Enhancing Deaf and Dumb Collaboration by Video Annotations

Zaina Taha, Roaa Ghanem, Amjad Hawash [1]

Department of Information & Computer Sciences,
Faculty of Engineering & Computer Sciences
An-Najah National University, Nablus, Palestine, P.O. Box: 7
zenataha20@gmail.com, roaa .arabasi@stu.najah.edu , amjad@najah.edu (ORCID: 0000-0001-5115-8404)

Abstract

Online annotations, which enable information to be added to existing data without altering it, have emerged as a crucial technique for engaging with web pages and user queries. Annotations become a tool for targeted collaboration between users (Annotators) with similar interests when the visibility of annotations is restricted to groups of users. Many strategies have been developed to improve online user conversations and collaboration. Websites are suitable media for that since they enable users to engage in online discussions by adding comments (also known as annotations) to page elements like texts, photos, and videos. There are numerous annotations, including written, spoken, visual, and graphic. Although textual annotations are well-known, users are increasingly adding their own video comments to the HTML web document components to have conversations and exchange ideas. This enables deaf and dumb people to have the chance to participate in online discussions. At the end of the work, we conducted a comprehensive experimental test in order to compute the collaboration percentages between users in which promising results were achieved.

Keywords: Web Annotation, Visible Annotations, Sign Language, Deaf & Dump Collaboration.

---

[1] Corresponding Author: amjad@najah.edu

1. Introduction

Expressing one's thoughts and exchanging ideas and experiences with others can be done in several ways. Starting from the traditional ways of communication between people up to the usage of the latest technologies represented in the utilization of the different social network applications [1]. Expressing ideas to others perfectly facilitates their understanding and decreases their confusion regards the issues under discussion. However, conducting online discussions increases people's experiences due to the universal communication between people from different countries, nations, and backgrounds.

Annotating websites has become an increasingly popular method for individuals and organizations to enhance their online experience. By adding annotations to websites, users can provide additional context, insights, and information about the content of a webpage [2]. This helps other users to better understand the content and its relevance, and can also facilitate collaboration and knowledge sharing between them.

Recently, several applications are used to create online annotations of different types and map these annotations to different resources of websites. Textual and vocal online notes are public types of annotations that are used to express one's thoughts in which users are able to create and attach them to any website content like text, images, and videos) [3]. Although the traditional way of creating new annotations is by selecting some website textual contents to attach the created annotations with them, drawing over the website contents some graphical shapes like free drawings, circles, and rectangles becomes a popular way of mapping between website contents and the created annotations [4].

Despite that textual annotations are considered a perfect way in sharing ideas with others, they are limited in expressing emotions which enhances the amount of sharing ideas between online users that vocal annotations can do. The user's voice tone embedded in vocal annotations can substitute a lot of textual words needed to express his/her emotions. Expressing emotions with annotations add extra information and directs others to the points of online discussions and minimizes efforts and time in declaring ideas and feelings [5].

People with disabilities in their speaking or hearing are not able to express their ideas and emotions by vocal annotations. Usually, they communicate using symbols of sign language represented in some actions using their fingers and/or their bodies' gestures. Although some of them can communicate textually, they still have some limitations in expressing their feelings and emotions perfectly using this kind of communication. However, these limitations in communications make it hard for them to conduct online discussions using annotations [6]. Visible annotations are used to better exchange ideas and feelings between people all over the world by attaching digital contents of websites with visible annotations created as video files. Moreover, they can be used by deaf and dumb people in order to participate in some communication over the web. This enables them to involve more in communities and eliminates their feeling of isolated [7]. The ability to have an AI system that translates sign language gestures to spoken language facilitates the ability to conduct communication between deaf and dumb people with others without the need for others to know or learn sign language.

This work is related to enhancing the deaf and dumb users to create visual annotations for digital contents of websites as a way to increase their communication and make it easy for them to exchange their ideas and feelings regards the contents under discussion as well as enhance their involvement in online discussions with other people who are familiar with shared sign language.

The rest of this paper is organized as follows: Previous work is proposed in Section 2. Section 3 discusses the system architecture of the tool while Section 4 is related to Video Annotations Creation and Retrieval of the tool. Section 5 demonstrates the experimental results. Finally, Section 6 concludes this paper.

2. Related Works

Creating different types of annotations and assigning them to different types of digital content can be done these days using several types of annotation tools. Each of which has its own facilities for enhancing the intended collaboration between users. Brat (https://brat.nlplab.org/index.html), doccano (https://doccano.herokuapp.com/), LightTag (,https://www.lighttag.io/) and TagTog (https://www.tagtog.net/) are all examples of text-based annotation tools.

Another well-known practice is annotating internet content by sketching shapes over it. Users are able to annotate certain website contents by drawing shapes over them and adding textual notes to these drawings. Users of these tools exchange these drawings in order to have a debate based on the drawings and their notes. On the web, there are several tools available to improve this form of annotation. Famous tools for this type of annotation include Sketchpad1, Sketch to Web2, and MADCOW[4].

There aren't many tools available online for vocal annotations because of the small number of works in this area. Among them, XODO (https://xodo.uservoice.com/) stands out because it is used to produce audible annotations that can only be appended to PDF files on mobile devices. Use VISITView (https://www.ssec.wisc.edu/visitview/tutorial/img7.html) to record voice annotations solely for educational reasons. The last option is Chrome Audio Capture (https://github.com/arblast/Chrome-AudioCapturer), a Google Extension that enables users to record voice memos independently of some web content. This program is similar to ours in that it is implemented by a Google Chrome extension, but it just functions as a recording tool and does not add vocal annotations to the text of websites.

A possible method of facilitating collaborative sessions around the globe is by using video to provide visual annotations. Here, users can use web-based collaboration strategies by producing their own movies and attaching them to various pieces of online information. The work is yet insufficiently mature and is missing a number of supporting services.

The research done in [8] looked at how to depict professional collaborative learning in dispersed online environments using voice annotations and video clips of practice. The initiative conducted two major studies with experts in learning technologies. The main findings of this research are that various types of video clips can assist the need for professional development. They also felt that integrating artifacts from the practice helped to "base" or anchor a representation in realistic ways for reflective learning. The

subjects acknowledged that watching oneself in video clips was still a somewhat unique experience. According to the study, voice annotations and video clips can be produced quickly. Particularly in fluid, developing fields like the learning technology industry, speed was considered a key asset and regarded as more important than producing a polished result.

The authors of [9] describe a novel architecture and its prototype tool for indexing and retrieving particular voice recordings captured during conversations concerning tangible objects like text documents, images, or 3D models. An ink-dot is used to identify a specific component of an object when it is mentioned, and an image of the dot at a tiny scale is taken to record the information in a database. An index of the recording fragment is simultaneously created and connected to the dot. The associated recording fragment can be retrieved for playing after a dot has been scanned and identified by comparing its microscopic image with the database. To make these tasks easier while the user focused on the conversation, a useful tool has been created. Dot identification's performance tests have produced accurate matches that are authentic. The program successfully supported the building of indexes with dots during speech recordings in demonstrations of a real-world usage scenario, and it accurately played back all the distinct recording fragments connected to the dots.

Despite the existence of the mentioned textual and vocal tools, they are limited in completely expressing the users' emotions represented by body gestures nor introducing a solution for deaf and dumb people to communicate and collaborate online. This work is related to developing a visual-based annotation tool that enables the creation of video annotations (as well as texts) in order to better conduct online communication between users by attaching these annotations to the textual contents of websites. This enables people to better communicate and express their emotions in a better way rather than enabling people with disabilities related to their hearing or speech to better communicate with each other and involve in an online collaboration for selected topics. Moreover, the ability to attach textual tags with annotations makes our tool more fruitful in terms of searching for the most suitable annotations and hence their annotators to better conduct online discussions with the most related topics for users.

3. System Architecture

Adding a translucent overlay on top of an annotated website is all that the concept of annotation on web content entails. The tier where annotations are present is represented by this unseen layer. The utility built in this study follows the web-based client/server architecture. In an HTML web page, users can add their own video annotations to texts, and when they save the page, the relevant data are kept in a special database. According to this technique, the tool is made up of three layers: presentation, processing, and database. The Presentation Layer handles the user-browser interaction in which annotations are submitted, placed as placeholders next to chosen texts, and then saved in the specific database. Between the Presentation and Database levels, the Processing Layer handles the processing required to colour the highlighted text, record the video annotation, and save it as a video file. Last but not least, the Database Layer is concerned with the actual storage of all information pertaining to the created annotations represented in

the annotated text, including the URL, the video file, and the annotator-related information. Finally, the Database layer is in charge of retrieving the annotations for a specific person and website. The framework of the tool is shown in Figure 1 below, where the request/reply protocol is utilised for saving and retrieving annotations, respectively.
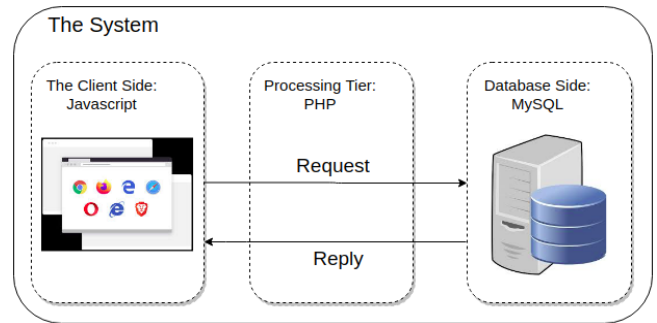

Fig. 1 The tool's architectural layout.

The tool's interaction is implemented via a Google extension. It comprises a series of JavaScript routines that are injected into each visited website in order to execute the functionality of text selection and video annotations recording. The extension also includes user login and account creation functionality.

The database's logical schema is shown in Figure 2 below as an entity-relationship diagram. The User, Video, and Text entities, as well as the entity relationships between them, are represented in the diagram. The text entity is used to preserve information about the annotated text, while the Video entity is used to save information about the video annotation itself. The user entity reflects all properties connected to users.
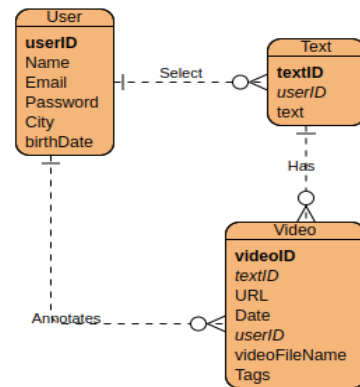

Fig. 2 The Entity Relationship Diagram.

The flow chart for the activity of creating annotations is shown in Figure 3. The action starts when a user chooses some text on a website. This occurs when the JavaScript object for the Browser object calls the *createAnnotation(selectedText)* function. The latter executes the r*ecordVideo()* method on the Video Recorder object, which returns the recorded video saved in the recordedVideo object and returns it to the JavaScript object. The Database object's function *Save(selectedText, recordedVideo, URL)* saves all relevant data in the system's database, and is then called by the JavaScript object. The *Notify()* function alerts the user when the vocal annotation has been successfully created before prompting them.
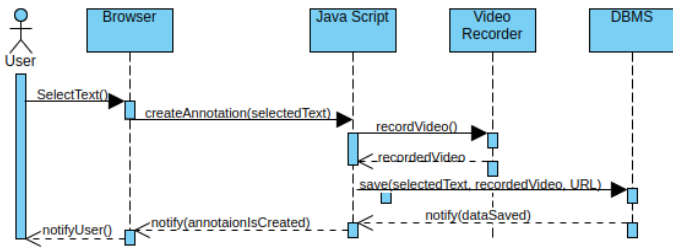
Fig. 3 Video Annotation Creation Sequence Diagram

The video annotation submission and retrieval are both summarised in the following pseudocode piece. The Google Extension is loaded and the aforementioned code is run as soon as the Google Chrome browser opens.

```
while(true){

    // List of Google Extension (GE) embedded listener functions

    GE.onChangeURL() = injectJavaScript(URL);

    GE.onLoadAnnotations() = loadAnnotations(URL);

    GE.onSelectText() = createAnnotation(URL, userID);

    GE.onHighlightTextClicked() = retrieveAnnotations(URL, clickedText);

}

function createAnnotation(URL, userID){

    placeHolder = selectText(selectedText);

    displayPopup(placeHolder, selectedText);

    recordedVideo = recordVideo(recordedFile);

    saveAnnotation(userID, selectedText, recordedVideo, placeHolder, URL);

}

function retrieveAnnotations(URL, clickedText){

    VideoFile Files[];

    popUpWindow = new popUpWindow();

    Files = getAnnotations(URL, clickedText);

    for(i = 0; i < Files.size(); i++)

        popUpWindow.insert(Files[i]);

}
```

By running the listeners mentioned inside the loop, the code keeps iterating in an unending loop. When a logged-in user navigates to a new URL, the listener function *GE.onChangeURL()* is called. This causes the method *injectJavaScript(URL)* to be invoked to inject the appropriate JavaScript code that will be in charge of processing video annotations. When a user clicks the Google Extension to obtain all annotations previously contributed to the current URL by calling listener function *GE.onLoadAnnotations(URL)* that causes the invocation of the function *loadAnnotations()*. This listener function additionally re-highlights all texts that are associated with annotations for the current URL.

When the user selects some text from the current URL, the listener function *GE.onSelectText()* is executed. The method *createAnnotation(URL, userID)* is called when this listener is activated, passing the URL and userID as inputs. The final

listener is *GE.onHighlightTextClicked()*, which is called when a user clicks a highlighted text in the active URL. This method calls *retrieveAnnotations (URL, clickedText)* to load all annotations associated with the clickedText.

The annotation creation process begins with the execution of the function *createAnnotation (URL, userID)* begins with the execution of the function *selectText(selectedText)* method that in turn returns the place-holder (the location information of the selected text) and stores it in the variable placeHolder. Upon the selection of some text, The function *displayPopup(placeHolder, selectedText)* inserts the selected text and associated placeHolder information into the pop-window itself. The controls required to construct the video annotation and save it into the database are included in this pop-up window. Figures 4 and 5 depict these contents. After that, the function *recordVideo(recordedFile)* is executed to begin recording the video annotation and upon completion, the recorded file is saved in the variable recordedVideo. The function *saveAnnotation(userID, selectedText, recordedVideo, placeHolder, URL)* is then executed to save all annotation related data inside the dedicated database.

The function *retrieveAnnotations(URL, clickedText)*, which accepts the current URL and the clickedText as inputs, is invoked to retrieve previously submitted annotations. This function generates an empty array of Files of type VideoFile that will be used to store all video files that are associated with the highlighted clicked text. A new pop-up window is then created and saved in the object popUpWindow by calling the constructor *popUpWindow()*. The next step is to call the function *getAnnotations(URL, clickedText)* to retrieve all associated video files, which are then saved in the Files array. The JavaScript code then loops the File array in order to insert the video files inside the pop-up windows created for retrieved annotations to be played back by the users.

4. Video Annotation Creation and Retrieval

After visiting an interesting website, a user can choose some text that matches his or her interest and add a verbal remark to it. The user then selects the system extension icon located in the Google extensions area of the Google Chrome browser which opens a pop-up window containing the selected text as well as a set of options to create and save a video file. After providing a series of optional tags that describe the tenor of the annotation, the user then starts recording his or her own video annotation. The search for annotations will make use of these tags. The program saves the annotation and all associated data in its database when the save button in the pop-up window is clicked. The pop-up window used to construct the annotation is seen in Figures 4 and 5 below, where the Gray-highlighted text is copied from the original webpage to the pop-up window.
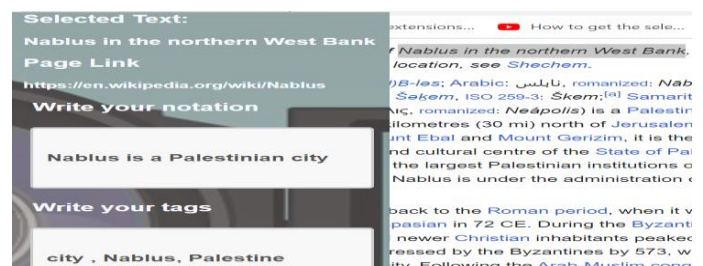

Fig. 4 Selecting a text to be attached with an annotation

The recording process begins when the user clicks the video area and stops when s/he clicks it again. By pressing the Save button, the embedded JavaScript code in the Google extension

communicates with the required PHP code to save the video file in a specific server folder with an automatically generated file name that is saved in the database, linking the saved annotation to the relevant video file. In this case, the name of the video file is saved in the database but the actual video file is saved in a separate server folder. When the video annotation needs to be recovered later, this generates a virtual mapping between the name of the file and the actual file in the server for file retrieval purposes.

Fig. 5 Recording a video for the created annotation

After logging into their portals, users can view a list of all of their own annotations and filter responses. They have the option to browse the websites related to these annotations and play the video recordings that are associated with them. The way the

annotations are shown in some users' portals is seen in Figure 6.

Fig. 6 Searching by tags.

The figure includes the annotated text as well as a collection of all relevant video annotations. However, the figure illustrates the ability to search for annotations by tags in which the users are able to search for some annotations depending on the tags associated with them in order to make it for users to search for the annotations they are interested in. Each annotation includes the video control for the video file it is linked to, as well as information on its owner, and the set of tags that are associated with it. The users have the ability to reply to these annotations by another video annotation giving them the possibility of conducting collaborations between them with respect to some interesting shared topics. Naturally, clicking on any of these annotations will take the user to the website where the corresponding annotation was created. When one of them is clicked, the tool opens the corresponding website and utilises its URL to obtain all annotations related to the website. To allow users to examine all filtered annotations connected to the clickable texts, all texts that were previously selected on that website will be re-highlighted and clickable. On the other hand, a logged-in user can ask the Google Extension to retrieve all annotations provided to this website (if there are any) if they surf a sample previously annotated website. With that request, the tool re-highlighted all annotated texts so that the user can click any of them to open a pop-up menu that contains all clicked text-related annotations. Users of the tool can also look for published annotations of interest using their portals.

5. Experimental Tests

We conducted an experimental test involving 20 participants in order to compare video, vocal and textual annotations. The participants are of two groups, each of which is composed of 10 participants. The members of the first group are people without deaf and dump disabilities, while the members of the other group are. We need to test the amount of collaboration between the members of each group and between the whole members of both groups (for those who practice sign language).

The supervised test lasts for 8 days in which a set of predefined websites are selected for different topics (Sports, Science, Education, Politics, Information Technology).

We divided the test into three apheses for the first group. In the first phase, we asked them to navigate the websites and create textual annotations only. In the second phase, we asked them to create vocal annotations (we asked them to switch off their cameras) while the last phase is done by using video annotations. At the end of the test, we noticed that 235 annotations were submitted (47 textual, 86 vocal, 102 videos), 32 distinct websites were annotated and 3 tags as average per annotation. For the second group, we asked them to submit textual and video annotations only in which the annotations contain sign language. At the end of this test, 135 annotations were submitted (29 textual, 106 videos), and 27 websites were annotated with 2 tags as average per each annotation.

In order to learn the thoughts of the two groups on the experiment and the cooperation acquired through the use of the three types of annotations, the participants of the two groups were requested to complete a questionnaire. The percentages below are computed by taking the average of the percentages gathered from the two groups. There are several questions on the survey, including:

1. What difference does the use of video annotations make in terms of how well participants understand the topics being discussed? reducing time and effort, with almost 75% of respondents choosing to spend less time and effort using video annotations, 15% for vocal and 10% for textual ones.

2. How does employing video annotations for collaboration compare to the other annotation types? Exceptional, Good, and Poor. Percentages computed were 80% for video 11% for vocal and 9% for textual annotations.

3. What degree of tool usefulness is there? Excellent, Very Good, Good, and Acceptable grades. The Percentage values collected were 85% for video, 10% for voice, and 5% for textual annotations.

4. Which is more effective for expressing your emotions in submitted annotations? (Video, voice, and text). We received 90% for video, 7% for voice, and 3% for textual).

5. Which is quicker to annotate? (Video, voice, and text). As speaking and showing body gestures is far faster than talking or writing and requires no manual labour like typing, we received a perfect score for video represented in the percentage 95%, while voice got 3% and 2% goes to textual annotations.

Figure 7 below depicts the previously computed percentages for the 5 questions in which the video annotations perform the other types.
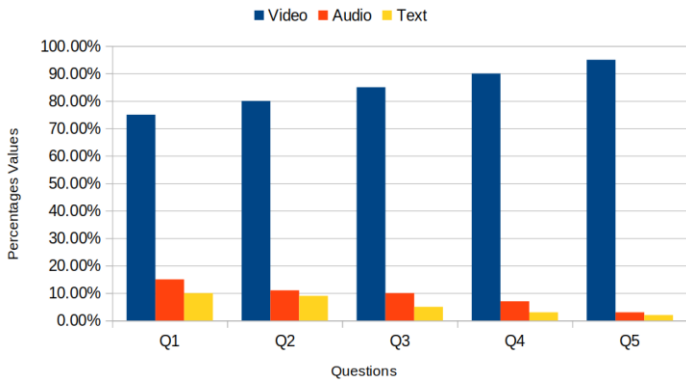
Fig. 7 Percentages values computed for the 5 questions.

6. Conclusion

In this work, a video-based annotation is put into practice and tested. People can hold online conversations utilising video-based annotations to collaborate and talk about topics of interest. For searching purposes, users can add a set of related textual tags to annotations. An experimental test under supervision was carried out, and encouraging outcomes were obtained.

For the next projects, we are planning to improve the search function by adding a video recognition system that allows recognising their signs and converting them to vocal notes in order to better collaborate disables users with others users.

References

1) Brinberg, Miriam, and David M. Lydon-Staley. "Conceptualizing and Examining Change in Communication Research."*Communication Methods and Measures* (2023): 1-24.

2) Garcia, Manuel B., and Ahmed Mohamed Fahmy Yousef. "Cognitive and affective effects of teachers' annotations and talking heads on asynchronous video lectures in a web development course." Research and Practice in Technology Enhanced Learning 18 (2023).

3) Neves, Mariana, and Jurica Ševa. "An extensive review of tools for manual annotation of documents." *Briefings in bioinformatics* 22.1 (2021): 146-163.

4) Antico, Matteo, et al. "An interactive tool for sketch-based annotation." JJAP Conference Proceedings 14th International Conference on Global Research and Education, Inter-Academia 2015. The Japan Society of Applied Physics, 2016.

5) E. Asmar, S. Salahat, F. Zubdeh and A. Hawash, "Enhancing Users Collaboration By Vocal Annotations," 2021 18th International Multi-Conference on Systems, Signals & Devices (SSD), Monastir, Tunisia, 2021, pp. 845-851, doi: 10.1109/SSD52085.2021.9429432.

6) Whitburn, Ben, and Rod Michalko. "Blindness/sightedness: Disability studies and the defiance of di-vision." (2020).

7) Alhassan, Fatemah, et al. "Technology-based services for deaf and dumb people." International Journal of Data Science 5.2 (2020): 160-167.

8) C. Steeples, "Voice annotation of multimedia artifacts: reflective learning in distributed professional communities," in Proceedings of the 35th Annual Hawaii International Conference on System Sciences, 2002, pp. 10 pp.–

9) T. van Bergen, M. Engineering, Y. Kudo, R. Ishiyama, T. Takahashi, K. Makino, and J. Goosen, "Indexing and retrieving voice recordings by instantly tagging mentioned objects with dots," 04 2019.