Designing Multimodal Video Search by Examples (MVSE) user interfaces: elicitation of UX requirements and insights from semi-structured interviews

Kyle Boyd
Ulster University
School of Art
Belfast, UK
ka.boyd@ulster.ac.uk

Patrick McAllister
Ulster University
School of Computing
Belfast, UK
p.mcallister@ulster.ac.uk

Maurice D Mulvenna
Ulster University
School of Computing
Belfast, UK
md.mulvenna@ulster.ac.uk

Raymond Bond
Ulster University
School of Computing
Belfast, UK
rb.bond@ulster.ac.uk

Hui Wang
Queen's University Belfast
School of Electronics, Electrical Engineering and Computer Science
Belfast, UK
h.wang@qub.ac.uk

Ivor Spence
Queen's University Belfast
School of Electronics, Electrical Engineering and Computer Science
Belfast, UK
i.spence@qub.ac.uk

Guanfeng Wu
Queen's University Belfast
School of Electronics, Electrical Engineering and Computer Science
Belfast, UK
g.wu@qub.ac.uk

Abbas Haider
Queen's University Belfast
School of Electronics, Electrical Engineering and Computer Science
Belfast, UK
a.haider@qub.ac.uk

Rob Cooper
BBC Research & Development
AI Research
London, UK
rob.cooper@bbc.co.uk

Andrew Wood
BBC Research & Development
London, UK
andrew.wood1@bbc.co.uk

ABSTRACT

In order to search for content from large video archives, it is typically undertaken via keyword queries using predefined metadata such as title and other tags. However, it is difficult to use keywords to search for specific moments in a video. Video search by examples is a desirable approach for this scenario as it allows users to search for content using one or more examples without having to specify a keyword. However, video search by examples is notoriously challenging, and performance is poor. To improve search performance, multiple modalities may be considered – image, sound, voice and text, multiple search cues could be used to identify more relevant content. This is multimodal video search by examples (MVSE), where users can search for content using multiple modalities. In this paper, typical end users - BBC archivists, programme support staff - are interviewed to identify how their search needs can be addressed with the technical capabilities of a MVSE tool. Such a search tool will be useful for organisations such as the BBC who maintain large collections of video archives and want to provide a search tool for their own staff as well as for the public. It will also be useful for companies such as Youtube who host videos from the public and want to enable video search by examples. The study's objectives explored in this paper were to inform the design and development of the UX workflows to gain a broader understanding of what opportunities and issues may arise from the proposed prototype tool. Results from the thematic analysis was highlighted 4 main themes: Opportunities, Time constraints, Activities, and Pain points. Further analysis highlighted key areas that should be considered for an MVSE-based system, such as scene recognition, face recognition, speed issues, and integration. .

ECCE '23, September 19–22, 2023, Swansea, United Kingdom

ACM ISBN 979-8-4007-0875-6/23/09.
https://doi.org/10.1145/3605655.3605665

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**.

KEYWORDS

Multimodal video search, Multimodal Video Search by Examples, Interfaces, UX, Requirements elicitation, Semi-structured interviews

ACM Reference Format:
Kyle Boyd, Patrick McAllister, Maurice D Mulvenna, Raymond Bond, Hui Wang, Ivor Spence, Guanfeng Wu, Abbas Haider, Rob Cooper, and Andrew Wood. 2023. Designing Multimodal Video Search by Examples (MVSE) user interfaces: elicitation of UX requirements and insights from semi-structured interviews . In *European Conference in Cognitive Ergonomics (ECCE '23), September 19–22, 2023, Swansea, United Kingdom.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3605655.3605665

1 INTRODUCTION

Videos are being generated in large numbers and volumes. Typically, videos are indexed by predefined metadata such as titles, tags, and viewer notes, making them searchable by keywords. Commercial video search engines such as YouTube, Vidrovr, Panopto, and IronYun are all keyword-based. Using these engines, we can search by any word spoken or displayed on-screen, or by traditional metadata. However, it is challenging to use keywords to search for specific moments in a video where a particular speaker discusses a specific topic at a certain location. Moreover, most videos have little or no metadata, and automatic metadata extraction is not yet sufficiently reliable. Therefore, video search by keywords has limitations. Video search by examples is a desirable alternative, as it allows users to search for content using content they already have. To improve search performance, multiple modalities should be considered, such as face, voice, context (or setting), and topic. Each modality provides a separate search cue, and multiple cues together should more accurately identify relevant content than any individual cue alone.

After a series of co-creation [6] meetings with the BBC Research & Development team, a general use case has been identified – "locate any half-remembered clip from any half-remembered TV programme in 30 seconds or less" and a query type is "locate clips with person X in setting Y talking about subject Z". This query is difficult to answer by keyword-based search, but could be answered by searching with a face image of X, a scene image of Y and a text phrase of subject Z. Our proposed solution is multimodal video search by examples (MVSE), where the modalities are person (face or voice), context and topic.

Presently, there is no tool or commercial service available for multimodal video search by example that is used by the BBC. For example, current search tools currently used by for BBC Archives rely on single-modality text-based search (either metadata, dates, or transcripts of programmes). This single modality presents problems, for example, users are often interested in finding footage of a notable speaker talking about a subject or searching for certain scenery (e.g. picture of a snow covered hill, or a particular scene). However, a keyword search of the transcript and metadata will often bring back examples of journalists or presenters talking about that subject rather than the person themselves. A multimodal search aims to address these problems through using a multimodal search mechanism.

2 RELATED WORK

To design and develop the MVSE system, co-creation approaches have been used to determine requirements, use-cases, and to identify appropriate UX approaches. Co-creation approaches have been extensively employed to inform the design and development of various systems [6]. Co-creation strategies offer several key advantages [8]. For instance, there is a greater chance of enhanced innovation, as involving the intended end-users in the design process can create an environment that promotes greater innovation, stakeholders can highlight diverse ideas and share experiences that designers or developers might not be aware of or understand. This invaluable insight ensures that the product is specifically designed and developed to meet the user's needs. Issues relating to misunderstandings can be discussed, and when coupled with agile approaches, the risk of creating a product that does not meet the end-user needs is significantly reduced.

Risk mitigation is an important advantage of using co-creation approaches. Potential end-users can identify risks and other issues that may affect the design and development of a system, particularly areas that the designer or developer might not be aware of. Access to this information can help prevent future problems, ultimately contributing to the project's overall success. In this study, our goal was to use semi-structured interviews [7] to illicit requirements. Using semi-structured interviews is a useful strategy to inform and support application design and development for several reasons, the most important being that it provides a flexible way to draw out rich detailed data from the interviewee, the conversation can be adapted depending on the interviewee's response to gain a deeper understanding of their perspective.

Semi-structured interviews are open-ended and there is the potential to gain a deeper understanding of the interviewee's perspective. Semi-structured interviews prove to be beneficial in dealing with intricate matters as they allow the use of probes to investigate, and clarify responses to inquiries [9]. In [5] researchers used semi-structured interviews with experts to develop a touch-based screening instrument for dementia. Expert participants who were knowledgeable in the areas of dementia and neuropsychological assessment examined 'DemSelf' prototype, they were able to identify various usability issues that were present [5]. Other works [3] used semi-structured interviews with refugee communities to inform the design and development of a healthcare app, this research used surveys, and semi-structured interviews to assess the effectiveness in adopting a human-centred design approach in an application. It is clear from the literature that semi-structured interviews is a valuable research technique that allows for flexible, in-depth exploration of a topic. Open-ended questions can encourage participants to share detailed experiences and perspectives, providing rich qualitative data. The format enables interviewers to probe further into interesting areas or clarify responses. Semi-structured interviews can also yield insights that may not emerge from structured interviews or questionnaires, in supporting app design and development [1-5].

3 AIM & OBJECTIVES

The aim of this study is to elicit requirements from journalists, archivists and programme production staff working at the BBC using 1 to 1 semi-structured interviews. These interviews also sought to understand the current approaches and tools journalists currently use. Additionally, during the interviews, design wireframes of the proposed system were used to gather user feedback. Several

objectives have been identified to guide this research. (1) To conduct user research with least 10 participants to gain insights into their current approaches for researching historical video archives. (2) To design UX wireframes for researching video archives that are intuitive and user-friendly. (3) To identify potential opportunities and issues that may arise from the semi-structured interviews. (4) To inform the design and development of the UX workflows and a user's 'happy path', whereby their search workflow is successful.

To further help guide this research study, several research questions were identified;

(1) What search approaches are currently used by BBC journalists, archivists, and programme makers in searching archives? The aim of this question is to elicit what BBC researchers and archivists currently use in researching for information or assets for programmes or feature articles. It was anticipated that the answers to this question would include workflows and the technology they use.
(2) What are the proposed approaches to be used by BBC journalists, archivists, and programme makers in searching archives using MVSE?
(3) How would the proposed MVSE system benefit end-users in carrying out their programme research activities/roles? After discussing the proposed MVSE system and what use cases, users are welcomed to discuss how the MVSE would benefit them personally in their duties and workflows. This question is important to highlighting further use cases and to feed these into the functional and non-functional requirements in future software development iterations.
(4) What usability features should be included in the developed solution? It is important to gauge intended user's opinions regarding the usability of a system and what UI components should be included to
(5) How useful and usable are the proposed prototypes?

The remainder of this paper discusses the approaches used to gather and analyse participant interviews. Results section highlights the themes and issues that were discussed during the semi-structured interviews and the discussion section highlights key take-home messages, functional and non-functional recommendations. Finally, the conclusion and future work summarises the results and presents additional future work to be undertaken.

4 METHODOLOGY

4.1 Participants

A convenience sample of participants from the BBC were invited by email to undertake the 1-to-1 user semi-structured interview. Interviews took place using Microsoft Teams and lasted between 45 minutes and 1 hour. Participants were interviewed by 1 or 2 researchers from Ulster University. The participants were given an information sheet and signed consent was obtained before the participant took part in study.

4.2 Semi-structured Interviews

Semi-structured interviews consisted of open-ended questions with the objective of understanding participants' current tools and approaches used to search media archives for assets. Participants were

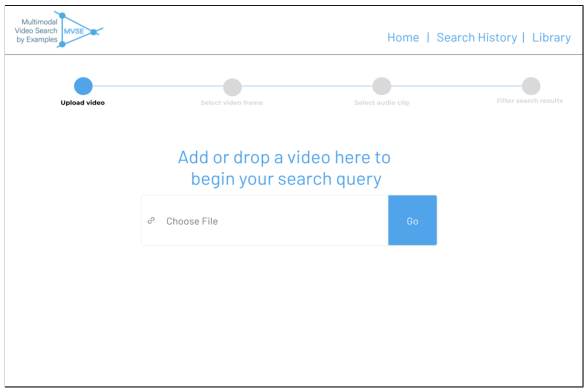

Figure 1: Wireframe of Multi-page: Adding a video.

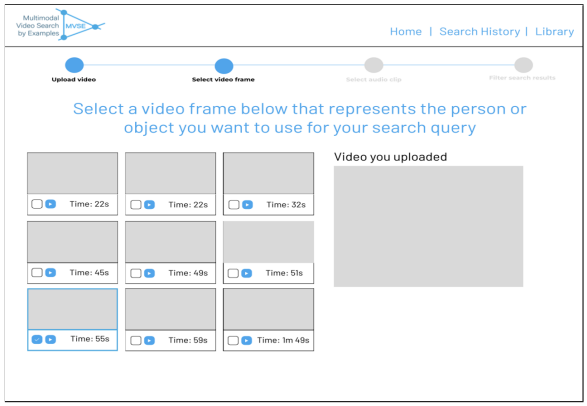

Figure 2: Wireframe of Multi-page: Select a video frame.

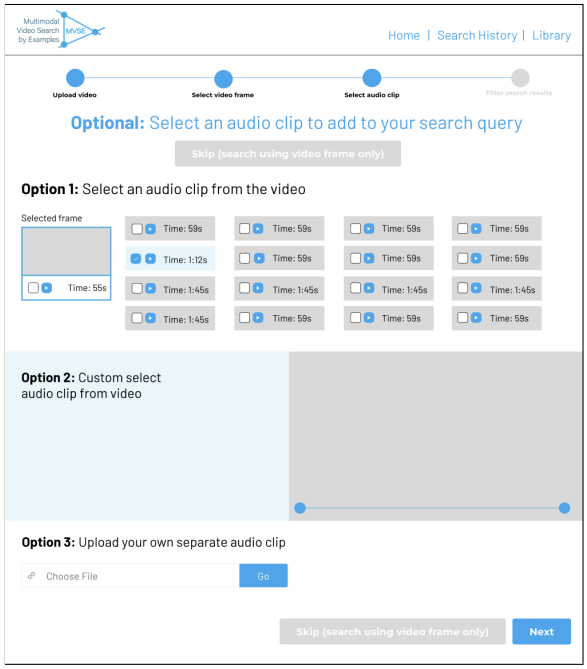

Figure 3: Wireframe of Multi-page: Selecting audio clip.

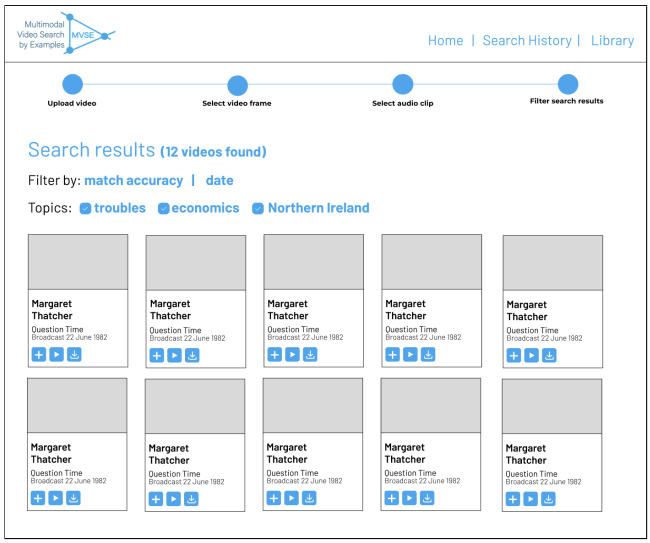

Figure 4: Wireframe of Multi-page: Viewing results.

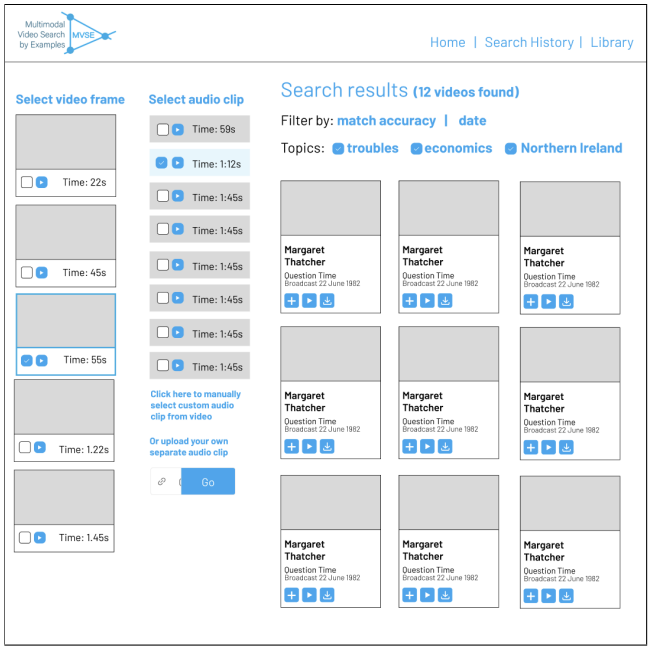

Figure 5: Wireframe of Single-page: Selecting video, audio and viewing results.

asked about their relationship with technology and their opinions of incorporating technology into their role (Table 1). Participants were asked about their current workflows, the stages they go through when searching for assets and media, the issues they have and if there are any opportunities to improve the service using a multimodal search approach. Purposed system wireframes were also showcased to promote discussion and to provide further context for the interviews. Each interview was video and audio recorded and transcribed for thematic analysis using Microsoft Teams software. Table 1 is a list of questions that formed the basis of the semi-structured interview.

4.3 Thematic Analysis

Thematic analysis (TA) is a qualitative method of analysing data that involves organising, describing, and reporting common themes within some text, such as interview transcripts. TA can be used to answer research questions that aim to explore people's views, opinions, and experiences on different topics. TA was used to analyse the interview transcripts to help answer the research questions. TA was beneficial when working with large transcriptions, as researchers could group statements and keywords together to uncover themes. Themes were generated based on the transcribed interviews, and were highlighted to showcase patterns to generate new insights to support the design and development of the MVSE system.

4.4 Wireframes

After the interview stage, we showcased two high-fidelity wireframes of the proposed MVSE system, both as a multi-page layout (Figure 1-4) and as a single page layout (Figure 5). Explaining how both would work and the functionality components that would be present on each page. Participants were also asked which wireframe they liked best along with their likes and dislikes and a discussion relating to the opportunities that the wireframe would present in achieving multimodal search. Figures 1-5 are examples of the wireframes showcased to the participants.

Table 1: A Selection of Interview Questions

Q1	What types of technologies (hardware or software) do you use to support your current job role?
Q2	What devices do you normally use in work to carry out duties?
Q3	What is the biggest pain point when carrying out your research/tasks under current practices in regards to the technologies you use? E.g. difficulties
Q4	How would you normally get information to support your role, what online systems do you internally use in the BBC, what search systems?
Q5	Are there any difficulties that may arise when searching for related content or using that system?
Q6	How long would it take to research related content using current practices?

5 RESULTS

Once the interviews were completed, these video files were uploaded to Dovetail[1] for thematic analysis. Dovetail allows videos to be transcribed and analysed by highlighting common themes in the interviews, these themes can be tagged to help focus key insights from the interviewees. Figure 6 showcases the main discussion points that were highlighted during the interviews after using Dovetail for analysis.

[1]Dovetail https://dovetail.com/

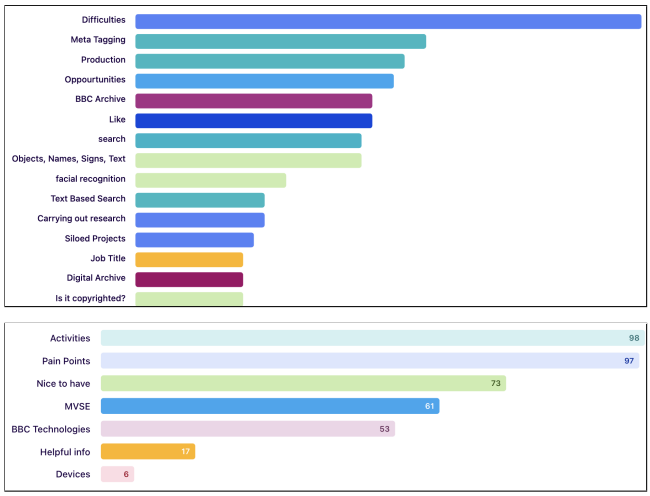

Figure 6: Dovetail analysis of transcripts of interviews.

After further analysing the transcripts, 4 themes began to emerge;

(1) Opportunities
(2) Time constraints
(3) Activities
(4) Pain points

Different examples and use cases within these themes were highlighted by participants during the interview. Participants frequently mentioned areas around production, meta-tagging, BBC archive, and objects, names, signs, and text. The core activities highlighted by the participants were meta-tagging, production, and asset search. Many of the participants' core activities are categorised into two areas: archiving material and finding material for production teams. Throughout the interviews we quickly realised that BBC Archive is a rich source of material, with a vast amount of assets available that heavily relies on previous meta-tagging and cataloguing processes. However, locating specific objects, signs, and locations (scene recognition) emerged as one of the most challenging tasks. This challenge was further heightened by the need to conduct copyright checks on the retrieved assets. According to participants, it is often difficult to ascertain the source of some materials and whether they can be used or not, or if they are available via Creative Commons licensing. The paperwork required for logging rights checks is extensive and can be challenging. The remainder of this section discusses how these 4 themes were highlighted in different areas and technologies.

5.1 Scene Recognition

A significant use case emphasised by participants was that, although many of the interviews can locate people, the more challenging requests involve finding objects, names, locations, and scenes. Scene recognition has the potential to make it easier to identify and locate specific scenes, objects, logos, or locations within the BBC archive. The use of scene recognition functionality can enhance content curation by easily identifying scenes, objects, or locations that are visually or contextually similar, making it easier to create thematic collections or compilations.

"…But then later on you're like, there's a particular building on that street that's maybe been renovated or knocked down or burnt down or whatever it is. And you're desperately trying to find that building, but no one ever tags that building cuz it's just some building…" **P9**

"You'd be surprised. We don't, news would want people a lot, but the things we get asked for a lot are like scene setting. So like, oh what model car is this or this high street, this kind of buildings, London skylines, empty shots like this seas or quite a lot of evocative moods stuff, which is a lot harder to do. Like eerie calm, right?" **P1**

"it's been able to find an object, object recognition, whatever that happens to be a face, a thing, you know, a car, a place." **P3**

5.2 Face Recognition

A reoccurring use case that several participants highlighted was the opportunity to incorporate face recognition. Incorporating facial has the potential to improve search retrieval processes, such technology would be able to locate specific people across different time periods and different locations.

"That's where it is. I think that's where the facial thing is more useful actually, is those lesser known people at the time who are now well known…" **P9**

"we've had pictures of people who aren't well known or wouldn't, wouldn't really have had their names tagged and things that we've kind of put into the system just to see what crops up and there has been the odd thing we found as a result of that…" **P9**

"I've always thought anything with sort of image recognition would be useful cuz we've never really had that…" **P4**

"…if it's got facial recognition, that's quite interesting as well because contributors are not always locked as well…." **P4**

"…It was more, what if we have footage of somebody who in the future became hugely important? What if we have them from the fifties or sixties talking about it?..." **P5**

5.3 Speed

Speed was also an issue that was highlighted by participants. Participants noted that the MVSE system should be fast and that they work on various projects that are require a quick turnaround in regards to finding appropriate assets. Participants stated that they work quickly to meet deadlines, such a system would need to operate quickly to effectively meet the needs of the project.

"And also speed, like how quickly will it produce the result? Cause obviously this is a time saving exercise to help people find assets, but if it's gonna take a long time to chug through, then obviously that's, that's, that's, that's not not ideal." **P5**

"So how quickly can your system search those million items to bring back what this thing is a kid with a gun standing in his hand in the

street. So, you know, how quickly can that happen?" **P3**

"Yeah, I could, I could do that in 10 to 10 minutes, half an hour…" **P4**

5.4 MVSE Integration

Another important suggestion was how the proposed MVSE system would interact with current archive systems. Rather than another new system, many interviewees stated that MVSE should sit within the current BBC Archive as a feature or other form of search. The MVSE system should be integrated into the current asset databases as currently there is already several search asset systems that participants have to use.

"I kind of like the idea of this system if a system like this sitting above all of them because sometimes if I want Tony Blair it, I might want to look at a broadcast from the news from 1997, whatever. But I also want a shot of him from a documentary that factual have made…" **P1**

"Can it do multiple siloed systems as well as our, can it look at Jupyter as well? Can it, can it look here, can it look there, can it look here? Kind like their account, look at the Scottish digital archive…" **P3**

"Basically that it's integrated into an archive search…" **P2**

5.5 Copyright Issues

Participants also highlighted the need to search archives based on copyright restrictions. Participants stated that having the ability to know what assets are copyrighted and where they have been used in previous searches or programmes would be useful as it would narrow what assets could be freely used or what assets need further approval. This could save time with the onerous task of copyright administration and determining if footage can be used in programmes.

" Or this is, this is, you know, this is free to use, this is, this is, you know, sort of creative common stuff, public domain or it's BBC no problem. But all the paperwork has to be done afterwards for a, for a factual production. So obviously all those details need to be passed along as well." **P3**

"I don't know if you could do anything around the copyright in terms of the metadata or whatever..." **P2**

"…there will be a way to look into the archive and it will also give you a sense of copyright holders, you know, where this material came from .... And who is the likely copyright holder..." **P5**

5.6 Wireframe Feedback

Participants also provided feedback on 2 high-fidelity wireframes; a multi-page wireframe and a single page wireframe. Participants stated that they would prefer to use the single page system to conduct multimodal video searches;

"I immediately thought I would use the single page view because I can see the results faster ..... Without having to go forwards and backwards between the pages…" **P1**

"I prefer the second this single page because you feel like the first page is obvious. Drop your video into it and then everything I need is on one…" **P1**

"So every time that's a second click through to a page, it, it, that's more time for me…" **P1**

"I think I'd prefer the single page approach. Okay. Or the based, rather than having to flip back and forth. Now I understand that some of the thing, you might get more detail on the other, but I think for this I would definitely prefer the one page approach…" **P2**

"I would prefer that [single page wireframe]. But then cuz I'm, you know, we, I dunno, I can't comment for all people in tv but you're always like, you're too busy and too lazy to do lots of multiple clicks…" **P4**

6 DISCUSSION

This paper discusses the results of interviews conducted with 10 participants who worked at the BBC and worked within archival asset management and retrieval. The interviews were conducted over Microsoft Teams and were video recorded for transcription. The video files were uploaded to Dovetail for thematic analysis. The findings highlights four themes that emerged from the interviews: Opportunities, Time constraints, Activities, and Pain points.

Within these themes, various discussion points were often repeated, such as production, meta-tagging, BBC archive, and objects, names, signs, and text. The participants' core activities were categorised into two areas: archiving material and finding material for production teams. The participants emphasised that the BBC Archive is a rich source of material, but locating specific objects, signs, and locations (scene recognition) is one of the most challenging tasks, further heightened by the need to conduct copyright checks on the retrieved assets.

Analysis of the interview transcripts highlighted 4 technical elements that could be incorporated into MVSE system to support users; scene recognition, face recognition, speed, and MVSE integration. Scene Recognition has the potential to make it easier to identify and locate specific scenes. Staff at BBC stated that scene recognition would be useful to locate specific scenes to use in programmes. Scene recognition has the potential to increase efficiency through automating the process of identifying assets to speed up the process, another related use-case that participants highlighted was the ability to use media from different sources and to use this as an example to search for similar assets in the BBC archive. Scene recognition models can also be used to accurately tag current BBC assets, the combination of using AI for automatic tagging models with manually meta-tagging already completed, along with speech-to-text capability has the potential to result in improved precision in locating resources.

The ability of using objective detection to find objects, logos within assets was also highlighted. Incorporating facial recognition has the potential to improve search retrieval processes, as technology would be able to locate specific people across different time

periods. Speed was also an issue that was raised with participants. Several participants stated that the MVSE system should be fast and that they work on projects that require immediate retrieval of assets. Many interviewees agreed that MVSE should sit within the current BBC Archive as a feature or other form of search, rather than another new system.

Participants also highlighted the need for the MVSE system to integrate within the current BBC archive system. It was highlighted during the interviews that the BBC currently have several asset databases and along with databases for different regions in the UK. Participants stated that the proposed system should be integrated into these current systems, this would promote ease of use and consistency due to familiarity of asset storage structure. Users would benefit from this interoperability as built-in AI functionality could be used to find unstructured media assets that are not currently catalogued. Furthermore automatic AI approaches could be employed to enrich current meta-tagging for assets for accurate descriptions and keywords.

In regards to the wireframes showcased (single and multi-page wireframes Figures 1-5), the majority of participants preferred the single page wireframe. Participants highlighted that having a single page application to search for assets would enable them to edit search criteria and see the results within the same page. This would increase efficiency as the users would not have to traverse multi-page to find relevant assets. Participants stated that they would need to find assets very quickly for programmes or news segments, therefore having a single page interface to interact with the BBC archive using MVSE technology would be preferential. Also in regards to usability, it is important to ensure the developed solution is user-friendly, it should include several usability features, for example, the ability to search for existing video content from mainstream outlets like YouTube would be beneficial, this would enable users to search for similar content already in the archive. Moreover, users should be able to edit video search result clips and download them, as well as have a stored search history which they can return to and edit. Furthermore, participants appreciated the ability to edit video search results clips and the option to have a stored search history. Table 2 lists the recommendations that should be considered when designing and developing a multimodal search system, including the ability to search for scenes or imagery based on different examples, the importance of incorporating facial recognition functionality to retrieve assets and to enhance metadata, as well as addressing issues relating to speed and optimising algorithms, considering user-centred design, copyright along with incorporating search history and asset caching.

Table 2: List of Recommendations for Multimodal Search by Examples

	Theme	Recommendation
R4	Scene Recognition	Having the ability to search for certain scenes or imagery based on examples from other sources. Using images and videos from various sources as an example to search for similar content.
R2	Facial Recognition	Being able to recognise individuals, facial recognition can enhance and add value to metadata. Facial recognition can support historical research and give archivists the opportunity to uncover assets that otherwise may not be found.
R1	Speed	Attention needs paid to the speed of a multimodal search engine such as MVSE, the complexity of search queries increase due to the combination of using different search modalities. Algorithmic optimisation mechanisms would need to be considered to ensure optimal performance. The process of caching data that user's have previously search for needs to be considered. Regular performance monitoring is needed to identify bottlenecks and to highlight areas of improvement. Executing multiple processes simultaneously so the user is not waiting for longer periods of time (e.g. executing AI functionality in the background as the video is being uploaded to be used as search input).
R3	Integration	Integrating a multimodal search system within the current archive would enable users to search for content using the content already catalogued and metatagged, and enable more accurate results due to the combination of modalities. The combination of different modalities would give the system greater contextual awareness in comparison to relying soley on textual input combined with boolean operators.
R5	User-centred design	Crucial to include key stakeholders and potential end-users in the design and development process to understand what types of modalities are being used. Important to understand the nature of the assets being retrieved, in regards to data size, length, and different use-cases undertaken by potential end-users.
R6	Copyright	Ability to determine the copyright rights of a particular asset.
R7	Search History	Ability to have a saved search history to repeat searches.

7 CONCLUSION & FUTURE WORK

The BBC's archive of audiovisual material is a valuable resource for journalists, archivists, and program makers. However, searching for specific content within this archive can be challenging and time-consuming. To address this issue, the BBC is exploring the use of deep learning in the form of a multimodal search system. This would enable staff at BBC to search for related content using facial recognition or scene/object recognition, MVSE propose to utilise state-of-the-art deep learning architectures such as convolution

neural networks and vision transformers which have shown excellent performance in image classification, scene recognition, and facial recognition. The integration of these deep learning models along with multimodal search queries seek to enable more accurate asset retrieval along with the ability to search for copyrighted material and access to content that has been catalogued in the BBC archive. These findings can be used to guide future research and development and UX requirements of the MVSE system and improve the efficiency and effectiveness of archival asset management and retrieval within the BBC, In regards to future work, it is anticipated that a focus group will be held with the same participants where the developed software prototype will be showcased. Participants will get the opportunity to provide further feedback on the functionality and usability. Further work will investigate the use incorporating other UX evaluation approaches to gain quantitative insights into the usability of the prototype such as system usability study (SUS), Eye-tracking (to measure user attention and engagement), metric analysis (task completion rates, time-on-task, and clickstream analysis).

ACKNOWLEDGMENTS

Multimodal Video Search by Examples (MVSE) is a three-year research project funded by EPSRC (EP/V002740/2) and undertaken by a team from Queen's University Belfast, Ulster University, University of Surrey, University of Cambridge, and the BBC. Many thanks to the BBC journalists, archivists, and programme makers who consented to be interviewed for this study.

REFERENCES

[1] Antonia Barbaric, Catalina Munteanu, Heather Joan Ross, and Joseph Antony Cafazzo. 2022. A Voice App Design for Heart Failure Self-Management: A Pilot Study. In *medRxiv*.

[2] Ana Maria Barberia, Moira J. Attree, and Christopher Todd. 2008. Understanding eating behaviours in Spanish women enrolled in a weight-loss treatment. *Journal of clinical nursing* 17 7 (2008), 957–66.

[3] Rebeccah Bartlett, Jacqueline Anne Boyle, Jessica Simons Smith, Nadia N Khan, Tracy Robinson, and Rohit Ramaswamy. 2021. Evaluating human-centred design for public health: a case study on developing a healthcare app with refugee communities. *Research Involvement and Engagement* 7 (2021).

[4] E. Blomkamp. 2018. The Promise of Co-Design for Public Policy. *Australian Journal of Public Administration* 77 (2018), 729–743. https://doi.org/10.1111/1467-8500.12310

[5] Martin Burghart, Julie Lorraine O'Sullivan, Robert P. Spang, and Jan-Niklas Voigt-Antons. 2021. DemSelf, a Mobile App for Self-Administered Touch-Based Cognitive Screening: Participatory Design With Stakeholders. In *Interacción*.

[6] Interaction Design Foundation. n.d.. *Co-creation.* https://www.interaction-design.org/literature/topics/co-creation Accessed: June 20, 2023.

[7] Interaction Design Foundation. n.d.. *Semi-Structured Interviews.* https://www.interaction-design.org/literature/topics/semi-structured-interviews Accessed: June 20, 2023.

[8] J. Torfing, E. Sørensen, and A. Røiseland. 2019. Transforming the public sector into an arena for co-creation: Barriers, drivers, benefits, and ways forward. *Administration & Society* 51 (2019), 795–825.

[9] Chauncey Wilson. 2013. *Interview techniques for UX practitioners: A user-centered design method.* Newnes.