Extracting information from audio data gathered in a virtual reality museum tour to determine personal preferences

E. REUVERS, University of Twente, The Netherlands

Abstract This paper explores the possibilities of identifying individual preferences of users through the analysis of audio recordings gathered in a virtual reality museum tour. The audio recordings contain user comments about their observations related to paintings at the museum, as well as inquires for further knowledge. Users were asked to think-out-loud during their visitation. Identification of user preferences would make it possible to design personalized museum tours in the future, for example. The analysis starts with transcribing the audios to texts using the Google Cloud speech-to-text framework. The text files are then analyzed using natural language processing techniques, in particular: part-of-speech tagging, noun phrases extraction and sentiment analysis. The results of the part-of-speech tagging and noun phrase extraction is that most comments contain objective information regarding the objects in the paintings. The sentiment analysis of the complete user comments confirms this. 59.4% of the user comments are neutral and 29.8% is slightly positive, meaning that the comments do not exhibit strong emotional polarity. This suggest that audio recordings based on a thinking-out-loud process may not be sufficient to reliably identify individual preferences. More research is needed to determine if the identification of preferences might be possible with techniques or approaches that have not been utilized in this study.

Additional Key Words and Phrases: Human computer interaction (HCI), Cultural heritage, Natural language processing, Sentiment analysis

1 INTRODUCTION

Museums hold vast amounts of artworks, such as paintings, sculptures, or antiques. Thus, visiting a museum can take a significant amount of time, especially considering grand museums such as *Musée du Louvre* in Paris or *het Rijksmuseum* in Amsterdam. These museums have a large surface area with numerous exhibits. Visitors may only be able to visit a number of them in the limited time of their visit. They would benefit from personalized museum experiences to guide them through the museum based on their personal preferences.

Before personalization is possible, it is necessary to identify the preferences of visitors. Previous studies have focused on extracting preferences using methods, such as star ratings of exhibitions [1] or a visitor quiz [2]. This study takes a novel approach by exploring the potential of extracting valuable user information from voice data. This opens up the possibility of developing personalized museum tours that dynamically adapt based on the visitor's spoken feedback. An example of an application could be a virtual guide that engages in conversation with visitors, extracting their sentiment towards specific paintings from their speech on which he bases his further tours. Therefore, this research paper addresses the challenge of extracting individual preferences from audio recordings collected during a virtual reality museum tour, with the aim of enabling personalized experiences.

The recordings consists of user comments related to paintings encountered in a virtual reality museum. These include remarks about their personal preferences as well as inquiries for further knowledge. The comments are valuable due to its contents and will help with achieving the objective of this study, which is to identify a user's sentiment towards these artworks and determine their preferences based on it. Therefore, the voice recordings are considered to be useful in determining if it is possible to extract individual preferences from voice data. This introduces the following research questions:

How can information about personal preferences be extracted from audio data collected in a virtual reality museum tour?

The research question has the following sub-questions:

1. How can natural language processing techniques be used to extract valuable information?
2. How can the analysis of the audio data be used to determine personal preferences?

These questions are answered by combining the voice and eye gaze data that has been previously collected in a different study. Initially, the voice data has been transcribed and subsequently linked to the correct paintings. The linked data has been analysed using natural language processing techniques, from which conclusions have been drawn upon the research questions.

The paper consists of 8 sections. First, the introduction is followed by a review of the related work in chapter 2. Then the data set is described in chapter 3, as well as the methodology of the research in chapter 4. Moreover, the results are presented and discussed in chapter 5 and 6. The conclusion in chapter 7 answers the research questions and chapter 8 gives suggestions for future work.

2 RELATED WORK

The related work focuses on research related the to personalization of museum experiences. Additionally, it examines personalization in other domains where natural language processing techniques have been employed to extract personal preferences. The related work aims at clarifying the gap in knowledge to which this research is devoted.

2.1 Museum personalization

Many researchers have researched how to personalize the museum experience for visitors. Several approaches will be mentioned in this section. Bohnert et al. [1] present the "GECKOmmender", a

TScIT 39, July 7, 2023, Enschede, The Netherlands

mobile system for personalized theme and tour recommendations in museums. It is based on a digital site-map representation and uses star ratings for seen exhibits to predict ratings for unvisited exhibits. These predicted ratings form the basis for the tour recommendations. Antoniou et al. [2] investigate the use of indirect profiling methods through a visitor quiz, in order to provide the visitor with specific museum content. Wang et al. [3] suggest connecting a museum's website with its physical museum space. In particular, they propose a Web-based museum Tour Wizard based on the user's interests and a Mobile Guide that converts these tours to a mobile device used in the physical museum space. Furthermore, Amato et al. [4] propose an agent-based approach for recommending cultural tours. They propose to integrate recommendation facilities with agent-based planning techniques in order to implement a planner of routes within cultural sites. Lastly, Tsiropoulou et al. [5] propose Quality of Experience-based (QoE) museum touring. The most influential QoE of visitors are used to provide them a customized and personalized experience.

2.2 Personalization based on natural language processing

This paragraph highlights various research conducted on natural language processing (NLP) for personalization in different domains than museum experience personalization. Huddar et al. [6] propose a multimodal sentiment analysis in which user sentiment is extracted from transcribed content, visual and vocal features. They aim at going beyond text based sentiment analysis and want to improve possible results. On the other hand, Zankadi et al. [7] aim at identifying and extracting topical interest from text content that has been shared on social media. They apply three NLP techniques to this textual data, namely: Latent Dirichlet Allocation, Latent Semantic Analysis and BERTopic. Reuver et al. [8] describe how NLP can play a central role in diversifying news recommendations for viewers. NLP helps with detecting different viewpoints of viewers, while also detecting individual latitudes of diversity. Furthermore, Gupta et al. [9] aim at extracting meaning from natural language speech. They propose a system that uses machine learning techniques to extract the intents and the named entities from users' spoken words. Finally, Paik et al. [10] describe a metadata extraction technique based on NLP to extract personalized information from email communications. This system enables automatic user profiling.

Existing literature covers museum personalization and NLP-based personalization in diverse domains. However, there exists a gap in knowledge regarding the application of NLP techniques for personalizing museum experiences through the analysis of transcribed audio recordings. Therefore, this research paper aims to address this gap by exploring how information about personal preferences can be extracted from audio data.

3 DATA SET

The data for this research has been gathered during a previous user study [11]. This study consisted of 31 participants that walked through a virtual reality exhibition called: “HERE: Black in Rembrandt's Time”, which was displayed at *Museum Rembrandthuis* in 2020. The participants wore VR glasses which showed the exhibition and were instructed to explore the museum at their own pace while expressing their thoughts in a "think-out-loud" process. These comments were recorded for analysis. Moreover, eye gaze data was collected to capture what paintings the users were looking at during their visitation. An overview of the general details of the audio files can be found in table 1. It can be noted that the average user's visitation lasted 17:59 minutes, with a visit of 7:26 minutes lasting the shortest and a visit of 36:01 minutes lasting the longest. The total duration of the audio recordings were 9 hours and 16 minutes. After the users were done with their visitation, they were asked in a survey to indicate on a scale from 1 to 5 how interested they were in each painting. Where a "1" means they were not interested at all and a "5" means they were very interested. The results of this survey have been stored in an Excel file.

Audio file details	**Time**
Average duration	17:59 minutes
Shortest duration	7:26 minutes
Longest duration	36:01 minutes
Total duration	9:16 hours

Table 1. Details regarding the audio files

4 METHODOLOGY

The methodology explains how this research is conducted. It provides a workflow, explains the transcribing process, and the usage of NLP techniques to analyze the data. The NLP techniques have been chosen based on their usability and their ability to identify sentiments from speech.

4.1 Workflow

A flowchart of the workflow of this research can be found in Figure 1. Input data is displayed with diamond shaped boxes and processes are displayed using the squared boxes. The initial step involves taking the speech data as input for the speech-to-text transcription process. Upon completion, the speech data and eye gaze data are linked to identify which paintings the users are discussing. Therefore, the eye gaze is also displayed as input data in the workflow. Once the linking process is finished, natural language processing techniques are used to analyze the data. These techniques include part-of-speech tagging, noun phrase extraction, and sentiment analysis. The resulting analyses are thoroughly reviewed and used to draw conclusions regarding the research questions.

4.2 Speech-to-text

To allow for the analysis of the user data, the audio files need to be transcribed into text files. Due to the length of each audio file manual transcription is too time consuming (see Table 1). Therefore, the transcription process has been accomplished using the Google Cloud Speech-to-Text framework [12]. This framework converts voice to text in over 125 languages and variants. Moreover, it is reliable, trustworthy and easy to use in combination with other technologies [13]. Therefore, this framework has been chosen as the best option to transcribe the audio files.

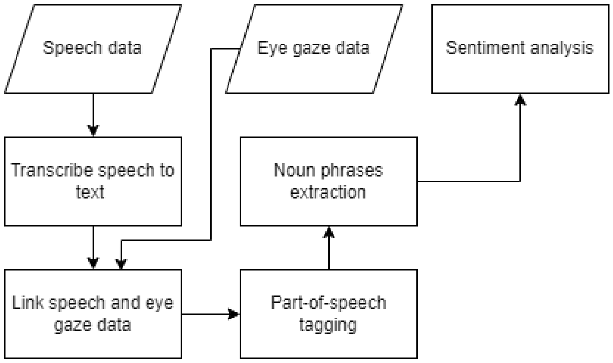

Fig. 1. A flowchart of the workflow of the research

The framework has been customized and modified using the programming language Python [14] to accommodate various requirements. One necessary modification involved storing the transcripts locally instead of within the Google Cloud framework. Storing locally allows for the association of the transcriptions with the corresponding artworks, which can not be achieved within the limitations of the original framework. For a detailed explanation of this association process, refer to section 4.3. Additionally, local storage is needed to facilitate the NLP processing since conducting such analyses is not possible within the Google Cloud framework.

The final result is a speech-to-text transcription that includes the transcription, the timestamp of the spoken sentence and the confidence of the transcript. Table 2 displays a number of examples of the results of the transcription process. As can be seen in the table, these transcripts are about the observations of a user at certain timestamps. These users talk about the Netherlands, Amsterdam, and a group of men. Moreover, it can be seen that the Google Cloud framework has a confidence of at least 84% for all paintings, indicating that it is fairly sure that its transcriptions are correct.

Transcription	Start time	End time	Object	Confidence
I really should know more about the Netherlands	275.400s	277.400s	A1 text	95%
okay so this is Amsterdam	316.800s	319.600s	A2 text	98%
it shows a group of men Eastern turbans	326.900s	330.700 s	A2 text	84%

Table 2. Examples of the speech-to-text transcription results

4.3 Linking the speech and eye gaze data

The audio recordings of each user cover their visit to the virtual reality museum. From only the audio or transcriptions it is unclear what painting the user are talking about. In order to extract meaningful information about a user's preferences, it is important to know what paintings belong to which user comments. Therefore, each user comment needs to be matched to the correct painting.

The transcriptions have been matched to the eye gaze data by using the timestamps value that both the transcripts and eye gaze data contain. Because of this linking, each specific sentence a user has spoken can be coupled to the painting the user has been observing at that time. The final results have been stored in a CSV file, wherein the headers represent the individual paintings and the users' comments have been inserted accordingly. Table 2 shows an example of how the transcriptions are linked to the paintings by showing the object that belongs to the transcription. It can be seen that the different transcriptions belong to A1 text and A2 text, which are the explanatory texts of painting A1 and A2.

4.4 Part-of-speech tagging

After collecting all of the audio transcriptions and matching them to the correct paintings, useful information needs to be extracted from this data. The most straightforward option is to achieve this by using natural language processing techniques on the data. A simple technique suited for the type of data used within this research is part-of-speech tagging. This is the process of marking up a word in a text as corresponding to a particular part of speech, such as nouns or verbs [15].

A motivation for using part-of-speech tagging is that extracting the nouns from the transcriptions can help with the identification of the type of objects or people in the paintings. This can give an impression of what a user notices in paintings and deems significant enough to comment on. It serves as the initial step in understanding the content, themes, and users' interests. Providing the foundation for further analyses, including sentiment analysis. By utilizing and combining these techniques a more comprehensive understanding of users' sentiments towards the artwork can be achieved.

All the data in the CSV file has been processed using part-of-speech tagging. For this, the widely used NLTK (Natural Language Toolkit) Python library [16] has been used because it processes text quickly and easily. Furthermore, NLTK is known for its capability to handle complex and intricate text data effectively [17].

The tagging process on all CSV files has only been conducted for nouns, adjectives and adverbs. Nouns are useful to find entities in paintings, whereas adjectives and adverbs are useful to identify a user's feelings and preferences regarding the paintings. Other word types do not clarify much about the paintings itself, or a user's sentiment related to the painting. An example of the tagging process is the sentence: "I like beautiful paintings." The sentence is tagged like this: [("I", PRONOUN), ("like", VERB), ("beautiful", ADJECTIVE), ("paintings", NOUN), (".", PUNCTUATION)].

Based on the results of this processing a word cloud has been generated for each painting and text belonging to the painting. It has been generated using the WordCloud Python library [18] and includes only the nouns that are stored during the process of part-of-speech tagging. The adjectives and adverbs are not displayed because nouns give the clearest visualization of a user's observations related to the objects in the painting. Words that are mentioned more often are displayed bigger in the word cloud, whereas words that are said less appear to be smaller. The complete word cloud gives an overview of the general consensus regarding the painting, what objects are visible, and what type of people can be seen in the painting.

4.5 Noun phrases extraction

Another type of natural language processing technique that has been used to analyze the CSV file is noun phrases extraction. Noun phrases are part of speech patterns that include a noun and showcase information about the context of that particular noun. This is useful because the context of the noun can give valuable information about a user's sentiment towards an artwork, whereas part-of-speech tagging would have only been able to give a general overview of the user's sentiment by identifying the most talked about objects.

To ensure a swift extraction of noun phrases the NLTK Python library [16] utilizes a regular expression (regex) pattern. The regex pattern in the NLTK library defines a noun phrase's structure as follows: <DT>?<JJ>*<NN>. The "<DT>?" resembles the determinator, these are words such as "an", "the", or "this". The question mark means that it is optional and does not necessarily need to be part of the noun phrase. The "<JJ>*" stands for the adjectives belonging to a noun phrase, for example "pretty", "ugly", or "good" could be adjectives. The asterisk indicates that there can be zero or more adjectives. Finally the "<NN>" specifies the noun itself, this is a mandatory element of the noun phrase.

An example of noun phrase extraction for the sentence: "I like beautiful paintings." would result in this: ["I", "beautiful paintings"]. Here, the word "I" is considered a noun on its own. Moreover, "beautiful" and "paintings" are grouped together as a single noun phrase because they form a meaningful unit within the sentence. The word "like" is left out, as well as the punctuation because they are not part of the noun phrases.

This type of natural language processing has been done on the entire CSV file to determine whether users have said specific phrases that can be valuable for determining a user's sentiment and preferences. In comparison to the part-of-speech tagging, word clouds have not been generated for the noun phrases. This is because the WordCloud library can only count single words, not phrases. Therefore, it can not display the noun phrases correctly within a word cloud. Instead the results have been stored in text files and manually analyzed to determine whether they contain valuable user information.

4.6 Sentiment analysis

Sentiment analysis is also applied to the user comments to determine whether data is negative, neutral, or positive. It allows for the verification of the data against the survey results. This has been done by verifying whether the outcome of the sentiment analysis aligns with how interested each user is in the paintings.

The NLTK Python library has been used once more for this analysis, since NLTK has a pre-trained sentiment analyzer called VADER [19]. Because it is pre-trained results will be available more quickly than with other analyzers. Moreover, VADER is particularly suited for short sentences which makes it suitable to analyze the user comments because these comments consist of one sentence only. The sentiment analysis has been conducted for each user separately.

VADER calculates a sentiment for every sentence a user has spoken about a painting. This results in a compound score between -1 and 1, where results close to -1 mean that the user is negative and results close to 1 mean the user is positive. After all the scores have been calculated, the average is taken to determine the user's sentiment towards the painting they are talking about. These results have been stored in a bar chart, providing a clear overview of the user's sentiment towards each specific painting. This has been done for each user, allowing for a check of the user's sentiment against the survey data.

5 RESULTS

This section is devoted to the results of the part-of-speech tagging, noun phrases extraction and sentiment analysis. It also gives thought to the variation between users.

5.1 Part-of-speech tagging

The part-of-speech tagging has resulted in the creation of word clouds based on the tagged nouns. A word cloud has been created for each painting. An example of a generated word cloud can be found in Figure 2. The words "painting", "information", "picture", and "thing" are the most mentioned words, however, they have been excluded from the word cloud because they lack specific relevance to the paintings and tamper with the representativeness of the word cloud.

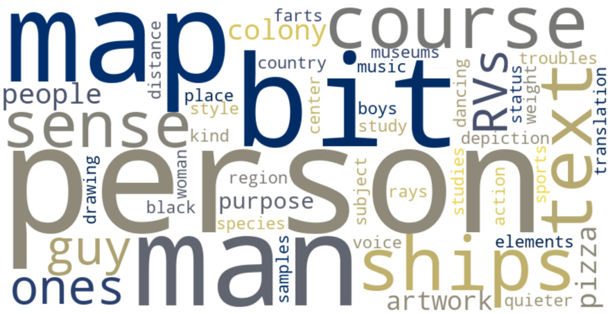

Fig. 2. A wordcloud of the nouns generated by the part-of-speech tagging for painting A1

The word clouds can be practical because they provide a general view of a user's perceptions about a painting, offering insights into what objects are present and the general emotions it evokes. However, it needs to be noted that word clouds alone can not be used to confirm the presence of specific objects in the paintings. To determine whether the object is present in a painting, it is needed to analyze the eye gaze data at the moment the user mentions the object and verify if there is a match between them.

In the example word cloud, words such as "square", "map", and "colony" can be found, suggesting that these can be found in the painting as well. Figure 3 displays the painting and while it may appear to be similar to a map, possesses square-like elements, and looks like a colony, these interpretations still need to be linked to the eye gaze data to verify their correctness.

From the comparison of each painting's word cloud, it can be observed that the word clouds consist predominantly of the objects or people in the painting. Hence, the results of the part-of-speech tagging is not useful in giving a comprehensive understanding of the overall impressions or emotions evoked in the users by the paintings.

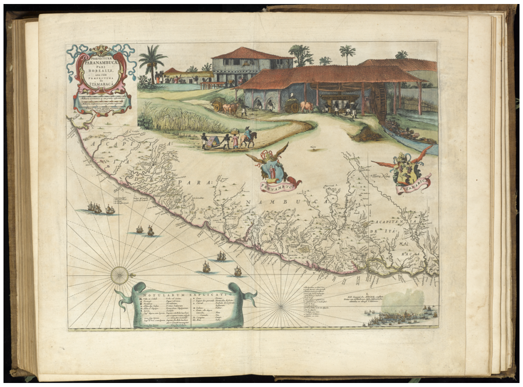

Fig. 3. Painting A1 (Blaeu, Johannes. *Kaart van Pernambuco.* 1665, painting. Museum Rembrandthuis, Amsterdam.)

5.2 Sentiment analysis

Noun phrase extraction has been conducted on the user comments too. The expectation is that noun phrases offer more information about a user's sentiment towards the painting. For example, that a combination of words such as "lovely square", "detailed map", or "interesting colony" occur in the user comments. These words give a better impression of the user's sentiment, and whether it is mostly positive or negative. However, the words spoken by the users do not contain these type of words. The noun phrases are more objective than subjective, containing words such as "Dutch colony" instead of "interesting colony". This description can still be valuable to determine what objects are present in the painting, however, they are not useful for identifying a user's sentiment towards the painting.

The part-of-speech tagging and noun phrases extraction demonstrate the limitations to determining the sentiment from isolated parts of user comments. It illustrates that only a noun is not enough to identify a user's sentiment. Therefore, a sentiment analysis has been conducted on the complete user comment as well. To achieve this, all remarks that belong to a certain panting and are spoken by a specific user have each been given a compound score. For each user's specific comments to a painting, the average of the compound scores has been calculated to determine the user's sentiment towards each different painting. These average compound scores have then been grouped into 5 categories, which are the same categories that have been used in the survey to determine a user's interest in the paintings.

Table 3 provides an overview of the different interest level categories and their associated compound scores. Since there are 5 categories and the compound score can range from -1 to 1, each category has a size of 0.4. An average compound score of 0.7 falls into category 5 and illustrates that a user is extremely interested.

Interest level categories	**Associated compound scores**
1. Not at all interested	-1 <= compound score < -0.6
2. Slightly interested	-0.6 <= compound score <= -0.2
3. Moderately interested	-0.2 < compound score < 0.2
4. Very interested	0.2 <= compound score <= 0.6
5. Extremely interested	0.6 < compound score <= 1

Table 3. Interest level categories of user's for the paintings and their associated compound scores

An example of a user's interest level per painting can be found in Figure 4. The goal of this bar chart is to visualize whether the survey results are similar to the results of the sentiment analysis. If this is the case, it can be concluded that a user's answer to the survey questions follow from their sentiment towards the painting.

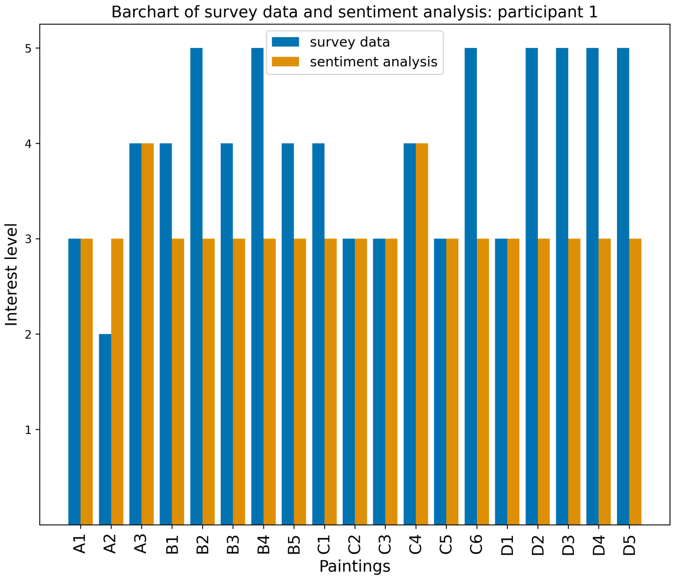

Fig. 4. A bar chart with the results of the survey data and sentiment analysis for participant 1 of the user study

5.2.1 Statistical analysis of survey and sentiment data. To determine if there is a significant association between the survey data and the sentiment analysis data, a chi-square test has been conducted. This type of statistical analysis examines the presence of a meaningful relationship between the survey data and the sentiment data. The assumptions for this analysis include independence of data, a sufficient sample size, and categorical data. These assumptions are satisfied in our study, because the data has been collected independently of each other, the sample size is equal to the population, and the data is categorical, since interest level categories have been used to group user's interest in the paintings.

In order to conduct the chi-square test, a null hypothesis and alternative hypothesis have been determined. These are as follows:

H0: Assumes there is no association between the survey data and sentiment analysis data

H1: Assumes there is an association between the survey data and sentiment analysis data

The chi-square test calculates the p-value. When the p-value is smaller than the significance level ($\alpha = 0.05$) the null hypothesis can be rejected. For those paintings, it means that there is a significant association between the survey data and the sentiment analysis data. This means that the sentiment analysis is not independent of the survey data, which means that there might be correlation between the two. When the p-value is higher than the significance level, the null hypothesis can not be rejected. For these paintings it means that no associations exist between the survey data and the sentiment analysis data. A complete overview of all the results of the chi-square test can be found in table 4.

As can be seen in table 4, all 18 paintings do not have a significant association between the survey data and the sentiment analysis data; the p-value is larger than 0.05 for each painting. Moreover, the chi-square scores are almost all between 0 and 10. This indicates the presence of a weak association, or no association at all. Hence, the survey data and sentiment data are independent of each other and further research is needed to understand why.

Painting	**Chi-square statistic**	**P-value**	**Significant difference**
A1	3.269	0.512	no, do not reject H0
A2	1.461	0.691	no, do not reject H0
A3	4.423	0.620	no, do not reject H0
B1	3.047	0.803	no, do not reject H0
B2	3.991	0.858	no, do not reject H0
B3	13.307	0.150	no, do not reject H0
B4	2.004	0.572	no, do not reject H0
B5	3.519	0.900	no, do not reject H0
C1	2.070	0.723	no, do not reject H0
C2	9.518	0.150	no, do not reject H0
C3	3.328	0.950	no, do not reject H0
C4	0.270	0.991	no do not reject H0
C5	8.837	0.717	no, do not reject H0
C6	7.939	0.439	no, do not reject H0
D1	5.498	0.240	no, do not reject H0
D2	0.565	0.904	no, do not reject H0
D3	5.101	0.745	no, do not reject H0
D4	1.926	0.749	no, do not reject H0
D5	6.588	0.360	no, do not reject H0

Table 4. The results of the chi-square test for the survey data and the sentiment analysis

5.3 Variation in users

Each user has made a different amount of comments and per user the amount of comments per interest level category varies too. Figure 5 contains the amount of comments each user has made per interest level category. These categories are the same as in table 3.

Figure 5 gives information about the strength of the emotional polarity per user. A strong emotional polarity indicates that the comments are predominantly positive or negative, while a weak emotional polarity would suggest a more balanced distribution of positive and negative comments. The presence of a strong emotional polarity indicates that users have distinct preferences or strong opinions, whereas a weaker emotional polarity suggests that users have a varied range of sentiments and opinions, with no single sentiment category dominating the others. For those users, it would be harder to extract personal preferences. Hence, this chart is beneficial for determining if a user's data is suitable for extracting personal preferences.

A glimpse at Figure 5 reveals that most of the comments are neutral, or positive for each user. The other interest level categories have significantly less comments. This suggests that the emotional polarity per user is weak, indicating that they do not have distinct preferences or strong opinions.

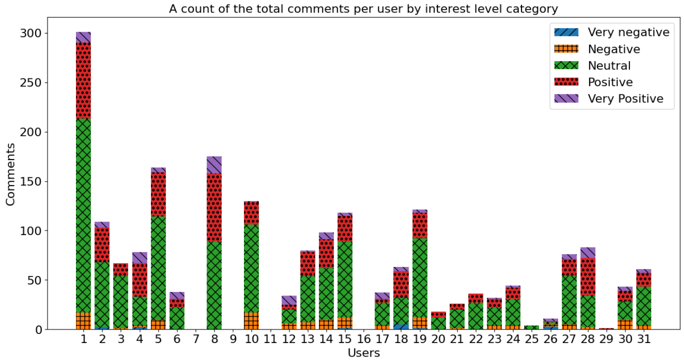

Fig. 5. A stacked bar chart that shows the amount of comments per interest level category

The pie chart in Figure 6 endorses this conclusion. It displays the distribution of the user comments per each interest level category. It can be noted that there are mostly neutral comments, a fair amount of positive comments and small amounts of very negative, negative, and very positive comments. This showcases that most users have a neutral, or slightly positive sentiment indicating a weak emotional polarity. Meaning that it is challenging to draw personal preferences from this data set.

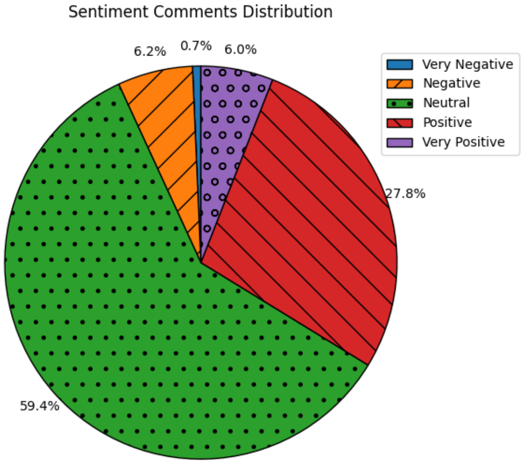

Fig. 6. A pie chart with the distribution of the user comments per category

6 DISCUSSION

The transcription process can result in transcription errors. This has the consequences that the input for the natural language processing (NLP) techniques might be faulty possibly leading to divergent results. For example, a word might have been transcribed the wrong way which can have the consequence that the *wrong* word is inputted into the part-of-speech tagging, noun phrases extraction, or the sentiment analysis. This affects the results of these NLP techniques and can affect the final results of this study. Therefore, it is important to consider these speech-to-text errors when looking at the results of the NLP techniques used in this study.

There are a number of reasons why none of the paintings have a significant association between the survey data and the results of the sentiment analysis (see Table 4). The first reason is that the correct sentiment cannot be extracted from the audio recordings because the user's have been asked to think-out-loud during the user study. This has the same outcome as discussed in the result for the part-of-speech tagging and the noun phrases extraction; users are more likely to talk about objects than they are to talk about their sentiments towards the painting. Moreover, the predominance of objective sentences causes the average compound scores of the sentiment analysis to be more neutral too, further emphasizing that more subjective user comments are needed to extract a user's sentiment correctly enough to determine associations between the survey and sentiment data. This is in line with the results in Figure 5 and Figure 6 that showcase that users are neutral 59.4% of the time and positive 27.8% of the time. This also suggest the lack of subjectivity in the user comments in the audio recordings.

These results mean that more research would need to be done on correctly extracting sentiment from audio recordings, since the results imply that this is not possible with the current data set and methodology. A potential solution to this issue would involve asking users a list of predetermined survey questions aimed at evoking users' preferences related to each artwork. These questions would elicit more subjective answers than the thinking-out-loud process, which would result in the natural language processing techniques to have results that represent a user's sentiment. This might lead to the survey data and the sentiment analysis data being more equal. The bar chart in Figure 4 would then have more bars that are of similar height. In turn, resulting in that the chi-square test of more paintings would reject the null hypothesis indicating that associations between the survey and sentiment data can be determined.

The second reason for the discrepancy lies in the limitations of VADER. Despite its contextual understanding capabilities, VADER might still struggle to comprehend the context present in the user comments. This misunderstanding can lead to sentences being given the wrong sentiment score upon analysis. In the case of this research, it is a plausible reason, because a significant amount of user comments are quite vague which makes them contextually complex. This is once again due to the thinking-out-loud process and the fact that the sentences are transcribed from audio to text. These inaccuracies that follow from this transcription can make the context of a user comment's context harder to understand in general, and will especially affect VADER's accuracy.

Moreover, VADER's analysis might suffer from inaccuracies because it uses a sentiment lexicon to analyze sentences. This lexicon might be incomplete, meaning that it does not cover all possible words or phrases. For example, the word "colony" might not exist in the lexicon, while it does occur in the transcriptions. If these words *are* in the user comments, VADER might have a hard time giving an accurate sentiment for these sentences because it does not know what sentiment score to give to these missing words.

Both of these reasons can result in an inaccurate analysis of user comments, which leads to deviations in the average sentiments of user's towards specific paintings. Eventually, leading to too much difference between the survey data and sentiment analysis data to determine associations between them. However, this is not an issue that can be resolved at the moment. VADER is the most accurate library for sentiment analysis without having to pre-training data yourself [20]. Therefore, it is still the best option for this research.

The main arguments presented in the discussion align with the related work primarily in terms of the study's objective. The related work shows studies that have addressed similar challenges in extracting personal preferences from data related to personalized museum experiences. Matshoff et al. [1] propose the extraction of preferences from star ratings, while Amato et al. [4] leverages multi-agents planning methods to generate the museum routes, and Tsiropoulou et al. [5] suggest Quality of Experience-based (QoE) museum touring based on visitors' profiles. While these approaches differ from the one within the current research, they have all faced challenges in extracting personal preferences from data.

Similarly, the related work on personalization based on natural language processing has explored sentiment analysis in various domains. Zankadi et al. [7] uses sentiment to extract topical interest, wheareas Reuver et al. [8] uses it to diversify news recommendations, and Paik et al. uses it personalize email communication [10]. The current research aligns with these papers because it also showcases the value of personalization based on natural language processing techniques, especially in the museum domain.

7 CONCLUSIONS

This paper assesses the possibility of extracting personal preferences from audio data through speech-to-text transcription and analysis using natural language processing techniques. The findings indicate that the audio recordings used in this study cannot be used to determine a user's individual preferences towards the paintings as of now. This is a result of the thinking-out-loud process during the users virtual reality museum visit, which has elicited more objective than subjective comments about the paintings. These objective comments cannot be used to determine a user's individual preferences because they do not give information about a user's sentiment towards the paintings.

While the data set in this study does not contain enough information to extract users' personal preferences, this research still contributes to the understanding of *how* valuable information can be extracted from audio data collected in virtual reality settings. However, the effectiveness of the proposed methodology in this study has been found to be limited. Further research is required to develop a more reliable approach for extracting personal preferences from audio recordings.

This study employs part-of-speech tagging, noun phrases extraction, and sentiment analysis to extract valuable information from the audio data. Although the results of these natural language processing techniques revealed the possibility of extracting valuable information, they are mostly limited to objective aspects within the current data set. However, it is crucial to acknowledge that the effectiveness of these techniques in capturing subjective aspects, such as sentiment and personal preferences, remains limited within the scope of this study.

Future investigations should focus on refining the methodology to enhance the extraction of personal preferences from audio data. Furthermore, it is essential to evaluate the methodology's effectiveness by testing it on different data sets. Additionally, exploring alternative approaches to determining user sentiment can potentially lead to more conclusive results. Moreover, increasing the proportion of subjective comments in the data set could improve the accuracy of the sentiment analysis and enable the extraction of personal preferences.

To conclude, this research has contributed to understanding users' preferences in virtual reality museum experiences by analyzing audio recordings using speech-to-text transcriptions and natural language processing techniques. It has proven that it is possible to extract valuable information from audio recordings. However, this information can not be used to determine personal preferences for the current data set since the audio recordings do not contain enough subjective data. Therefore, further research is required to develop methods for effectively extracting personal preferences from audio data. Several potential fields of research are discussed in the future work section (chapter 8).

8 FUTURE WORK

The future work that follows from this research is to develop a more reliable methodology to extract users' personal preferences from audio recordings. One approach to achieve this is by eliciting subjective comments from users during their virtual reality museum tour. An option would be to use a virtual tour guide, that prompts users with questions that are designed to evoke subjective comments and opinions.

Moreover, the methodology of this study needs to be tested on different data sets to assess its effectiveness and generalizability. This will provide more insights into the applicability and accuracy of the methodology used in this study.

Furthermore, it would be good to further research alternative methods to extracting information from audio recordings too. This can involve using more advanced natural language processing techniques, or using machine learning models to improve the accuracy and depth of the analysis. By exploring different approaches, a more conclusive result can be given on determining whether it is possible to extract personal preferences from audio data.

This future research will allow the development of a more reliable process for extracting users' personal preferences from audio recordings. Improving the understanding and application of audio data analysis in virtual reality museum experiences.

REFERENCES

[1] J. Masthoff, B. Mobasher, M. C. Desmarais, R. Nkambou, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, and G. Weikum, Eds., *User Modeling, Adaptation, and Personalization: 20th International Conference, UMAP 2012, Montreal, Canada, July 16-20, 2012. Proceedings*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, vol. 7379. [Online]. Available: http://link.springer.com/10.1007/978-3-642-31454-4

[2] A. Antoniou, A. Katifori, M. Roussou, M. Vayanou, M. Karvounis, and L. Pujol-Tost, "Capturing the Visitor Profile for a Personalized Mobile Museum Experience: an Indirect Approach."

[3] Y. Wang, N. Stash, R. Sambeek, Y. Schuurmans, L. Aroyo, G. Schreiber, and P. Gorgels, "Cultivating Personalized Museum Tours Online and On-Site," *Interdisciplinary Science Reviews*, vol. 34, pp. 139–153, Sep. 2009.

[4] F. Amato, F. Moscato, V. Moscato, F. Pascale, and A. Picariello, "An agent-based approach for recommending cultural tours," Jan. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865520300027

[5] E. E. Tsiropoulou, A. Thanou, and S. Papavassiliou, "Quality of Experience-based museum touring: a human in the loop approach," *Social Network Analysis and Mining*, vol. 7, no. 1, p. 33, Jul. 2017. [Online]. Available: https://doi.org/10.1007/s13278-017-0453-2

[6] M. Huddar, S. Sannakki, and V. Rajpurohit, "A Survey of Computational Approaches and Challenges in Multimodal Sentiment Analysis," *INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING*, vol. 7, pp. 876–883, Jan. 2019.

[7] H. Zankadi, A. Idrissi, N. Daoudi, and I. Hilal, “Identifying learners’ topical interests from social media content to enrich their course preferences in MOOCs using topic modeling and NLP techniques,” *Education and Information Technologies*, vol. 28, no. 5, pp. 5567–5584, May 2023. [Online]. Available: https://doi.org/10.1007/s10639-022-11373-1

[8] M. Reuver, N. Mattis, M. Sax, S. Verberne, N. Tintarev, N. Helberger, J. Moeller, S. Vrijenhoek, A. Fokkens, and W. van Atteveldt, “Are we human, or are we users? The role of natural language processing in human-centric news recommenders that nudge users to diverse content,” in *Proceedings of the 1st Workshop on NLP for Positive Impact*. Online: Association for Computational Linguistics, Aug. 2021, pp. 47–59. [Online]. Available: https://aclanthology.org/2021.nlp4posimpact-1.6

[9] N. Gupta, G. Tur, D. Hakkani-Tur, S. Bangalore, G. Riccardi, and M. Gilbert, “The AT&T spoken language understanding system,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 1, pp. 213–222, Jan. 2006. [Online]. Available: http://ieeexplore.ieee.org/document/1561278/

[10] W. Paik, S. Yilmazel, E. Brown, M. Poulin, S. Dubon, and C. Amice, “Applying natural language processing (NLP) based metadata extraction to automatically acquire user preferences,” in *Proceedings of the 1st international conference on Knowledge capture*, ser. K-CAP ’01. New York, NY, USA: Association for Computing Machinery, Oct. 2001, pp. 116–122. [Online]. Available: https://doi.org/10.1145/500737.500757

[11] D. Javdani Rikhtehgar, S. Wang, H. Huitema, J. Alvares, S. Schlobach, C. Rieffe, and D. Heylen, “Personalizing Cultural Heritage Access in a Virtual Reality Exhibition: A User Study on Viewing Behavior and Content Preferences,” in *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*. Limassol Cyprus: ACM, Jun. 2023, pp. 379–387. [Online]. Available: https://dl.acm.org/doi/10.1145/3563359.3596666

[12] Google Cloud, “Speech-to-Text: Automatic Speech Recognition.” [Online]. Available: https://cloud.google.com/speech-to-text

[13] J.Y. Chan and H.H. Wang, “Speech Recorder and Translator using Google Cloud Speech-to-Text and Translation | Journal of IT in Asia,” Dec. 2021. [Online]. Available: https://publisher.unimas.my/ojs/index.php/JITA/article/view/2815

[14] Python Software Foundation, “Python Programming Language,” Jun. 2023. [Online]. Available: https://www.python.org/

[15] Sketch Engine, “POS tags and part-of-speech tagging | Sketch Engine,” Mar. 2018. [Online]. Available: https://www.sketchengine.eu/blog/pos-tags/

[16] NLTK, “NLTK :: Natural Language Toolkit.” [Online]. Available: https://www.nltk.org/

[17] M. Wang and F. Hu, “The Application of NLTK Library for Python Natural Language Processing in Corpus Research,” *Theory and Practice in Language Studies*, vol. 11, no. 9, pp. 1041–1049, Sep. 2021, number: 9. [Online]. Available: https://tpls.academypublication.com/index.php/tpls/article/view/1400

[18] WordCloud for Python, “WordCloud for Python documentation — wordcloud 1.8.1 documentation.” [Online]. Available: https://amueller.github.io/word_cloud/

[19] NLTK, “NLTK :: nltk.sentiment.vader.” [Online]. Available: https://www.nltk.org/_modules/nltk/sentiment/vader.html

[20] C. Hutto and E. Gilbert, “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text,” Jan. 2015.