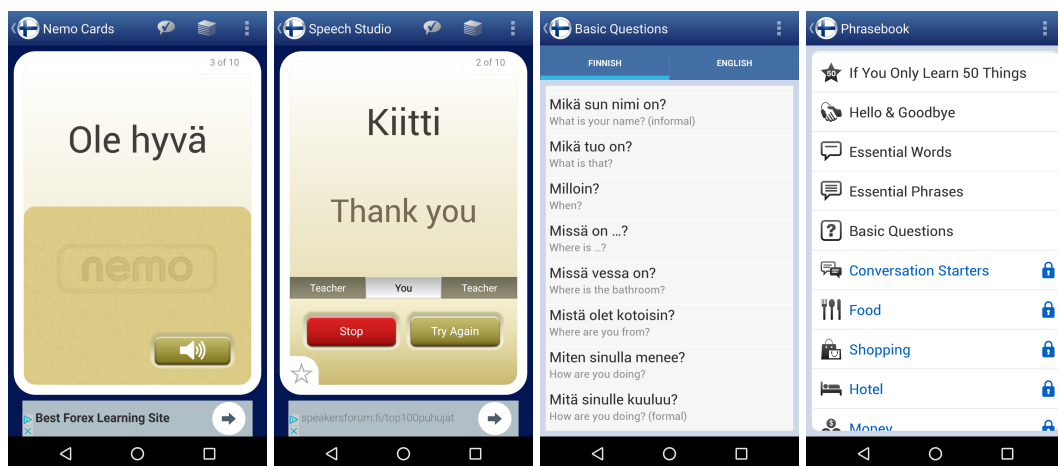


Figure 4. Finnish by Nemo flashcards, ‘speech studio’ and topic phrases (Screens captured from mobile app)

Feature wise, Finnish by Nemo takes a flash-cards approach and splits the vocabulary into topics to help the user progress. Notably, it utilises the recording feature of mobile phones in so-called “Speech Studio” that will help you compare your pronunciation with teacher’s. This approach is different from WordDive since it manages to enable one additional side of interaction and lets the user create his or her own material (their own voice).

However, the app only stops at recording and playback then leaves the heavy part of the feedback to the user - he or she will have to judge their own pronunciation by comparing their own voice with that of the teachers'. One would argue that one's own judgement is subjective and would not promise 100% accuracy or reliability to be one complete trusted source of feedback. Of course, this traditional approach of learning a new language that involves recording and hearing one's voice has proven itself with time, but it gets better: technologies have evolved much further with speech recognition so that it can now supports an extra source of judgement: artificial intelligence. We shall get back to that in later sections.

Despite utilizing their advantages as mobile applications, the current available mobile apps offered on Google Play are still lacking something that really differentiates them from learning from textbooks. The problem with Finnish by Nemo and WordDive is that those applications are currently built from exercises and lessons that is generic for everyone who wants to learn. This would lead to the situation that students often need to study topics and skills that are unrelated to their interest and level, which brings us to Anki, an 'intelligent and friendly' flashcard application.

Anki is a versatile application that assists users to memorize a various range of topics, from language learning to guitar chords. Anki uses a flashcard system, aimed at encouraging language learners to make connections, repeat and interpret the cards. During the user's learning experience, Anki figures out if and how well one memorized each flashcard and it will not list again those the user already knew. This is an example of how an application can change and adapt to the user's skills based on the usage of the user.

### **2.3 Applä's Proposal**

Employees working in Finnish companies, for instance cleaners, restaurant waitresses or developers, often already have a basic understanding of the Finnish language. For developing their language skills, it is predicted that full-time employees need something, e.g. a mobile application, that can be adapted to their working life and eventually optimize their process of learning. While helping them learn the language, the application could be used in various ways to integrate the employees into the Finnish working life. Also, foreign employees working in their non-native language environment could sometimes find it challenging to communicate with their colleagues and customers. In



complicated situations, there could be misunderstandings if either party mishears or does not fully understand the other, leading to inefficiency in communication.

A MALL application is intended to help the learners to progress in language learning using their mobile phones, along with other means of study like going to courses or learning from textbooks. To help solve the possibilities of miscommunication in a working environment, Applä is an application that helps users to record their conversations and display the transcript simultaneously on their mobile phone. To understand better about Applä's concept, let us through some scenarios in which Applä can be useful.

During phone calls, while talking to the other party, the user can see the texts regarding their talk displayed on the phone in real time. Supervisors and managers can give instructions to their employees on the phone – the transcripts would be saved in case the employees want to refer to the verbal instructions. In presentations or meetings, the user can get help from Applä by reading the transcripts of the discussion simultaneously displayed on the phone's screen. These texts could later be reviewed together with a teacher or a colleague to reflect and improve the users' skills.

Choosing a solution that is powered by speech recognition, Applä wants to help the users to build up their knowledge based on real-life situations that they encounter daily at work. Speech recognition technology is decided to be the core feature behind the app because of its versatility and scalability. Applä can be applied in different settings in professional situations like meetings, phone calls, presentations, text interpretation from interview recordings, etc. While working, the conversations between an employee and other native speakers can be invaluable materials for the student (employee) to look back and learn from.

The core functionality of Applä is recording speeches and transcribing them into texts. The learner will use their mobile application to record the conversation and simultaneously see texts on the mobile screen. In addition, other users (teacher, colleagues, supervisors) can also inspect the ongoing discussion via the website. Each discussion can have its own session (room), allowing multiple users to communicate at the same time. These are the main features that were implemented in Applä.

Applä targets to help the users to store these materials they generated by themselves that are relevant to each individual user. In future development, the available data can be crunched and analyzed for results that will aid the user's learning process. For in-



stance, the user can see his or her language knowledge growth by checking the variety of vocabulary he or she used in daily conversation.

As a result, it would be considerably easier for users to track the progress while being alert for valuable and instant feedback gathered from real world communication. The user is expected to produce the data, gather the results from analytics and from them continue to improve their skills. Applu helps users to view the conversations in text format, allowing them to improve both reading and listening skills, as well as building a better vocabulary foundation. It also motivates the users to correctly pronounce words because the speech recognition performance relies on the accuracy of the user's pronunciation. The results in texts would be poor if the user pronounces incorrectly.

Appla took an alternative direction in assisting Finnish language learning by utilizing user-generated content, in comparison to the traditional approach which embedded lecture content inside MALL applications. This is by no mean trying to replace the traditional way of Finnish teaching (by going to language courses and learning from textbooks). Opposite to that, it is an application that utilizes available technologies to assist the traditional learning by providing an additional medium for users, which can be used to continuously learn while spending time at work. Applu targets to be an optimal solution for those who work full-time and have trouble finding available slots for language classes in their packed schedule.

Before the application technical overview, it is important to have a general look of the current scene of the core technology of Applu, speech recognition. In the next section, the speech recognition technologies from different providers will be briefly introduced and discussed.

## **2.4 Speech Recognition**

It is an interesting time in 2017. Currently, a junior high school student can get access to widely available A.I. services to build his own chatbot or an AI that controls his lights in his bedroom by voice commands. Interesting time indeed. This advancement in technology has brought interesting perspectives to the research and development of MALL applications.

Powering the most important core feature of Appla, the speech recognition engine is the one critical component that will either make or break the application. The development of speech recognition has been evolving as quickly as ever, when major companies start to introduce their own products in this area. There are plenty of options, but to keep it short, we shall now go through the three major available options to discuss their advantages/disadvantages and reasons behind the choice of Google Cloud Speech API.

### **IBM Speech to Text**

IBM Speech to Text offers advanced speech recognition technology that employed machine intelligence to associate data about grammar and language arrangement with knowledge of the composition of the audio signal. The service constantly returns and updates the text as speech is streamed. The technology is available in many interfaces for example WebSocket, HTTP REST and asynchronous HTTP interface for different purposes. They also offer other various options like labelling speakers to differentiate the speakers in *one* input source, spotting keywords which is helpful in customer service case to route to certain help category or smart formatting of dates, times, phone numbers and currency.

It is very interesting to see how IBM develops in a few years' time. The one deal-breaker which led to the decision of Appla not pursuing IBM technology was that Finnish language was not currently, as of April 2017, in the list of supported languages. This may change soon, of course.

### **Google Cloud Speech API**

Coming from Google, the company that has access to large user information ranging from their user voices repeating "Ok Google" to their location and email conversation history. Having great advantages by owning YouTube that has considerable amount of audio inputs or Google Translate that constantly learns how the language is used in real life, Google introduced Google Cloud Speech API in June 2016 as a Beta version so developers can try out and contribute feedback.

Supporting over 80 languages, Google Cloud Speech API can be considered as one of the most powerful speech recognition engine in 2017. The engine employs the most advanced deep learning neural network algorithms to the user's audio for accurate

speech recognition. Google claimed that their API's accuracy shall improve over time as Google matures the internal speech recognition technology used by Google products.

By the time Appla started, the services offered by Google are unparalleled in quality. Few simple tests made in the Finnish language appeared to impress native speakers, given the ability to distinguish the differences between the spoken and written language. The longer phrases and rare nouns like Finnish company names are also recognized correctly with an impressive short period of time. Streaming in real-time to the Google server is easy with the detailed documentation for developers.

What Google lacks, compared to other speech recognition providers, is the ability to deliver speech intents and topic extraction to their service. In the attempt to match their competitors, api.ai was introduced with richer features like Natural Language Understandings and the easy process to integrate to different platforms.

### **Bing Speech API**

The Microsoft Cognitive Services provide multiple Service APIs that cover a long range of technologies, including their own speech recognition service: Bing Speech API. Being a major corporation with rich resources, Bing Speech API is a strong competitor in the field with speech-to-text, text-to-speech and especially speech intent recognition. So rather than only limited in speech recognition, the service also allows one to parse the tone of the speech for example approval, attention or prohibition.

Undeniably, Bing Speech API intends to provide a whole stack of services so developers can build their own machine which reacts to voice commands. They had been recently expanding supporting to many different languages: as of April 2017, the total number of languages was 28, including Finnish, which was not available in October 2016 when Appla started. No tests were made to check the accuracy of Finnish speech recognition of Bing Speech API.

In addition, the first few tests on the home page with the English language show promising advantages such as the addition of punctuation marks and intentions; which is missing from Google Cloud Speech API. However, streaming real-time audio for recognizing speeches only allows a maximum of 2 minutes long – a workaround might be needed to stream in longer time. On top of everything, there should be an in-depth

study of the Bing Speech API documentation in the future to consider the possibilities and benefits of it over Google Cloud Speech API.

### 3 Appla Overview

#### 3.1 Overview of the Appla Application

Appla is a mobile application that transcribes ongoing discussions into texts and displays them on mobile and web screens in real time. As a MALL application, Appla possesses the characteristics such as spontaneousness and instantaneousness so the user can quickly use it in different situations to help their purposes in language learning.

For example, while on a phone or Skype conversation, the user can use the application to aid their listening by having a transcript of the dialogue in real time. In a meeting, Appla can be used to record the discussions, deliver the texts accordingly and give users the context of the meeting. To keep the conversations private and clean, each session can have their own 'rooms' – only those inside the rooms can see the conversation. The texts are saved on the database so the user can have a review session with their teacher later to talk through the mistakes and difficulties that he or she encountered in daily conversations in a professional environment.

The users can turn on the Appla application on their mobiles to start recording and have the texts of the ongoing discussing displayed on their phone. During the conversation, the application will be displaying the texts on the screen for the user to follow. These texts will also be streamed to a web server that displays real-time information for others (e.g. teachers) to read. These contents, which is saved into a database, can later be visualized, enriched, explained and repeated for language learning purposes.

In Appla, there are three core user cases:

- The learner can record speech and see the transcript accordingly.
- The reviewer (teacher) can inspect ongoing conversations and see the transcript.
- The learner can create and join different rooms.

## AppLa User Interfaces

The screenshots of the application on Android are shown in figures 5, 6 and 7.

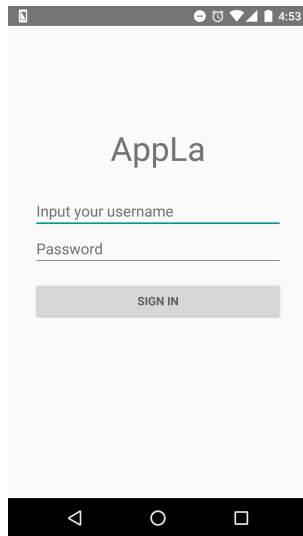


Figure 5. AppLa login portal on Android

Figure 5 shows the login portal of AppLa. Users need to authenticate themselves before using the application's features.

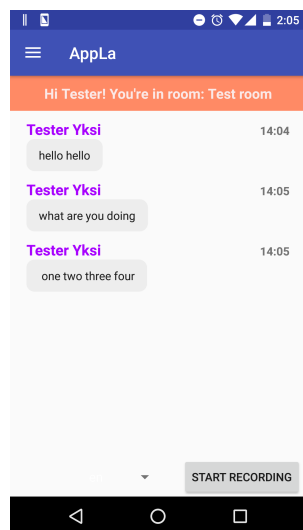


Figure 6. AppLa's conversation interface

The conversation recorded by AppLa is illustrated in figure 6.



Figure 7. AppLa's room list

Figures 5, 6 and 7 show the three main screens of the AppLa Android application: a login portal, conversations inside of a room and a room list. In the room when the speech recognition 'happens', there is a notice banner to prompt the name of the user and the room he/she is in, a drop-down list for switching languages, a button to start recording and of course, the conversation with user's full name and the message details. They are quite simple and straight forward that everyone can start to use right away without much help.

The Web client has color schemes similar to the Android application: blue and orange. The screenshots are shown in figures 8 to 11.

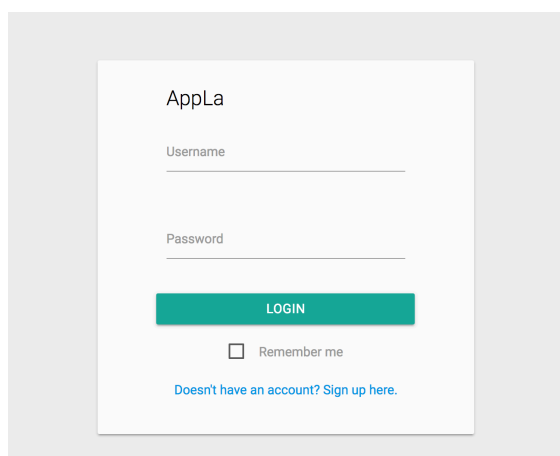


Figure 8. AppLa login registration portal

Figure 9. AppLa registration portal

The sign in portals in figure 8 have only a simple form that asks for the username and password. In figure 9, there is a sign-up form that allows the user to register for a new account. These steps are quite standard – minus the need to verify as a human or confirm the email.

Figure 10. AppLa Home Page

As depicted in figure 10, AppLa has three components: the navigation bar, the user's home room and the room list dashboard. The navigation bar has a link to get the latest Android apk installation file, so the user can download and install it quickly using their phone.

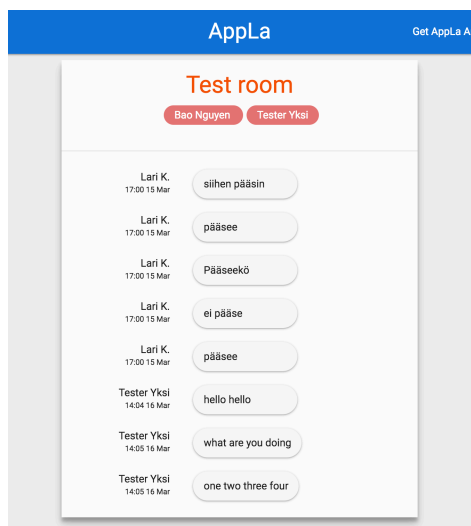


Figure 11. Inside a room that shows the conversation and active users

After clicking to a room on the room list, the user will then be immediately connected to a room and shown the room interface in figure 11. The room has two main components: on top is the room title with a list of active users currently connected to the room and below is the conversation board that displays the messages. The messages contain the information like timestamps, the sender's full names and the messages content.

The sequences and functional requirements of each user case and user interaction of AppLa will be investigated in detail in the next section.

## 3.2 User Cases

### 3.2.1 Recognizing Speeches

#### User Cases

The backbone of the application relies on the feature of recording the conversations from the mobile application, transcribing them into texts and listening to the audio if needed. The user cases concerning this core feature are shown in figure 12.



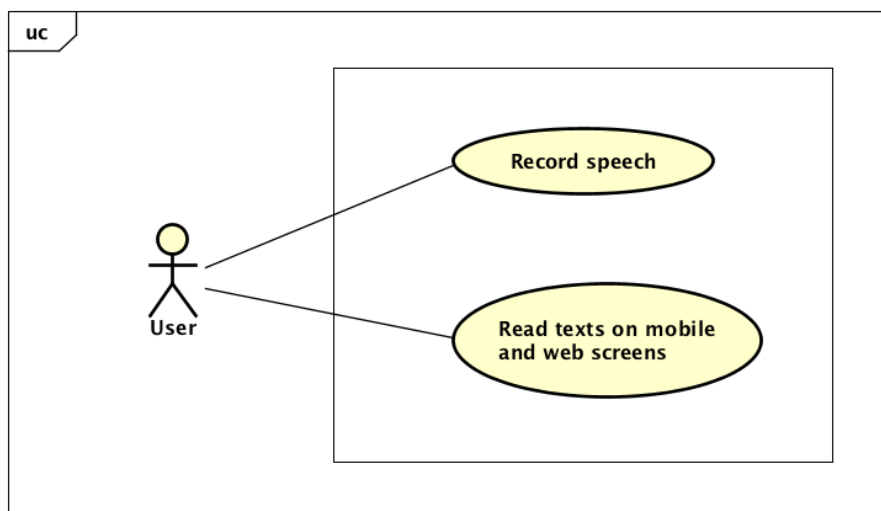


Figure 12. User cases regarding recording speeches in Appla

One of the most important functionalities of Appla is to record speech. As discussed in the previous section, Appla will assist the users with their language learning by bringing the discussions around them into something beneficial to their study. The user story is straight-forward: the user opens the application, presses the recording button then starts talking to the phone's microphone. The application should listen to the speeches and record everything to an audio file. The user will then see the transcripts of their speech appearing on the phone screen. If there are others inside the same room as the user, the texts will appear on all listening clients. This logic is related to how chatting in a room would operate, but instead of manually typing, the input is from your voice.

Appla will utilize speech-to-text technology that, in a simple way of explanation, makes real-time subtitles for users' conversations. Of course, it would be an ideal approach if the engine can separate and differentiate the sources of speech within the same stream of audio. However, as a limitation in current technology, the speakers in a room using Appla should be prepared for separate sources of input to achieve the best result.

Accessing the application from the web, the user will have a separate interface that displays all the rooms of Appla. It is not possible at the time of writing to record directly from the web, as we favour the possibility of a mobile application and it is not an urgent need for web users.

## Recognizing Speech Implementation

Speech recognition is the core feature that supports and builds up the foundation of Appla. All actions of recordings and transcribing texts happen first from the mobile application client side.

After signing in, the user will then be joined into a private room called “Home” that is only available to their own account. Each user will have his or her own private pre-created room for quick and instant recording, which is intended for users to try out the speech recognition feature without joining to any specific room. The data exchanging by the current user within this room will not be transferred to anywhere else. The managing of Rooms will be discussed in detail in the next section.

In detail, recording from mobile devices is achieved in a fragment that will be “pulling” (reading) the data through the physical hardware input (microphone). The audio bytes recorded via the Android phone will then be stream to Google Cloud Speech API using the gRPC API provided by Google. When the recording starts, a new thread will be opened that contains a `StreamingRecognizeClient` class which handles the bi-directional streaming of audio bytes and returning result transcript. Inside this class, requests will be made continuously while receiving responses at the same time.

Sending messages requires the usage of `WebSocket` to achieve the real-time aspect of the application. While streaming, the mobile will constantly be sending (emitting) message information to the server. The loop of receiving audio and emitting to server is depicted in Figure 13. The server will process the data by saving it and attaching the messages to the correct room. After that, the mobile application will receive a processed message sent from the server.

To distinguish between the messages from different sources, each message will be assigned an ID that will remain unchanged until a signal from the Speech API returns a result with a final status.

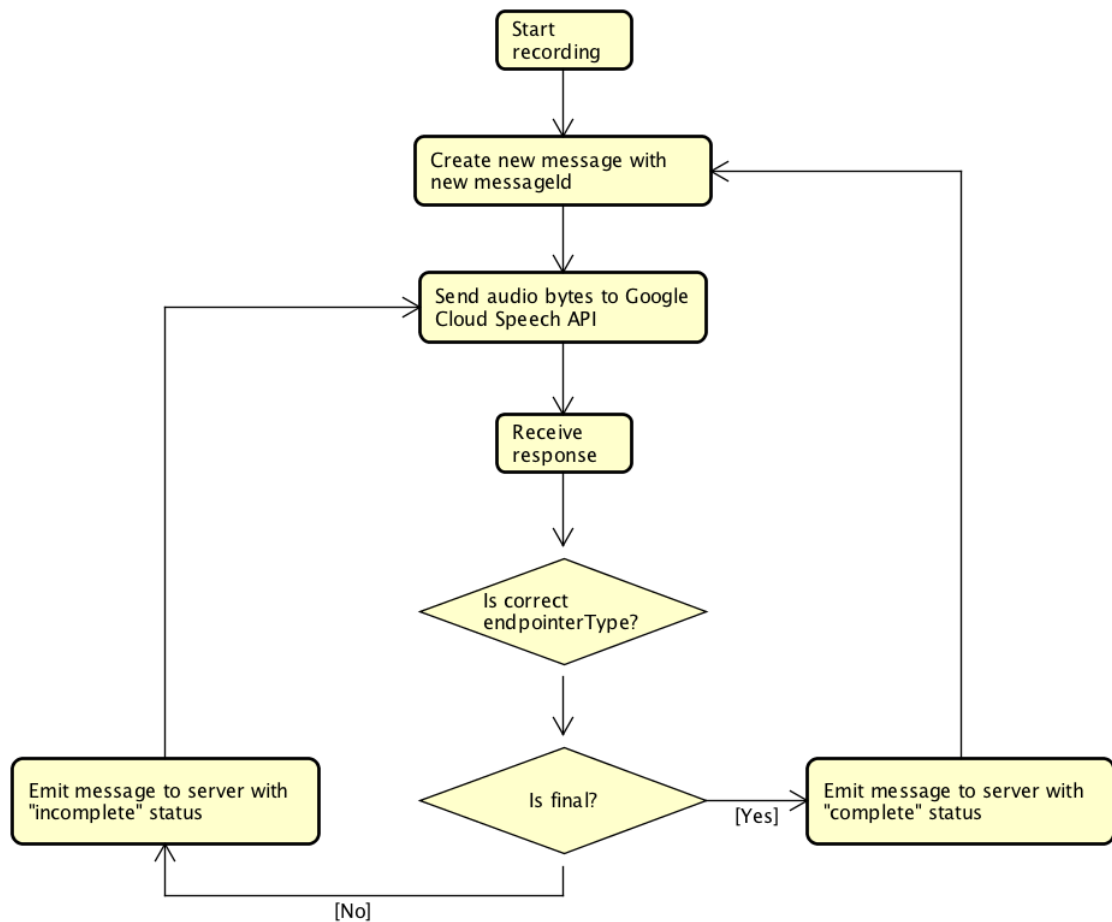


Figure 13. Flowchart illustrating the process of sending transcripts to server

When Appia receives a response that is final, as shown in figure 13, the application will emit a message with a “completed” status to the server, which means the message will be saved into the database and a new ID will be attach to the next message. This process helps to achieve the conversation with the “subtitling” effect as the speech is being recorded. This has proved to be a great advantage in user experience, as the user can see the texts coming as they speak with little delay.

### Limited Streaming Time

For a streaming connection with Google Cloud Speech API, it was explicitly stated in the official documentation that the connection is limited to only one minute of streaming. Not being able to record for a long period of time is a major setback for an application with ambitious goals which is to work in conversations or meetings.

As a workaround, in Appla there is an internal clock that begins to tick as soon as the user starts recording. In the thread of recording audio bytes, there is a flag that checks for a mark of about 55 seconds into recording. When it comes close to the 60 seconds cut, the old connection stream will be shutdown, following by an initialization of a new connection stream. As a result, Appla has been able to record for a long period of time without being limited to the API's restriction. After a period of testing, there was no noticeable errors concerning the gap between the shutdown and initialization of two connections.

### 3.2.2 Authenticating User Account

#### User Cases

As a standardized process, the user authentication of Appla has three basic cases that are sufficient for a small-scale application. The cases are demonstrated below in figure 14.

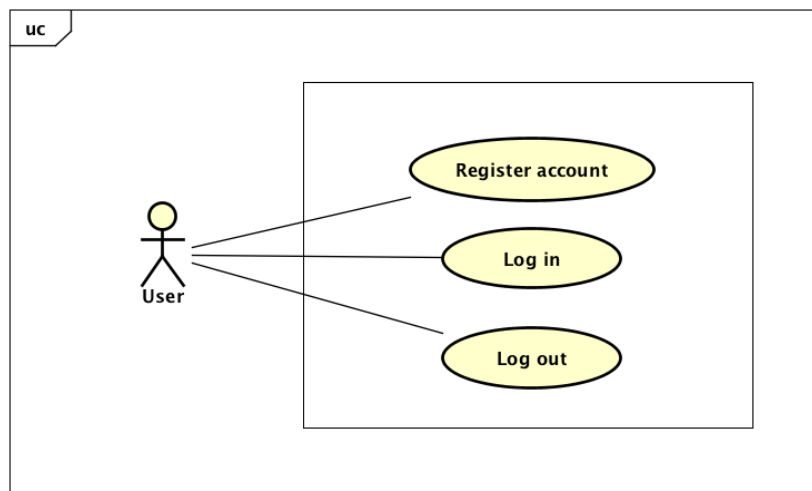


Figure 14. User authentication user cases

When accessing the dedicated address for the application ([discourse.metropolia.fi](https://discourse.metropolia.fi) at the time of writing), the user will be redirected to a login page, prompting for username and password. The application will display error messages in case of invalid authentication. As seen from figure 14, the user can log out and will not be remembered for the next session.

If not registered before, the user is able to register for a new account with personal information such as full name, username and password. The username is a unique key in database thus no duplication is allowed. During registration, the user will be prompted with messages concerning error or successful events in case of invalid inputs; this is a simple process of check-types like invalid first/last name and username/password. For the first prototype version, email confirmation and password reset is not yet a concern and shall be considered when developed in a bigger-scale production version.

### **Authentication Implementation**

The authentication process is handled entirely by Node.js back-end server. No major processing is being done on the client-side except for simple client-side check-types. On the client-side, a routing Angular component was built and customized depending on the “ui.router” module. This component handles all the URL routing functions when a user accesses the home page. The logic behind this routing is depicted in the flow chart in figure 15.

When the user first accesses the application from either a web or mobile application, Appli will immediately check for the availability of the access token in the local storage of each client. If available, the access token will then be sent to the server for validating. If token is not verified, the user will be prompted a message error and then required to input the correct username/password to sign in. Otherwise if a valid token is available in the local storage, the user will be ‘remembered’ and be signed in without any effort.

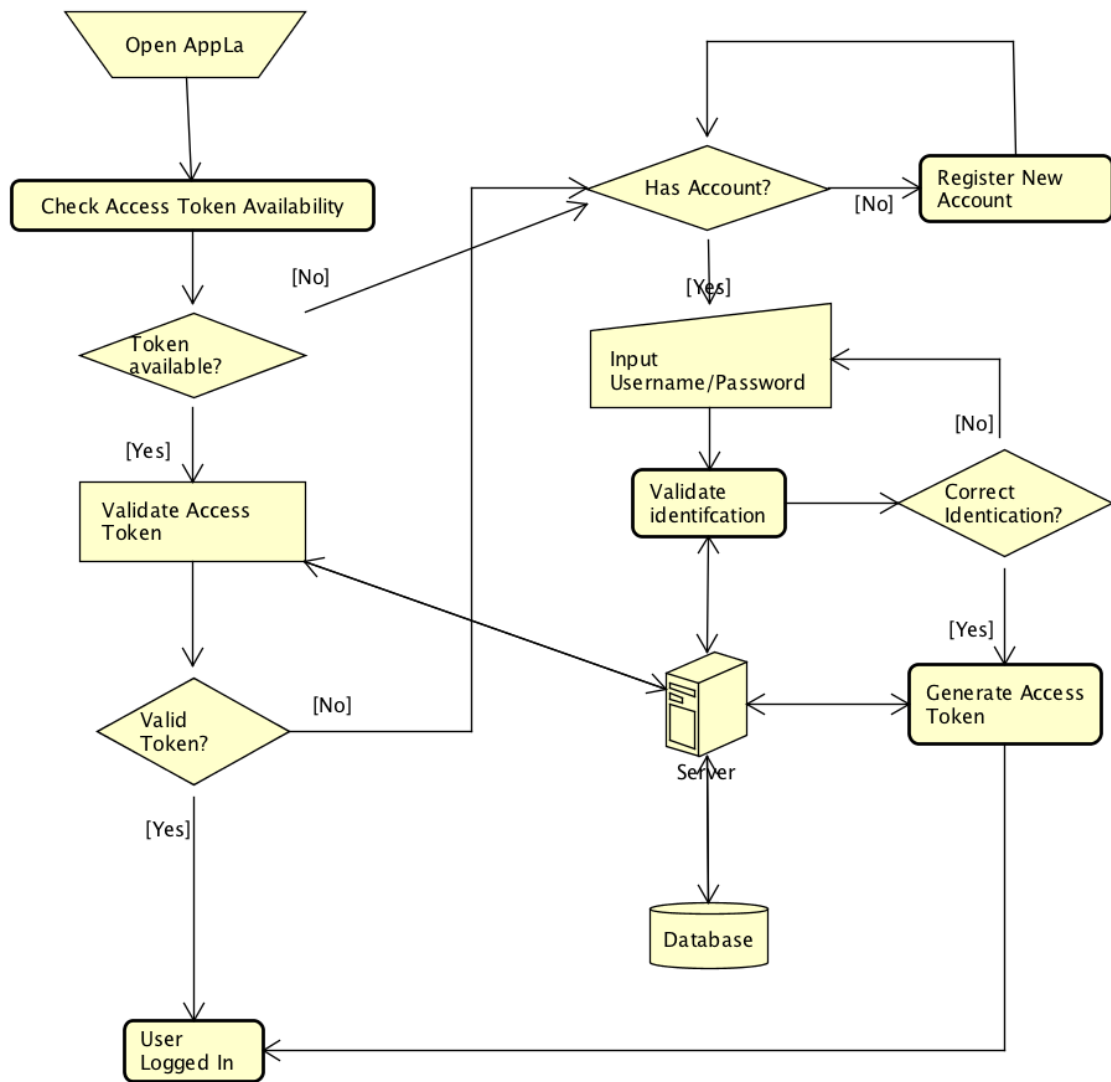


Figure 15. Flowchart showing the authentication process of AppLa

If the token is not available from the local storage or invalid, the manual input from user is required to continue. After the user's submission, Node.js server will be checking the username and its corresponding password. The password is compared with the registered password saved in database by bcrypt, a password hashing function. If any error occurs, for instance wrong password or unknown username, Node.js will immediately return a JSON containing a success status (false) and the error message to the requested front-end client. If successful, a new access token with the expiration length of 24 hours (customizable) and a success value of true will be attached to the result JSON.

Angular.js routing on the client side has been configured with three states, following the address of AppLa, below referred as *{address}*:

- Authenticated states (log-in required)
  - **Home:** *{address}/*
  - **Rooms:** *{address}/rooms/room-name*
  
- Unauthenticated state (no log-in required)
  - **Login portal:** *{address}/login/*

All connections to Home and Rooms route will need to present a valid access token or else they will be redirected to the Login portal. The same logic applies to the mobile application: when user opens Appla on Android, a SharedPreferences object will be checked for any available and valid access token. This token is saved across user session, which means the data will persist even if the application is killed.

### 3.2.3 Managing Rooms

#### User Cases

The ability to split connections into different channels for conversations is standard in communication applications. Users accessing Appla will be able to see all the available rooms and join the conversation anytime. The user cases are shown in figure 16.

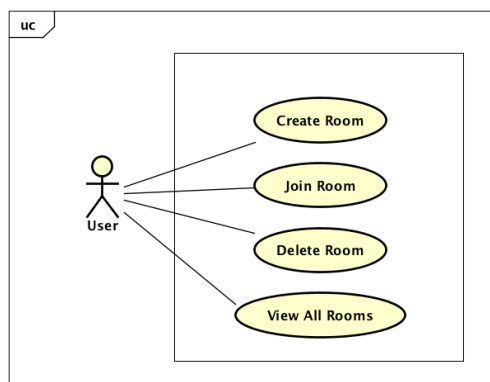


Figure 16. User cases of managing rooms in Appla

Managing rooms is a straight forward process. Figure 16 shows the four cases:

- **Create Room:** From the web client, the user will be able to create a room with title and brief description of the room. All other users will see the room appear as soon as they are created.
- **Join Room:** When joining a room, the user's full name will appear on the room's name list and only be removed after the user leaves. It does not matter if the user has multiple devices connected to the room, the name only shows once.
- **Delete Room:** A room can only be removed by the one who created it. Deleting rooms does not completely remove its information from the database.
- **View All Room:** A list of rooms will always be displayed on the Home page of Appla as well as in the mobile application's room section. The list is updated as soon as any changes happen, for instance creating or deleting room.

### Room Management Implementation

Room managing implementation relies heavily on the features offered by WebSocket: there are built-in options to use rooms and namespaces to split to different channels and customize them accordingly to each application's needs.

Currently there are two namespaces available in Appla:

- **Home:** This namespace will handle anything happens in the home page of Appla: data communication in users' Home room and managing rooms.
- **Rooms:** When the user joins a room, the address will be changed to this format: {address}/rooms/{room name}. WebSocket will recognize the change and immediate calls the Rooms namespace, which is defined as "/rooms", after disconnecting the Home namespace.

Once the connection between server and client is established, it is possible to stream data between two points in full-duplex mode. Basically, the server will have many pre-defined endpoints that will be listening to all other clients. If a client send (emit) a message with a name (not payload or content) that matches any endpoint, the server will immediately react accordingly. Inside each server point, there will be also a 'broadcast'



message that sends the processed information to all other listening clients. The processes happen instantly in real-time with no additional connection request, apart from the original established WebSocket connection.

### **Room Address**

If the user wants to access or join a room, he or she can click directly to the room on the room list view on the Home page. The direct address of the room is defined as: *{address}/rooms/{room name}* where room name is an auto-generated slug from the room's title given by the user when creating the room. This usage of room slug makes sure the address is user friendly and never duplicated from another room with a same name.

Above is the clarification of different Appla user cases and the business logic implemented from those cases. Next section will investigate the architecture of the Appla application to obtain a general look at the system.

### **3.3 Architecture Design and Diagram**

The architecture of Appla system comprises many components, each holds its dedicated features, as illustrated in figure 17 below. The diagram shows the most fundamental core feature that is the process of speech recognition.

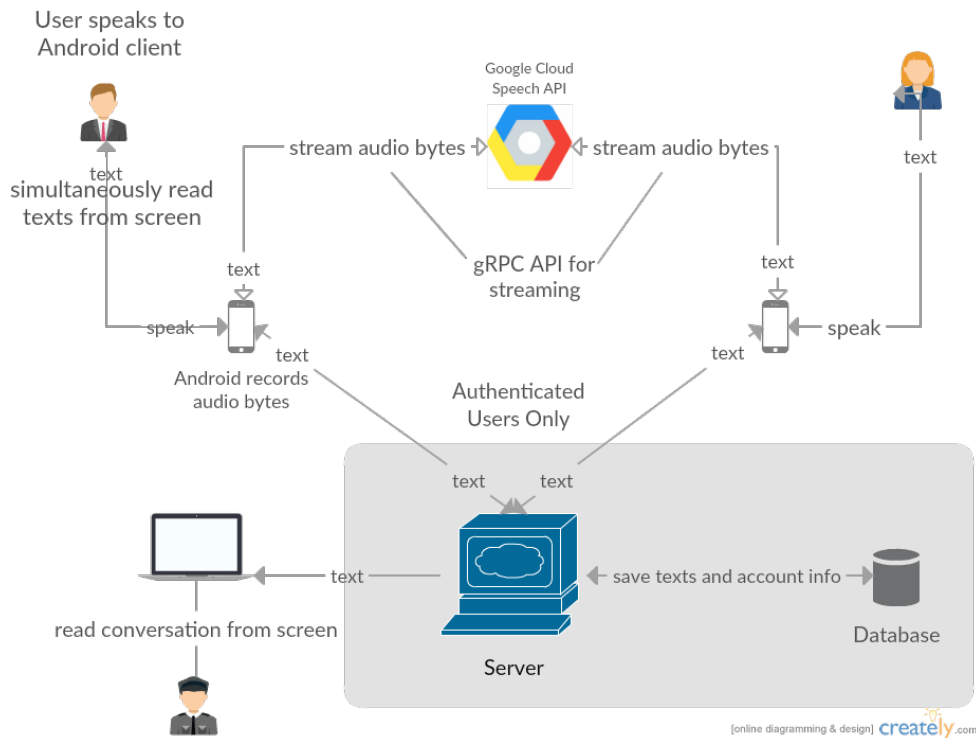


Figure 17. Appla architecture diagram

Figure 17 roughly shows a combination of *four* main components that build up the foundation of Appla:

- Mobile Clients:** The component that is responsible for handling most of the interactions with the users. The Appla application on mobile (in this case, Android clients) will act as an interactive interface: pulling input data from users, sending them to other components and receiving the result data. All these actions happen in real time with little delays (hence, streaming), depending on the network connection strength. These mobile clients will send the audio bytes they recorded from the physical input device (microphone) and stream them directly to the Google Cloud Speech API from which they will be receiving the results. After such process, the users shall see the texts from all other mobile clients changing as they speak, finding the correct phrases from speeches.
- Google Cloud Speech API:** The engine that powers the speech recognition feature of Appla. The API will process the audio bytes provided by the mobile client and return the results in JSON.

- **Web Server:** This component comprises of two other smaller components. The first component is a server that handles WebSocket and the data exchanging process between multiple mobile/web clients. The server will also communicate with the second component, the database, which stores user information as well as rooms and messages content. All text information exchanged through Appla will be saved and protected for any further data investigation.
- **Web Clients:** User can read see the transcribed conversation on the website in real time. The website will pull data from the back-end server, which fetches data (sent previously by mobile clients) from the database and display them in a user-friendly interface.

The processes happening between these components are only available for authenticated users. The reason for authenticated user is for Appla to associate them into separated meetings or sessions. Users having their own account will allow them to secure their data (conversation history), track the personal progress and build a personal vocabulary in the future. The authentication process will be explained in-depth in *Authenticating User Account* part under *User Cases* section.

It is also worth to note the split of *rooms* when users communicate with each other via Appla. For an application that handles multiple clients talking at a same time, this is necessary for managing different sessions. Splitting the connections into channels is good for keeping each conversation clean and easy to follow, while not being distracted by random interruptions from others. Undoubtedly, this is a standard step for an application that involve multi-session communication and data exchange.

In summary, the application involves different ends of parties to support the case of real-time transcribing of the speech into texts and transfer them to other listening clients. The next section will inspect in more detail the database schemes and how they all connect to each other.

## 3.4 Database Schema

### 3.4.1 Android Models

Since the database is hosted virtually on the server, no extensive usage of the local storage is implemented except for authentication token saved in the SharedPreferences object of Android. However, three model classes are still necessary to implement technical requirements of Android, as shown in figure 18.

Each model class will have their own static Builder class that handles the *set* operations of the class. Other *get* operations are public methods inside of parent class. Investigating, the three model classes are as follow:

- **Room:** The room model has the room basic information such as room name, description and room id. There is also a time stamp of the room to tell the time of room creation.
- **Message:** The message model represents each message transcribed from users' speeches. A message contains its sender information, a message id assigned randomly during run-time as explained previously in *Recognizing Speeches* section, message content and the time stamp when the message is created.
- **User:** Containing username, user ID and other personal information like first name and username. No password is archived in the client application.

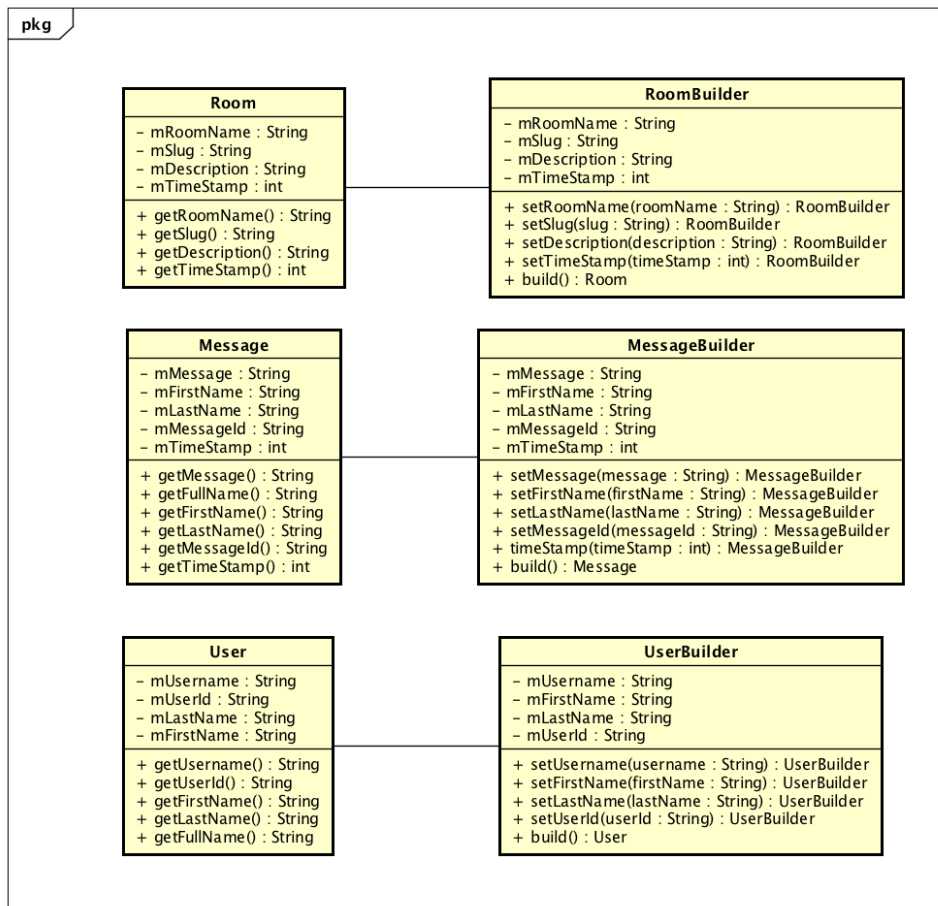


Figure 18. Java classes of Appla Android application

The Appla Android class diagram, with three models containing essential information to display on the mobile application, also represents a similar logic from the database schemes originated in the server, which will be investigated in the next step.

### 3.4.2 Models and Mongoose Schemas

#### Models

Storing the majority of its information on the server, Appla server has the same three core schemes as the Android application for the its first working prototype version. However, as showned in figure 19, the models on the server have different goals and methods such as communicating directly to the database and receive the requests from the clients.

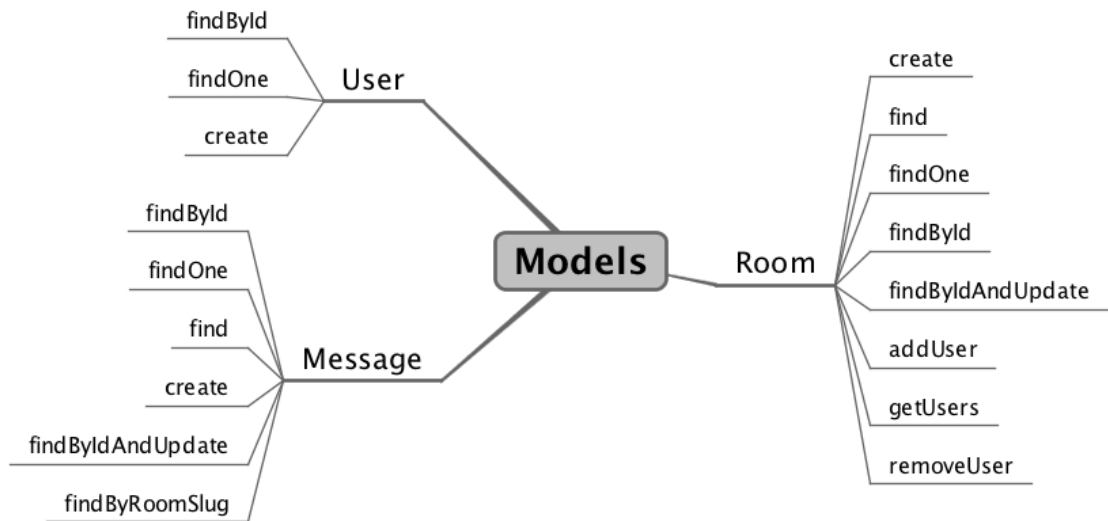


Figure 19. Appla models and the implemented methods

The models have basic find and update methods implemented along with the built-in methods functions provided by Mongoose. In figure 16, it is easier to spot extra methods to support Appla's causes include methods like "findByIdAndUpdate" in Message Model. This method returns all the messages that were exchanged in the room by different users. Calling this method by room slug will return the conversation history of that room.

In the Room model, apart from standard methods, the customized methods include managing the user list of those clients who are currently connected to the room. The room has a connection array that saves objects with the information of socket ID and user ID. A notable case that needs to be taken care of is when a user has multiple devices (which means multiple socket IDs) connected to the room – the "getUsers" method will only return the user list and eliminate duplicated user IDs.

Communications between routers and models are bridged through the usage of Controllers, which is irreplaceable in the whole architect of the application. The router will only interact will controllers and then controllers will make requests through models, which is the only component that can make changes directly to the database. This prevents unknown connections made to the database via the REST API designed in the router. The controllers are configured as figure 20 below.

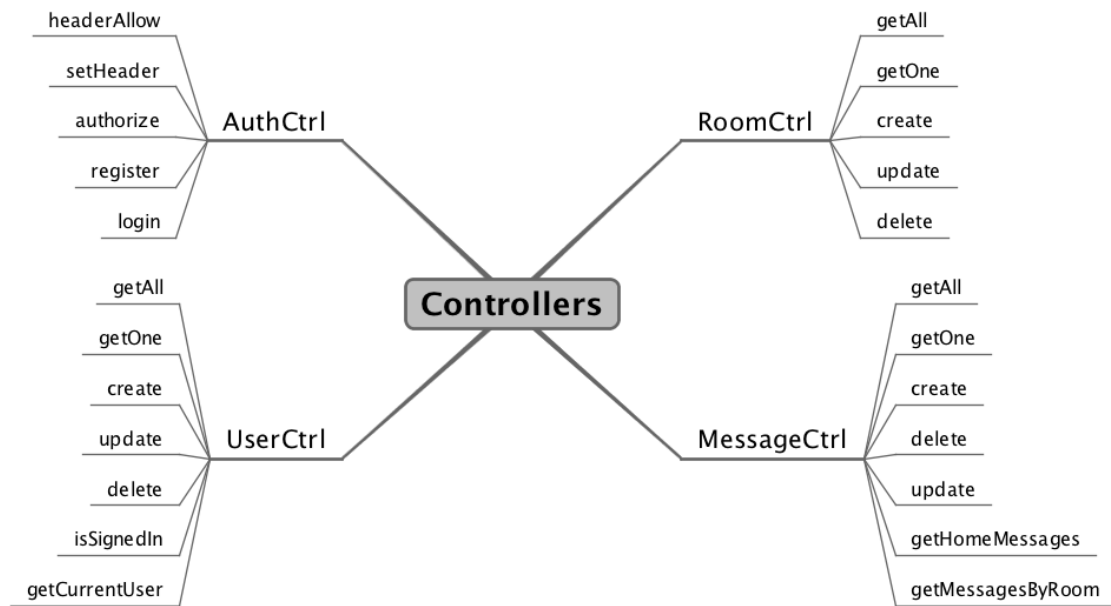


Figure 20. Node.js server controllers and methods implemented

All controllers possess standard CRUD operations with customized methods to fit the business logic of Appla. To break it down:

- **AuthCtrl** is the standard controller that bridge all authorizations request like register and logins. More importantly, there is a middle ware component that will handle the API authorizations and only allow authenticated request to make certain API calls.
- **UserCtrl** has authenticated routes that query the user login status as well as get the current user information based on the access token.
- **RoomCtrl** queries the room list and handles delete/update operations from the Room model.
- **MessageCtrl** has the function “getHomeMessages” to retrieve the user’s messages inside of their private home room. In addition, “getMessagesByRoom” return all the messages by the requested room.

The models are built based on the Mongoose Schemas that define the structure of a MongoDB document, which will be introduced and investigated thoroughly in the next section.

## Mongoose Schemas

In MongoDB, there can be many collections inside a database. Each collection will have their own documents with a same scheme, defined beforehand when the collections are built. These concepts are similar to the concepts of tables in SQL databases; however, the collections do not have primary keys and the relations between tables.

Figure 21 attempts to explain the structure as well as the relations between different schemas.

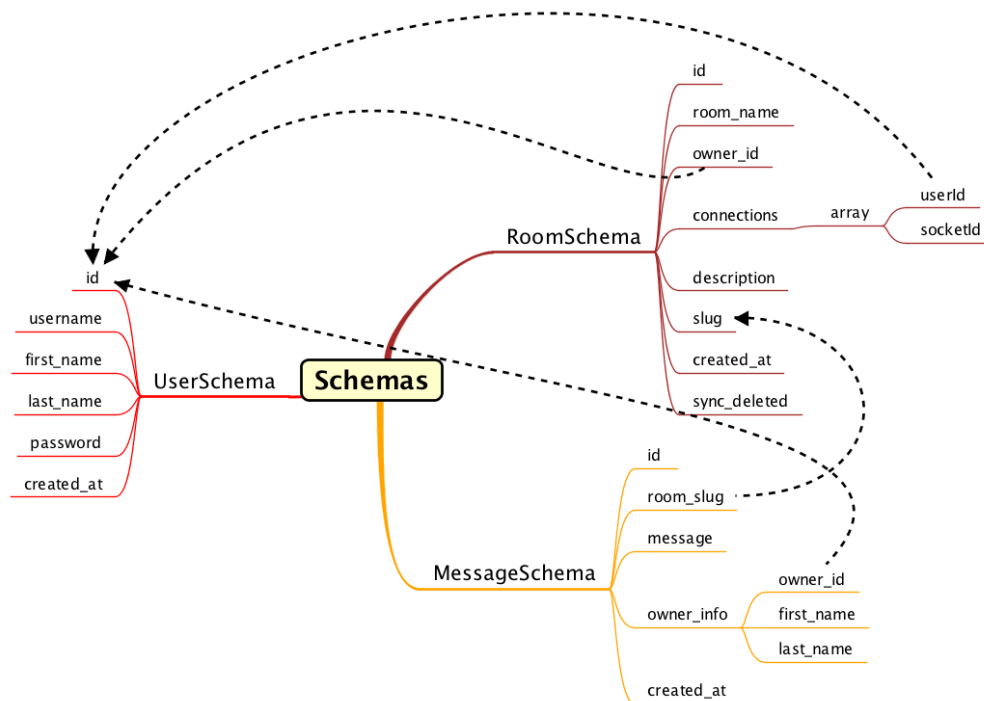


Figure 21. Schema structures and their relations in Appla application

All the schemas have a unique index “id” key (ObjectId type) that is auto-generated every time a new document is inserted in to the database. It is worth to note the “created\_at” key is available to all the schemas because it shall be helpful in case the database need to be upgraded and scaled.

To explain the relations between the schemas, referring to figure 21: a message will have two connections depicting the owner information (user\_id, first\_/last\_name) and the room it belongs to (room\_slug). A room shall have the information of the creator



(owner\_id) and a connection array, which contains the ids of the connected clients (combinations of userId and socketId). This array is updated every time a user joins or leave a room. An element in this array may share the same user ids, but not the socket ids. This is helpful when broadcasting message from the room to all its clients is needed.

## **4 Evaluation and Future Development**

After approximately 6 months of research and development, spanning through the phases of planning, researching, development, testing and documenting, Appla gradually and successfully turned the idea into a working prototype. The product of the project was well-received as a versatile approach that can be adapted and expanded into various solutions. In this section, Appla's results and limitations will be discussed, which will eventually open to some suggestions to the future development of the application.

### **4.1 Results**

#### **4.1.1 Evaluation**

It is good to evaluate the project results based on the targets and requirements set in the beginning. Though not carved in stones, following a set of checkpoints can deliver a great sense of how good in practice the project result is.

As an ambitious project targeted to explore and discover new solutions in language learning, the bars were set high. A mobile application, which smoothens the employees' lives at work by helping them with the language aspects, was expected to deliver with exceptional accuracy and reliability. However, with limited time and resources, Appla turned out to be a lean and agile application, which certainly still needed more dedicated development and research.

As for the core feature, recognizing speech in Appla works well. The surroundings can have some effects on the application's performance, e.g. network connection, background noises: too quiet/noisy, incompatible devices, etc. The application can be used in phone conversations and the texts are archived directly on the website in a 'chat' interface. The user can sign in to the web to check the conversation history in his or her

personal dialogue room. No plans in the beginning were set for account authentication and rooms but after discussions, it was clear that personal user account / room conversation was necessary.

Above is the assessment of the project outcome and how it matched the target set in the beginning. During the testing phases at various university and language course locations, the users' feedback remained very positive. The students who were given the application to test have given many thoughts and ideas of how to improve and suggested other the possible user cases for their personal situations: lecture transcribing, professional interview, etc. The application however also showed flaws and limitations that should be overcome in future development.

#### 4.1.2 Limitations

Talking from a technical viewpoint, Appla is a wide system that comprises many different components that ranges from Android to Web development. While each component has their own issues, a combination of them could cause potential problems that cannot be seen on the surface of the application. Below are some known issues on Appla and the predictions of their causes:

- Sign-in bug: When signed out and sign in with a different account, Appla will shows the message "Hi ..." incorrectly with the signed out user's names. Prediction: when old activity is stuck on the front because of the two-activity switching logic. This logic should be renovated to a different mechanism and a better solution of authentication process.
- Screen orientation changes will lead to the activity to be reset. This should be an easy fix for which solutions can be found online.
- When switching to different rooms, the socket initializations were not working correctly in some cases, which led to the connection cut some times. This bug is predicted to be caused by the fragment switching logic of Appla Android.

From the technology and concept point of view, Google Speech Cloud API did well in processing the speech and transcribing them in real time. The text result showed promising possibilities. However, being a beta version, the API is occasionally unstable

and its strength in the Finnish language is weaker compared to the English language. Text results returned in English is considerably better in accuracy. The speech recognition in numbers works best with simple number cases, however in large numbers with complex structure, the engine fails to return a decent result. Google also lacks the feature to distinguish different speakers from the input source, making conversations in a silent room from two close inputs difficult and prone to errors.

Although errors, bugs and limitations in Applä are observable, it is undeniable that the project has gone a long way from an idea to a working prototype. No development is ever finished as there is always room for improvement in the future.

## **4.2 Future Development**

### **Gamification**

Learning on mobile devices has been following the direction to be fun and spontaneous. The project's idea board was filled with plenty of ideas on Applä's gamification. The games should not be complicated and focused on how to maximize the use of speech recognition in the application. Certainly, recognizing speeches brings new perspectives into improving pronunciation skills while using MALL apps, which were only stopped at playbacks and recording voices with no concrete feedback and judgement. Speech recognition brings objective views into measuring the accuracy of the users' pronunciations, so they do not have to listen to judge by themselves like with traditional MALL apps.

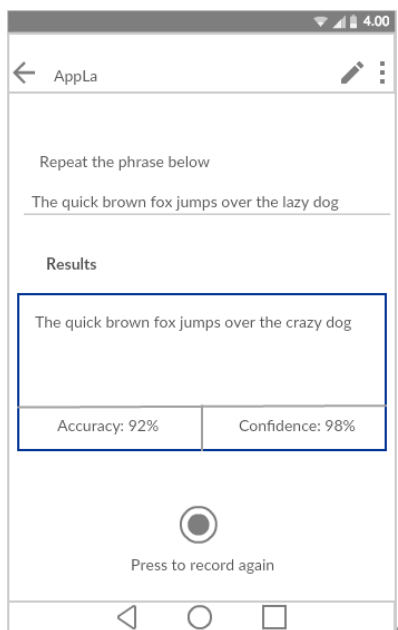


Figure 22. UI mockup for AppLa gamification

One suggestion is to either let users create a phrase they would like to practice pronouncing, or a mechanism to search for difficult sentences extracted from their dialogue recorded previously by AppLa. The users will then record themselves and have the texts transcribed to them on the screen, with an accuracy meter, or “confidence” level of the results. Text-to-speech technology can be used to help users with customized phrases so they have a model to follow, just like usually in traditional learning. A mockup to present the idea is drawn in figure 22.

### Topic Keyword Extraction

In meetings, non-Finnish speaking employees working in Finland usually find themselves in meeting situations that were only or partially in Finnish. It is no doubt that even employees with advanced skills in Finnish can sometimes be lost in the conversation. AppLa can be used as a note-taking application that records everything and in turn extracts the most important keywords as well as the topic of the conversations. The topics can help the users to follow and have a better sense to expect to listen to what is being discussed.

## Other ideas

The Cognition services are becoming more and more available. Although still under the development phase, they still show great promising possibilities to help Appla achieve the novelty of an application in the language learning fields. Also, iOS development is also a good step ahead to be taken if production of Appla is considered.

Progress tracking and data analysis is truly an important topic to be investigated and explored to maximize the benefits to the user. The data is available in the database concerning how people converse in real life situations. If a progress tracking system is implemented based on the analysis of the user's vocabulary and interest topics, it will bring exceptional benefits to the end users of Appla.

## 5 Conclusion

To begin with, the Appla project had an abstract idea and an ambition to explore and propose a new approach to using a mobile application as an assisting tool for language learning. The idea originated from the current demands from a rising labour market in Finland. Those who seek work often found it confusing and difficult to find a job without proper skills in communication. Companies were spending resources from their own pockets to open Finnish classes and hire teachers to improve their employees' language skills, especially those in social and services sectors. Eventually, employers were always looking for new solutions that would help reduce the cost and time resources.

After months of researching and developing, the concept behind Appla started to form its shape. Many prototype versions were introduced and constantly being refined as discussion and feedback went on. Appla has gone through many changes and improvements while employing cutting-edge technologies and services. It followed the trends of using Artificial Intelligence over the cloud and customized according to the needs of the application.

As for results, the deliverables included a full-stack web system backed by a database and an Android application that handles real-time speech recognition. It has been perceived as a solid and versatile foundation for further implementation and research to be taken place in the future.

The route to overcome the barriers of languages is challenging and exciting. We still have a long way ahead, but as in an interesting time we are living, I could not wait to see where technologies take us.

## References

- 1 Key Figures on Immigration [online]. European Migration Network; Finnish Immigration Service; 2016.  
URL: [http://www.migri.fi/download/65019\\_Tilastograafit\\_2015\\_EN\\_final.pdf?9142ca54ab72d488](http://www.migri.fi/download/65019_Tilastograafit_2015_EN_final.pdf?9142ca54ab72d488). Accessed: 2 March 2017.
- 2 Duman G, Orhon G, Gedik N. Research trends in mobile assisted language learning from 2000 to 2012. ReCALL. Cambridge University Press; 2015;27(2):197–216.
- 3 Aga Palalas JO. Mobile-Assisted Language Learning. Georg Brown Coll [Internet]. 2011;3(1):309–19.  
URL: <http://teachonline.ca/pockets-innovation/mobile-assisted-language-learning>
- 4 Klopfer E, Squire K, Jenkins H. Environmental Detectives: PDAs as a window into a virtual simulated world. Proc - IEEE Int Work Wirel Mob Technol Educ WMTE 2002. 2002;95–8.
- 5 Stockwell G. Investigating learner preparedness for and usage patterns of mobile learning. 2008;20(3):253–70.
- 6 Heil CR, Wu JS, Lee JJ, Schmidt T. A review of mobile language learning applications: trends, challenges and opportunities. J Palliat Med. 2002;5(1):32–50.
- 7 Godwin-jones R. Emerging Technologies Mobile Apps for Language Learning. Lang Learn Technol. 2011;15(2):2–11.
- 8 Chinnery GM. Going to the MALL : Mobile Assisted Language Learning. Lang Learn Technol [Internet]. 2006;10(1):9–16.  
  
URL: <http://www.llt.msu.edu/vol10num1/pdf/emerging.pdf>
- 9 Expat-finland, Finnish Language Schools and Courses [online]. Stuart Allt Web Design; 2017.  
URL: [http://www.expat-finland.com/living\\_in\\_finland/language\\_training.html](http://www.expat-finland.com/living_in_finland/language_training.html). Accessed: 21 March 2017.
- 10 Sukimob Project  
URL: <https://skimob.wordpress.com>. Accessed: 24 March 2017.
- 11 Smartphone OS Market Share, 2016 Q3. International Data Corporation.  
URL: <http://www.idc.com/promo/smartphone-market-share/os>  
Accessed: 08 April 2017
- 12 International Immigration Report 2015, United Nation, 2016  
URL:  
[http://www.un.org/en/development/desa/population/migration/publications/migrationreport/docs/MigrationReport2015\\_Highlights.pdf](http://www.un.org/en/development/desa/population/migration/publications/migrationreport/docs/MigrationReport2015_Highlights.pdf) Accessed: 2 March 2017
- 13 Krashen SD. Principles and Practice in Second Language Acquisition [Internet]. Vol. 67, The Modern Language Journal. 1982. 168 p.

- 14 Vygotsky L. Interaction between learning and development. *Mind and Society*. 1978. p. 79–91.
- 15 Scarino A, Liddicoat AJ. *Teaching and Learning Languages: A Guide*. 2009;1–116.