White Paper for NEH grant PR-263888-19

Accessible Civil Rights Heritage Project (2018-2023)

This white paper details the development and current NEH-funded research goals of The Media Ecology Project (MEP), directed by Prof. Mark Williams and designed by Dr. John Bell at Dartmouth. The virtuous cycle of access, research, and preservation that MEP realizes is built upon a foundation of technological advance (software development) plus large-scale partnership networks with scholars, students, and institutions of historical memory such as moving image archives. Our Onomy vocabulary tool and NEH-funded Semantic Annotation Tool (SAT) [discussed below] have been applied in this NEH advancement grant from the NEH regarding accessibility to archival civil rights newsfilm.

MEP is fundamentally 1) a sustainability project that 2) develops literacies of moving image and visual culture history, and 3) functions as a collaborative incubator that fosters new research questions and methods ranging from traditional Arts and Humanities close-textual analysis to computational distant reading. New research questions in relation to these workflows will literally transform the value of media archives and support the development of interdisciplinary research and pedagogy/curricular goals (e.g., media literacy) regarding the study of visual culture history and its legacies in the 21st century.

Our moving image heritage is at enormous risk. Moving image archivists and digital repository advocates are developing solutions to these problems, but we cannot sustain interest in “preservation” without a better sense of the historical value of these materials. “Access” is not enough; new knowledge production is required in order to connect archival materials with audiences and accelerate preservation efforts. The Digital Humanities must move concertedly forward to engage visual culture with the same dedication and technological ingenuity it has brought to the study of word culture.

With internal support at Dartmouth and especially support from the National Endowment for the Humanities, MEP has developed several digital tools that support and sustain the creation of new networked scholarship and pedagogy about archival moving image materials. These include:

> The Semantic Annotation Tool (SAT), which enables the creation of time-based annotations for specific geometric regions of the motion picture frame.

> Onomy.org, which is a vocabulary-building tool that helps to grow and refine shared vocabularies for tags applied to time-based annotations.

Together, these two tools support close contextual analysis of moving pictures based on time-based annotations. Annotations denote a start time and stop time for a subclip, a description and tags related to that clip, and attribution for its creator. This granular approach to media literacy and scholarly annotation is flexible enough to be applied to many types of research and analysis.

MEP promotes the study of archival moving image collections, enhances discovery of relevant corpora within these archives, and develops cross-disciplinary research methods. These efforts help ensure the survival of these collections via new published scholarship, plus contributions of
metadata and research on studied corpora back to the archival community — what we refer to as a virtuous cycle.

While developing MEP as a rather distinctive Digital Humanities project, we have learned first-hand several key lessons about this important and emerging field. Because we are building MEP from an Arts and Humanities perspective, we recognize that our goals must always be framed to raise awareness about the significance of cultural-critical perspectives within the various institutions that we have engaged (archives, libraries, universities, grant resources, etc.).

Everyone who engages in MEP is at some level working outside their comfort zones: across disciplines, across expertise, across vocabularies. In a very real sense we are engaged in “translation” work, the great benefit of which can be experimentation regarding methodologies of study but also in infrastructural designs of work-flow and output. New research questions in relation to these workflows will literally transform the value of media archives and support the development of interdisciplinary research and curricular goals (e.g., media literacy) regarding the study of visual culture history and its legacies in the 21st century. These goals have grown to be especially timely: the conceptual and ethical significance of re-imagining our collective purchase on historical imagination has been axiomatic to the emergent blur of the production and reception processes defined by what is called the “attention economy”.

In 2018 Prof. Williams and Dr. Bell were honored to receive two advancement grants from The National Endowment for the Humanities for The Media Ecology Project. These grant projects had each been initially developed as demo pilots for MEP and are now poised to realize significant advances in Digital Humanities scholarship via further developments of SAT, both technologically and conceptually.

Our MEP pilot on historical news materials (newsreels, news telecasts, newsfilm, and other associated footage) was dedicated to new scholarship on news materials from multiple archives. We gradually honed this topic and its participants into a focused address to Civil Rights content, with a double purpose of developing new scholarship plus a dedicated line of research and development to enable access of these semantically rich and complex materials for blind and visually impaired (BVI) users.

The term “newsfilm” has evolved in relation to historical changes in media technology and media formats. Television newsfilm evolved after decades of motion picture newsreel and news magazine production that was intended for exhibition in motion picture theaters on a weekly or bi-weekly basis. Local television newsfilm—often shot on site by local television station news crews that only broadcast a fraction of what they recorded—is a largely untapped source of local and national history that captured powerful moments throughout the emotionally and politically charged American civil rights era. Television newsfilm was produced in a different media industry context and was intended for “exhibition” to domestic audiences on an almost daily basis in the U.S. for many decades. It was regularly produced by both local television stations and network television news divisions.

A key context significant to most television newsfilm collections is local television itself, which is a conspicuous, persistently ignored aspect of U.S. media history, even though the local station is the backbone and the condition of possibility for the dynamics of U.S. network television.
Local television history is a common site of disavowal regarding many media histories, especially work on U.S. media history. As such, it can be the site of significant resultant capacities for historiographic depth and complexity. What is often truly compelling about sophisticated historical research is the relationship between the already-understood and what we think we know, versus the capacity to interrupt given history, perhaps even to intervene in that history. Local and network television newsfilm features the full spectrum of these historiographic capacities.

Television newsfilm was a primary form of extended news coverage for roughly forty years, from the evolution of television in the 1940s to the gradual adaptation to video formats in the 1980s. Much of the footage to be considered for this study is what is sometimes termed “raw” newsfilm: the footage that existed in-camera before it was edited and repurposed by the news professionals at a station or network. “Newsfilm” also describes local and network “finished” news stories and news programs and documentaries that utilized footage from multiple sources, aired for domestic audiences and sometimes distributed via film prints to local, national, and international markets. Collections of television newsfilm are being preserved and curated today in many archival collections across the U.S. and around the world. Most of these collections are as yet unavailable for critical and historical consideration.

It is important to underscore that historical newsfilm from different eras and industrial contexts have become digitized and available for study and time-based annotation only very recently. Many of the newsfilm clips and raw footage materials engaged for this study will be seen by scholars for the first time. This is a burgeoning area for both scholarly and public interest, a site for critical awareness about the (mediated) past. Also distinctive to this project is that much of the newsfilm to be studied was never screened publicly. That is, in many instances this newsfilm footage has never before been part of the public sphere, and thus has never been considered by even casual historians in any field. It has existed outside of critical inquiry and scholarship that is devoted to social history and media history. From a cultural perspective, these are indexical materials that have not yet been in a position to be remembered, let alone forgotten.

The conditionally absented or fugitive aspects of these civil rights materials inspire awakenings of the historical imaginary, and we expect this material will become especially relevant within our very contemporary 21st century context of global demonstrations and calls for social justice. For example, the outcry over the brutal murder of George Floyd within the U.S. context of the “Black Lives Matter” movement occurred during the weeks and months of the genesis of this project, and represented what Raymond Williams referred to as a major shift in “structures of feeling” [Williams 1977]: affective relations between consciousness and social institutions that are strongly emergent and inflect palpable pressure on commonplace rationalizations and actions.

The historical and historiographic potential of these materials is both vast and substantial. The Accessible Civil Rights Heritage (ACRH) project will help to expand the discoverability of these historical materials for critical consideration, by developing scholarly practices in relation to archival practices that will enhance searchable access to these historically rich items that would otherwise continue to be isolated in archival and data silos and virtually unavailable for search of any kind.
The thousands of newsfilm elements made available for study in this project will serve as a representative sample of the ocean of television newsfilm collected in archives and historical societies across the U.S. The study of the newsfilm era and subsequent eras of news coverage (i.e., post-celluloid eras) will be significantly enabled by the development of workflows, protocols, scholarly methods, and augmented vocabularies/ontologies to be developed in this proposed study.

In order to consolidate these diverse materials we engaged archivist Becca Bender (Rhode Island Historical Society) to advise on the creation of a common metadata spreadsheet format, and veteran professional moving image cataloguers Kathy Christensen and Laura Treat to help develop an Onomy vocabulary specific to the purpose of annotating civil rights news footage. These new cataloging and access procedures will assist in the parallel development of innovating high-quality, meaningful experiences of the collection to BVI users.

Given the special concerns of close textual analysis and its importance to humanities researchers, it is critical that any toolset designed to support humanities research be developed with that specific application in mind. However, any existing collection of materials and scholarship would carry with it the limitations of the tools that were originally used to create it—vocabulary and metadata that was designed to fit into a particular schema.

For the purposes of ACRH research, annotations featuring close reading analysis of civil rights newsfilm will be merged with extant metadata from the contributing archives and scholarly essays that feature newly-generated time-based descriptions. The result will be a re-animation of sorts for these historical media documents where specific events and images are contextualized in relation to known descriptors and vocabularies. ACRH's research into articulating the hermeneutics of moving images using annotations will result in a synthetic process that engages archivists, scholars and students to share their experience of these key cultural heritage texts with others—hopefully even those who cannot legibly “see” the original texts.

ACRH will repurpose a selection of assorted newsfilm to produce a corpus of material that is uniquely challenging to describe: historically charged footage laden with contextual meaning but limited extant metadata. This sub-corpus is being repurposed for research into adaptive technology for BVI users, a fundamental example of cross-disciplinary opportunities that MEP is designed to enable and investigate. Although the potential historical value of newsfilm materials for BVI access is evident, accessible delivery of online video is a challenge that higher education has struggled to meet, leading to thousands of hours of video instruction being taken offline because the schools that created it could not provide equal access to all users.

The state of BVI accessibility on the web is, in short, disastrous. Web browsers in general are riddled with inconsistent implementations of reference specifications and vendor-exclusive features. Adding accessibility features that are often poorly understood and costly to implement to that unstable environment has resulted in–at best–inconsistent efforts to ensure web content meets accessibility guidelines, e.g. [Clossen and Proces 2017]. The type of content that ACRH is targeting, online videos with time-based annotations, is so new that accessibility has not yet been thoroughly considered in this context. By researching key guidelines and technologies, ACRH has an opportunity to direct the accessibility conversation about time-based annotations in positive directions.
As the online market has matured, the penalties for organizations that fail to make content accessible online have grown. In 2015 the Department of Justice settled with online education giant EdX for failure to comply with the Americans with Disabilities Act and forced EdX to implement a number of accessibility standards including WCAG 2.0 and WAI-ARIA. UC Berkeley decided to remove thousands of hours of open educational audio and video content because it did not have the resources needed to make it ADA compliant, touching off a heated back and forth between university administrators and faculty. Accessibility problems are not limited to higher education either, as in the case of a 2014 lawsuit brought against Seattle School District 1 that resulted in a consent decree that was estimated to cost the district in the range of three-quarters of a million dollars.

BVI students in particular may have trouble fully understanding primary video sources because the text or audio descriptions associated with them rarely convey the full meaning and context of the images on screen. BVI users cannot see content filled with small clues that may be critical to its interpretation. Humanities scholars pore over information-dense resources like video to closely read it as a primary historic text at a level of detail that goes far beyond the ability of traditional accessibility adaptations like captioning to capture. ACRH proposes that time-based annotation techniques can provide support for humanistic interpretation of video far better than existing adaptive technology. Beyond the BVI community, though, researching best practices for time-based annotation will provide scholars with a new perspective on how to integrate data-centric digital heuristics with deeply cultural hermeneutics.

Existing accessibility guidelines for online video usually focus on creating secondary audio or caption tracks that synchronize playback with the video itself. The closest these recommendations come to SAT's inaugural methodology for this grant was Mozilla/A11y's recommendation to embed a timed text track into Ogg video. Setting aside the browser restrictions introduced by using Ogg video, timed captions have a number of drawbacks in an educational setting when compared to full annotations: they are only delimited by time, not geometric space in the frame; they do not carry additional metadata like tags that are useful for cataloging and search; and they do not include authorship information that is important to convey in a scholarly context. Additionally, SAT's separation of annotation data from the video file provides opportunities to readily query that data using external tools—a key feature that streamlines the workflow of digital humanities scholars.

ACRH studied how to write video annotations that convey the rich content of evocative videos and create adaptive technology that supports playing back those annotations audibly. The resulting guidelines and technology is published as an open resource that all schools, museums, and archives can use to make their own video collections more accessible to BVI users. The grant brings together scholars, archivists, cataloging experts, and cognitive neuroscientists to research best practices for these requirements. The result is an evidence-based set of guidelines for creating accessible video annotations, documentation on how to implement those guidelines using open-source software, and a demonstration corpus of civil rights newsfilm showing humanities scholars how to apply these guidelines to their own research. Just as there is a concept of resources that are “born-digital,” ACRH proposes to inaugurate a humanities corpus that includes video, annotation, and metadata so it is, as a body, “born-accessible”.
Composing annotations of moving image culture that are meant to assist blind and low vision viewers redefines certain basic assumptions about visual culture among the sighted, and demands careful attention to details otherwise taken for granted. It is surprisingly difficult to capture the basic information of a shot. Our methods are experimental, in that we completed the compilation of a sufficiently large data set to provide our colleagues at VEMI for their qualitative social science research.

The participating archives have generously provided core descriptive metadata for thousands of civil rights newsfilm clips, and assisted in selecting the dozens of clips for which we provided more extensive time-based annotations via SAT. The methodology, process, and culminating metadata is published and made available for open access and use. Much of the archival media will also be available for scholarly and public access from the participating archives, dependent upon archival protocols and online capacities.

The participating archives are:

- American Archive of Public Broadcasting (collating local and national PBS archives)
- National Archives and Records Administration (NARA) (curated selection of materials)
- Library of Congress (All-American News collection)
- Smithsonian Institution Collection (across various Smithsonian Museums)
- Bay Area Television Archive (San Francisco, CA)
- Media Burn Archive (Chicago, IL)
- The Mississippi Department of Archives and History (station WLBT materials -- Jackson, MS)
- The Pryor Center (Arkansas)
- The Wolfson Archives (Florida)
- UCLA Film and Television Archive: KTLA Newsfilm Collection (Los Angeles, CA)
- University of Georgia: The Walter J. Brown Media Archives (multiple local station materials)
- University of South Carolina Moving Image Research Collection (multiple local station materials)

The special issue of *The Journal of e-Media Studies* that will be published in relation to the assembled grant materials will feature 12 original essays inspired by and utilizing the assembled grant materials.

**The Long Duree of a Grant Interrupted by COVID**

This grant was unduly impacted by the tragedy of the COVID era, which supremely delayed many of our technological endeavors and continually thwarted our ambitions for social scientific
quantitative and qualitative scholarship. But we persevered and actually produced a considerably larger set of data regarding civil rights news film, dwarfing our initial estimates.

The original plan for the grant was much more modest than what we achieved across the participating archives. Here is the original set of goals and desired procedures:

Develop processes and guidelines supporting the delivery of annotated archival video to the higher education community with a particular focus on blind and visually impaired (BVI) users. In an effort to move quickly and explore new directions, innovators in online education have often delayed or ignored the details required to make mature, universally-adoptable technologies. In particular, the problem of adaptive access in higher education has been highlighted by recent compliance orders that have led, for example, to UC Berkeley removing tens of thousands of hours of free online educational content because the university was unable to make it accessible. Failure to consider the needs of all users not only harms those who cannot participate but also represents a substantial threat to the large and growing world of digital education resources.

The ACRH project will research the creation, curation, and consumption of online humanities collections by developing a test corpus of culturally significant American civil rights newsfilm dating from the 1950s-1980s. ACRH will then explore new cataloging and access procedures that deliver high-quality, meaningful experiences to BVI users, resulting in:

• Building and publishing a test collection of approximately one hundred hours of culturally significant, but little studied, civil rights newsfilm from the 1950s-1980s that has significant value to both BVI and sighted users.

• Running social science experiments to determine best practices for conveying both textual and contextual meaning to BVI users.

• Bringing together a group of consultants that includes civil rights newsfilm experts, archival catalogers, and human cognition researchers who examine the realistic scope of cataloging methods that support BVI users.

• Publishing developed accessibility software and guidelines as open resources for use in higher education and communities of archivists.

• Deploying a public demonstration corpus consisting of the civil rights newsfilm, annotations describing the newsfilm, and documentation explaining how the new evidence-based accessibility guidelines were applied to the annotated materials.

We were delayed by the government shut-down for the first several months of the project, which raised the possible need for an extension of the grant period. Delays from the 2020 COVID pandemic then confirmed that we would require a cost-free extension for all of 2021. These events were significant to slowing participation of Dartmouth's Research Computing on the project, and also the participation of the VEMI Lab at the University of Maine. In order to fulfill the technical design obligations of the grant, we asked for an extension of several months from the NEH to improve The Semantic Annotation Tool, to build bridges to the Scalar publishing platform, and to pursue the research on BVI users of the materials.
Nevertheless, we have considerably over-delivered on the number of civil rights newsfilm titles to be listed, and the number and range of resources to be networked is far richer than initially proposed. The interest in this grant has grown substantially in light of episodes of murderous police encounters, public outrage, and social engagement in our fundamentally charged civil rights environment in the U.S. We have added to the number of participating archives, and expanded the number of individual archival entries to over 8000 media texts, with more to come.

The rise of the pandemic in March, 2020 significantly altered all of our work flows, especially the work that needs to be done by the Research Computing team at Dartmouth. Part of the impact was due to an immediate COVID-related hiring freeze at Dartmouth, which halted significant searches then in process.

Prior to COVID, the RC team had completed a complex upgrade to the ONOMY tool and also made progress on ensuring inter-operability with other key digital tool sets as part of our ongoing cooperation with the CLARIAH group of associated DH scholars across the EU. We regularly corresponded with our key collaborators at USC, regarding the development of technological bridges from our tools and annotations into the Scalar digital publishing platform. We have continually processed serial additions of civil rights newsfilm titles.

In the spring we were delighted to learn that we would be able to work with a gifted summer intern from the Moving Image Archive and Preservation program at NYU, Shiyang Jiang. We were careful to ensure that Jiang did not directly contribute to any stated deliverables for this grant (for example time-based annotations or film metadata, nor Scalar contributions) and instead helped to advance basic organizational aspects of MEP in general by helping to achieve collection legibility via Airtable sheets for a number of MEP pilot projects.

We have been continuously industrious pursuing the Civil Rights newsfilm research, have made excellent progress acquiring titles to be listed in the public demonstration corpus, the first draft of the related Onomy vocabulary, and the initial body of annotations.

Technical developments on ACRH:

• Over 650 video files from eight archives were collected and entered into our web-based video annotation system, which is also the prototype for the system we have integrated into Scalar. A significant amount of effort went into normalizing the metadata associated with these videos since each of the archives keeps a different set of metadata internally.

• Students and experts at Dartmouth annotated many of these videos for testing purposes. Per the research plan, this initial round of annotations was done “naively,” i.e. without much guidance beyond the general goal that annotators should describe the content of the videos.

• This first round of annotations was sent to the VEMI Lab. Trials with blind/low vision users were deployed to evaluate the effectiveness of the annotations among that population. In this experiment, users played videos and had annotations read to them at the appropriate timecodes during playback. At the end of the videos, they were asked a number of questions created by MEP’s civil rights experts to test their comprehension of the video. The results were compiled and the resulting guidelines were returned to MEP for further annotation development.
• Integration of the Semantic Annotation Tool (SAT) into Scalar. The SAT has been updated to resolve several known interface bugs and to read annotation data stored in Scalar. Scalar has been modified to act as a data source by adding a new set of output encodings. The next step will be to modify Scalar to accept new annotations created by SAT, which is in progress. We worked closely with Scalar’s Information Architect, Craig Dietrich, to move the integration forward.

• Related to this project, MEP’s Onomy.org collaborative vocabulary building tool was updated. The new version of the tool can be used to formalize vocabularies developed for ACRH and publish them as linked data resources that can be integrated with SAT. Those terms will be used as controlled vocabularies when tagging ACRH videos. Among Onomy.org’s new features relevant to this projects is the capacity to link project-scale vocabularies like the one developed for ACRH to terms in external vocabularies such as Library of Congress subject headings.

Since the COVID outbreak in March, 2020, VEMI made many revisions to the Accessible Civil Rights Heritage Project to accommodate for the COVID-19 pandemic. Our efforts to run human-subject participants in the ACHR studies entirely online led to many unforeseen challenges and further delays. An unanticipated issue we encountered during piloting was mental fatigue of the participants. Since we could not eliminate the necessary instructions and steps for how to engage with the experimenter and how to access the platform for the study, we decided to cut out two of the five experimental clips. This reduced the average time of the study from nearly two hours to approximately one hour, which is more in-line with most of the behavioral studies run in the lab (and consistent with the time participants are able to focus undivided attention). The biggest hurdle we faced with the remote studies was the unpredictability of participants’ internet access. High speed internet was a requirement for eligibility, but many participants were concurrently sharing bandwidth with other household members. Streaming the experimental film clips and monitoring participants in real-time on Zoom oftentimes resulted in substantial audio and video lag. This unfortunate issue forced the majority of the seventeen attempted experimental sessions to be terminated early before the data sets were complete.

Once the University of Maine allowed human-subject research to continue, VEMI applied for and was granted IRB approval for a new in-person iteration of the ACHR study. Although the earlier data from the remote participants was not viable for analysis, it was used to revise protocols and data collection methodologies. All the issues they could not overcome online, including audio and video lag, have since been resolved and controlled. Since participants no longer had to set up the experiment with remote guidance from the researcher (i.e., everything was prepared in the lab upon arrival), another film clip was added back into the study without increasing the overall duration of the experiment. Additionally, in-lab setup included dual-monitors to allow participants to watch the videos on one screen and type descriptions on another. The intention was to reduce cognitive load from scrolling between the videos and the form fields for descriptions. Participants from the remote study were consulted on this decision and their feedback indicated that it would be easier to navigate the videos and type concurrently with the dual-monitor setup compared to watching videos embedded in the same window as the description prompts. VEMI has piloted the in-lab experiment and was on track to run half of the Phase 1 participants (control group and one experimental group) before the end of the year. The rest of the Phase 1 participants were intended to run and complete the study by December 31, 2020. Given the changing dynamics of the pandemic, circumstances beyond our control were
expected to impact our projections. (See documents “VEMI-ACRH Report-May 2021” and “VEMI-ACRH Interim Report 11_10_2020”.)

Late 2020 and 2021:

We made some progress on building out the Semantic Annotation Tool (SAT), though the focus within Research Computing at Dartmouth was still on fighting COVID. Overall, we are continuing to build the SAT tool in relation to 1) the goals for enhanced manual time-based annotations (especially developing the capacity for geometric time-based annotations); 2) the potential for iterative dynamic work-flows between manual annotations and computer vision/machine learning tools; and 3) the implicit design aspects to be realized as output to the Scalar digital publishing platform (ACRH online resource and accompanying scholarly essays).

Due to complications and limitations to research environments due to COVID, we made limited progress working with the VEMI Lab at The University of Maine to develop the capacity for our manual annotations via SAT to enable “access” for blind and low-vision users. We anticipated this research would proceed much more smoothly in summer, 2021.

In November, 2020, Williams and Bell participated in an associated NEH tool development meeting: “Distant Viewing Tool (DVT)” Advisory Board Meeting (online), hosted by Lauren Tilton and Taylor Arnold, University of Richmond.

March, 2021:


They were invited to lecture to international colleagues via a symposium hosted at Oxford: Mark Williams and John Bell, “New Directions in DH Visual Culture Studies: The Media Ecology Project”, International Multimodal Communication Centre, University of Oxford (March, 2021).

Prof. Williams was the co-host and a presenter at a two-day international online workshop conference entitled “Excavated Footage: U.S. Archives and Alternative Historiography”. This event was sponsored by a National Research Foundation of Korea grant received by Prof. Williams and Prof. Han Sang Kim at Ajou University, where the event was hosted. Prof. Williams’ presentation was entitled “Notes on Networked Scholarship, Fugitive Archives: The Media Ecology Project” (March, 2021).

For the annual Society for Cinema and Media Studies conference, held online in March 2021, Prof. Williams was a presenter and participant on a roundtable entitled “Below ‘Below the Line’: Establishing Inclusionary Cataloging for Equity and Cultural Competence in Moving-Image Collections”. The other roundtable participants were colleagues from The Academy of Motion Picture Arts and Sciences Archive, The UCLA Film and Television Archive, and The American Film Institute. Prof. Williams presented on “The Media Ecology Project: New Vocabularies to Describe Civil Rights Media”.

10
In April 2021, Prof. Williams was chair of a panel updating our progress on MEP for The Association of Moving Image Archives online conference, and was co-presenter with John Bell for “Media Ecology Project Update: Collaborative Synergies to Produce New Research in Visual Culture History” (April, 2021).

After months of discussion and deliberation, we were also successful in establishing additional research directions via the development of a new archival partnership: we signed an MOU for a remarkable exclusive agreement for access and collaboration regarding the historic Sherman Grinberg Library, which features the full run of two major newsreels (Paramount Newsreel and American Pathé newsreel). The materials from these newsreels will supplement the collections featured in the online ACRH resource.

The online ACRH resource was already scheduled to feature over 8000 entries, and we anticipate that the historical materials from the Grinberg Library may add dozens of items. All of these materials will be placed into relation with a set of search keywords derived from the item descriptions provided by the respective participating archives plus the manual annotations produced within the Semantic Annotation Tool (SAT).

We have solicited twelve sponsored essays to accompany the online ACRH resource via a special issue of The Journal of e-Media Studies, which is being updated by The Dartmouth Library to an online Scalar format. The issue will feature major scholars in the field but also represent new generations of scholarly practice.

VEMI ultimately completed Phase 1 in Spring/Summer, 2021 and recognized the need to transition Phase 2 of the study online due to the virus.

Williams and Bell were invited to make a presentation about the ACRH project to the EBSCO scholars at The University of Alabama in September. “New Directions in DH Visual Culture Studies: The Media Ecology Project” with John Bell, EBSCO Scholars, School of Library and Information Studies, University of Alabama, (September, 2021).

In fall, 2021, we recognized that COVID delays would necessitate a grant extension to Fall, 2022.

Our work with Scalar programmer Craig Dietrich continued in order to build a significant bridge between the mature Semantic Annotation Tool and the latest version of the Scalar digital publishing platform.

In September, 2021, VEMI supplied their report on Phase 1, which included a spreadsheet of Phase 1 data (see attached documents “VEMI-ACRH Report - September 2021” and “VEMI-ACRH Data (Phase 1)_Sep2021”).

In October, 2021, we expressed our struggle in trying to translate the VEMI data into suggestions we could give student annotators to shape their descriptions in ways that will support BVI use. We requested guidelines suitable for providing an annotator useful information to more successfully generate the next round of test data. VEMI provided a set of guidelines later that month (see document “VEMI preliminary captioning guidelines_ Oct2021”).
These guidelines were shared with two accomplished student annotators, who each applied them independently to four videos selected by VEMI. These annotations were then judged according to “inter-rater reliability” — essentially correlating evaluations of the annotations to arrive at a ‘best fit’ description for VEMI to use. The result would provide credibility to the descriptions for VEMI to use based on a principled ranking of the output of the two sets of annotations.

In 2022 and 2023, work toward completing the ACRH project continued and several related projects also came to fruition.

In January, 2022, we were asked by the NEH grant program Research and Development for permission to use our grant narrative for project PR-263888-19, Accessible Civil Rights Heritage Project, as a sample grant narrative to help other potential applicants to understand the qualities of a successful R&D application. We happily agreed!

Due to complications regarding the matriculation patterns of the quarter system at Dartmouth but primarily the cumulative stress from COVID that was experienced by our students in the Fall, 2021 term, we had to wait until the mid-Winter, 2022 term to engage the student annotators.

In February, 2022, both of the selected students were provided the VEMI guidelines and the technical instructions to create annotations for the set of four VEMI videos. The annotations were completed in March, 2022.


At the SCMS Conference in April, Williams chaired an online “Roundtable on ‘The Media Ecology Project: Key Outcomes from Two Major NEH Grant DH Endeavors’ Chair and Presenter, Society for Cinema and Media Studies 2022 Conference, Online (April, 2022).

In April, 2022 we evaluated the results and agreed that the annotations by one student seemed consistently superior: detailed attention to the visual track without seeming cluttered, providing identifications of notable historical figures, allowing the vocal track to represent itself (with some necessary attention to details of who is talking, especially if partially heard or overlapping voices), consistency of approach to the material across varied content. This led to our suggested rubric to VEMI:

- Close implementation the 11 points VEMI identified in their guideline doc - Detail of description of visual content - Identification of named entities and significant places - Lack of duplication of info already conveyed by audio

VEMI began working with this rubric and made progress on their research in the spring and summer of 2022.
The ACRH Grant was featured in Prof. Williams’ invited inaugural lecture for the Ebert Center for Film Studies at The University of Illinois: "Re-Enchanting Our Relationship to Film and Media History: A Brief Survey of The Media Ecology Project" Inaugural Ebert lecture, University of Illinois, Urbana/Champaign (October 2022).

Williams and Bell delivered a report about their two major NEH grants at the AMIA Conference in December: “Media Ecology Project NEH Grants Report: Early Cinema; Accessible Civil Rights” Association of Moving Image Archivists Conference, Pittsburgh, with John Bell (December, 2022).

2023 was devoted to refining the Airtable list of civil rights elements and working through a surprising number of difficulties in refining a new Scalar backbone at The Dartmouth Library for publishing The Journal of e-Media Studies, where the original essays related to the ACRH will be published in mid-2024. Dartmouth student Lauren Spencer has been a critical asset to working through these difficulties and enabling a functional journal special issue to be produced.

Here is the Table of Contents for the Special Issue:

Mark Williams (Dartmouth College), “Special Issue Introduction: The Accessible Civil Rights Heritage Project”

Jay Barth (Hendrix College), “Remembering the Little Rock Central High Crisis: The Pryor Center’s Distinctive Insights”

Stephen Charbonneau (Florida Atlantic University), “Insurgent Leisure, Aquatic Angst: Postwar Newsfilm, Civil Rights, and Coastal Imaginaries”

Joseph Clark (Simon Fraser University), “The Newsfilm Archive and the Struggle for Civil Rights”

Bobby Donaldson (University of South Carolina-Columbia), “’The Most Frightening Thing I Have Ever Seen’: Moving Images and the South Carolina Civil Rights Movement”

Rodolfo Fernández (University of Connecticut) and Deborah L. Jaramillo (Boston University), “Mexicans in Your Town: Histories of Mexican Migration to the United States in Local Television Documentaries”

Desiree J. Garcia (Dartmouth College), “In Living Color: Chicano Televisual Media at the Dawn of the Movement”

Jeffery Hirschy (University of Southern Mississippi), “Historical Distance and the Mississippi Civil Rights Movement: Opening Windows into Mississippi Civil Rights History”

Dimitrios Latsis (University of Alabama), “Revisiting Newsfilm of the 1970s Jackson State Killings: Digital Humanities as Antiracist Praxis”

Curtis Marez (Ethnic Studies Department, University of California-San Diego), “TV News and the Origins of Black Studies and Ethnic Studies”
Brandy Monk-Payton (Fordham University), “’Atlanta’s Image is a Fraud’: Fragments of Black Protest in Local TV Newsfilm”


Appendices from VEMI Lab (University of Maine) follow:
Since March, VEMI has made many revisions to the Accessible Civil Rights Heritage Project to accommodate for the COVID-19 pandemic. Our efforts to run human-subject participants in the ACHR studies entirely online led to many unforeseen challenges and further delays. An unanticipated issue we encountered during piloting was mental fatigue of the participants. Since we could not eliminate the necessary instructions and steps for how to engage with the experimenter and how to access the platform for the study, we decided to cut out two of the five experimental clips. This reduced the average time of the study from nearly two hours to approximately one hour, which is more in-line with most of the behavioral studies that we run in the lab (and consistent with the time participants are able to focus undivided attention. The biggest hurdle we faced with the remote studies was the unpredictability of participants’ internet access. High speed internet was a requirement for eligibility, but many participants were concurrently sharing bandwidth with other household members. Streaming the experimental film clips and monitoring participants in real-time on Zoom oftentimes resulted in substantial audio and video lag. This unfortunate issue forced the majority of the seventeen attempted experimental sessions to be terminated early before the data sets were complete.

Once the University of Maine allowed human-subject research to continue, VEMI applied for and was granted IRB approval for a new in-person iteration of the ACHR study. Although the earlier data from the remote participants was not viable for analysis, it was used to revise protocols and data collection methodologies. All the issues we could not overcome online, including audio and video lag, have since been resolved and controlled. Since participants no longer have to setup the experiment with remote guidance from the researcher (i.e., everything is prepared in the lab upon arrival), another film clip was added back into the study without increasing the overall duration of the experiment. Additionally, in-lab setup includes dual-monitors to allow participants to watch the videos on one screen and type descriptions on another. The intention is to reduce cognitive load from scrolling between the videos and the form fields for descriptions. Participants from the remote study were consulted on this decision and their feedback indicated that it would be easier to navigate the videos and type concurrently with the dual-monitor setup compared to watching videos embedded in the same window as the description prompts. VEMI has piloted the in-lab experiment and is on track to run half of the Phase 1 participants (control group and one experimental group) by November 30. The rest of the Phase 1 participants are intended to run and complete the study by December 31. While this is our goal, given the changing dynamics of the pandemic, circumstances beyond our control may impact our current projections.
Guideline 1: The highest level guidance is to always identify who is speaking as this can be hard for a person who is blind/visually impaired (BVI) to determine but is often critical for accurate interpretation of the exchange or for correctly parsing the scene. This should ideally be done by providing the person’s name but when not possible, the description should use some form of unique identifier. Good alternatives to names include gender coupled with clothing (e.g., the woman in a striped shirt) or age coupled with gender (an older man). These may use additional descriptors as identifiers that pull from more of the following elements, if it helps to uniquely identify the person. The rule of thumb is to determine what characteristic is uniquely descriptive of that person versus anyone else in that scene and to then be consistent in using that descriptor.

Guideline 2: The second major high-level guideline is to always adopt a consistent manner of describing people or scenes. If you use wording in one description, use the same wording in another, as users will learn to interpret what is meant but will be confused if this description logic changes across scene or clip. As with above, multiple elements may be required in a description to make the content clearly distinguishable or meaningful.

Guideline 3: make descriptions as short as possible while still conveying meaning. Try to use short crisp phrases without filler words. Consider what is visually salient/important in the scene and whether this information is also available from context or nonvisual sensing, e.g. hearing. If not, convey only the key information that gets across the requisite visual element(s). In sum, strive to describe visually salient elements through a short, meaningful nonvisual description.

We provide a ranking of importance for the information conveyed by the 11 elements we coded from the clips below, accompanied with some further guidance and explanation. Keep in mind that if the context of the scene or clip is dependent on any of these elements, their priorities need to be recalculated (e.g., race is greatly elevated in importance when describing a race riots).

1. **Proper names**
   a. Important because often needed to understand the clip and associated characters but voice recognition is highly variable and often difficult without a visual reference.
   b. Person recognition is generally tied to associated visual recognition (which is why we have dedicated face recognition regions in the brain) but we have no such specialized neural hardware for voices. Thus, explicitly naming any key characters in a scene is critical.
   c. Often times in movies and clips you just hear the voice and without the name, lose information.
   d. Clearly articulating the name consistently is hugely important

2. **Physical features**
   a. These Can be an important identifier; they also provide an important aspect of parity to what visual viewers access. Should use words that are descriptive but are not judgmental (e.g., morbidly obese vs. disgustingly obese). Use terms that provide descriptions without
breaking down all the elements -- hipster and hippie vs. describing all the elements. This increases Efficiency while reducing verbiage (i.e., allows use of cognitive processes like schemas to fill in gaps).

3. **Perceived gender**
   a. Relative importance depends on if there is audio
   b. Generally known from voice, or assumed from voice
   c. If not clear from an audio track (e.g., person is not speaking), this should be mentioned, perhaps combined with other elements to uniquely identify the person (see guideline 1).

4. **Facial Expressions**
   a. Description should include aspects that are not easily conveyed by the auditory track. Expression is generally such an element. It is not always important, but it can be in some scenes.
   b. Most important to clarify reactions to people who are talking (e.g., a person is speaking and the silent listener is frowning, obviously not happy with what is being said)

5. **Number of People**
   a. Generally, only important to set the context
   b. Say it once and then move on

6. **Perceived Age**
   a. Really hard to get from voice
   b. Can be a reference if there is no name
   c. Only need to describe it once

7. **Clothing and Accessories (should be higher if no other unique descriptor)**
   a. Unique identifier for person who is otherwise unnamed
   b. Often not important beyond a surface feature but sometimes clothes are a big part of a specific aesthetic of the scene or highly characteristic of a person or group, and in these instances, they should be consistently described.

8. **Race and Ethnicity**
   a. Importance is extremely dependent on the context of the clip. It is a visual element that is salient but also easy to misattribute. For instance, in the black panther clip, it is important to know this information to understand the dynamic. Its importance is highly contingent on the scene.

9. **Gaze**
   a. Important if gaze is conveying information that is salient to a visual viewer, e.g. the person is looking at something in the scene that is not described but is important. Here again, the decision to describe is tied to the importance of the visual information to correct interpretation of the scene.

10. **Cinematographic Elements**
    a. Depends on the purpose of the media (e.g., news vs art films).
    b. These are attributes that are visually salient and can be used to show different things but are not generally tied to core content that must be described nonvisually. This is more important for artistic films or movies, where these elements may be strongly tied to the underlying plot.

11. **Title Screen**
    a. Important at the beginning to set the context
Notes:
We acknowledge that this project, as with all our behavioral research involving human participants, continues to be running behind due to COVID-19. Pandemic-related University restrictions over the last 15 months have prohibited (or greatly limited) most research involving recruitment and conduct of human subjects. Studies that were able to be run remotely were less impacted but those that required having participants be physically in the Lab, such as this project, were significantly delayed. In addition, until the second week in May, in-person experiments required the participants to be from the university community (which had an established COVID-19 testing protocol). Participants from outside the University were prohibited. As we don’t have a film program, this was a problem as we were planning to recruit experts for some of the experimental conditions from outside institutions. The good news is that we have been able to run 10 participants, which are described here. As of the second week of May, the University has loosened some of the existing restrictions on running in-person human participants and we are now able to schedule more ‘outside’ participants, allowing us to recruit experts and finish up this phase of the project. This process is now underway.

Objective:
Identify general trends in descriptions of various archival film clips with the intent to provide guidance on generating new descriptive audio annotations to test with BVI users. We are also investigating if descriptions vary significantly based on the audience and purpose for viewing the film.

Method:
10 participants have completed this study so far. 5 in the Control Group (ages 19-31, 3 male and 2 female) and 5 in the General Experimental Group (ages 19-35, 5 male). All participants reported being white/Caucasian and none reported having a scholarly background in U.S. history or media/film.

Prior to completing the study, each participant received informed consent, filled out a form for demographic information, and completed a pre-survey. The survey aimed to assess individuals’ prior knowledge of notable people and events that would be presented in the experimental stimuli (i.e., the archival film clips). The questions on the pre-survey were open-ended and the order was randomized. This randomization minimized potential association bias (i.e., having someone remember something about one person or event from reading about another). For instance, if the questions about the Delano Strike and related people were grouped, the participants may have guessed by chance that Dolores Huerta was associated with that event and/or the National Farm Workers Association.

Each participant watched all four film clips and was instructed to describe the clips to someone else based on a prompt (see below). The order of the clips was randomized per participant to reduce potential bias from factors such as learning order or cognitive fatigue. The participants watched/navigated the videos on one monitor and used a second monitor to type out their descriptions. Pilot testing revealed that participants gave richer and more useful data (with regards to analyzing and translating to descriptive audio annotations) when they were given a form with a table that required them to describe the clips in segments (as opposed to giving a single open-ended prompt to describe the clip in its entirety with one long response). The attached spreadsheet (file: VEMI – ACRH Data (Phase 1)) shows how each clip was segmented. Since the duration and content of each clip varied considerably, we chose to segment based on breaks in audible dialogue.

Only the description prompts varied between the control and experimental conditions:

- Control: “How would you describe this film clip to someone else? Your description should include enough detail so the person reading could understand the clip without watching it.”
• Experimental Group (General education) – “How would you describe this film clip to someone else who wants to use it as a source for a research project? Your description should include enough detail so the person reading could understand the clip without watching it.”

We hypothesized that indicating an academic/educational purpose for viewing/using the film would result in different descriptions than not indicating a specific use or intention for accessing the clips.

Experimental stimuli/clips:
- National Farm Workers Association (NFWA) - https://youtu.be/yzzL9W04q-g
- Black Panther Party - https://youtu.be/i4lQmSyGe9k
- Apollo 11 - https://youtu.be/_GkZxJ45Um0

Preliminary Analysis:
The following are general results:
- After identifying the factors (listed below) and conducting a descriptive analysis, we found no discernable significant differences in the descriptions between the control group and the first experimental group. This indicates that the descriptions generated by people with general knowledge (i.e., not scholars in U.S. history or media/film), may not be influenced by an academic purpose/use of the film. Since there were no identifiable differences, we collapsed the data across groups. More detailed analysis is below.
- Only 1 participant (“Experimental 1 – 002”) consistently summarized the dialogue that was included in the given transcript for each film segment (i.e., did not follow instructions). There were occurrences of 4 other participants summarizing speech, but it happened sporadically (5 times for 1 subject, twice for 1 subject, and once for 2 subjects). This data should be omitted when considering descriptive audio annotation generation.
- Since the given transcripts in the study only included spoken words from primary speakers in the films, participants described elements such as background noise, silence/breaks in audio, and tone of voice of speakers. These audio elements could be omitted when creating the annotations. Descriptions about elements such as “applause” and “laughter” need more careful consideration because they could be strictly audio or visual elements on screen.

The following factors were identified in the descriptions and analyzed. Most of this data cannot be generalized, but some trends were revealed.
- Proper names
  - All 10 participants demonstrated prior knowledge of Neil Armstrong, Edwin “Buzz” Aldrin, and Muhammad Ali. Only 1 person knew of Michael Collins, 3 knew of Cesar Chavez, and 0 knew of Dolores Huerta. Stokely Carmichael was not included on the pre-survey but should have been.
  - The following is a list of the number of participants who described each notable person by name and when the naming occurred:
    - Dolores Huerta: 0/10 mentioned her name at all. Her identity was not revealed in the clip. This was to be expected based on pre-survey results.
    - Cesar Chavez: 2/10 mentioned his name before it was revealed in clip and only 5/10 (including the initial 2) referred to him by name after it was revealed.
    - Muhammad Ali: 7/10 referred to him by name before he said it in the clip, and 9/10 described him by name after.
    - Stokely Carmichael: 2/10 mentioned “Stokely Carmichael” or “Carmichael” before his last name is mentioned. Only 1 additional person mentioned his name once when it was disclosed and then reverted to calling him “the man.”
- Neil Armstrong: 1/10 recognized him and described him by name before it was disclosed, 8/10 total ended up describing him by name after it was explicitly stated in the clip.
- Edwin “Buzz” Aldrin: 0/10 mentioned his name at all. His identity was never revealed. This was not entirely expected because all 10 participants had demonstrated prior knowledge of him on the pre-survey.
- Michael Collins: 0/10 mentioned his name at all. His identity was never revealed.

Data interpretation: This variability indicates that just because someone has prior knowledge of a person, does not necessarily mean they will visually recognize them.

Consideration for future work: To identify whether proper names are in fact important/necessary to describe, we would need to control for whether the participant visually recognizes the individual(s) in the films in the first place.

- Gender
  - When individuals are not mentioned by name, they are most often referred to using gender terms, such as “man” or “woman” (9 out of 10 participants). Only 1 subject used “male” or “female” to describe people but those terms were not use consistently across all stimuli (they also said “men”).

- Ethnicity, race, and/or skin color
  - All of our participants reported to be white/Caucasian. The Apollo 11 clip and the Muhammad Ali clip show men at a press conference. 2/10 participants described the Apollo 11 crew as “white”, whereas 5/10 initially describe Muhammad Ali and/or the men next to him as “black” or “African American.”

  Consideration for future work: Do individuals tend to explicitly describe the ethnicity, race, and/or skin color of those different from themselves? And if so, how does this affect the aim to construct generalized best practices for descriptive audio?

- Physical features of person (other than skin color)
  - Hair (on head). Individuals’ hair was mentioned in some of the descriptions but there was no apparent pattern. Analyzing across all the stimuli, 4 participants made at least one reference to hair.
  - Facial hair. 4/10 participants mentioned the facial hair of people in the Black Panther Party clip. About half of the descriptions stated, “facial hair” and the others specified the type of facial hair (e.g., mustache).

  More analysis will be conducted to investigate other physical characteristics.

- Clothing and accessories
  - The most consistently described clothing and accessories across stimuli and participants were suits, hats, and glasses (including sunglasses). To make any generalizations about clothing and accessories, it would be necessary to choose and evaluate more clips with predetermined common elements and investigate the incidence of each element in the descriptions.

- Perceived age
  - In the NFL clip, 2 participants referenced the age of a single person once. And in the Muhammad Ali clip, 4 participants referenced age. Most of these instances had comparative language (e.g., “older”). There were no age references in the other two clips. Overall, there appear to be no discernable trends of mentioning perceived age of people in the current data set.

- Gaze
  - 8/10 mentioned the gaze/focus of people at some point in their descriptions. The gaze was oftentimes described directionally (e.g., “looked down”) or by specifically naming something/someone in or out of frame.

  The participants generally described the directionality of where a person was looking from their own perspective. In only 2 singular instances, participants mentioned the gaze
of Muhammad Ali from his perspective (i.e., said ‘looks to his…’). One of those participants incorrectly stated he “looks to his right” when it was really his left.

- **Title screen**
  - Only the National Farm Workers Association clip had a ‘title screen’ at the beginning. 8/10 participants mentioned that part of the clip at all and 5/10 stated specifics of the text on screen. More clips with these ‘title’ elements would need to be evaluated to determine if the information is considered generally important to describe.

- **Number of people in frame/on screen**
  - There is a lot of variability to describe how many people are visible on screen at a given moment. All 10 participants specifically said, “three men” in their description of the Apollo 11 clip. However, in the other clip descriptions some participants chose to number those on screen as in the Apollo 11 clip, using general terms, such as “group” or “crowd,” or they didn’t give any quantifier at all.

- **Facial expressions (objective descriptions)**
  - “Neutral” expressions are not mentioned at all. The most common expression mentioned was smiling or some synonymous word.
  - It appeared that when someone’s facial expression was not “neutral” or “smiling” the participant was apt to write a description that was their interpretation of the person’s emotions.
  - Consideration for future analysis: What are other “objectively classifiable” facial expressions beside smiling?

- **Cinematographic elements**
  - The most common terms used for camera/scene changes in the descriptions were: zoom (6/10 participants), pan (7/10), and cut (6/10). When camera/scene changes occurred and participants did not use the terms mentioned above, the elements were still often described in their responses. For instance, instead of saying “cut”, a participant said, “the camera switches…” In our more rigorous final analysis, we will investigate the omission of cinematographic elements (e.g., most participants did not mention all the individual cuts in the Muhammad Ali clip) and check for errors in the usage of terminology (e.g., a participant described a cut as “panning to a new scene.”)

**Next Steps:**

We are finally able to start recruiting participants to the lab with less restrictions and have thus started recruiting scholars for the following experimental group:

- **Experimental Group 2 (Media/film scholars)** – “How would you describe this film clip to someone else who wants to use it as a source for a media or film research project? Your description should include enough detail so the person reading could understand the clip without watching it.” Because of the specific knowledge of this expert group, we plan to recruit 5-10 people in this condition. We will complete the analyses after data collection with these participants. We may also collect data from history scholars but this is not our primary focus and given the time constraints and recruiting challenges, coupled with the outcomes of the current data, this group is lower priority than the film experts.

We hypothesize that the scholars’ descriptions will be significantly different from those of the non-scholars. We expect a higher frequency of cinematographic language/terminology.

**Further analyses:**

- Reexamine the aforementioned factors when we get more data and look for any new outcomes or parameters of interest.
- Affective analysis for interpretive/subjective descriptions (e.g., “man looks overwhelmed and slightly distressed”).
- More thorough cinematographic analysis. We anticipate the media/film scholars will include more terminology in their descriptions.
- Process analysis for actions on screen (e.g., hand gestures and other visible movements).
- Descriptive analysis of non-human elements on screen (e.g., microphones were described in several clips). Since we are using a small number of film clips and they do not have many common non-human elements, we do not anticipate finding generalizable results.
Notes:
As we described in our previous report, this project, as with all our behavioral research involving human participants, has proceeded much slower than we had planned and expected due to COVID-19. Pandemic-related University restrictions over the last 18 months have prohibited (or greatly limited) most research involving recruitment and conduct of human subjects. Studies that were able to be run remotely were less impacted but those that required having participants be physically in the Lab, such as this project, were significantly delayed. After running 5 more participants in our final experimental group since the last report, this report adds to and extends what we previously submitted to you in May.

Objective:
Identify general trends in descriptions of various archival film clips with the intent to provide guidance on generating new descriptive audio annotations to test with BVI users. We are also investigating if descriptions vary significantly based on the audience and purpose for viewing the film.

Method:
In total, data was collected from 15 individuals who participated in three conditions in the study. There were 5 participants in the Control Group (ages 19-31, 3 male and 2 female), 5 participants in the General Experimental Group (ages 19-35, 5 male), and 5 participants in the Film Background Group (ages 37-69, 2 male and 3 female). 1 participant, who was in the Film Background Group, reported being African American; the rest of the participants reported being Caucasian. It should be noted that none of the participants in the Control Group and General Experimental group reported having a scholarly background in media/film, whereas every participant in the Film Background Group reported some experience in studying film or any video-based media in a collegiate or professional setting.

Prior to completing the study, each participant received informed consent, filled out a form for demographic information, and completed a pre-survey. The survey aimed to assess individuals’ prior knowledge of notable people and events that would be presented in the experimental stimuli (i.e., the archival film clips). The questions on the pre-survey were open-ended and the order was randomized. This randomization minimized potential association bias (i.e., having someone remember something about one person or event from reading about another). For instance, if the questions about the Delano Strike and related people were grouped, the participants may have guessed by chance that Dolores Huerta was associated with that event and/or the National Farm Workers Association.

Each participant watched all four film clips and was instructed to describe the clips to someone else based on a prompt (see below). The order of the clips was randomized per participant to reduce potential bias from factors such as learning order or cognitive fatigue. The participants watched/navigated the videos on one monitor and used a second monitor to type out their descriptions. Pilot testing revealed that participants gave richer and more useful data (with regards to analyzing and translating to descriptive audio annotations) when they were given a form with a table that required them to describe the clips in segments (as opposed to giving a single open-ended prompt to describe the clip in its entirety with one long response). The attached spreadsheet (file: VEMI – ACRH Data (Phase 1)) shows how each clip was segmented. Since the duration and content of each clip varied considerably, we chose to segment based on breaks in audible dialogue.

Only the description prompts varied between the control and experimental conditions:

- Control: “How would you describe this film clip to someone else? Your description should include enough detail so the person reading could understand the clip without watching it.”
• General Experimental Group (General education) – “How would you describe this film clip to someone else who wants to use it as a source for a research project? Your description should include enough detail so the person reading could understand the clip without watching it.”

• Film Background Experimental Group – “How would you describe this film clip to someone who wants to use it as a source for a media or film research project? Your description should include enough detail so the person reading could understand the clip without watching it.”

To guide the research process, the following hypotheses were developed:

• H0: When tasked with describing the clips for a research project (as was the case with both experimental groups) compared to a general purpose (as with the control group), the descriptions will be more detailed, as measured by the overall frequency of identified codes.

• H1: The film background experimental group will be more verbose in their descriptions, use more detail and nuance than the other groups, and also prioritize cinematographic elements more so than the other groups.

Experimental stimuli/clips:

• National Farm Workers Association (NFWA) - https://youtu.be/yzzL9W04q-g
• Muhammad Ali - https://youtu.be/ZBO9KaEDpYo
• Black Panther Party - https://youtu.be/i4lQmSyGe9k
• Apollo 11 - https://youtu.be/_GkZxJ45Um0

To ascertain the significance of the participants’ descriptions, we selected the following eleven elements from coded trends in the data. Below is a list of the elements and an explanation for how they were coded:

**Cinematographic Elements:**
Any description of camera movement or commentary on the camera’s level of focus was coded for this factor. Additionally, if a participant was clearly describing a complete change of scene and/or fade to black, this was also coded. Examples include: “zooms in”, “pans”, “cut”, and “new shot”.

**Clothing and Accessories:**
The code for this factor was applied to any mention of a person’s clothing or accessories. Examples include “hat” and “glasses”. Additionally, items that could have some fashionable significance, such as cigarettes and umbrellas, were also included. If two or more articles of clothing created a singular ensemble, such as “suit and tie”, this was coded as one mention of clothing.

**Ethnicity, race, and/or skin color:**
Any mention of an individual’s ethnicity, race, or skin color was coded for this factor. As race was a topic of discussion in both the Muhammad Ali clip and the Black Panther Party clip, it is interesting to observe how participants used it in their descriptions. Examples include: “Latin American descent”, “black”, “white”, and “darker skin”.

Facial Expressions (objective descriptions):
This factor only includes mentions of facial expressions or movements that can be clearly imagined. For example, “smiles”, “frowning”, “raises his eyebrows”, and “moves his jaw a little” are all objective movements that were coded for this factor. Other descriptions that resulted from the participants’ subjective interpretations of facial expressions, such as “frustration on their faces”, were not included.

Gender:
Any time a person was identified by gender it was coded for this factor. Examples include “man” “woman”, “female”, and “boy”. As “man” and “woman” were one of participants’ primary methods of describing people, this factor is very prevalent.

Gaze:
Any time someone is described as looking in a particular direction or at a particular thing or person, it is coded for this factor. Examples include: “looks to the right”, “looking at the camera”, and “makes eye contact with him”.

Number of people in frame/on screen:
When participants described the people in a video in terms of how many of them there were, this factor was applied. Examples include: “two men” and “three African American men”. Larger phrases that indicated multiple groups of numbered people were also coded once, as they constituted a singular instance of the participant seeking to number people. An example includes “Ali at a table with 6 other people two standing behind”. Also, the word “one” was not coded, as it often served as more of a grammatical article than a number.

Perceived Age:
Phrases were coded for this factor when they commented on the apparent age of a person in the video. Examples include: “older”, “younger”, and “adults”. We considered including “man” and “woman” for this factor, but decided these words were primarily indicative of gender and not of any particular age. Since phrases like “young man” could mean a male child, “man” is not itself indicative of a particular age.

Physical Features of Person (other than skin color):
Phrases were coded for this factor when they described a physical characteristic, other than skin color, of a person on screen. Examples include: “dark hair”, “tall” and “full eyebrows”. Mere mentions of common body parts, like “hand” or “face”, were not coded.

Proper Names:
Anytime a participant indicated a person on screen by name, the name was coded for this factor. It was coded whether the participant used an individual’s first name, last name, or both. Examples include: “Muhammad Ali”, “Chavez”, and “Neil”.

Title Screen:
The only video to include a title screen at the beginning was the National Farm Workers Association video, which began with some information about the video archive. Any mention of this information, whether a word-for-word transcription or a comment that there are words on the screen, was coded for this factor. Examples include: “Bay Area TV Archive Film Reference number KQN 3558, April 10th 1966” and “Opening Statements and visual credits”.
Note: Anytime that a participant paraphrases or directly quotes a person’s speech from the video, none of that data was coded as it contains no relevant visual information. The same is true for descriptions of any audio elements.

**Observed Trends and Results:**
To begin our analysis, of interest was to determine whether the number/frequency of elements (codes) was significantly different between conditions. As hypothesized, participants in the Film Background Group provided descriptions that included a greater overall frequency of identified elements (723) than the General Education group (560) and the control group (531). Given the categorical (non-parametric) nature of these data, a chi-squared test for independence was utilized to determine if this difference was the result of a reliable statistically different relationship between condition and code frequency. A contingency table (Table 1) was constructed to compare observed values to expected values, with the result (p<.001) leading us to reject the null hypothesis that the conditions would result in the same frequency.

<table>
<thead>
<tr>
<th>Table 1. Observed and (Expected) Frequency of Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Condition</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>Cinematographic Elements</td>
</tr>
<tr>
<td>Clothing and Accessories</td>
</tr>
<tr>
<td>Ethnicity, race, and/or skin color</td>
</tr>
<tr>
<td>Facial Expressions (objective descriptions)</td>
</tr>
<tr>
<td>Gaze</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Number of people in frame/on screen</td>
</tr>
<tr>
<td>Perceived Age</td>
</tr>
<tr>
<td>Physical Features of Person (other than skin color)</td>
</tr>
<tr>
<td>Proper Names</td>
</tr>
</tbody>
</table>
The following describes trends of elements with frequency totals that differed significantly between groups and offers possible explanations for these trends.

**Cinematographic Elements**
As expected, participants in the Film Background Group identified significantly more cinematographic elements than the Control Group, specifically 1.89x as many (119 identified by the Film Background group, 63 by the Control group). This may be due to the Film Background Group incorporating a greater focus on cinematics, not only because of their experience in the subject, but also because of their prompt to describe the clips as a source for a media or film research project.

**Clothing and Accessories**
Participants in the Film Background Group identified 102 elements of Clothing and Accessories compared to the 44 identified by the Control Group (2.32x less), and the 55 identified by the General Experimental Group (1.85x less). These data are congruent with our hypothesis that the Film Background Group would, in general, prioritize more cinematographic elements. Clothing and accessories are often an important aspect of “setting the scene” in film through costume design and thus may seem more relevant, and solicit more attentional focus by these participants, which thereby manifests in their descriptions.

**Ethnicity, Race, and/or Skin Color**
One surprising result was that the Control Group identified 75 elements of ethnicity, race, and/or skin color, whereas the General Experimental Group identified 12 elements (6.3x less), and the Film Background Group identified 47 elements (1.6x less). Although there is no way to verify, we interpret these findings as being strongly influenced by cultural events at the time the study was run. In other words, the fact that the control group participated in the experiment the week after George Floyd’s death likely resulted in heightened attention to race, which would have increased focus on this parameter during the study, and by extension, in their descriptions. Other potential factors may relate to experience with diverse groups, or the task at hand (i.e., providing descriptions for a general purpose and not for research).

**Perceived Age**
An unexpected result was that the Control Group identified 12 elements of perceived age—3x as many as the General Experimental Group (4 identifications), and 2.4x as many as the Film Background Group (5 identifications). This may have occurred since the Control Group was not given a specific audience (i.e., for research) to provide descriptions for and since age is especially salient to a general audience, they may have been more inclined to include this in their description.

**Physical Features of Person**
Participants in the Film Background group identified 25 elements of physical features of a person, compared to 13 identified by the Control Group (1.92x less), and 10 identified by the General Experimental Group (2.5x less). We interpret this outcome as supporting our hypothesis that the Film Background Group would focus on cinematographic elements, which consist of physical features as they can often determine aspects of films such as the frame of shots and camera angles.

**Gender**
What was noted from this element was not the difference in the number of identifications between each group, but rather the high total number of identifications. The control group identified 173 elements of gender, the General Experimental Group made 190 identifications, and the Film Background Group made 224 identifications. We suspect that these high cross-group counts may be caused by the inclusion of the
words “he”, “she”, “man”, and “woman”, as primes for identification for this element; these are common descriptors to avoid repeating the subject’s name, or if a participant is unaware of the subject’s name, and would therefore lead to an increase in identifications.

Also Noted:
When there was a common proper name that participants were familiar with, like Muhammad Ali, there were fewer descriptors as a whole (e.g., clothing, gender, and facial expressions). This is likely because when someone is familiar with a name, they are less likely to use generic or detailed descriptions of the person. For example, “Muhammad Ali did X” opposed to “The African American man wearing the white shirt did X.”

The following trends were previously observed in the Control and General Experimental Groups:

- Proper names
  - All 10 participants demonstrated prior knowledge of Neil Armstrong, Edwin “Buzz” Aldrin, and Muhammad Ali. Only 1 person knew of Michael Collins, 3 knew of Cesar Chavez, and 0 knew of Dolores Huerta. Stokely Carmichael was not included on the pre-survey but should have been.
  - The following is a list of the number of participants who described each notable person by name and when the naming occurred
    - Dolores Huerta: 0/10 mentioned her name at all. Her identity was not revealed in the clip. This was to be expected based on pre-survey results.
    - Cesar Chavez: 2/10 mentioned his name before it was revealed in the clip and only 5/10 (including the initial 2) referred to him by name after it was revealed.
    - Muhammad Ali: 7/10 referred to him by name before he said it in the clip, and 9/10 described him by name after.
    - Stokely Carmichael: 2/10 mentioned “Stokely Carmichael” or “Carmichael” before his last name is mentioned. Only 1 additional person mentioned his name once when it was disclosed and then reverted to calling him “the man.”
    - Neil Armstrong: 1/10 recognized him and described him by name before it was disclosed, 8/10 total ended up describing him by name after it was explicitly stated in the clip.
    - Edwin “Buzz” Aldrin: 0/10 mentioned his name at all. His identity was never revealed. This was not entirely expected because all 10 participants had demonstrated prior knowledge of him on the pre-survey.
    - Michael Collins: 0/10 mentioned his name at all. His identity was never revealed.
  - Data interpretation: This variability indicates that just because someone has prior knowledge of a person, does not necessarily mean they will visually recognize them.

- Facial expressions (objective descriptions)
  - “Neutral” expressions are not mentioned at all. The most common expression mentioned was smiling or some synonymous word.
  - It appeared that when someone’s facial expression was not “neutral” or “smiling” the participant was apt to write a description that was their interpretation of the person’s emotions.

- Cinematographic elements
  - The most common terms used for camera/scene changes in the descriptions were: zoom (6/10 participants), pan (7/10), and cut (6/10). When camera/scene changes occurred and participants did not use the terms mentioned above, the elements were still often described in their responses. For instance, instead of saying “cut”, a participant said, “the camera switches…”
Conclusions and future work:
Both experimental conditions resulted in a higher overall frequency of elements than the control condition. This finding is in line with the hypothesis that the audience for which descriptions are being provided influences the detail included in the descriptions. It is also clear from these data that increased knowledge of cinematography creates a more verbose description of a scene, with a higher likelihood of cinematographic elements included, which supports the second hypothesis. This is evident from the total number of elements that experts provided in their descriptions and their proportionally more frequent descriptions of the scene (i.e., cinematographic elements, clothing and accessories, and number of people in frame) over more generally salient elements (i.e., gender, race, perceived age, and other physical features). For further analysis of these data, it might make sense to submit the descriptions and coding rubric to formal linguistic analysis that captures more than frequencies, e.g. relational pairings, spatial continuity, and other semantic coupling. It would also be relevant to ascertain whether these descriptions provide better access to the material for people who are blind and visually impaired (BVI) than what might be generated by a traditional video captioning process. The effectiveness of each description could also be compared between the original Control Group, General Experimental Group, and Film Background Group, based on how the BVI participants assess the content. By doing so, the effectiveness of specific elements and the level of detail therein could be measured and better quantified for future guidelines.