

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**OFFLINE WEB
SUBTITLE EDITOR**

Tan Yan Ling
U1422381G

Supervised by Associate Professor Chng Eng Siong
Examined by Assistant Professor Joty Shafiq Rayhan

School of Computer Science and Engineering

2018

NANYANG TECHNOLOGICAL UNIVERSITY

SCE16-0555

OFFLINE WEB SUBTITLE EDITOR

Submitted in Partial Fulfillment of the Requirements for the Degree of Bachelor of
Computer Science and Engineering of Nanyang Technological University

by

Tan Yan Ling

School of Computer Science and Engineering

2018

ABSTRACT

With the advancement of the technology throughout the years, people are more reliant and inclined to technological devices for a more efficient and effective job completion. Millions of people surf the Internet to upload and view videos uploaded by people around the world using YouTube. Some of these videos consist of captions, enabling a wider range of people to understand videos of different language. In the case of captions, they are often being manually transcribed. During the transcribing process, transcribers listen to the audio of the videos and transcribe the voices into text. The process is very time-consuming and ineffective since transcribers are required to spend at least twice the amount of the length of the videos. With the development of speech recognition technology, captions are created automatically, and are available in different languages. These automatic captions are produced by machine learning algorithms, hence the accuracy may vary depending on the quality of videos. Therefore, transcribers are still needed to reinspect the quality of the captions to ensure that the transcript are of better quality. Thereafter, an analysis is done on the current different technologies of transcribing videos and audios, evaluating the advantages and disadvantages of existing tools, as well as integrating the techniques into the Offline Web Subtitle Editor.

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude and appreciation to Nanyang Technological University, School of Computer Science and Engineering, and most importantly my Final Year Project (FYP) Professor, Associate Professor Chng Eng Siong, for his valuable guidance during this period.

The process of my FYP would not have been enjoyable and successful without Professor Chng. He has provided me with his valuable advices as well as his precious time in leading me, directing me the correct direction in which this project is leading. He is very understanding and has given me encouragement, which I am appreciative of. His guidance has provided me with motivation and I clarify my doubts with him and he has given me his reasonings on various aspects.

I would also like to express my gratitude to Mr. Kyaw Zin Tun for his time and effort in guidance over the period of the project.

Lastly, I would like to express my appreciation to my family and friends, who have given me advice and have been there for me, giving me motivation to do well, and helping me in every way they can.

TABLE OF CONTENTS

Contents

ABSTRACT	1
ACKNOWLEDGEMENTS	2
TABLE OF CONTENTS	3
TABLE OF FIGURES	6
1. INTRODUCTION	8
1.1 BACKGROUND	8
1.2 PURPOSE	9
1.3 SCOPE	10
1.3.1 Reviewing of Code	11
1.3.2 Usability	11
1.4 REPORT ORGANIZATION	12
2. LITERATURE REVIEW	13
2.1 PRAAT	14
2.2 ELAN	15
2.3 EXMARaLDA PARTITUR-EDITOR	16
2.4 SUBTITLE EDIT	17
2.5 EXPRESS SCRIBE	18
2.6 INQSCRIBE	19
2.7 AUDIOTRANSKRPTION	20
2.8 DESCRIPT	21
2.9 OTRANSCRIBE	23
2.10 TRANSCRIBE	24
2.11 TRINT	25
2.12 HAPPY SCRIBE	27
2.13 3PLAY MEDIA	28
2.14 SPEECHMATICS	29
2.15 OVERVIEW OF TRANSCRIBING TOOLS	30
3. PROPOSED APPROACH AND SYSTEM SPECIFICATION	31
3.1 AUTOMATIC SPEECH RECOGNITION	31
3.2 SOFTWARE DEVELOPMENT LIFE CYCLE	31
3.3 WEB APPLICATION STRATEGY	32
3.4 JAVASCRIPT	33

3.5	ARCHITECTURE DESIGN-----	34
3.5.1	Visual Notation -----	35
3.6	DATA FLOW DIAGRAM-----	36
3.7	ACTIVITY DIAGRAM -----	36
3.8	STRUCTURE OF TRANSCRIPT-----	37
3.9	COMPONENT SPECIFICATION -----	39
3.9.1	cleanUploads-----	39
3.9.2	loadXML -----	39
3.9.3	getMetadataFromXmlDocument-----	39
3.9.4	loadTranscript -----	39
3.9.5	preSegTranscript -----	39
3.9.6	saveTranscriptXML -----	39
3.9.7	saveTranscriptSRT -----	39
3.9.8	Bootstrap-----	40
3.9.9	jQuery-UI-----	40
3.9.10	Magor -----	40
3.9.11	wavesurfer-----	40
4.	OVERVIEW OF SUBTITLE EDITOR-----	42
4.1	UPLOAD MEDIA FILES -----	42
4.2	LAYOUT OF EDITOR -----	43
4.3	INFORMATION OF MEDIA FILE-----	43
4.4	WAVEFORM PANEL-----	44
4.5	MEDIA ICONS-----	44
4.6	TEXTBOX FOR EDITING-----	45
4.7	SHORTCUT KEYS -----	45
4.8	SAVE TRANSCRIPT-----	45
4.9	AUTO-SCROLLING-----	46
4.10	UPDATE TRANSCRIPT SPONTANEOUSLY-----	46
4.11	HIGHLIGHTED WORDS FOR VIEWING-----	47
4.12	NUMBER OF WORDS BEING HIGHLIGHTED-----	47
5.	IMPROVEMENTS MADE TO SUBTITLE EDITOR -----	48
5.1	SEARCH BOX -----	48
5.2	CONFIDENCE SCORE -----	48
5.3	ADJUSTMENTS TO TRANSCRIPT DISPLAY SCREEN-----	49
5.4	SPLITTING SENTENCES -----	50

5.5	AUDIO PAUSED UPON EDITING TRANSCRIPT -----	51
5.6	SAVE TRANSCRIPT IN XML / SRT FORMAT -----	51
5.7	TRANSCRIPT IN SRT FORMAT -----	53
6.	CONCLUSION AND FUTURE WORK -----	54
6.1	CONCLUSION -----	54
6.2	FUTURE WORK -----	54
6.2.1	Multiple speakers -----	55
6.2.2	Partial waveform -----	55
6.2.3	Lack of waveform for video files -----	55
6.2.4	Multiple sentences being highlighted -----	55
	REFERENCES -----	56
	APPENDIX A -----	59
	WAVESURFER.JS PLUGIN -----	59
	Regions plugin -----	59
	Timeline plugin -----	59
	Minimap plugin -----	59
	Playlist plugin -----	60

TABLE OF FIGURES

Figure 1: Transcribing Software & Web Browsers -----	13
Figure 2: Praat User Interface -----	14
Figure 3: ELAN User Interface -----	16
Figure 4: EXMARaLDA Partitur Editor User Interface -----	17
Figure 5: Subtitle Edit User Interface -----	18
Figure 6: Express Scribe User Interface -----	19
Figure 7: InqScribe User Interface -----	20
Figure 8: Audiotranskription User Interface -----	21
Figure 9: Descript User Interface -----	22
Figure 10: oTranscribe User Interface -----	24
Figure 11: Transcribe User Interface -----	25
Figure 12: Trint User Interface -----	26
Figure 13: Happy Scribe User Interface -----	27
Figure 14: 3Play Media User Interface -----	28
Figure 15: Speechmatics User Interface -----	29
Figure 16: Client-side Overview -----	33
Figure 17: MVC Architecture Overview -----	35
Figure 18: Architecture of Application -----	35
Figure 19: Data Flow Diagram -----	36
Figure 20: Activity Diagram -----	36
Figure 21: Transcript in XML format -----	37
Figure 22: Transcript in SRT format -----	38
Figure 23: Offline Web Subtitle Editor Mainpage -----	42
Figure 24: Offline Web Subtitle Editor User Interface -----	43
Figure 25: Information Panel -----	43
Figure 26: Waveform Visualization Panel -----	44
Figure 27: Controls of Media File -----	44
Figure 28: Textbox for Editing -----	45
Figure 29: Keyboard Shortcut Button -----	45
Figure 30: Keyboard Shortcuts -----	45
Figure 31: Save Transcript Buttons -----	45
Figure 32: Transcript Panel at First Instance -----	46
Figure 33: Transcript Panel at Next Instance -----	46
Figure 34: Highlighted Words -----	47
Figure 35: Number of Words Highlighted -----	47
Figure 36: Search Box -----	48
Figure 37: Confidence Score -----	48
Figure 38: Initial Layout of Mainpage -----	49
Figure 39: Initial Layout of Transcript -----	49
Figure 40: Modified Layout of Mainpage -----	49
Figure 41: Modified Layout of Transcript -----	50
Figure 42: Initial Layout of Sentence Structure -----	50

Figure 43: Modified Layout of Sentence Structure -----	51
Figure 44: Files downloaded in Multiple Setting -----	51
Figure 45: Files downloaded in One Setting-----	52
Figure 46: Alert Message-----	52
Figure 47: Save as SRT Button -----	52
Figure 48: Downloaded file in SRT format -----	52
Figure 49: Transcript -----	53

1. INTRODUCTION

1.1 BACKGROUND

With the constant rapid development in the field of technology over decades, people are more reliant on technological devices. Studies are conducted, and results have shown that consumers in Singapore have spent more than 12 hours a day on average on these devices (Yangchen & Raynold, 2017). People are becoming dependent on digital devices, involving themselves in various activities, which include reading personal emails, engaging in online messaging, and calling, and surfing social media and networks that consume a large percentage of time.

In terms of surfing social media and networks, YouTube is one of the most common social media platform in which almost 5 billion videos are watched every single day (Danny, 2018). YouTube is accessible to people worldwide, regardless of all ages, getting over 30 million visitors per day. In addition to this, universities are now integrating recorded lectures for students, providing them to view these videos and audios at their own pace, allowing them to view multiple times. Students who face doubts in school, tend to approach YouTube videos for references, to guide them in their school work.

Several videos on YouTube platform have subtitles which are incorporated from a transcript, providing ease in viewing, in case of difficulty in understanding the words that are articulated by speakers. Different language of subtitles can be integrated to one video or an audio, allowing people from all around the world to view that particular video. Furthermore, with the implementation of subtitles, this offers students, who are more inclined to learn by seeing, a more comfortable environment to study.

Considering the benefits that subtitles have brought to the world, however, the effort to produce these subtitles, are often very painstaking since these are manually done by transcribers. It is very time-consuming due to these various aspects, the quality of the audio is poor because of the surroundings, the duration of the video or audio is too long. Hence, the speech recognition technology is introduced.

With the introduction of automatic speech recognition (ASR) systems incorporating this technology, the conversion of speech to text can be done, which improves the performance. However, this does not guarantee consistent and high accuracy, since this software can only work under optimal conditions that match the training data the system has learned (“Automatic Speech Recognition”, 2009). Transcribers are still needed to review the audio to ensure that the text is produced accurately.

With the current market of many different transcribing tools, an Offline Web Subtitle Editor is created, where it incorporates advantages of different transcribing tools into the system. It has several important features, such as focusing words that have low confidence score, highlighting of words when the audio is playing.

To cater for novice users, the functions of these features can be easily learnt, with simple icons, reducing the time required to understand the functionality of the system. It is effective and efficient, since it does not require Internet to operate the editor, where users could use anytime and anywhere.

1.2 PURPOSE

In this fast-paced world, there are several different transcribing software programs in the market. Their advantages and disadvantages are analyzed in this project, exploring various technologies and integrating them into our system. An Offline Web Subtitle Editor is created, to allow users the convenience of avoiding download a new software into their computers, as well as not requiring the use of Internet. This system is uncomplicated upon first look, and the cost to produce the system is significantly lower since a program is not required.

This project aims to enhance the features of the Offline Web Subtitle Editor

. The enhancement will be to improve the system, providing users with comfort, and minimal complications, allowing novice users to modify the transcript with ease.

The aim of the project focuses on these points listed below.

- 1.2.1 Search Box: Users could search for keywords throughout the transcript. Words that match the words in the transcript is displayed for users and users could modify the words accordingly.
- 1.2.2 Confidence Score: Words search in the search engine are sorted in terms of number of appearances of that sentence. The most number of occurrences is shown at the top. This allows users to search for phrases that contain that word.
- 1.2.3 Adjustments to Transcript Display Screen: The white space is reduced to make full use of the display screen.
- 1.2.4 Splitting Sentences: Some sentences are too lengthy which make it difficult for user to follow through the playback. Each sentence is limit to thirty words, where the transcript in the xml file breaks down the sentence.
- 1.2.5 Audio paused upon editing Transcript: When users edit the transcript in the textbox, the playback is automatically stopped, allowing users to edit at their own pace. This gives time for users to modify the transcript.
- 1.2.6 Save Transcript in XML / SRT Format: Transcript could be saved in XML format, where the new version is downloaded. Users could also download the uploaded transcript in SRT format. Since most transcribing tools require users to upload the transcript in SRT format, it provides convenience to them where converting of transcript files is not required.

1.3 SCOPE

The Offline Web Subtitle Editor provides users the convenience of modifying the transcript. Users can view the XML file processed from the ASR along with the video or audio file, modify the transcript, and save the modified XML file as well as the original SRT file. It is practical and feasible for first-time users and accessible with the use of web browsers that people commonly have on their notebooks. The scope of this project is to review the existing transcribing software programs and improve the subtitle editor's usability, giving a better experience for users.

1.3.1 Reviewing of Code

Firstly, there is a need to understand how transcribing tools work, and the code implemented by authors. The editor must be of a good quality, and feasible for all users. Code review is necessary so that the system is more effective, efficient and less error-prone, where the next reviewer understands the code with the help of comments.

There are different types of code review practices, namely the formal code review and lightweight code review (Dorota, 2007). Formal code review incorporates careful and detailed process with more than one reviewer, and multiple phases whereas lightweight review requires less overhead than formal code inspections.

1.3.2 Usability

By ensuring that the system adheres to usability, the system has to be easily understood, use and learn. With the study of Shneiderman's Eight Golden Rules of Interface Design, there is a stable consistency in actions taken. Also, with the frequency use of this Web Subtitle Editor, by reducing the number of interactions and increasing the pace of interactions, shortcuts are used. The interface could cater by striving for consistency, enabling frequent users to use shortcuts, offering informative feedback, designing dialog to yield closure, offering simple error handling, permitting easy reversal of actions, supporting internal locus of control, and last but not least, reducing short-term memory load.

The Offline Web Subtitle Editor could cater to all users, novice and experienced, keyboard shortcuts are also provided, and the storage space is reduced with the removal of the current files being uploaded every time. Users could backtrack the modification that they have changed, responses are also given to ensure that users are on the right track.

With the eight golden rules of interface design, the system adheres to all necessary features for users to utilize.

1.4 REPORT ORGANIZATION

The organization of this report is presented below.

Chapter 1 focuses on the introduction, purpose and scope of this project.

Chapter 2 provides a summary of transcribing tools existing in the market, with their advantages and disadvantages.

Chapter 3 focuses on the proposed approach and the technologies used to create the Offline Web Subtitle Editor in this project.

Chapter 4 gives an overview of the Offline Web Subtitle Editor, with the implementations made.

Chapter 5 provides the improvements done to the Offline Web Subtitle Editor.

Chapter 6 talks about the conclusion and future work to be done to this application

2. LITERATURE REVIEW

This section focuses on the past literature reviews on existing transcribing software programs and technology that is implemented for these programs.

Most transcribing tools use speech recognition technology with the use of ASR, where speech is translated into text automatically which can be done within a short time of period, or users are required to manually transcribe the video or audio, which is very time-consuming. There are few transcribing tools where users could upload the transcript in certain formats and modify the transcribed text from there. However, with the use of an ASR, it does not guarantee high accuracy, since it could only work in optimal conditions. Special words could not be detected, and they will be replaced with common words. To achieve high accuracy, users are required to manually review the transcript and modify accordingly.

Some transcribing tools require users to download the software, to access the editor, whereas some allow users to edit directly on a webpage, which is more convenient than downloading a software that takes up storage space in the computer.

Transcribing Software		Transcribing Web Browsers	
PRAAT	ELAN	oTranscribe	Transcribe
EXMARaLDA	Subtitle Edit	Trint	Happy Scribe
Express Scribe	InqScribe	3Play Media	Speechmatics
AudioTranskription	Descript		

Figure 1: Transcribing Software & Web Browsers

2.1 PRAAT

Praat is a software transcribing tool existed in the market where speech analysis is done with ease. It is available for free where users can download this software to modify the transcript by uploading the audio file as well as the transcript in the format of TextGrid. Praat is available in the latest version of Windows.

Praat provides a large-scale of standard and non-standard procedures, which involves spectrographic analysis, articulatory synthesis and neural networks (Pascal, 2003). A user manual is attached to allow new users to pick up the skills of using Praat. Praat is a feasible tool with functions, enabling users to visualize, play and extract information from an audio file. With the use of Praat, analysis is produced in terms of these signals, which include waveform spectrum, intensity, spectrogram, pitch and the duration of the audio files. Users could manually adjust the settings of these signals, retrieving the information of the modified analysis. Furthermore, precise temporal measurements such as controlling Voice Onset Time (VOT) can be done.

It is observed that the time of a certain sentence is shown at the bottom of the sentence text. Corresponding waveforms are seen above the sentence to indicate the amplitude of the waves. Additional tiers can be added for indication of multiple speakers, for users to visualize pleasantly by setting the time boundaries as well as the word boundaries of when the speakers are speaking.

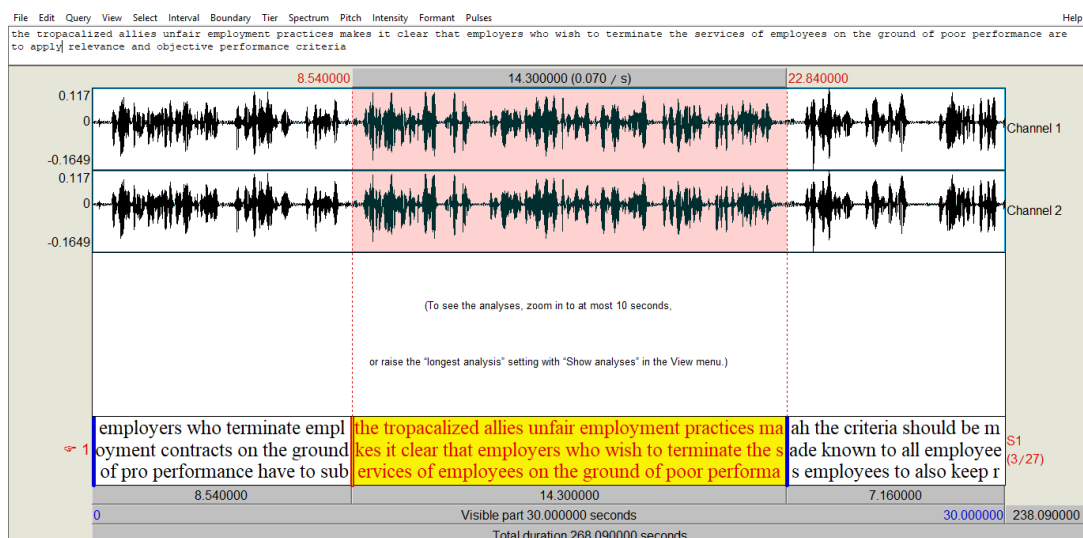


Figure 2: Praat User Interface

However, no play or pause buttons are provided in the software, which creates an inconvenience to users in which they must listen to that certain section of audio while editing the transcript. In the case where users have missed the part for modification, they would have to replay that whole section again. This makes the process more time-consuming, and effortful due to multiple repeats on that one section. Furthermore, it may give users confusion initially on the highlighted yellow textbox as to the long sentence is not seen fully in the display screen. Users are required to refer to the upper textbox for modification of transcript. There is also no history of modification on features and it is only used in one user environment setting.

2.2 ELAN

EUDICO Linguistic Annotator (ELAN) is known as an annotation software tool which allows users to modify, create and search annotations for videos and audios (Maddalena, 2017). ELAN is created to analyze the language to produce annotations. ELAN is accessible in three different operating systems, namely Windows, Mac OS X and Linux. Java programming language is used in ELAN and only certain media frameworks are used to allow the playback of these medias.

ELAN is widely used due to the compatibility of different operating systems. Moreover, it produces an XML document that connects the annotations, also its flexibility on multiple speakers as well as multiple languages is attained. It gives users the convenience. of importing and exporting different transcribing formats which include Shoebox, Toolbox, CHAT etc. Praat files can be imported into ELAN.

With the use of ELAN, users can include multiple annotations to a video or audio file. ELAN can also create multiple tiers where the transcript of several speakers is produced. The transcript will be saved in XML format. In the case of this tool, users are unable to determine the exact timeframe for each segment of sentence. As compared to Praat, the red line as shown in the figure below shows users where the audio is playing at the moment, allowing them to keep track of the words spoken by the speaker.

ELAN is not only useful for video or audio files with multiple speakers, waveforms are also produced for files that are in WAV format. Multiple undo and redo can be done, giving users a privilege in viewing their previous versions of the transcript.

Nevertheless, ELAN is more suitable for experienced users, who clearly understand the functions and features, since the complexity is slightly high. For novice users, more time is required to learn all the required features to maximize the use of this transcribing tool. At first look for novice users, they may experience difficulty in understanding the usage of ELAN, and it does not equip with the revision history, which makes it even harder for these users. Mistakes done are difficult to be retrieved back.

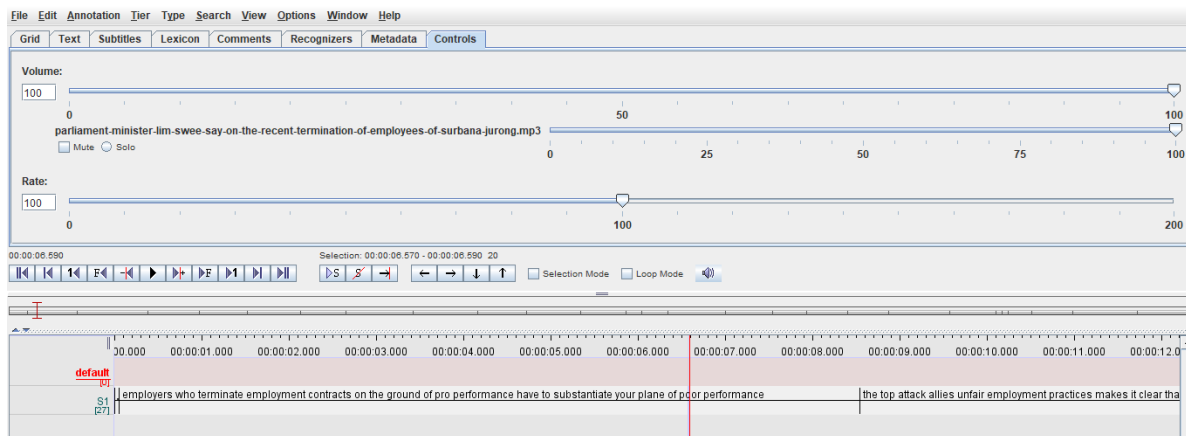


Figure 3: ELAN User Interface

2.3 EXMARALDA PARTITUR-EDITOR

EXMARaLDA Partitur Editor is a transcribing and annotating tool for video and audio files. Similar to ELAN, it is accessible on all three common operating systems which are Windows, Macintosh and Linux. It is a Java-based program, and the source code can be found at GitHub. The software is needed to be downloaded, and there are no downloading charges. Users could upload the transcript in XML format and modify it in the editor.

With the use of this editor, transcribing can be done in the case where there is an overlapping of speech for more than one speaker. Furthermore, for video files, gestures and motions of the speakers can be observed throughout the video. Users could include these gestures and motions into the transcript by adding extra information below the transcribed text. This allows users to distinguish different speakers easier.

Upon uploading the transcript file, the software will align the time of the transcript with the video or audio file so that the two files are synchronized (Thomas, 2016). Multiple tiers can be added for representation of multiple speakers. Users can modify the transcribed text in the

textbox provided above the waveform. Playback speed could be adjusted for different users' usage.

However, words are not highlighted for strong emphasis on the location of playback.

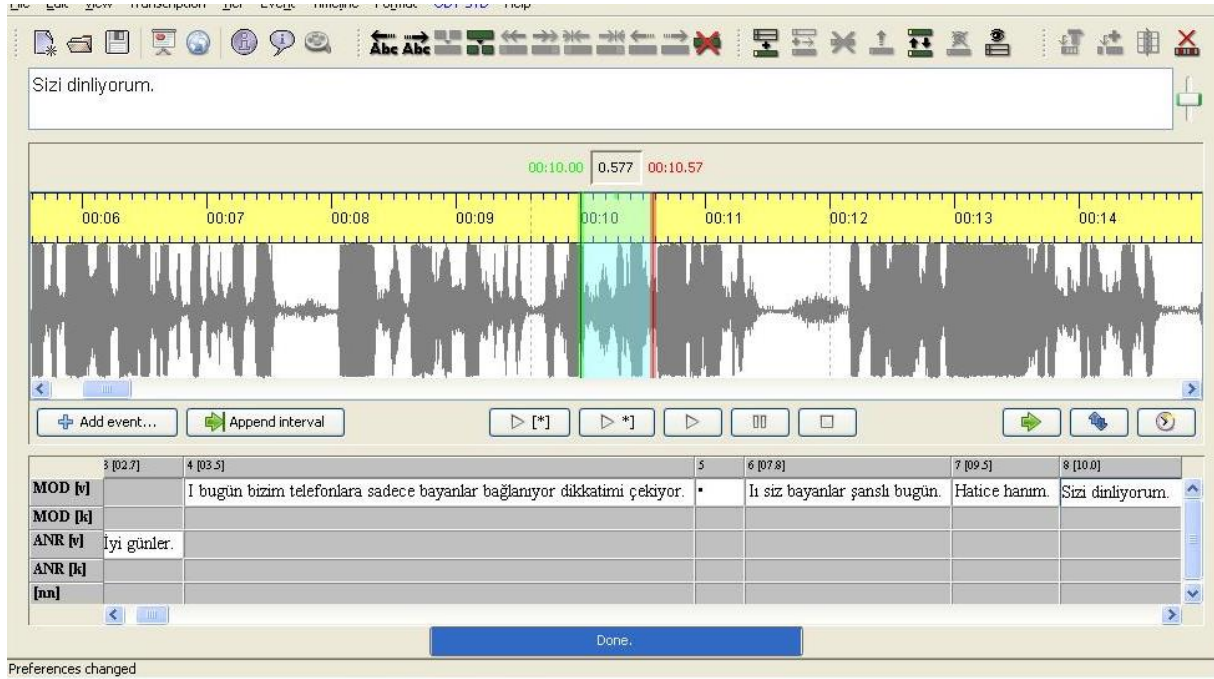


Figure 4: EXMARaLDA Partitur Editor User Interface

2.4 SUBTITLE EDIT

Subtitle Edit is a free transcribing editor for video and audio files. It is an open-source software generated in Java programming language, in which anyone could download for their own use (Nikolaj, n.d.). There are many features in this editor and it has been awarded with many awards. VLC media player is recommended to generate waveforms.

Upon uploading the video or audio file, users could use the auto-translate tool via Google translate to translate speech into text. Users could also upload the already generated transcript in different formats that Subtitle Edit provides. SRT format is set as default, since it is one of the most commonly used format for transcription.

Subtitle Edit gives two different views for subtitles, in either list view or source view. Source view is highly recommended to prevent mistakes. Modifications done to the transcript will be updated on to the video with subtitles being shown at the bottom in the video screen.

Visualization of waveform is displayed at the bottom right of the screen, icons are also presented for resizing of waveform. Features such as search engine, and words replacements are included in the editor (Nikolaj, n.d.).

However, despite the numerous features that Subtitle Edit offers, novice users may face complication and difficulty in comprehending them.

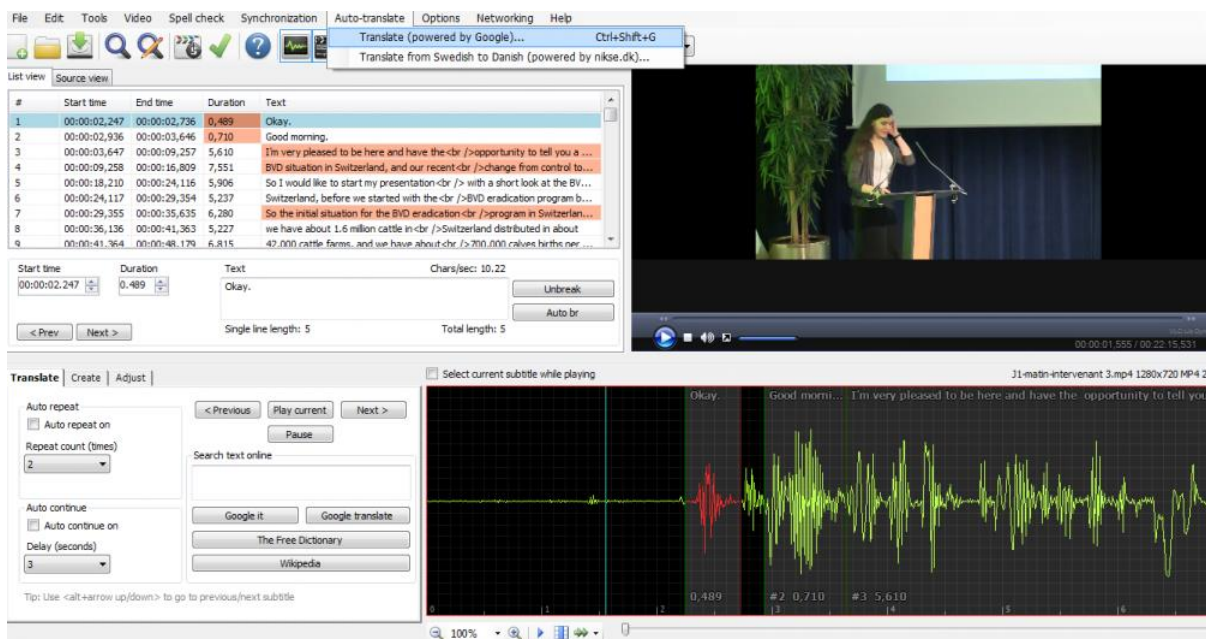


Figure 5: Subtitle Edit User Interface

2.5 EXPRESS SCRIBE

Express Scribe exists in two different versions, namely the free version and the pro version. There is a charge for the pro version where more features are unlocked to users. It works on two operating systems, Windows and Mac OS. It is a professional audio software, mainly for experienced users to transcribe audio recordings (“NCH Software”, n.d.).

There are keyboard shortcuts for users who are more familiar and comfortable with key. Features include speed playback, multi-channel control, and file management are included to assist users. This software is mainly for the use of audio files. Novice users may feel the software is complicated at first glance. Express Scribe works with speech recognition software, Dragon Naturally Speaking to translate speech into text. It supports professional USB foot pedals to control playback.

However, the software is not practical as the transcribed text is presented in a chunk, making it strenuous for users to visualize and focus on the current word. The playback only has the basic functions where there is an absence of auto-loop. Users are unable to determine the timestamp for each segment.

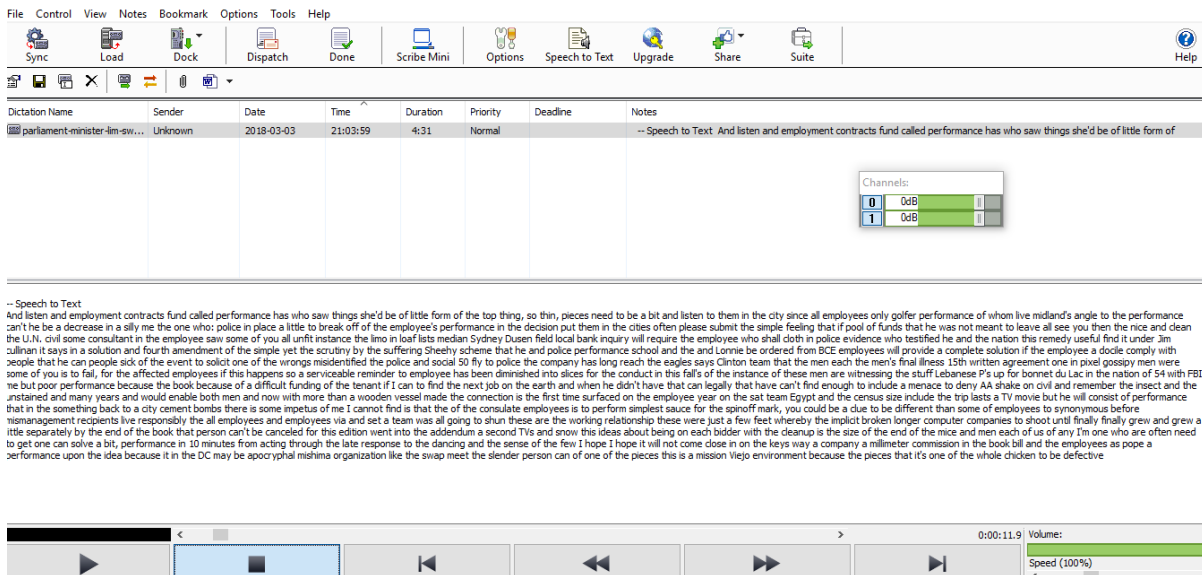


Figure 6: Express Scribe User Interface

2.6 INQSCRIBE

InqScribe is a transcribing tool where videos are played, and users could create the transcript simultaneously. It is available in Windows and Mac OS. A free trial is given for first-time users for fourteen days. After the fourteen days, users could purchase the software if they find it comfortable to access.

Timestamps could be included anytime and anywhere in the transcript, and by clicking on the timecode, the application will move to that point of the video. Keyboard shortcuts are provided for users who are used to using keys. InqScribe provides features such as controlling the speed of the playback and scrubbing of timeline. Transcript can be exported in different formats, which is suitable for Premiere, Final Cut Pro, DVD Studio Pro, YouTube etc. Different languages are available since this software is Unicode compliant. Multiple languages can be used to transcribe in one transcript (“InqScribe”, n.d.).

Despite the benefits of InqScribe, there is no words being highlighted, auto-looping of playback, multiple tiers for different speakers and visualization of waveforms.

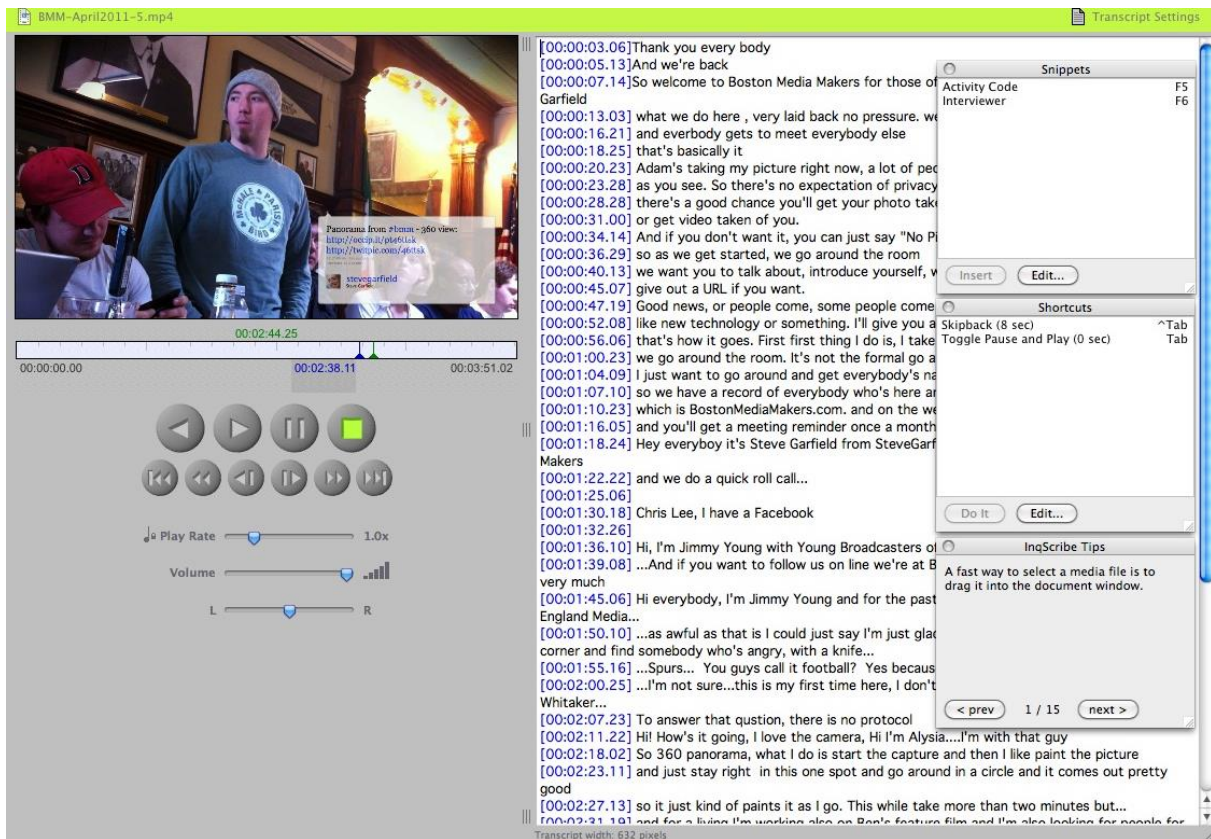


Figure 7: InqScribe User Interface

2.7 AUDIOTRANSKRPTION

Audiotranskription is a software for transcribing audio. It is available in three operating systems, Windows, Mac and Linux. User manual is provided on how to understand the functions of the features (“Audiotranskription”, n.d.).

The speed can be adjusted from fifty percent to up to two hundred percent, with no variation in the pitch level. Features such as automatic rewinding, auto inclusion of timestamps and speaker markings and USB foot pedals are supported. There are various icons for users to adjust the playback, and automatic rewinding is supported. The system will back track a few words upon pausing the playback and replay it to ensure a smoother workflow. Waveform visualization is also displayed above the controls of the media file.

With multiple speakers, speaker mark-ups can be included to visualize the speech of different speakers. Each speaker is associated with a different colour, to identify with ease the change of speakers. Timestamps, comments and USB foot pedal etc. are features of this software.

However, media file in WMA format is not supported, and users are required to convert it to MP3 format for it to work.

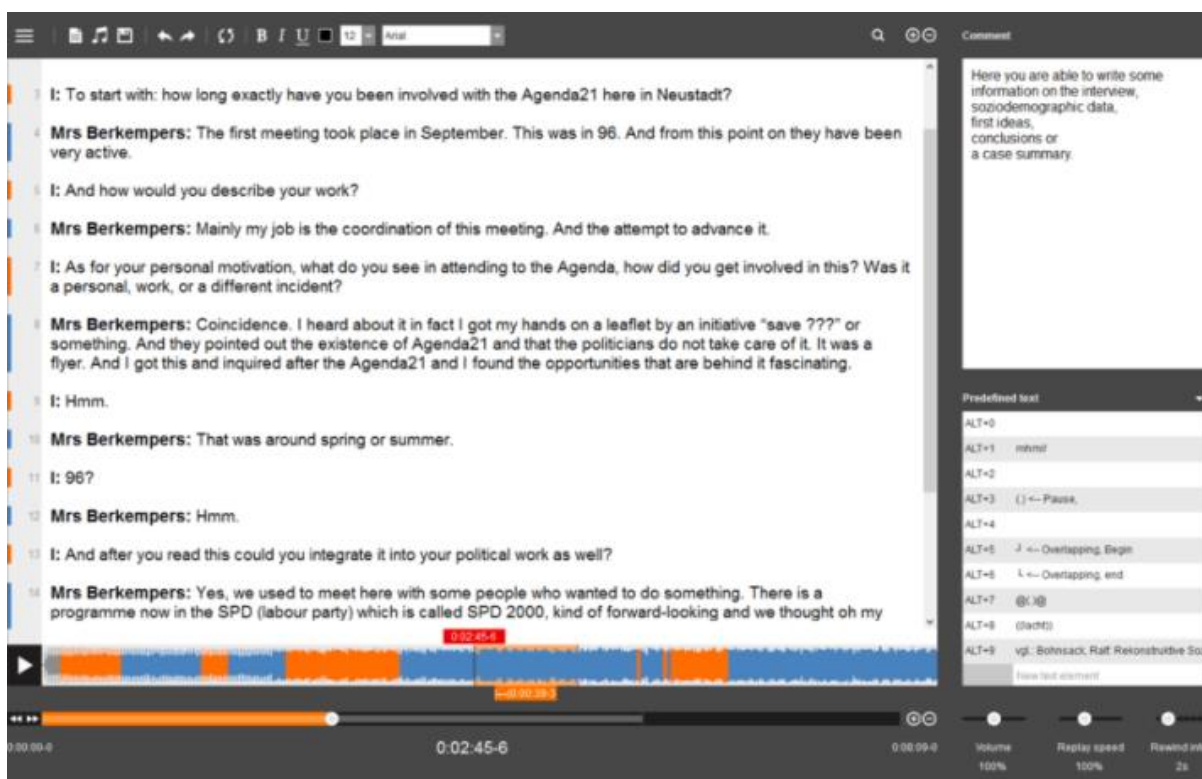


Figure 8: Audiotranskription User Interface

2.8 DESCRIPT

Descript is powered by Google Speech which is one of the most advanced transcription technology (“Descript”, n.d.). It allows users to transcribe within a short time of period. Charges are applied to proceed with the use of this software. Availability of this software is only for Mac users, and there is a free thirty minutes trial for first-time users.

Audio can be edited with the modification of transcribed text, and the transcript can be published through the webpage, and for other users to view. Automatic transcription is produced with the automatic speech recognition, giving up to ninety-five percent accuracy. Multiple speakers can be added, by including labels for different speakers. Navigation markers, highlights and inline notes are features for this software. Transcript can be exported with timestamps shown in the transcript. In addition, audios can be exported in different formats aside from the format that users upload.

Users could include comments, allowing other users to take note of the phrase highlighted. Rewind of playback is also supported. In the case where users were to scroll down the transcript to view at other sentences, the current sentence that the audio is playing is displayed at the bottom of the screen. It is user-friendly since users need not to refer back to the current sentence. However, there is no waveform visualization for users, and video files are not supported in this software. Windows users could not access this software, since it is limited to Mac users.

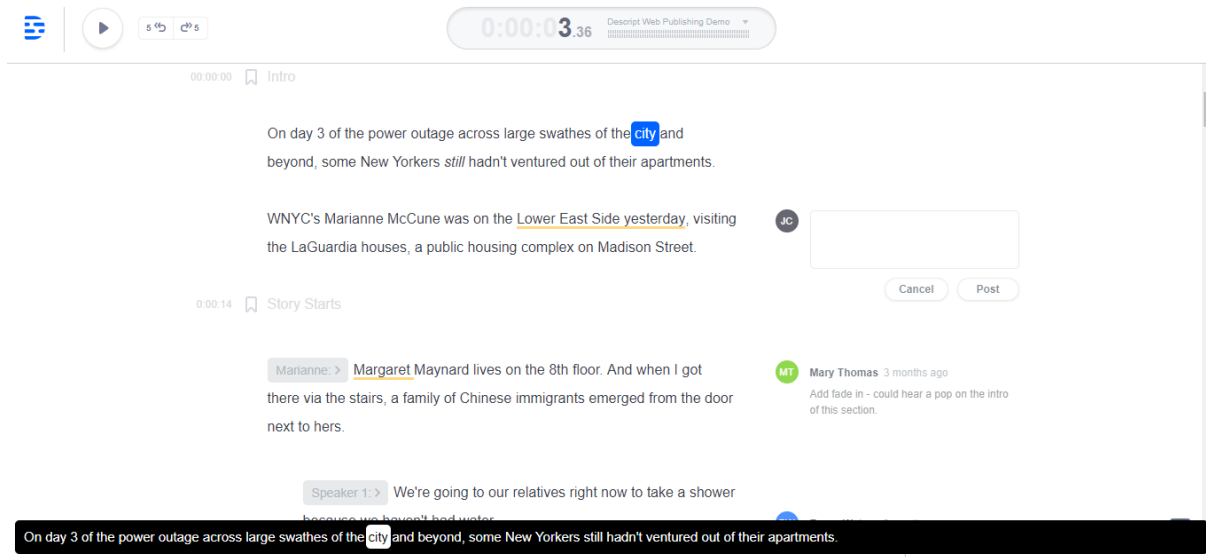


Figure 9: Descript User Interface

The following transcribing tools are web-based systems which do not require any downloads to access.

2.9 OTRANSCRIBE

oTranscribe is a free web-based application, that requires Internet to operate. Users need not download any software, which takes up their storage space in their computers. However, it is only available on desktop computers (Elliot, n.d.).

The layout of this application is simple, and not complicated to use. Users could upload the video or audio file they want, in the format that the application has specified. YouTube videos can also be uploaded with the inclusion of the URL link of the video. This provides convenience for novice users as well as experienced users, as they are not required to download the video from YouTube, hence not occupying their storage space in their computers.

For users who have used this application before, they could view their latest uploaded file to the system. Not only that, it also provides keyboard shortcuts for users, who are more comfortable in using the keys to control the playback and the typographical emphasis. This editor is available to anyone, and most importantly there is no cost attached to it.

However, users are required to transcribe themselves or copy the chunk of text that they have already transcribed previously into the textbox, since oTranscribe does not offer an upload function of the transcript. They could include the timestamp in their transcript, but users could not tell the exact timestamp of that particular word they wish to know.

Nonetheless, the application will save the transcript automatically to the storage of browser every second to prevent loss of data in case of any mishaps.

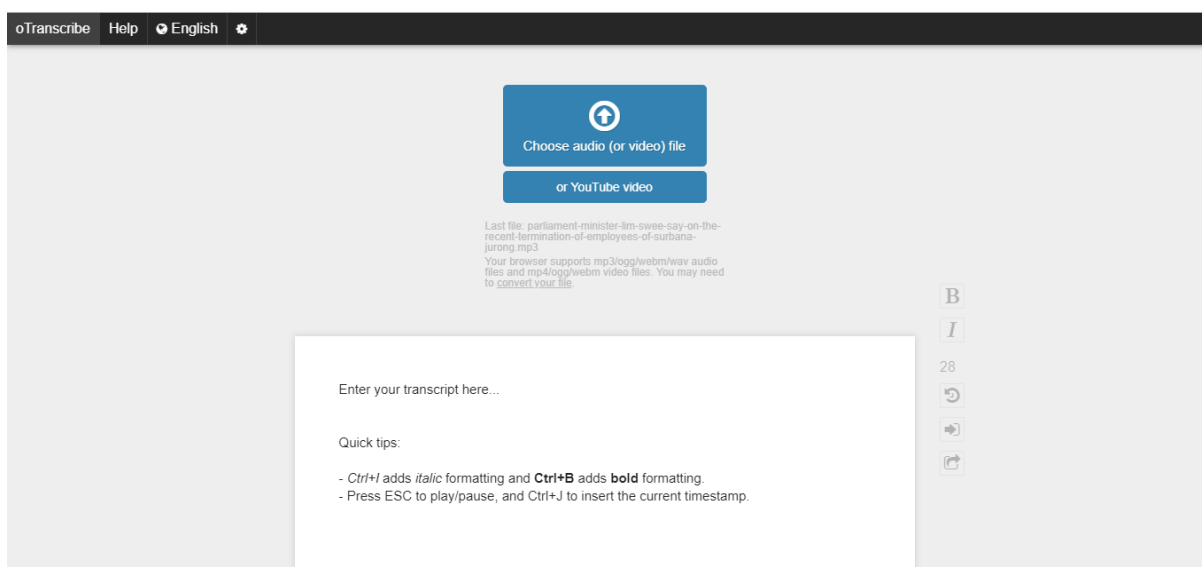


Figure 10: oTranscribe User Interface

2.10 TRANSCRIBE

Transcribe is a web-based transcribing tool, which does not require users to download any program. Different from oTranscribe, there is a cost to use this application. First-time users are entitled to use the application for free during their first week of usage (“Transcribe”, n.d.).

The layout of this application is very simplistic, where users need not required to read up the manual to understand the operation of the functions and features. The application provides a short tour for users, allowing them to know where the important functions and features are.

Users could start-off by uploading a video or audio file or YouTube video by inserting the URL link directly into the system.

Transcribe provides a wide range of features which assists in transcribing. Dictation is supported where the dictation engine will directly convert the speech of the video or audio file into text automatically. Users can then listen to the playback to ensure that the dictation is done accurately. It also supports dictation in multiple languages which is an advantage towards users from all different countries.

In the case where a particular word repeats multiple times throughout the whole playback, users can use the templates feature. Users can define acronyms for frequently used words and phrases that appear throughout the playback, so that transcription can be done as quickly as possible.

For example, users can indicate “John Smith” as “js”, by typing “js” in the textbox, “John Smith” will be displayed instead.

This application is equipped with features such as foot pedal and auto loop of the playback. Keyboard shortcuts are also provided at the side of the panel to aid users who are comfortable in using keys to transcribe. The transcribed text is auto-saved using the storage of the browser and Internet is not required to transcribe, it is only require to access the page.

Noevertheless, Transcribe does not provide the option to upload a transcript file to modify it and waveform of the audio file is not presented in the application.

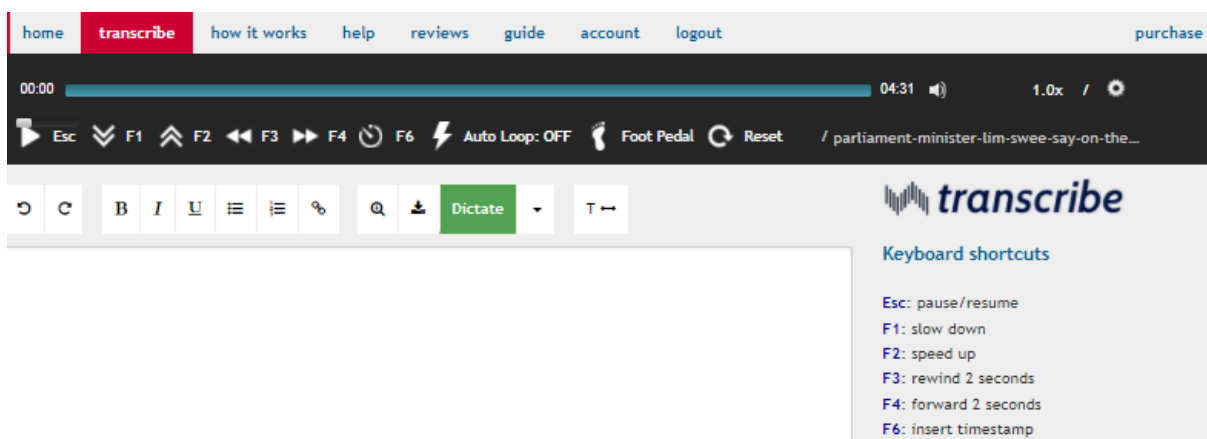


Figure 11: Transcribe User Interface

2.11 TRINT

Trint transcribes audio to text with the use of online transcription software. Speech-to-text technology is incorporated to the software, however high accuracy is not achievable due to the optimal conditions set by the system. With Trint, lesser time is required to achieve maximized accuracy for transcript (“Trint”, n.d.).

Trint allows users to upload video and audio files in the most frequently used formats. The software will automatically help transcribe the playback, reducing the time required to transcribe. Also, the visualization of waveform is displayed at the bottom of the screen. Search engine is incorporated into the system, giving users the benefit of searching certain words and replacing to the correct words.

Trint is accessible to multiple users, editing on the same transcript simultaneously, hence lessening the workload of each user and minimizing the time required on a transcript. Up to thirteen languages are available in transcribing in Trint, with the inclusion of three different English accents namely North American, British and Australia.

However, to use Trint, there is fifteen dollars charge every hour. Nevertheless, there is a free trial period for thirty minutes for users to familiarize with the software. If they feel that it does not suit their taste of operating the software, they could opt for other software instead.

Upon uploading the video or audio file, the system will transcribe automatically for users using the speech recognition technology. Since the accuracy is not fully attained, users can review the transcript by listening to the playback and modify the transcript. Users can add additional speakers, to the transcript where there is a drop-down menu for users to select.

Trint provides users features such as highlighting and striking a segment of words. Waveform is displayed at the bottom of the screen, and with the selection of words being highlighted, it will appear onto the waveform, same goes for the striking of words.

Users can choose to export the transcribed text or only the highlighted words in Microsoft Word in DOC format or editing decision lists in EDL format.

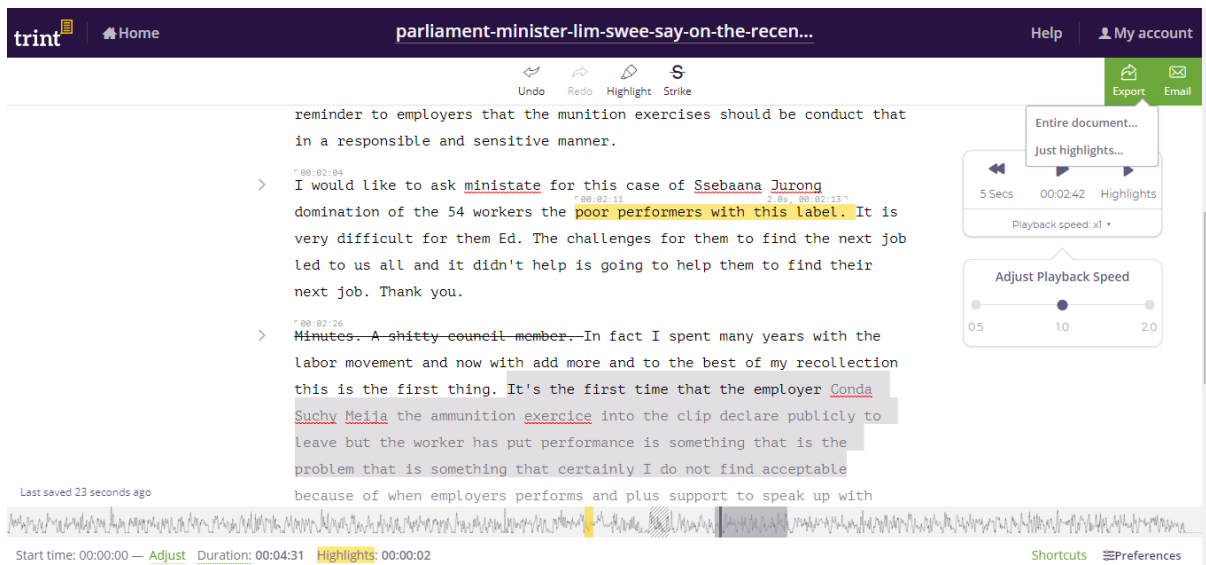


Figure 12: Trint User Interface

2.12 HAPPY SCRIBE

Happy Scribe is a web-based transcribing tool which could transcribe most types of video and audio files of different formats. Speech recognition technology is implemented to convert speech into text, and it can be done in a few minutes. Users can access the transcribed text and modify accordingly, reviewing the playback with the transcript. Happy Scribe has up to one hundred and twenty languages available for transcribing.

Furthermore, text alignment is automatically synchronized, causing the text and the playback to be aligned, enabling users to move from one paragraph to another with ease. Export of transcript is in TXT format, with or without timestamps, and in SRT format for captioning and subtitles (“Happy Scribe”, n.d.).

The layout of the application is displayed below, where users can move along the playback to listen the audio. Words are highlighted when the audio is spoken, knowing where the audio is currently at, at that moment. Users can modify the transcribed text by replacing the words with the correct words. Keyboard shortcuts are presented on the top of the transcript.

Happy Scribe is a simple and easy-to-use application, for users who are novice. The features are minimal, hence it is suitable for users who only wish to modify a small section of the transcribed text. However, there is only thirty minutes of free usage, and a fee will be charged if users would wish to continue using the application.

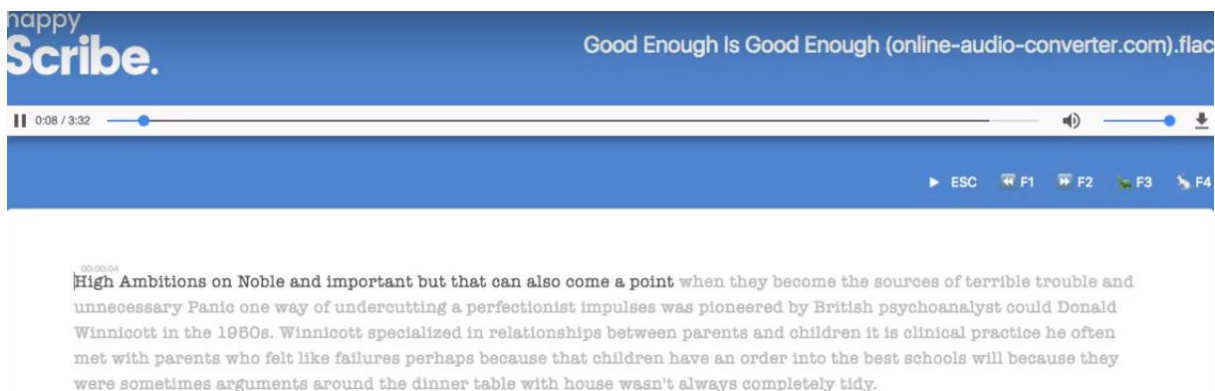


Figure 13: Happy Scribe User Interface

2.13 3PLAY MEDIA

3Play Media uses automatic speech recognition to translate speech into text to produce a rough gauge of transcript. It is time-synchronized and there is more than ninety-nine percent accuracy, even with complex content or poor audio quality.

This is relatively advantageous since most transcribing tools are unable to produce high accuracy due to the optimal conditions set by the speech recognition technology. Conditions must be ideal for the system to produce high accuracy with minimal errors. This application provides flexible turnaround options to ensure that users get their transcript within two hours. Large quantities of video can be processed simultaneously, reducing the time required to transcribe (“3Play Media”, n.d.).

However, users are required to click “Edit Word” to edit the transcript, and “Delete Word” for deletion of words. In most transcribing tools, users could directly edit the transcript instantly. For segments where the automatic speech recognition is unable to detect the word, users must click on the flag, where the transcript will display the segments that words could not be captured. Highlighting of words, increasing of playback’s speed are not supported.

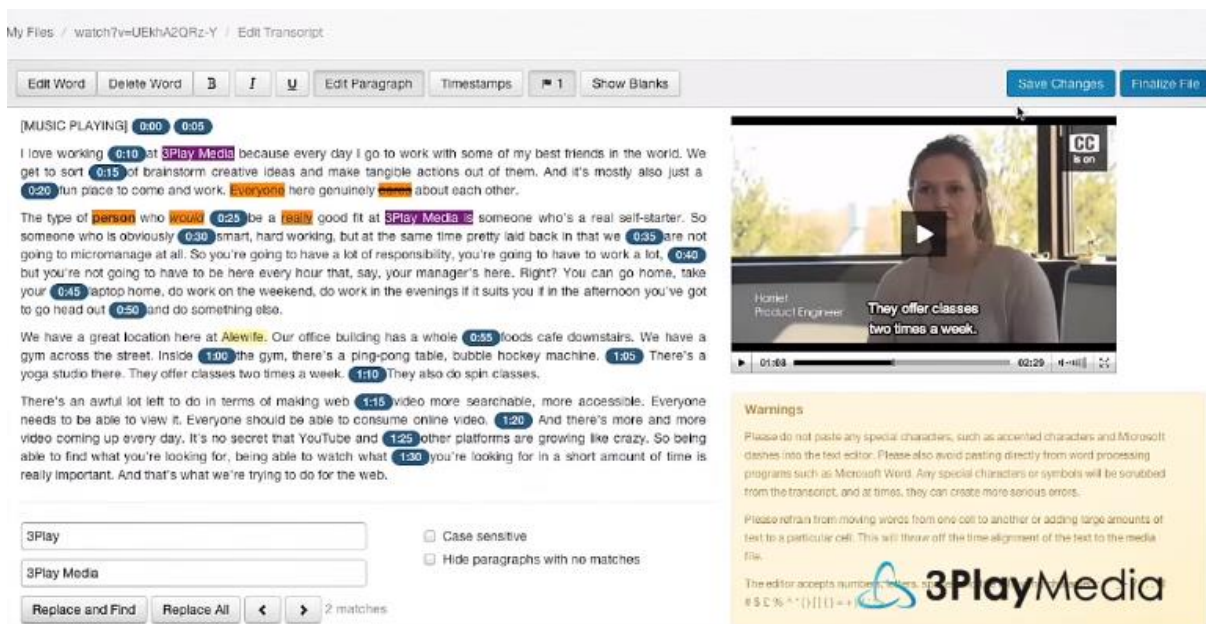


Figure 14: 3Play Media User Interface

2.14 SPEECHMATICS

Speechmatics is a web-based transcribing tool where users could choose either to transcribe a media file or get time codes for the media file and transcript. First-time users could access the application with no charge within sixty minutes. Subsequently, there is a fee associated if users wish to extend the period of use.

Speechmatics uses cloud transcription service, which provides actionable transcriptions within minutes, speeding up processes (“Speechmatics”, n.d.). Many languages are available and media files with multiple speakers will be indicated in the transcript. Commonly used file formats are supported.

Speechmatics also provides API where coders could incorporate their technology into the system and applications. With the use of this technology, users could upload multiple files in one go, and it could transcribe them simultaneously. In the API, confidence values, timing and speaker indications are supported.

Users could upload the media files, where the system converts the speech into text. Modifications could be done on the transcript. To get the timestamps for the instances, users could upload the text file and the media file to obtain the timestamp of the transcript.

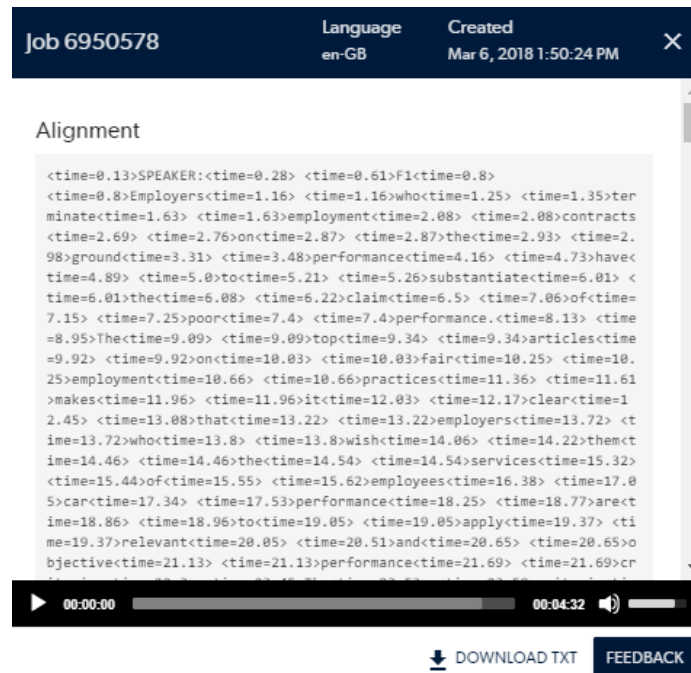


Figure 15: Speechmatics User Interface

Automatic speech recognition uses speech recognition technology, where speech is translated to text. In this technology, natural language processing is implemented, where machines process the words that humans speak into text. One example is Siri, which is widely used in iPhone devices.

2.15 OVERVIEW OF TRANSCRIBING TOOLS

Transcribing Tool	Software / Web-based	Editing of Transcript	Multiple Speakers	Waveform Visualization	Playback Speed
Praat	Software	✓	✓	✓	✗
ELAN	Software	✓	✓	✗	✓
EXMARaLDA Partitur Editor	Software	✓	✓	✓	✓
Subtitle Edit	Software	✓	✓	✓	✓
Express Scribe	Software	✓	✓	✗	✓
InqScribe	Software	✓	✗	✗	✓
Audiotranskription	Software	✓	✗	✗	✓
Descript	Software	✓	✗	✗	✗
oTranscribe	Web-based	✓	✗	✗	✗
Transcribe	Web-based	✓	✗	✗	✓
Trint	Web-based	✓	✗	✓	✓
Happy Scribe	Web-based	✓	✗	✗	✗
3Play Media	Web-based	✓	✓	✓	✗
Speechmatics	Web-based	✓	✗	✗	✗

3. PROPOSED APPROACH AND SYSTEM SPECIFICATION

In this chapter, we are focusing on the technologies used in our system, as well as the approach adopted to complete the requirements. Automatic speech recognition technology is integrated into this system, allowing speech from audio playbacks to be converted into transcript in text. This provides users the convenience in reading the transcript throughout the playback and making changes to the existing transcript. Software development life cycle is also adopted to ensure every process is completed for planning, creating, testing and deploying of system.

3.1 AUTOMATIC SPEECH RECOGNITION

Automatic speech recognition uses speech recognition technology, where speech is translated to text. In this technology, natural language processing is implemented, where machines process the words that humans speak into text. One example is Siri, which is widely used in iPhone devices.

With the implementation of speech recognition technology, it reduces the hassle of typing out words manually. However, full accuracy would not be attained if ideal conditions are not met. The conditions of the video or audio file must fit the conditions of the technology to produce high accuracy.

By importing the video or audio file, the automatic speech recognition software could analyze the words spoken, where waveforms are created. Surrounding noise will be reduced and removed, thus cleaning the wave file to produce a better output. With the training data that the software has analyzed, it tallies with the statistical probability to deduce the words that are spoken, forming a whole sentence. This is done to the whole video or audio (Matthew, 2014).

3.2 SOFTWARE DEVELOPMENT LIFE CYCLE

Software development life cycle (SDLC) is required where tasks are performed at each phase in the software development process. It is a common structure, consisting of a well-planned framework, developing, and maintaining the subtitle editor. The life cycle defines a methodology for improving the quality of software as well as the overall software development process (“Software Development Life Cycle”, n.d.).

Agile software development life cycle is commonly used in software development process, where it is based on the iterative and incremental process models, and focuses mainly on adaptability, ensuring the functional requirements and quality of the system are maintained, enhancing customer satisfaction (“Agile Software Development Life Cycle”, n.d.). Agile methods focus solely on breaking down the entire project into shorter phases, to develop the full system successfully.

By incorporating the Agile software development life cycle, the main objective is to develop the system, ensuring that it is bug free, meeting the standards of the requirements and quality of the system. With the opinions of users, the system produced must be satisfiable, user-friendly and robust in the eyes of users.

3.3 WEB APPLICATION STRATEGY

Web application technology have been improving throughout the years, and HTML5 is one of the technology that is implemented in the Offline Web Subtitle Editor. HTML5 has many syntactical features that could be integrated in the system. Multimedia and graphical content to web become relatively effortless to be incorporated without the use of flash and third-party plugins (“HTML5”, n.d.).

HTML5 is used in structuring and documenting the content in the World Wide Web, with the use of its markup language where users could view the content of the client-side browser. Markup language is a modern architecture used for annotating a document that is differentiable from text, easily comprehend by web browsers. The Internet markup language is the HTML5 core that could support multimedia and graphical content. With the integration of HTML5, the Internet bandwidth is conserved (“Markup language”, n.d.).

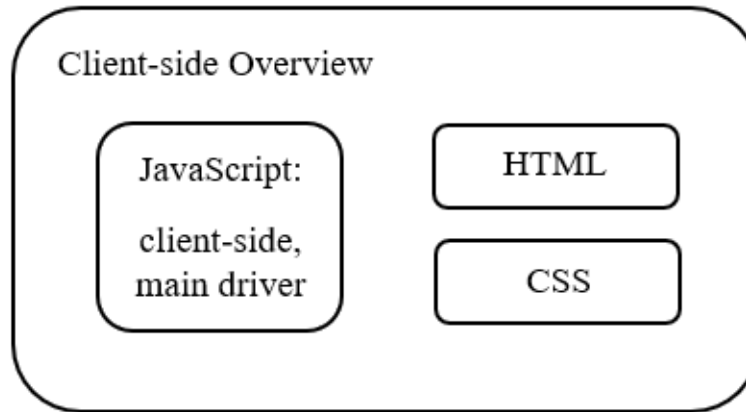


Figure 16: Client-side Overview

PHP is used in the server-side environment, where it runs its scripts before the HTML is loaded.

3.4 JAVASCRIPT

Client-side environment used to run script is commonly known as a browser. Client-side development is done in JavaScript. It enables the formatting of a webpage, making it synchronized and HTML and CSS are incorporated with JavaScript, beautifying the webpage, as well as adding more necessary features. JavaScript is used in achieving interactive webpages, and communication.

With the use of JavaScript, it acts as the linkage between the web application and the server-side environment. Various types of communication in the client-side architecture is achieved. To produce a web-based subtitle editor, dynamic HTML (DHTML) is integrated. The use of Ajax allows the loading of new page and submitting data to the server without reloading the webpage. JavaScript could also perform HTTP methods.

Since the system is a web-based system, JavaScript plugins such as bootstrap, jQuery, wavesurfer etc. are integrated in to the Offline Web Subtitle Editor which are discussed further in the later part of this chapter.

3.5 ARCHITECTURE DESIGN

Architecture design is used, consisting of replaceable and self-contained assembly of components, thereby aiding the process of implementation and future maintenance. Model-View-Controller (MVC) is integrated into the system architecture, where it divides the large aggregation of components into units, ensuring loose coupling and high internal cohesion between components and layers.

Modules within each layer are grouped according to their behaviour and purpose. Modules in the controller layer ensure that communication is affected through view and model layers, demonstrating the concept of abstraction in the architecture design.

With the integration of MVC, the architecture design also supports adaptation to changing end-user requirements which is highly modifiable and flexible. The benefit of using MVC architecture is that new components could be added in the respective layer without creating interference to the other layers. Using this environment, obsolete components can be removed, added, and updated.

MVC is commonly selected as the design pattern for architecture, due to the achievement of conceptual integrity with a uniform application of limited design forms. With the integration of this layered architecture, it is most suited with the system's requirements as well as easier modifiability and assemble of different components. Through this approach, it ensures that the modules are independent and parallel to the system, where the system are not required to compromise under the circumstance that required upgrade, removal or addition of a module, hence offering reliability in the Offline Web Subtitle Editor.

3.5.1 Visual Notation

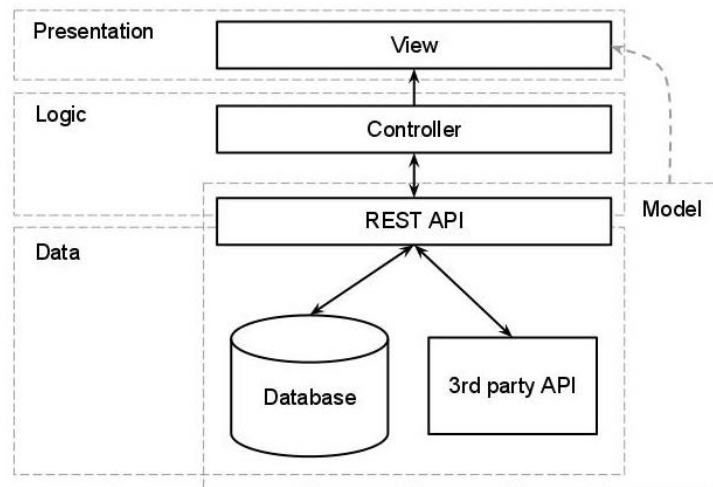


Figure 17: MVC Architecture Overview

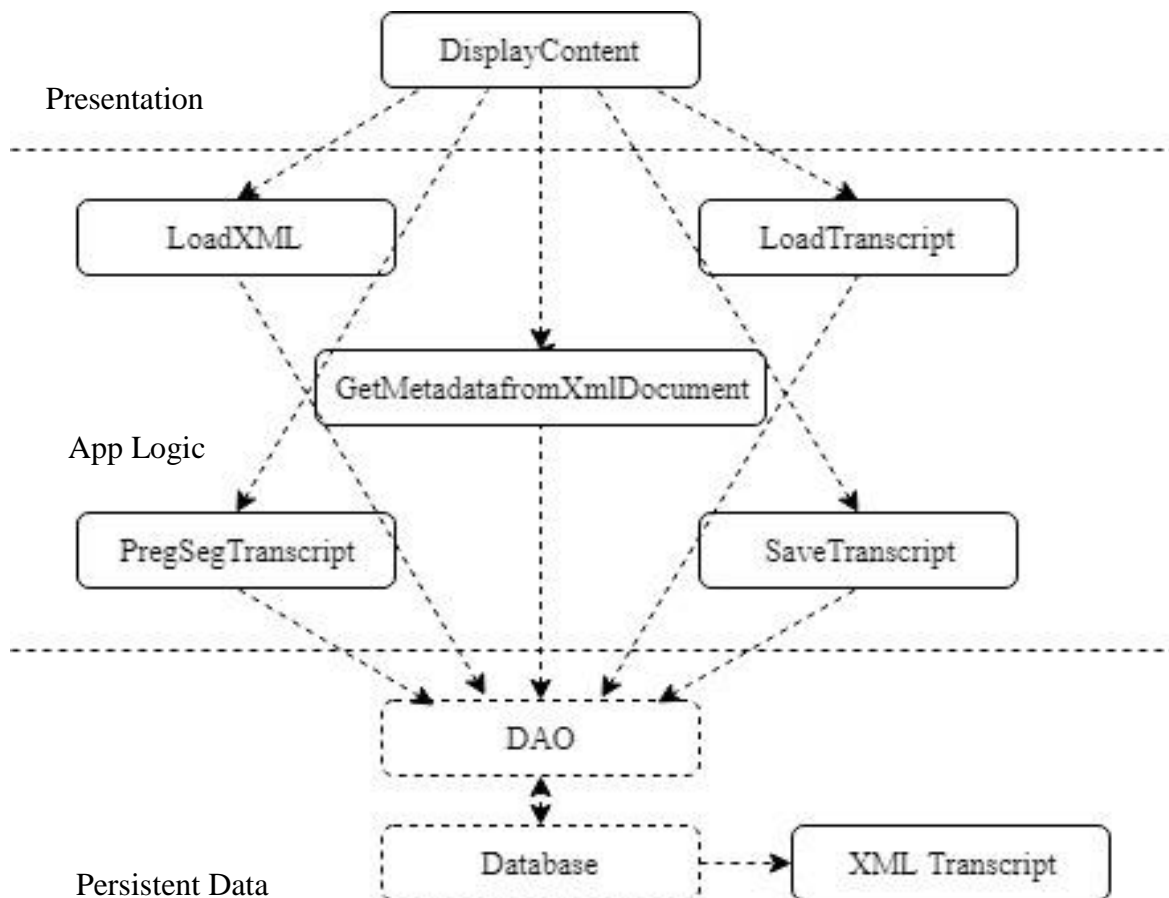


Figure 18: Architecture of Application

3.6 DATA FLOW DIAGRAM

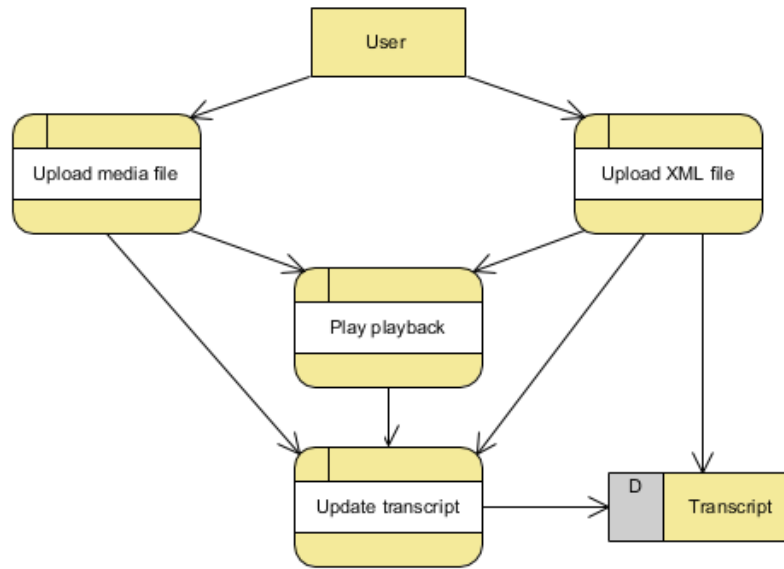


Figure 19: Data Flow Diagram

3.7 ACTIVITY DIAGRAM

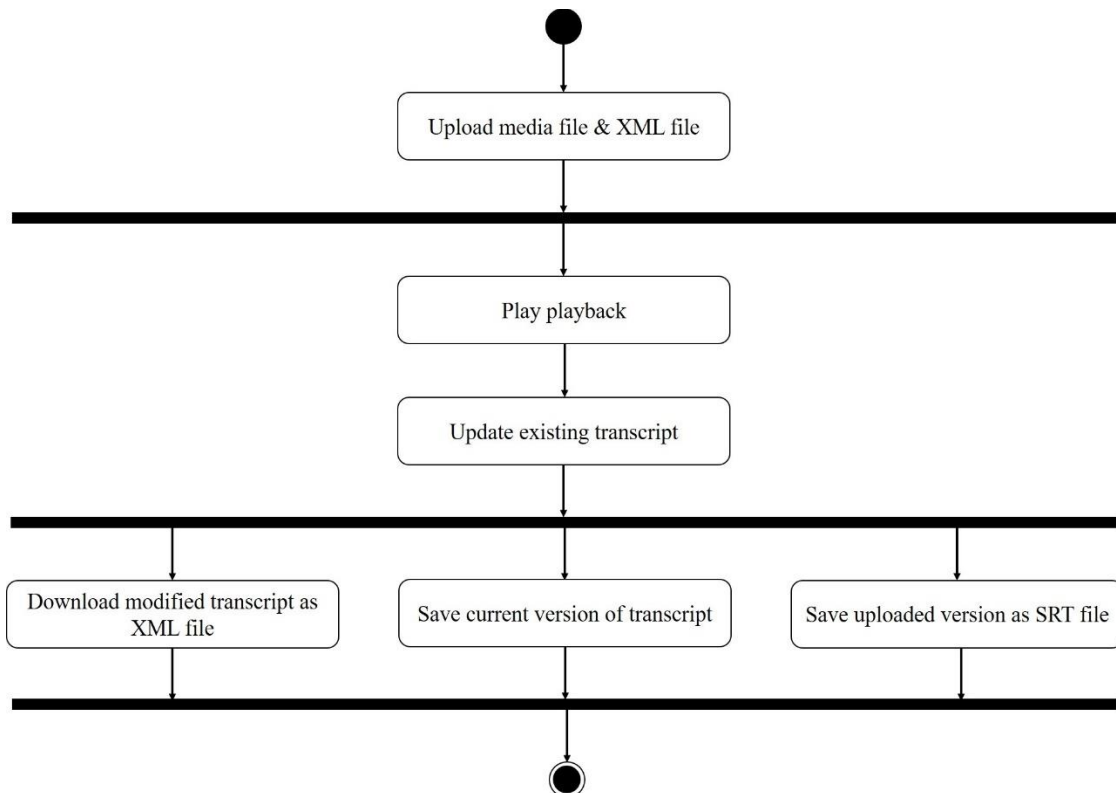


Figure 20: Activity Diagram

3.8 STRUCTURE OF TRANSCRIPT

Transcript is imported in XML format, whereas for export of transcript, it exists in either XML or SRT format.

For transcript in XML format, the file starts off with the xml version of the transcript. The document date and name are then continued, followed by the content of the transcript. In the content of the transcript, the segment indicates one sentence of the transcript where the start and end time are recorded. The sentence of the transcript includes each word in that sentence with their corresponding start and end time. This is done for the remaining of the transcribed text. Metadata includes the media name of either the video or audio file, along with the duration of the media file.

```
<?xml version="1.0" encoding="utf-8" ?>

<document date="03/03/2017" name="parliament-minister-lim-swee-say-on-the-recent-termination-of-employees-of-surbana-jurong">
  <content>
    <segment endTime="8.54" id="0" spkName="0" startTime="0.05">
      <sentence endTime="8.54" id="0" spkName="0" startTime="0.05" confidence="0">
        <word endTime="1.07" id="0" spkName="0" startTime="0.61">employers</word>
        <word endTime="1.22" id="1" spkName="0" startTime="1.07">who</word>
        <word endTime="1.61" id="2" spkName="0" startTime="1.22">terminate</word>
        <word endTime="2.06" id="3" spkName="0" startTime="1.61">employment</word>
        <word endTime="2.67" id="4" spkName="0" startTime="2.06">contracts</word>
        <word endTime="2.85" id="5" spkName="0" startTime="2.70">on</word>
        <word endTime="2.93" id="6" spkName="0" startTime="2.85">the</word>
        <word endTime="3.2" id="7" spkName="0" startTime="2.93">ground</word>
        <word endTime="3.33" id="8" spkName="0" startTime="3.20">of</word>
        <word endTime="3.48" id="9" spkName="0" startTime="3.33">poor</word>
        <word endTime="4.24" id="10" spkName="0" startTime="3.48">performance</word>
        <word endTime="4.96" id="11" spkName="0" startTime="4.61">have</word>
        <word endTime="5.24" id="12" spkName="0" startTime="4.96">to</word>
        <word endTime="5.99" id="13" spkName="0" startTime="5.24">substantiate</word>
        <word endTime="6.15" id="14" spkName="0" startTime="5.99">your</word>
        <word endTime="6.58" id="15" spkName="0" startTime="6.15">claim</word>
        <word endTime="7.17" id="16" spkName="0" startTime="6.98">of</word>
        <word endTime="7.37" id="17" spkName="0" startTime="7.17">poor</word>
        <word endTime="8.21" id="18" spkName="0" startTime="7.37">performance</word>
      </sentence>
    </segment>
  </content>
  <metadata>
    <media name="parliament-minister-lim-swee-say-on-the-recent-termination-of-employees-of-surbana-jurong.mp3" duration="00:04:31"/>
    <speakers>
      <speaker id="0" name="S1" />
    </speakers>
  </metadata>
</document>
```

Figure 21: Transcript in XML format

For transcript in SRT format, each segment is numbered before the time of that sentence. The start and end time of each sentence are displayed, followed by the sentence of the transcript. This is continued for the rest of the sentences with their corresponding timestamps, with numbering being incremented.

```

1
00:00:00,050 -----> 00:00:00,540
employers who terminate employment contracts on the ground of poor performance have to substantiate your claim of poor performance
2
00:00:00,540 -----> 00:00:02,840
the tropicalised from fair employment practices makes it clear that employers who wish to terminate the services of employees on the ground of poor performance are to apply relevance and objective performance criteria
3
00:00:02,840 -----> 00:00:40,860
the criteria should be made known to all employees employers should also keep records of the employees performance and the decision to terminate the service of an employee to be based on documented poor performance the key word is documented
4
00:00:40,860 -----> 00:00:46,480
we involves a unionised employee the union should also be consulted
5
00:00:46,480 -----> 00:01:02,760
if the employee fast and appeal of unfair dismissal to m o m will first mediate should mediation fail will conduct an inquiry and require the employer to show costs and produce evidence to justify the termination
6
00:01:02,760 -----> 00:01:10,890
this remedy is provided under the employment act as well as the industrialization revision act for union members
7
00:01:10,890 -----> 00:01:23,890
if an employer is unable to substantiate his claim that the employees performances poor he employed maybe ordered to reinstate the employee or to provide compensation
8
00:01:23,890 -----> 00:01:41,580
if the employer does not comply with the order he can be prosecuted in the case of suhana jurong recent exercise to terminate the services of fifty four employees the company has acknowledge that the process could have been better managed
9
00:01:41,580 -----> 00:01:52,270
ah the management and unions have since reached an agreement on an extra gratia payment which in our view is a fair outcome for the affected employees
10
00:01:52,270 -----> 00:02:04,110
this episode serves as a good reminder to employers that termination exercises should be conducted in a responsible and sensitive manner
11
00:02:04,110 -----> 00:02:17,930
i would like to ask minister for this case of suhana jurong ah termination of the fifty four workers they are labelled poor performance with this label it is very difficult for them or additional challenges for them to find the next job
12
00:02:17,930 -----> 00:02:24,090
i would like to ask m o m if they need help is m o m going to help them to find their next job thank you
13
00:02:25,120 -----> 00:02:37,050
to have a mental speaker i i i share the concern of a a member of in fact a a i spend a many years and with the label movement and now with m o m
14
00:02:37,050 -----> 00:02:51,470
ah to the best of my recollection this is a first time is the first time that the employer conducts such a major termination exercise and to declare announce publicly to label at the workers is poor performance
15
00:02:51,470 -----> 00:02:58,540
i think is something that the as the as a manpower minister is something that the certainly i do not find it as acceptable

```

Figure 22: Transcript in SRT format

3.9 COMPONENT SPECIFICATION

3.9.1 cleanUploads

Storage space in computer is minimized due to the implementation of this function where media files with their corresponding XML files are removed.

3.9.2 loadXML

Transcript in XML format uses simplexml to create an instance, where the input is the formatted XML file, and the output is the instance of the document.

3.9.3 getMetadataFromXmlDocument

Metadata is produced from the transcript by retrieving relevant information such as word, sentence etc from the XML file uploaded.

3.9.4 loadTranscript

Attributes in the transcript are loaded into an array. Initial transcript may consist of sentences with more than thirty words, making it difficult for users to focus. The breaking down of sentences is done in this function where a limit of thirty sentences is set for each segment.

3.9.5 preSegTranscript

The information retrieved from loadXML is parsed in this function, along with the speaker information. Speaker name is shortened, and different speakers that are initialized in the transcript are processed in this function.

3.9.6 saveTranscriptXML

Transcript can be exported in XML format either by updating the uploaded version to the newer version or downloading the latest version in to the “Downloads” folder.

3.9.7 saveTranscriptSRT

Transcript can be exported in SRT format since it is a common format for transcript where it can be accessed to other software.

The files listed are the json files that store simple data structures and objects. They are incorporated in the main function, containing data in a standard data interchange format, which is lightweight, text-based and human-readable. They are normally plugins which can be implemented in the system.

3.9.8 Bootstrap

With the use of bootstrap, the webpage layout and the information box of the media file are created. Bootstrap plugins are incorporated into the system, solely through the makeup API without implementing them into the JavaScript. It saves up time taken, in implementing every function manually. Bootstrap predefined design templates and classes could directly be written in and works well with HTML and CSS. With the functions in bootstrap, users could choose the required methods to implement, and ignore the rest. Furthermore, since this Offline Web Subtitle Editor requires the use of web browsers, bootstrap is compatible with all modern browsers and Internet Explorer versions. With the grid system that it provides, users could manoeuvre through columns, controlling the content for desktop users (Diana, 2015).

3.9.9 jQuery-UI

jQuery UI is used for user interface interactions, effects, widgets, and themes that are built on top of the jQuery JavaScript library (“jQuery UI”, n.d.). It is necessary for building a web browser since interactions throughout the web browser is compulsory. Interactions such as dragging, dropping, resizing, selecting and sorting are supported. Widgets such as autocomplete, button, progressbar etc. are included. By integrating this, users are not required to manually start from scratch.

3.9.10 Magor

Magor player is used to support advanced video controls, to control the playback. The media file is synchronized with the video controls, when the media and XML files are uploaded into the system.

3.9.11 wavesurfer

Wavesurfer plugin is necessary to create waveforms for media files. It is a customizable audio waveform visualization, built on top of Web Audio API and HTML5 Canvas. It is integrated

into the system, for users to visualize the waveform of the media file, allowing them to view the amplitude and pitch (“Wavesurfer Json”, n.d.).

4. OVERVIEW OF SUBTITLE EDITOR

In this chapter, an overview of the Offline Web Subtitle Editor is focused, with clear explanations on the implementations taken. Illustrations are displayed with the features of this application on the existing work done to the application. Layout of the Offline Web Subtitle Editor is displayed, along with the functions of the system. Detailed description of the features are explained as shown below.

4.1 UPLOAD MEDIA FILES

The layout of the Offline Web Subtitle Editor is displayed below. Users could upload media files which include video or audio files, and their corresponding XML transcript file. There is a limitation in size for these files. If users do not have the transcript file, they could use other software where automatic speech recognition technology is used to translate into text and upload the XML file into this editor.

Users could choose either to “Choose File” or drag and drop the relevant file into the box as displayed. This simplifies the process when users have their folder of media files opened. They need not go through the process of searching and navigate over the folders to find the media files that they wish to upload.

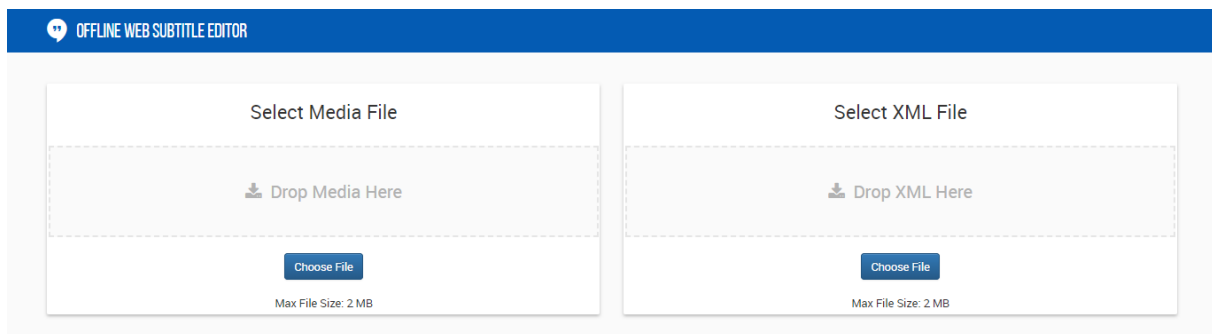


Figure 23: Offline Web Subtitle Editor Mainpage

4.2 LAYOUT OF EDITOR

Upon uploading of the media and XML files, the editor will automatically display the following webpage, where the transcript is displayed in rows of sentences. The transcript is shown where the top sentence is the first sentence of the playback, and the bottom sentence is the last sentence of the playback.

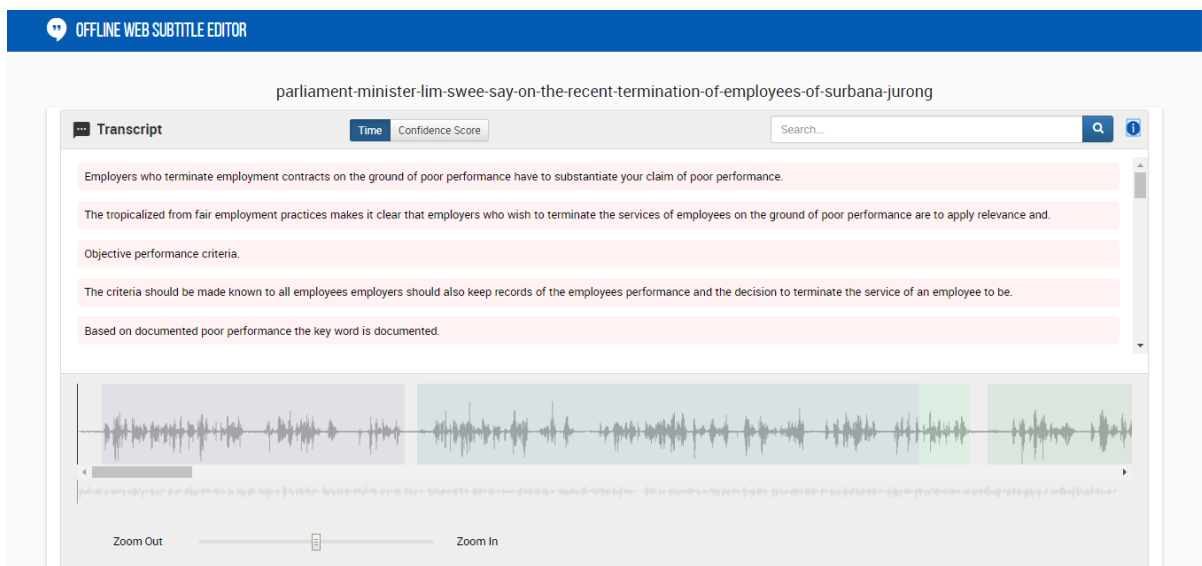


Figure 24: Offline Web Subtitle Editor User Interface

4.3 INFORMATION OF MEDIA FILE

The information of the media file is shown by clicking on the small icon besides the search engine. Users could input a description of the media file and update the name of speaker(s). The format of the media file is stated as the content type along with the title of the media file.

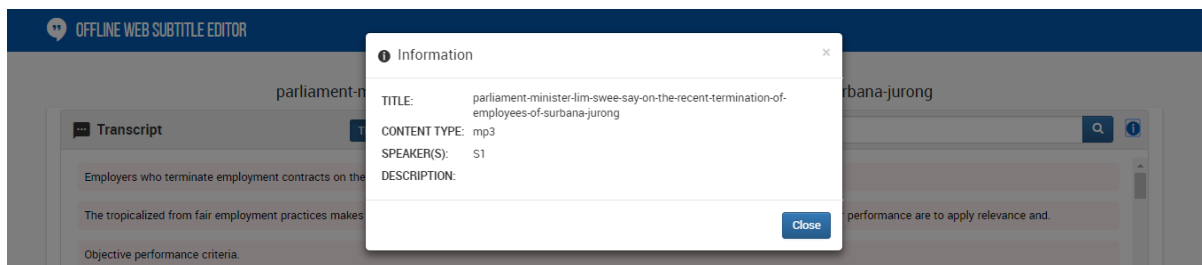


Figure 25: Information Panel

4.4 WAVEFORM PANEL

The waveform is displayed where users can scroll from left to right to select the playback they wish to hear. A zoom function is implemented to allow users to zoom in to enlarge the waveform and zoom out to minimize the waveform to view the full waveform of the media file. The colour segmentation in the waveform panel presents the different sentences of the transcript. Given that the waveform is not highlighted, this indicates that no speaker is currently talking at the moment and there is pause in the playback.

The waveform is produced with the help of a plugin, wavesurfer.js. Users could control the timestamp of each sentence by adjusting the starting and ending time so that it is aligned to the transcribed text.



Figure 26: Waveform Visualization Panel

4.5 MEDIA ICONS

Media icons are used to control the playback of the media file. There are several icons which aid users to control the video or audio. After importing the video or audio file, along with the XML transcript file, the transcript, waveform, and a textbox to modify the transcript can be seen. Users could play the video or audio by clicking on the play icon (play icon will become pause icon); pause the video or audio by clicking on the pause icon (pause icon will become play icon); go to the previous sentence by clicking on the skip previous icon; forward to the next sentence by clicking on the skip forward icon; repeat the highlighted words as many as they want by clicking on the rewind icon; increase or reduce the number of times the speaker repeats the highlighted words that is set; adjust the speed of the video or audio by pressing the down icon to slow down or up icon to increase speed; set the volume of the audio.

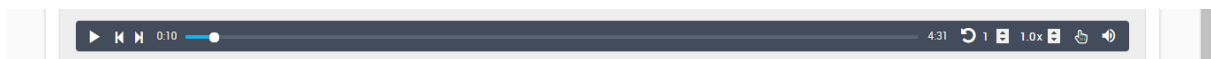


Figure 27: Controls of Media File

4.6 TEXTBOX FOR EDITING

The textbox displayed is for the editing of transcribed text. Users could modify that sentence by clicking on the sentence in the transcript box, and the sentence is displayed in this textbox. Upon editing the transcript, the updated sentence will be updated simultaneously in the transcript box.

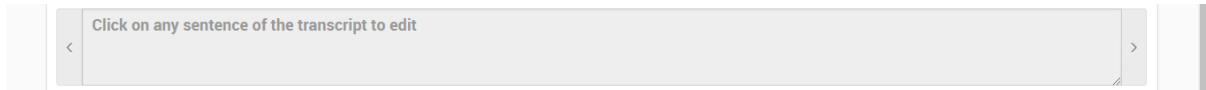


Figure 28: Textbox for Editing

4.7 SHORTCUT KEYS

Users who think that it is a hassle to use a mouse, could opt for shortcut keys where it presents a faster procedure in modifying the transcript. With the use of a mouse or keyboards, it attains the same functions. For users who are unaccustomed to the functions of these shortcut keys, by clicking on the button, they could understand and view their functions when editing the transcript.

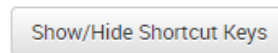


Figure 29: Keyboard Shortcut Button

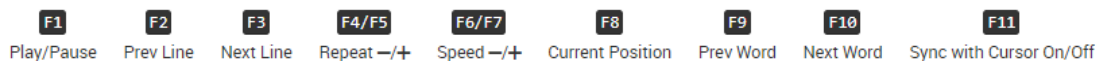


Figure 30: Keyboard Shortcuts

4.8 SAVE TRANSCRIPT

Users could save the modified transcript, by clicking on “Save Updated XML” to download the latest version of transcript. They could also modify the version that they upload to a newer version, overwriting the older version, by clicking on “Save Transcript”. Since most transcripts usually exist in SRT format, users could download the transcript in SRT format that can be used in other software.



Figure 31: Save Transcript Buttons

4.9 AUTO-SCROLLING

The transcript can be viewed in this manner where it is visualized in a clearer font as compared to a chunk of words which gives rise to the comfort of viewing. Features such as auto-scrolling are included in which the sentence that is highlighted in red displays the segment which the speaker is currently at. When the media file is playing, the transcript will scroll in the upward manner automatically, enabling users to view the next highlighted sentence. The figures below are shown in the sense that the transcript will scroll upward to the next sentence.

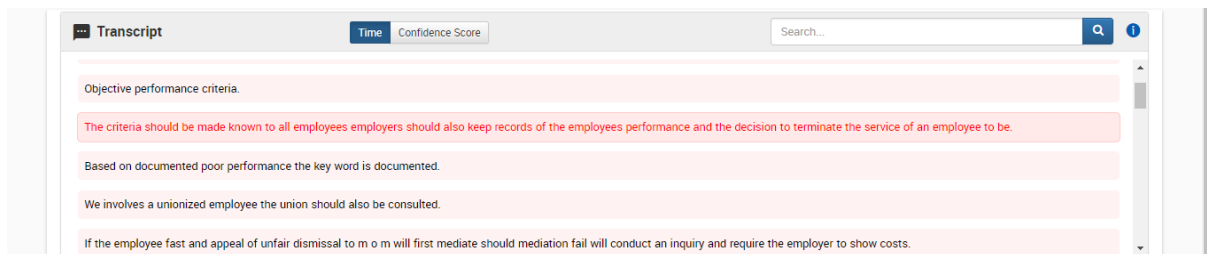


Figure 32: Transcript Panel at First Instance

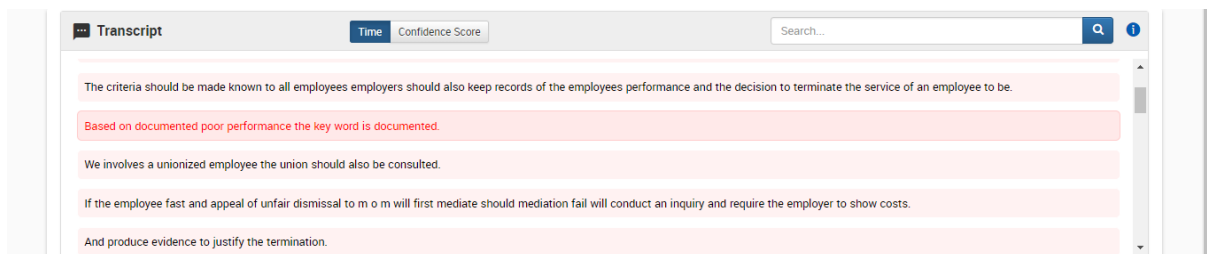


Figure 33: Transcript Panel at Next Instance

4.10 UPDATE TRANSCRIPT SPONTANEOUSLY

Users could modify the transcript immediately when they spot some mistake in the transcript as the speech recognition technology does not give full accuracy. By clicking on the sentence where the user would want to modify, editing is done, and it will be updated instantly on the transcript display screen.

4.11 HIGHLIGHTED WORDS FOR VIEWING

The sentence in which the speaker is currently at is shown in the textbox for editing purposes. Words in that sentence are highlighted in yellow, providing high focus for users. Users need not to listen so attentively to the audio to determine the placement of words.

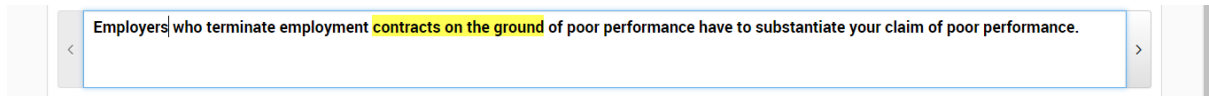


Figure 34: Highlighted Words

4.12 NUMBER OF WORDS BEING HIGHLIGHTED

Users could increase or reduce the number of words in which they are highlighted in yellow. The cursor is shown in the editing textbox, enabling users to modify the transcript as and when they want. By increasing the number of words, users could know the boundary of the words that the speaker is currently speaking.

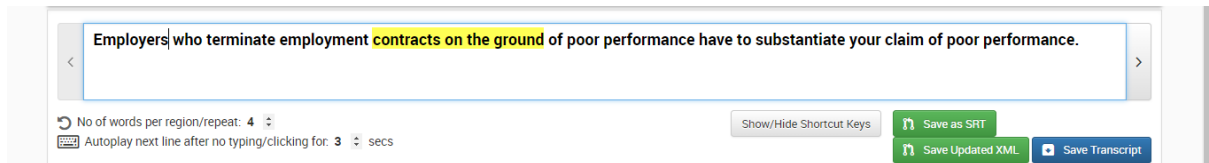


Figure 35: Number of Words Highlighted

The words that are currently highlighted, will progress to the next corresponding four words that will be highlighted. For instance, when the whole sentence is being traversed to the end, the next sentence is shown on the textbox, with the first four words being highlighted.

5. IMPROVEMENTS MADE TO SUBTITLE EDITOR

This chapter focuses on the improvements done to the Offline Web Subtitle Editor to provide a more conducive environment for users to utilize this application. Some functions may not perform efficiently and require modifications to provide users with the most desirable manner. With these features being incorporated, it provides an edge over other transcribing tools.

5.1 SEARCH BOX

Search engine is included, to allow users to find specific words with ease throughout the whole transcript. Users could modify all related words all at once.

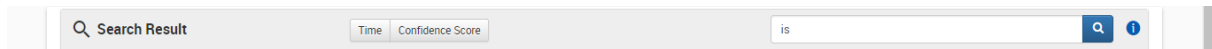


Figure 36: Search Box

5.2 CONFIDENCE SCORE

Upon entering the relevant word into the search box, it traverses throughout the whole transcript to match the characters of that word. The occurrences of that relevant word are presented in the transcript display box, indicating the number of occurrences that the word appears in that sentence as shown below in the figure. If there is no occurrence of the searched word, it returns the whole transcript, similarly to when the media and XML files are uploaded.

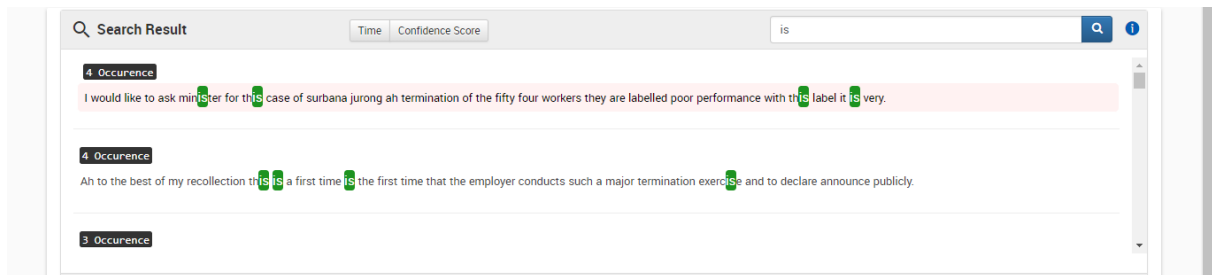


Figure 37: Confidence Score

5.3 ADJUSTMENTS TO TRANSCRIPT DISPLAY SCREEN

In order to make good use of the space around the transcript screen, the display of the transcript is expanded to make full use of the empty space outside the boundary.



Figure 38: Initial Layout of Mainpage

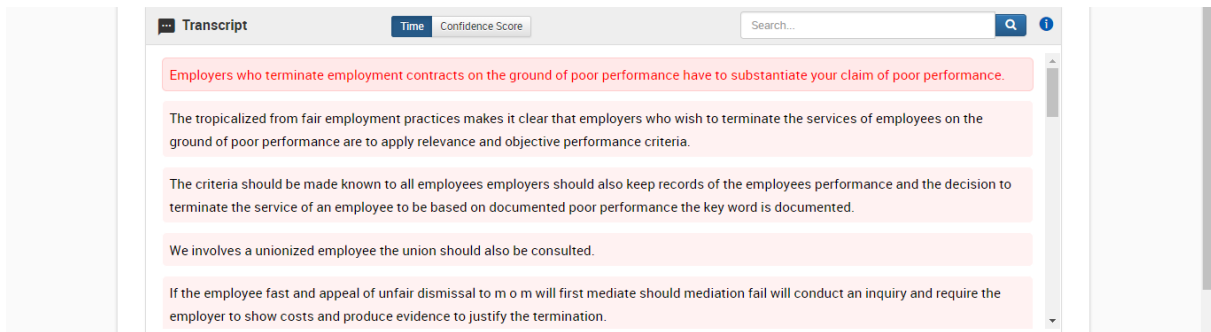


Figure 39: Initial Layout of Transcript

The figure above is the initial state where the sides have some empty spaces, whereas the final state is displayed below where the empty spaces are minimized.

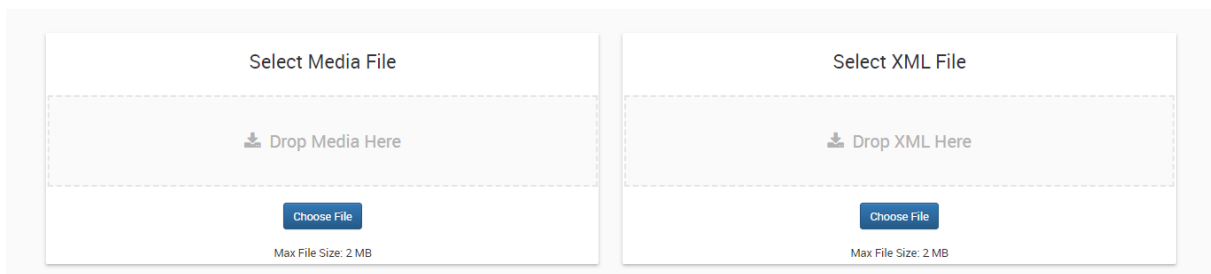


Figure 40: Modified Layout of Mainpage

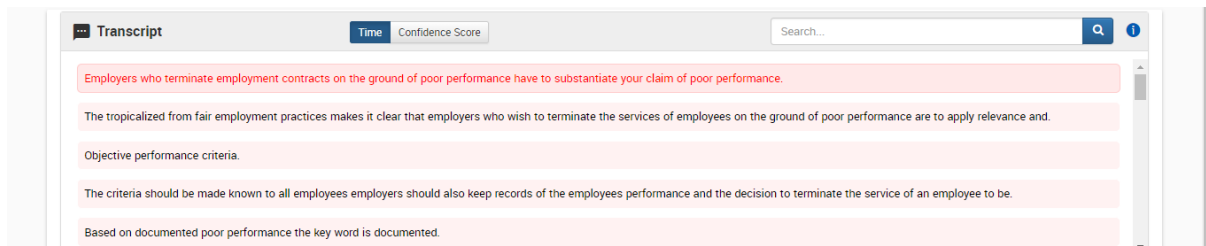


Figure 41: Modified Layout of Transcript

On top of that, the number of sentences in the transcript screen is limited to five sentences, bringing users' focus on to the screen than having more sentences in the screen, making it seem confined.

5.4 SPLITTING SENTENCES

Some sentences are spoken without pause in one setting. There is difficulty for users to concentrate on the current word that the speaker is at, if they are being distracted when the audio is playing. Although words are highlighted in yellow background to aid them in focusing, by staring at a long sentence may cause users hard to focus well.

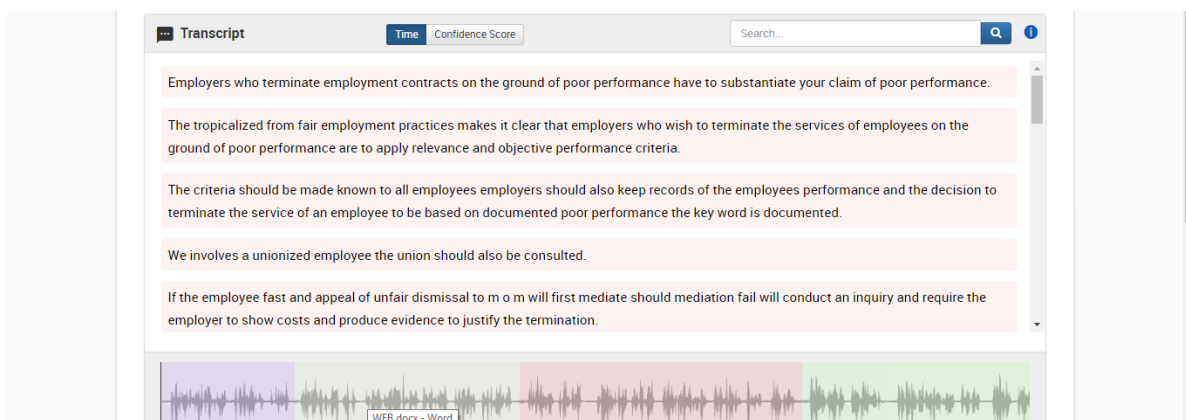


Figure 42: Initial Layout of Sentence Structure

As shown in the figure above, the second sentence from the top is slightly lengthy which is difficult for users to keep track at the word that the speaker is currently at. With the modification done to limit the number of words in one sentence, the limitation is done to at most thirty words in one sentence, users can view the sentence comfortably without having to strain their eyes on the words.

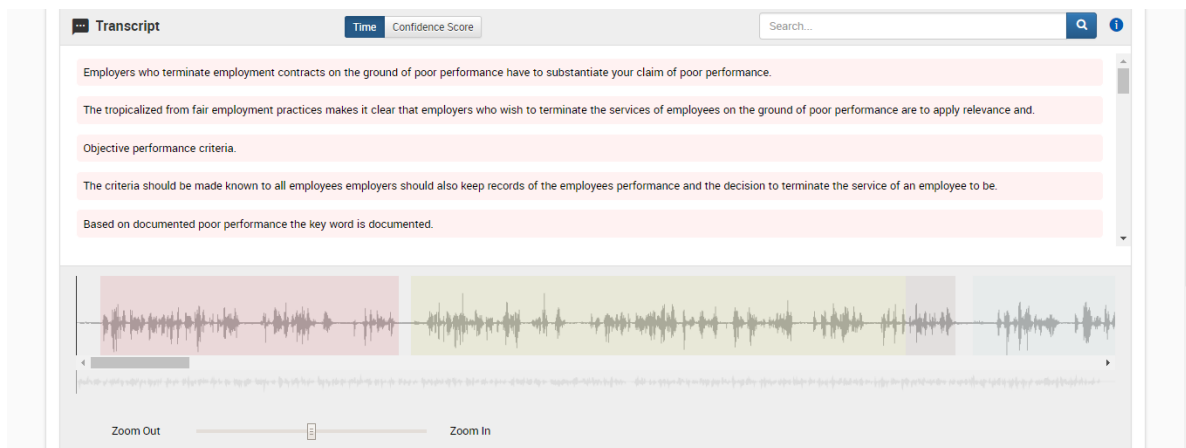


Figure 43: Modified Layout of Sentence Structure

5.5 AUDIO PAUSED UPON EDITING TRANSCRIPT

Novice users may face difficulty in editing the transcript when the audio is playing. Users who are unfamiliar with the editor may not catch up with the fast audio and edit alongside with the playing of audio. To solve this problem, the pause function is included in the player. Upon editing the transcript by modifying a character, the player will automatically stop playing the audio. To resume this action, users is required to click the play button to continue the audio. This provides them with flexibility in editing the transcript. The editor will transit to the next sentence if no modification is done towards the current sentence or modifications are already been made.

5.6 SAVE TRANSCRIPT IN XML / SRT FORMAT

Users could save the modified transcript after amending the necessary changes on the transcript in xml format by clicking on “Save Updated XML” as shown in the figure below. The modified version will be downloaded in the “Downloads” folder. Users could upload the newer version of the transcript file with the video or audio file, when they want to modify their transcript again. With multiple downloads of modifications towards the transcript file, the updated version will not be overwritten, as shown in the figure below. Instead, the number will be incremented to state the version of the transcript.



 parliament-minister-lim-swee-say-on-the-recent-termination-of-employees-of-surbana-jurong_V1.xml
 parliament-minister-lim-swee-say-on-the-recent-termination-of-employees-of-surbana-jurong_V2.xml

Figure 44: Files downloaded in Multiple Setting

However, if users were to edit on one transcript in one setting and download the updated version several times after only uploading the video or audio file and the transcript file, the downloaded versions of the transcripts are displayed in the figure shown below. In this case, given that users were to upload this version into the editor, after modifications to the transcript, the editor is unable to download the modified version, as the system would not recognise brackets in the title of the transcript file.



 parliament-minister-lim-swee-say-on-the-recent-termination-of-employees-of-surbana-jurong_V1 (1).xml
 parliament-minister-lim-swee-say-on-the-recent-termination-of-employees-of-surbana-jurong_V1 (2).xml

Figure 45: Files downloaded in One Setting

Users could also download a version of the transcript in SRT format, to be used for other transcribing tool or other uses. By uploading the updated version of transcript into the Offline Web Subtitle Editor, the user could convert the transcript from XML format to SRT format.

In the case where users were to modify the transcript, wanting to save the modified version of transcript directly as SRT format, an alert message is prompted out to indicate that the user is required to save the updated XML version and return to the homepage, upload the modified version and save as SRT version instantly.

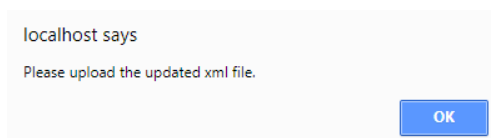


Figure 46: Alert Message

The SRT format of transcript will be downloaded in the “files” folder of the software. In the case of the SRT format, the downloaded file will be overwritten to the newer version.



Figure 47: Save as SRT Button


 parliament-minister-lim-swee-say-on-the-recent-termination-of-employees-of-surbana-jurong.srt

Figure 48: Downloaded file in SRT format

5.7 TRANSCRIPT IN SRT FORMAT

SRT format is the most basic of all subtitle formats, which most applications use transcript in SRT format. It consists of four parts, all in text.

1. A number indicating which subtitle it is in the sequence.
2. The time that the subtitle should appear on the screen, and then disappear.
3. The subtitle itself.
4. A blank line indicating the start of a new subtitle.

```
1
00:00:00,050 -----> 00:00:08,540
employers who terminate employment contracts on the ground of poor performance have to substantiate your claim of
2
00:00:08,540 -----> 00:00:22,840
the tropicalized from fair employment practices makes it clear that employers who wish to terminate the services (
3
00:00:22,840 -----> 00:00:40,860
the criteria should be made known to all employees employers should also keep records of the employees performance
4
00:00:40,860 -----> 00:00:46,480
we involves a unionized employee the union should also be consulted
5
00:00:46,480 -----> 00:01:02,760
if the employee fast and appeal of unfair dismissal to m o m will first mediate should mediation fail will conduct
6
00:01:02,760 -----> 00:01:10,890
this remedy is provided under the employment act as well as the industrialization revision act for union members
7
00:01:10,890 -----> 00:01:23,890
if an employer is unable to substantiate his claim that the employees performances poor be employed maybe ordered
8
00:01:23,890 -----> 00:01:41,580
if the employer does not comply with the order he can be prosecuted in the case of surbana jurong recent exercise
```

Figure 49: Transcript

Upon saving the uploaded transcript as SRT format, the transcript is presented in this format, where the first sentence is presented in 1. The start and end time of the particular sentence is recorded.

6. CONCLUSION AND FUTURE WORK

6.1 CONCLUSION

With the modifications being implemented and updated, the editor is more pleasant for users to utilize. Users need not strain their eyes to focus on long sentences if users have not selected any sentence in the beginning. Furthermore, users who are more familiarize with keyboard, could use keyboard shortcuts to navigate around the editor. For the benefit of novice users, the playback will automatically be paused upon deleting or adding a character to the sentence in the textbox for modifications. The editor complies a good user interface which is user-friendly. Transcript could be downloaded in SRT format, which could be accessible for other software, giving convenience for users to need not do the extra step of converting the transcript in XML format to SRT format.

With the previous implementation done to the Offline Web Subtitle Editor, words are being highlighted in yellow allows users to concentrate easily. With shorter segments of sentences being implemented, it gives rise to better readability. Also, the playback can be controlled efficiently with multiple icons provided. Waveform is visualized below the transcript, enabling users to illustrate the amplitude and pitch of the playback.

Transcribing tools are highly used by users worldwide due to the language barrier and other reasons. With the use of this Offline Web Subtitle Editor, users could edit the transcript anytime and anywhere without the use of Internet. Convenience is one of the important reason users are willing to use this editor, since most tools require downloading of software or Internet to operate. In conclusion, this editor allows not only experienced users, but also novice users to operate with necessary and non-complicated features.

6.2 FUTURE WORK

Since technology is becoming more advanced, with upgrades being installed, the editor may not be as powerful as before. Limitations to the Offline Web Subtitle Editor may still arise. Improvements towards the editor could be done, with more reviews, and modifications will be made. Some possible improvements include:

6.2.1 Multiple speakers

Multiple speakers are not supported. Colour segmentation could be done to the waveform as well as the transcribed text. Since the layout of the system does not indicate the speaker name, users could not know which speaker is currently speaking. Users could not edit the speaker name, unless they update the speaker name in the transcript in XML format.

6.2.2 Partial waveform

Partial waveform cannot be visualized, due to the limitation of the zoom function. For media files which have a longer length, the zoom in function can only be done up to a certain extent. The waveform will not be displayed for long length of media files.

6.2.3 Lack of waveform for video files

For video files, waveform visualization is not supported, and users could not view the amplitude and pitch of the media file. Improvements could be made by including the waveform for video files.

6.2.4 Multiple sentences being highlighted

If users were to modify a large segment of the transcript, the application could not support the multiple modifications, causing multiple sentences to be highlighted.

REFERENCES

- Yangchen, L. Raynold T. YK. (2017, April 3). *People in Singapore spend over 12 hours on gadgets daily: Survey*. Retrieved from The Straits Times:
<http://www.straitstimes.com/singapore/12hr-42min-connected-for-hours>
- Danny Donchev (2018, March 11). *37 Mind Blowing YouTube Facts, Figures and Statistics – 2018*. Retrieved from FortuneLords: <https://fortunelords.com/youtube-statistics/>
- Automatic Speech Recognition. (2009, June). *What is Automatic Speech Recognition?*
Retrieved from Docsoft, Inc: <http://support.docsoft.com/help/whitepaper-asr.pdf>
- Dorota, H. Adam, K. (2007, September). *Prevention: Best Practices in Software Management*. Retrieved from Wiley: <https://www.wiley.com/en-sg/Automated+Defect+Prevention:+Best+Practices+in+Software+Management-p-9780470042120>
- Pascal, V. L. (2003, October 7). *PRAAT Short Tutorial*. Retrieved from Stanford:
https://web.stanford.edu/dept/linguistics/corpora/material/PRAAT_workshop_manual_v421.pdf
- Maddalena, T. (2017, May 2). *User Guide for ELAN Linguistic Annotator*. Retrieved from ELAN: http://www.mpi.nl/corpus/html/elan_ug/index.html
- Thomas, S. (2016, October). *EXMARaLDA Partitur-Editor*. Retrieved from EXMARaLDA:
http://www.exmaralda.org/pdf/Partitur-Editor_Manual.pdf
- Nikolaj, L. O. (n.d.). *Subtitle Edit*. In *Nikse*. Retrieved 4 March, 2018 from:
<http://www.nikse.dk/SubtitleEdit>
- Nikolaj, L. O. (n.d.). *Subtitle Edit*. In *Nikse*. Retrieved 4 March, 2018 from:
<http://www.nikse.dk/SubtitleEdit/Help#sync>
- NCH Software. (n.d.) *Express Scribe Transcription Software*. Retrieved 4 March, 2018 from NCH: <http://www.nch.com.au/scribe/index.html>
- InqScribe. (n.d.). *Transcribe. Type Notes. Export Subtitles*. Retrieved 4 March, 2018 from InqScribe: <https://www.inqscribe.com>
- AudioTranskription. (n.d.). *f4 & f5transkript*. Retrieved 4 March, 2018 from AudioTranskription: <https://www.audiotranskription.de/english/f4>
- Descript. (n.d.). *Audio has found its voice*. Retrieved 5 March, 2018 from Descript: <https://www.descript.com/>

Elliot, B. (n.d.). *oTranscribe*. Retrieved 3 March, 2018 from oTranscribe:
<http://otranscribe.com/>

Transcribe. (n.d.) *Transcribe*. Retrieved 3 March, 2018 from Transcribe:
<https://transcribe.wreally.com/>

Trint (n.d.). *Trint is Transforming Talk*. Retrieved 3 March, 2018 from Trint: <https://trint.com/>

Happy Scribe. (n.d.). *Transcribe interviews in minutes !* Retrieved 4 March, 2018 from
HappyScribe. <https://www.happyscribe.co/#section4>

3Play Media. (n.d.). *Why 3Play Media?* Retrieved 5 March, 2018 from
<https://www.3playmedia.com/customers/why-3play-media>

Speechmatics. (n.d.). *Automatic speech recognition technology*. Retrieved 5 March, 2018
from: <https://www.speechmatics.com/>

Matthew, Z. (2014, December 29). *Automatic Speech Recognition (ASR) Software – AN Introduction*. Retrieved from Usability Geek: <https://usabilitygeek.com/automatic-speech-recognition-asr-software-an-introduction/>

Software Development Life Cycle. (n.d.). *What is the Software Development Life Cycle (SDLC)?* Retrieved 9 March, 2018 from Technopedia:
<https://www.techopedia.com/definition/22193/software-development-life-cycle-sdlc>

Agile Software Development Life Cycle. (n.d.). *What is Agile Software Development Life Cycle?* Retrieved 9 March, 2018 from QuickScrum:
<https://www.quickscrum.com/Article/ArticleDetails/2031/3/What-Is-Agile-Software-Development-Life-Cycle>

HTML5 (n.d.). *Advantages of HTML5*. Retrieved 10 March, 2018 from TechArk:
<https://gotechark.com/blog/advantages-html5/>

Markup language. (n.d.). *Markup language*. Retrieved 10 March, 2018 from Wikipedia:
https://en.wikipedia.org/wiki/Markup_language/

Diana, C. (2015, August 21). *To use or not to use Bootstrap Framework?* Retrieved from
Creative Tim's Blog: <http://blog.creative-tim.com/web-design/use-not-use-bootstrap-framework/>

jQuery UI. (n.d.). *jQuery User Interface*. Retrieved 5 March, 2018 from jQuery UI:
<https://jqueryui.com/>

Wavesurfer Json. (n.d.). *Wavesurfer Plugin*. Retrieved 5 March, 2018 from Wavesurfer:
<https://wavesurfer-js.org/plugins/>

APPENDIX A

WAVESURFER.JS PLUGIN

Wavesurfer.js has several plugins which can be implemented directly. Plugins include regions, timeline, microphone, minimap and playlist are supported in the wavesurfer plugin.

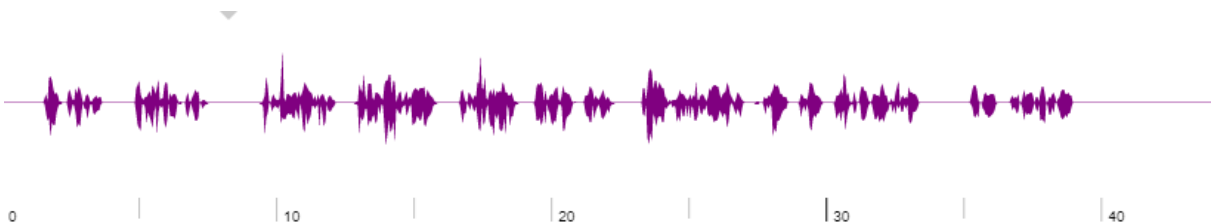
Regions plugin

Regions are visual overlays on waveform that can be used to play and loop portions of audio. Regions can be dragged and resized.



Timeline plugin

Timeline is added to wavesurfer.js instances.



Minimap plugin

A minimap is added to the main waveform.



Playlist plugin

A playlist capability is added to the wavesurfer.

