UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Digitale
Kulturwissenschaften

Projektarbeit

Dialectics of Liberation Congress/Digital Archive

nnaumann@posteo.de

June 2022: This is the fourth update of the original paper. However, some information is outdated due to ongoing research. The archive holding the estate of Joseph Berke was renamed PET Archives and Special Collections. A team of professional archivists work on modernizing the service and providing digital access. It is recommended to check regularly, since a follow-up paper will soon be available.

Table of Contents

The folders that are referred to in *4.4. Transcription of Audio Files with YouTube*

(Praxisprojekt_Stokely_Carmichael_18_July_1967_audio_files_1_3_16_1_3_15)

(Praxisprojekt_Counterculture_compared_1_2_24_VTS_01_3_VTS_01_4)

(Praxisprojekt_audio_file_1_2_30)

can only be shared with restrictions and on demand due to copyright issues. Info available at nnaumann@posteo.de

List of Acronyms

ASR	Automatic Speech Recognition
CLARIN	Common Language Resources and Technology Infrastructure
CMS	Content Management System
DCMI	Dublin Core Metadata Initiative
DFG	Deutsche Forschungsgemeinschaft
DH	Digital Humanities
DOL	Dialectics of Liberation Congress
MIT	Massachusetts Institute of Technology
NLP	Natural Language Programming
PBS	Public Broadcasting Service
PETT	Planned Environment Trust
TEI	Text Encoding Initiative
UK	United Kingdom
WDR	Westdeutscher Rundfunk
WER	Word Error Rate
XML	Extensible Markup Language

1. Purpose and Objectives

The present report is the assessment of a subproject of a project aiming at establishing the *Congress Dialectics of Liberation Digital Archive*. The subproject deals with audio files generated at the congress. It aims at giving first insights into the role audio files could play in generating the archive. It further examines possibilities to prepare the files in such a way as to make them usefully accessible for further research. The object of this report is to document the subproject's genesis and progress to date. It should generate findings and observations culminating in research-based recommendations to inform future related planning and decision-making.

It is expected to:

- describe the project goal and evaluate the work accomplished thus far
- document the explorative process of the project
- present the main features of the tools used
- discuss practical questions emerging during the process
- evaluate the findings of the project
- provide recommendations on tools and short- and medium-term work on remaining audio files
- identify lessons learned and good practices

This will be done throughout the report and summarized at the end.

2. Methodological Approach

By choosing the audio files as the starting point, a first multilayered approach to handling the complex archive project is undertaken. The first step is to get an overview of the situation by documenting the preliminary work regarding the audio files and sifting through the compiled written and video data that might be useful in dealing with the audio files.

A subset of audio files is chosen for this first trial. The audio files are positioned in relation to other sources. Digital tools are tested in an explorative approach to gain new insights on which tools are needed to best meet the project's needs. The most promising results of the process will be added to the report.

Suggestions are made regarding the practical implementation of the archive. Recommendations for the further approach are given based on the results of the theoretical research and the first screening of digital tools.

3. Dialectics of Liberation Congress, Digital Archive Project Overview

3.1. Project Goal

The research project *Dialectics of Liberation Digital Archive* is managed by Dr. Alexander Dunst, currently at the University of Paderborn (Germany). The project aims to establish an open-access, online, multimedia archive of the *Congress on the Dialectics of Liberation (DLC).* The digital space offers new possibilities of archiving: By compiling already published material, hitherto unpublished sources, retrospective interviews with participants, and contemporary research results the archive is supposed to serve a wide range of interests. Preliminary work has been accomplished by Dunst, Bolu Akinro, and Cecilie Rasmussen.

3.2. Dialectics of Liberation Congress

The DLC took place at the Roundhouse in Camden/London, United Kingdom (UK) from 15 to 30 July 1967. It was convened by the anti-psychiatrists Joseph Berke, David Cooper, Ronald Laing, and Leon Redler of the *Institute of Phenomenological Studies.* Joseph Berke had originally come up with the idea[1] and organized it, mainly assisted by Leon Redler and a group of dedicated helpers (Berke, 1969, p. 410). A poster advertising the congress announced it as "*a unique gathering to demystify human violence in all its forms, the social systems from which it emanates, and to explore new forms of action*" (Davis, 1968b). For two weeks, a large number of contemporary intellectuals, activists, artists, and an interested public gathered at the Roundhouse in Camden to discuss ways to change violence and destruction. The 300 to 3000 daily participants listened to lectures and poetry, attended music performances and a theatre happening, engaged in question-and-answer sessions, and met for informal discussions in between the scheduled events (Sherman, 2007, pp. 134–136).

3.3. Documentation of the Congress

The congress was covered by the contemporary press and reporting should be traceable in respective archives[2]. According to Berke, a big feature film of the conference was

---

[1] According to Berke Berke, 2005.

[2] No newspaper articles have been researched yet by the project team. Question-and-answer sessions with the press are documented on the audio files, e.g. with Stokely Carmichael on audio file 1_4_10. Coverage of the congress in newspapers is referred to e.g. in 1_2_24 by Roy Sawh and Michael de Freitas. The PETT contains queries from dozens of British and international television and radio stations, and the BBC hosted a live discussion on TV between Carmichael, Laing and Paul Goodman on the Panorama program.

made by Roy Battersby, for the BBC. Considered inflammatory, the film was censored, never shown, supposedly locked up in a BBC safe, and eventually lost (Berke, personal communication, 2005). The documentary filmer Peter Davis filmed at the congress. A 27-minute compilation of the material titled *Anatomy of Violence* was broadcasted by the precursor to the Public Broadcasting Service (PBS) in the United States[3]. The project acquired four separate films with cut-together parts of lectures and question-and-answer sessions, as well as several interviews with participants.[4] Robert Klinkert and Ian Sinclair made a 30-minute documentary (O Sonnenblume, Ah! Sunflower) for the Westdeutscher Rundfunk (WDR). It documents Allen Ginsberg's visit to London, which included attending the DLC, and was aired in June 1968 (imdb).[5]

The main events of the congress were tape-recorded by Roy Battersby and Ben Churchill (Battersby, 2010). The PET Archives and Special Collections in Toddington/Gloucestershire, UK (The Mulberry Bush Limited, 2022) keeps Berke's estate related to the congress, including the audio tapes (PET Archives and Special Collections, 2021). The Institute of Phenomenological Studies published the recordings of all the main events on 23 LPs. They are no longer available for purchase, but accessible at the British Library and US-American university libraries[6]. Digitized versions of the LPs are available on-site at the British Library in London.[7] A complete set should be in Joe Berke's estate (Berke, personal communication, 2005). Some of the LPs are accessible as sound files on a website named *Antihistory*[8] (Jakobsen, 2012).

In 1968, Penguin published transcripts of eleven speeches based on the recordings (Cooper, 1968).[9] The book has been republished by Verso (Cooper, 2015).

---

[3] The video is available on Youtube and via LibrayHub. Davis, 1968a; Library Hub Discover.

[4] The British Library holds a "collection of original 16mm negative, workprint, audio track, approximately 4 hours in total, from the Dialectics of Liberation Congress, (Shelfmark C1748)". More details are available on the website http://sami.bl.uk, Sound and Moving Image Catalogue/British Library. This seem to be the videos this project has copies of plus additional material.

[5] In 2007 a DVD *Debriefing* was published together with the original film complemented by recollections of Ian Sinclair, Chris Oakley (an associate of R.D. Laing and David Cooper) and Gareth Evans. *Debriefing* was filmed at the refurbished Roundhouse in November 2006. It is no longer puchasable, but available e.g. at the British Library in London/UK. British Library, 2021.

[6] A follow-up of a WorldCat search showed that the libraries of the Princeton and Northwestern University as well as the Beloit College own the complete set of 23 LPs. WorldCat

[7] The first version of this paper stated that the National library held a digitized version of the audio tapes. This is not true. According to Stephen Cleary (Lead Curator, Literary and Creative Recordings of the British Library), the British Library created CD-R format playback copies of the LPs. Cleary, Stephen, British Library, Lead Curator, Literary and Creative Recordings, 2022.

[8] Jakobsen claims that only 17 of the 23 LPs have been published. He doesn't cite a source. This claim is not true, since complete sets of LPs exist.

[9] The original version was also available in Danish, Swedish, Brazilian and Japanese translations. Dunst, 2017.

One of Stokely Carmichael's/Kwame Ture's[10] speeches is printed in the book *Counterculture* (Berke, 1969, pp. 182–194).A report of the congress was published in October 1967 in the IKON magazine by participant Susan Sherman and republished in 2007 in her autobiography (Sherman, 2007, pp. 134–141). The most extensive source of information online is a website created in 2010 in the run-up to an event about the DLC which was finally staged in London in 2016 (Cafe Oto, 2016). It combines stills from Davis' films, a list of well-known participants, and memories of participants (Davis, Ivimy, & Levy, 2010). It serves well for a first overview, but is far from exhaustive and does not satisfy scientific requirements. Next to that, there are bits and pieces of scattered information about the congress and its impacts to be found on the internet.[11]

3.4.   Concise Overview of Work Accomplished Before the Start of this Project

Dunst e-mailed the three people responsible for the Dialectics of Liberation website[12] to gather more information and met with Joe Berke in 2016. He visited the predecessor organization of the PET Archives and Special Collections in 2016, before the digitization of the archive had begun, and browsed through the unsorted boxes and took notes. The project's main source, for the time being, are the audio tapes stored at the PETT.

4.   Audio File Project
4.1. Significance of the Audio Files for the Digital Archive

Digitization offers the possibility to create an open-access, interconnected, and interactive multi-media online space. The audio file project used an explorative approach. It was guided by the intent of, as Tanya Clement cites Jerome McGann *"imagining what you don't know"* (Clement, 2016, p. 350) to leverage existing digital tools and possibly come up with ideas for new ones to make sound artifacts accessible for use and analysis.

According to Susan Sherman *"the Dialectics was in every sense a congress in motion. Sometimes invigorating, often confusing, at times depressing, but always filled with energy - a microcosm of ideas and theories, a hodgepodge of messianic leaders and followers, a gathering of individuals and of groups, condensed into one space, for*

---

[10] He changed his name from Stokely Carmichael to Kwame Ture in 1978. He is called Stokely Carmichael in this document when referring to the speeches at the congress, because it was his name in 1967.

[11] A more detailed account of this information will be given in another paper.

[12] Peter Davis, Jacky Ivimy, Martin Levy

one short period of time. (p.134)(...) Everybody had ideas and opinions. Everybody talked. Sometimes there were just too many voices to hear anything at all. (p.135-136)(...) It wasn't so much what the speakers said. Anyone familiar with their work (which many at the conference weren't) would have been better off, in terms of any real analysis of their ideas to read their books or to wait for a quieter moment to examine the contents of the lectures and discussions. The most interesting part of the conference came in the confrontation between the speakers, the organizers, the rest of the participants. (p.136) (...) Carolee Schneemann fought hard for her space and respect of her work and ironically the totality of the conference was best represented in her kinetic theater happening" (p.140) (Sherman, 2007, pp. 136–141). She further describes how Julian Beck, co-founder of the *Living Theater* in New York, spoke outside the settings of the large meetings, and according to Sherman experienced a community and moments of clarity he would have not imagined before.

Joseph Berke remembers that the Roundhouse "*was occupied 24+ hours a day for 16 days by hordes of people meeting, talking, fucking, fighting, flipping, eating and doing nothing, but all trying some way to 'make it' with each other and together seek ways out of what they saw to be a common predicament – the horrors of contemporary existence*" (Berke, 1969, p. 410).

Angela Davis depicts the DLC and Carmichael's speeches as follows: "*The gathering was an unlikely conglomeration of Marxist theoreticians, philosophers, sociologists and psychologists, radical political activists, hippies, and Black Power advocates. In the enormous barn-like structure, its floor covered with sawdust, the air reeked heavily of marijuana, and there were rumors that one speaker, a psychologist, was high on acid. [...] As I listened to Stokely's words, cutting like a switchblade, accusing the enemy as I had never heard before, I admit that I felt the cathartic power of his speech*" (Davis, 1989, p. 149).

These reports are cited to illustrate how mere textual representations of the events at the DLC create a bias that does not account for what took place. However, due to the inaccessibility of the audio files to a wider public, the print version of the so-called *main speakers* currently shapes the main image of the DLC. Granted, the audio files can't testify what happened in between official gatherings. Transcripts of all the audios and not only the already published ones would be a great asset to using the transcripts for combined close-reading and distant-reading research to gain more knowledge which in turn could enhance the archive.

The next step was to look into transcription.

4.2. What is Transcription and How Does it Work?

A transcription provides the text for further use, such as making soundtracks available to hearing-impaired or deaf people, text-based close-reading research, or as the basis for text-based analysis with digital tools.

In the linguistic sense, *transcription* means the representation of language in written form. There are different possibilities of transcribing oral language depending on the requirements of the foreseen usage. Phonetic transcriptions are used for linguistic studies, whereas orthographic transcriptions focus on the content of the text. Oral language can be transcribed exactly as it sounds, including all utterances of the speaker(s) or cleared of e.g. speech errors, false starts, or filler words and nonspeech elements.

There are countless transcription services available online, based on Automatic Speech Recognition software (ASR), human transcription services, or a combination of both. Before ASR was widely available, transcription relied solely on human power.[13] ASR is the technology that allows human beings to use their voices to interact with a computer and dates back to 1952. There is a growing need for applications based on voice-driven technologies: ASR is used in education, legal, finance, government, military, health care, and media contexts as well as in the corporate world. Conversations often need to be tracked and transcribed for reasons of record-keeping, as the basis for media-transcription and captioning, and for human-machine communication. A lot of money and research is directed towards this technology. It has developed at a fast pace during the last years and a multitude of commercial as well as open-source tools are available.

The basic steps of a standard ASR pipeline:

1. Cleaning of the file

   Background noise is removed and the volume is normalized.

2. Voice activity detection

---

[13] The commercial transcription provider Rev offers various services. The transcribing through humans is outsourced to free lancers. The transcribers need to pass a test, work from home and choose from transcription jobs popping up on the website on a first-come, first-served basis. Rev pays 60 US cents for a minute of transcribed audio file. The client pays USD 1,25 to Rev. According to reviews of people transcribing for Rev on YouTube video the pay amounts from USD 3 to USD 8 per hour of transcribing, not counting the time used to find a suitable project. sugarydelight.

The software identifies when speech is present during the recording and cuts the soundtrack into segments.

3. Diarization

Different speakers are identified and grouped into segments.

4. Decoding

Each segment is decoded on the phoneme level. A long list of possible phonemes rated with a score is generated for each segment. Each phoneme is like a link in a chain. Statistical probability analysis is used to deduce whole words.

5. Rescoring

From there, the computer calculates which words are likely to form the most accurate sentence.

The ASR software can be trained in two ways.

a)   By human tuning: Words that the software came across in an audio file, but which were not in the pre-programmed vocabulary, are added manually to the library.

b)   By active learning: The software itself is programmed to autonomously learn new words and thereby constantly expand its vocabulary.

ASR machines in transcription tools are designed to provide captions and transcriptions for audio and video. They utilize adaptive algorithms and different models to enhance the ASR, e.g. linguistic models to identify specific terminology and recognize accents and dialects or contextual events models that incorporate current events and news (ASR, sources cf. ubiqus, 2021; cf. Zajechowski, 2020).

ASR machines are not perfect, because of errors caused, among others, by accents, complex vocabulary, and background noise. One method to measure the accuracy of a transcript is by calculating the Word Error Rate (WER). The number of substitutions, insertions, and deletions compared to a perfect transcript is divided by the number of spoken words. However, this method does not weigh the influence of the errors on the meaning and readability of the text. Another method checks for text similarity by weighing the differences of the transcript compared to a perfect transcript according to a range between e.g. *identical* to *wrong word/missing word.* A recent comparison of the three main players Google, Amazon, and Rev.ai using these two methods for three podcasts resulted in WERs between 2.2% and 5.6% and word similarity between 94 % and 97.83% (Cloud Compiled, 2020). These results are not generalizable though since the reference podcasts were contemporary English without major accents.

A transcript allows for machine-based exploration and usage with Natural Language Programming (NLP) tools in a multitude of settings. In academic contexts, a mere transcript of the text often does not suffice, though. Research questions of e.g. acoustic, anthropological, psychological, ethnographic, journalistic, or other types require further annotation which includes the necessary contextual features and is commonly accomplished by tagging the text in Extensible Markup Language (XML). The Text Encoding Initiative (TEI) guidelines provide a wide range of elements dedicated to this purpose (cf. Text Encoding Initiative, 2021).

4.3. Test: Transcription of Selected Audio Files

At the moment the first tests were run this project had not been defined yet, therefore the results weren't saved. Two short clips of audio file 1_2_12 were transcribed with four different free 15-minute-test versions of data analysis providers.[14] The results were rather discouraging, the tools seemed not apt to accurately transcribe the mix of Caribbean/American-English and Caribbean/British-English accents and specific vocabulary Carmichael and Michael de Freitas used in their speeches in 1967. The transcripts needed quite a bit of human editing. The purchase of any of these tools was ruled out after this experiment. Commercial human-based transcription services were ruled out for ethical reasons.[15] Google, Amazon, and Rev.ai offer free test versions for their ASR. At this point of the project, Drechsler stepped in with the idea of testing the open-source tool Mozilla DeepSpeech for transcription. His research will show if Mozilla DeepSpeech is an option for the DLC project based on its accuracy and if there are options for accuracy improvement. Since the two projects ran parallel and first transcripts were needed immediately to continue this project the author searched for an alternative: Video is often watched without sound, e.g. on mobile devices, and legislation in many countries compels institutions to provide captions for their videos. There are many free and commercial services available to generate captions for videos based on ASR. An inquiry among colleagues revealed that YouTube's Google transcription/caption tool is a good choice. It is supposed to be very powerful, due to the enormous amount of data it is trained on and it is free.[16]

---

[14] Among others Nvivo and voicedocs

[15] See footnote 13

[16] The audio can be manually transcribed and added to the video or it is first transcribed by ASR and can then be edited. In both cases YouTube/Google uses the data for training and provides captions for the video, which adds to YouTube's assets. In exchange, the service is free.

4.4. Transcription of Audio Files with YouTube

As a first test, the complete audio file 1_2_12 was converted into five videos with Windows Movie Maker by uploading an audio file and a photo and extending its display to the duration of the audio file. Carmichael's long speech was split into two parts, the other three shorter speeches are by Michael de Freitas, Obi Egbuna, and Winston Branch. The split into five shorter parts was meant to make these first tests easier to handle. The videos were uploaded to a private YouTube channel via YouTube studio. YouTube provided transcripts after a while of processing. The transcripts are in lower case letters without any punctuation and not diarized. YouTube provides an online tool to proofread/correct the transcript while listening to the audio. This tool allows uploading corrected text files (txt format) as well. The transcripts are available via copy/paste either as text or as text with timestamps which mark the segmentation used for the captions, and accessible via download as caption files (srt, vtt, sbv formats). The test transcripts needed less editing than the first test version with other tools. This was a distinct, but subjective evaluation that has not been verified by calculations of WER or similarity rate, since this was not the goal of this project. These promising results led to the transcription of three sets of audio files.

A: 1_2_12, 1_2_24, **B:** 1_3_15, 1_3_16, **C:** 1_2_30

According to the first assessment by Akinro/Rasmussen they represented three different speeches of Carmichael, whereas only two speeches have been printed. The processing takes varying amounts of time, depending on the length of the video and maybe server capacity. For a video of 60 minutes, it can take two to three hours, fifteen-minute clips sometimes are processed within ten minutes. The captions were proofread by using the YouTube interface and then copied. They were compared with the Peter Davis videos and the published transcripts, as well as cross-referenced with written sources and photos. This research provided several insights:

a)   Originally there was only one speech scheduled with Carmichael for July 18, 1967 (Institute for Phenomenological Studies, 1967). 1_3_16 is the first part of this first speech, which on LP 6 is labeled *Black Power*[17], 1_3_15 has the second part and is

---

[17] Although it is not clear if the LPs have the complete uncut speeches, because neither original LPs nor verified digitizations have been checked.

published in Cooper's *Dialectics of Liberation.*[18] Sherman refers to the content of this speech as given on *Tuesday* (Sherman, 2007, p. 138).[19]

(Praxisprojekt_Stokely_Carmichael_18_July_1967_audio_files_1_3_16_1_3_15)

b)   The Carmichael text in Berke's *Counterculture, You had better come on home* is the speech from July 22, 1967 (full speech and session on 1_2_24, parts on 1_2_12), LP 7 (*Address to the Black Community).* The layout got mixed up, the introduction is missing and there are minor faults in the transcript. Parts of Carmichael's speech are featured in the video VTS_01_3 and VTS_01_4.

(Praxisprojekt_Counterculture_compared_1_2_24_VTS_01_3_VTS_01_4)

c)   The speakers Michael X, Obi Egbuna, Winston Branch and George Ware were identified on audio file 1_2_12/1_2_24.

d)   1_2_30 features Allen Ginsberg chanting and Emmett Grogan, Carmichael, and Laing giving short speeches followed by a question-and-answer session with the audience. It should correspond to LP 13 and LP 14 *Public Discussion including David Cooper, Ronald Laing*, *Stokely Carmichael and Allen Ginsberg (Berke, 1969, p. 413).* Carmichael's complete speech (minutes 13:10 – 22:28 on audio file 1_2_30) is featured in VTS_01_1 (minutes 0:00 – 10:35, no cuts, camera on Carmichael most of the time, followed   by   a   part   of   the   question-and-answer   session). *(Praxisprojekt_audio_file_1_2_30)*

e)   The recordings on the tapes are not in chronological order, cut-off pieces are distributed across several files. It is not clear if the digital files represent physical tapes. If so, some parts of the tapes must have been copied and cut.

f)   A general problem with transcription became very clear by comparing the audio files with the published texts and the transcripts: a textual representation is an interpretation of the spoken word. Punctuation, paragraphs, editing of the text related to word choice or grammar, omission of parts of the speech, of nonspeech utterances or reactions of the audience, add bias to the original.

---

[18] The introduction is missing in the print version and there are minor changes in the wording.
[19] A scan of a plan for the congress schedules Carmichael for Tuesday, July 18, 10:30 am, "Black Power" cf. Jakobsen, 2012.

g)   Carmichael's speeches are, much more than the others, interrupted (or intentionally paused) by laughter (by him and/or the audience), applause, interjections.

5.   Evaluation of the Test Results

5.1. Transcripts

The hypothesis formulated in 4.1. proved correct. The transcripts helped

a) in identifying and sorting the files.

b) in gaining more knowledge which enhances the archive by using them for combined close-reading and distant-reading research.

The YouTube captioning tool served well for providing transcripts. The feature of captions had not been envisioned at the beginning of the project but emerged during the explorative process. It proved to add valuable functions to the audio files: Captions not only assist the hearing-impaired but also help the listener with unintelligible speech (e.g. because of heavy accents). The captioning function even yielded unexpected results which will be elaborated on in 5.2. The YouTube tool might not be the ultimate solution for the transcription of all the audio files, since it is not open source. Besides, questions about copyright issues remain, even if the videos are transcribed in private mode and not published, since nobody knows if Google stores and tags the data. On the other hand, it provides fairly good transcripts for free and the editing can be done online via a user-friendly interface. The transcripts can be copied with and without timestamps, the captions can be downloaded in several different formats, the videos can be shared via private links with view-only or editing rights, which makes collaborative work on a transcript easy.

5.2. Forced Alignment, Captioning, and Searchability

While watching a captioned YouTube video with the transcript opened it is possible to jump in the video by clicking on the words in the transcript, because the text is timestamped. The *subtitles* tool on YouTube studio allows for convenient proofreading and correcting: It is possible to jump in the audio/video by clicking on words in the transcript. This led to the next question: Is there a similar tool to search in an audio file via the text, like reverse captioning? The tool envisioned should display a text which, via a click on a word, plays the respective audio and makes the audio file searchable. It would have to be based on an alignment of text and audio.

The CLARIN Common Language Resources and Technology Infrastructure

(CLARIN)[20] provides a broad range of tools for DH. There is a Forced Alignment Tool (CLARIN D) that aligns the transcript to the audio on the base of phonemes. The output of the process is a TextGrid file; an online visualization of the related phonemes and audio wave representations is accessible. It was tested with a transcript and the respective audio file with good results. Unfortunately, a tool on the ClarinD website that was supposed to provide captions based on the TextGrid file functioned unsatisfactorily (BAS Web Sevices).

To date, the author has only found commercial tools that search the soundtrack in videos. The probably most sophisticated and widely applied toolset to create captioning at scale was developed out of MIT. The services are purchasable at the 3Play Media company (3Play Media, 2021). The transcripts created by the system have timestamps for each word. They are the basis for interactive transcripts that accompany video content. They scroll along with the video while highlighting the spoken text, and let users click words to lead them to the respective moment in the video. This soundtrack/text alignment additionally allows for interlinked archival search functions via keywords in videos and is used by educational institutions such as MIT, Princeton University, and major online learning platforms (Matheson, 2015).[21]

These functions would be a great asset to the project. For example, the audio files could be searched by keywords. Thereby related audios could be grouped and the user could be directed to the respective spots in the individual files. This would allow a more lively experience for the broad public and provide academic users with easier access to information for further research.

Since the 3Play Media tool is based on the alignment of text and spoken word the technology should be easily applicable to solely audio sources. However, such a tool does not seem to be available, neither commercially nor open source. A request at CLARIN-D whether the desired tool existed was negated. (Schiel, Florian, ClarinD Supportsystem, personal communication, 2021)

6.　Recommendations

For the time being it is recommended to

---

[20] *"CLARIN is a research infrastructure that was initiated from the vision that all digital language resources and tools from all over Europe and beyond are accessible through an online environment for the support of researchers in the humanities and social sciences."* CLARIN.

[21] A video tutorial demonstrates the functions of the interactive transcript and video search tools on the MIT150 Infinite History oral history website. 3Play Media, 2011 The website has a vast collection of videos. Massachusetts Institute of Technology, 2021.

- continue with the YouTube captioning tool to produce transcripts. This decision should be reviewed dependent on Drechsler's research outcome.

- analyze audio files and video files with the help of transcripts, photos, and secondary literature and by comparing the different cuttings to create complete transcripts and audio files of the respective lectures/sessions.

- present these complete sessions converted to videos with captions and, if available, matching photos, additionally to the audio files and transcripts.

- continue with the audio files featuring Carmichael, Sherman, Egbuna, Branch, Ware, and Herbert Marcuse, since the publication of transcripts on the website depends on the permission of the copyright owners to display the content.[22]

- continue to search for contact information of other speakers (or their respective heirs) to attain permission to display their speeches

- contact Mr. Davis and ask for more video material which is of great help for identifying people, location/setting, and dates of speeches/events.

- look into the possibility of creating an interactive audio search tool in collaboration with the Institute of Phonetics and Speech Processing at the Ludwig-Maximilians-University, Munich, Germany.

Concerning the form of the transcripts it is recommended to use the YouTube transcripts with minor adjustments:

The transcripts provided by YouTube are in lowercase letters without punctuation. Many, but not all utterances of filler words like *uh* and *uhm* are captured in the transcript. Nonspeech noises like applause etc. are captured as well, but not entirely.[23] Since the transcript mainly serves for captioning and as a basis for various differing research purposes there is no need for punctuation and exact rendition of filler words and nonspeech noises. For better readability, the personal pronoun *I,* names, and the word *Black* whenever it refers to people of African ancestry should be capitalized.[24]

---

[22]To date, permission has been granted by Carmichael/Ture's son Bokar Ture, Obi Egbuna's widow Bridget Egbuna, George Ware's widow Octavia Danielson, Winston Branch, Horace Ové's daughter Indra Ové, Robert Priddy, Herbert Marcuse's grandson Harold Marcuse and Susan Sherman

[23] This is owed to the inconsistent captioning of these noises by the ASR (probably related to volume). The inconsistency does not matter for our purposes. If the transcript is use for research purposes it will have to be preprocessed according to the research project anyway.

[24] The decision not to capitalize *white* when referring to people with European heritage should be reviewed medium-termed. The societal discourse on this subject is ongoing with contradictive opinions.

7.   Lessons Learned and Good Practices

The explorative approach led to unexpected productive results and in this sense proved to be a good choice. On the other hand, it produced a great amount of interlinked data which fast raised the question of how to organize the information in an overseeable manner. The complexity of the overlapping cut and edited parts of events/speeches documented on audio files, videos, and published in print versions demands a creative approach. This accounts already for the first step of keeping account of the research results and all the more for the next steps of displaying the material on the archive's website. The data presented in this report is the first try to communicate some of the project's results. Since it is a work in progress it is highly recommended to discuss the system of data storage and presentation in a timely manner to make it accessible and usable for others than the author who herself struggles to keep an overview.

It proved good practice to cut long audio files into shorter pieces before converting them to videos and uploading these to the YouTube channel for transcription. This accelerates the transcription process. Besides, it reduces loading and processing errors while working on the files online. Since this can be done by segments of the events (e.g by speakers or lectures/question-and-answer sessions) it makes long files easier to oversee and handle. These advantages are worth the extra time spent.

In a broader sense, the project made clear the importance of free, well-performing digital tools adapted to the humanities' needs. It underscored the need for further multi-disciplinary professional networking and collaboration to build such tools.

8.   Conclusion

This report

- described the project goal and evaluated the work accomplished thus far in *3. Dialectics of Liberation Congress, Digital Archive Project Overview.*
- documented the explorative process of the project throughout *4. Audio File Project.*
- presented the main features of the tools used in *4.2. What is Transcription and How Does it Work?* and *4.4. Transcription of Audio Files with YouTube.*
- discussed practical questions emerging during the process throughout *4. Audio File Project.*

- evaluated the findings of the project in *5. Evaluation of the Test Results*

- provided recommendations on tools and short- and medium-term work on remaining audio files in *6. Recommendations.*

- identified lessons learned and good practices in *7. Lessons Learned and Good Practices.*

Reference List

3Play Media (2011). *Video Tutorial: MIT150 Infinite History Archive Search & Interactive Transcript.* Retrieved June 12, 2021, from https://www.youtube.com/watch?v=jjH_B-dx9O8.

3Play Media (2021). *3Play Media.* Retrieved June 12, 2021, from https://www.3playmedia.com/.

BAS Web Sevices. *BAS Web Services Version 3.4: Subtitle.* Retrieved June 12, 2021, from https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/Subtitle.

Battersby, R. (2010). *The congress taped.* Retrieved May 16, 2022, from https://web.archive.org/web/20220516123118/http://www.dialecticsofliberation.com/1967-dialectics/memories/.

Berke, J. (Ed.) (1969). *Counterculture.* London: Owen.

Berke, J. (2005, January 13). Personal Communication.

British Library (2021). *Beats and Friends: A Check-List of Audio-Visual Material in the British Library.* Retrieved June 22, 2021, from https://britishlibrary.typepad.co.uk/files/beat.pdf.

Cafe Oto (2016). *Dialectics of Liberation Reconvened.* Retrieved May 09, 2022, from https://web.archive.org/web/20220509093157/https://www.cafeoto.co.uk/events/dialectics-of-liberation-reconvened/.

CLARIN. *CLARIN in a nutshell.* Retrieved June 12, 2021, from https://www.clarin.eu/content/clarin-nutshell-0.

CLARIN D. *Alignierung von Transkription und Sprachsignal.* Retrieved June 03, 2021, from https://www.clarin-d.net/de/auswerten/alignierung-von-transkription-und-sprachsignal.

Cleary, Stephen, British Library, Lead Curator, Literary and Creative Recordings (2022, March 02). Dialectics of Liberation Congress 1967 (Email).

Clement, T. E. (2016). When Texts of study are Audio Files; Digital Tools for Sound Studies in Digital Humanities. In S. Schreibman, R. G. Siemens, & J. Unsworth (Eds.), *Blackwell companions to literature and culture: Vol. 93. A new companion to digital humanities.* Chichester, West Sussex, Malden, MA, Oxford: Wiley Blackwell.

Cloud Compiled (2020). *Transcription API Comparison: Google Speech-to-text, Amazon, Rev.ai.* Retrieved May 09, 2022, from https://web.archive.org/web/20220509093715/https://cloudcompiled.com/blog/transcription-api-comparison/.

Cooper, D. G. (Ed.) (2015). *Radical thinkers. The dialectics of liberation.* London, New York: Verso.

Cooper, D. G. (1968). *The Dialectics of Liberation. A Pelican Original.* Harmondsworth: Penguin.

Davis, A. Y. (1989). *An autobiography* (Reprint). New York: Internat. Publ.

Davis, P. (1968a). *The Anatomy of Violence.* Retrieved May 09, 2022, from https://www.youtube.com/watch?v=DCsGT5Gc8p8.

Davis, P. (1968b). *The Anatomy of Violence.* Retrieved June 03, 2021, from https://www.youtube.com/watch?v=88M60oBU-Ms.

Davis, P., Ivimy, J., & Levy, M. (2010). *Dialectis of Liberation, Events in 2012.* Retrieved June 04, 2021, from http://www.dialecticsofliberation.com/2012-dialektikon/events-in-2012/.

Dunst, A. (2017). *Dialectics of Liberation.* Retrieved June 12, 2021, from Department of History, Classics and Archaeology: http://www7.bbk.ac.uk/hiddenpersuaders/blog/dialectics-of-liberation/#_edn6.

imdb. *O Sonnenblume.* Retrieved May 09, 2022, from https://web.archive.org/web/20220509094007/https://www.imdb.com/title/tt1877553/releaseinfo?ref_=tt_ov_infand.

Institute for Phenomenological Studies (1967). *Dialectics of Liberation, Schedule of Lectures.* Retrieved June 05, 2021, from https://antihistory.org/tagged/Dialectics-of-Liberation.

Jakobsen, J. (2012). *Antihistory.* Retrieved June 04, 2021, from https://antihistory.org/tagged/Dialectics-of-Liberation.

Library Hub Discover. *Anatomy of Violence- Dialectics of Liberation and the Demystification of Violence.* Retrieved June 05, 2021, from https://discover.libraryhub.jisc.ac.uk/search?q=Dialectics%20of%20Liberation%20&rn=2.

Massachusetts Institute of Technology (2021). *https://infinite.mit.edu/collections/: Video Collections.* Retrieved June 04, 2021, from https://infinite.mit.edu/collections/.

Matheson, R. (2015). *Captioning at scale: Company's more-efficient captioning system processes hundreds of video-hours per day.* Retrieved May 09, 2022, from https://web.archive.org/web/20220509104033/https://news.mit.edu/2015/3play-media-more-efficient-video-captioning-0408.

The Mulberry Bush Limited (2022). *About the PET Archives.* Retrieved June 27, 2022, from https://mulberrybush.org.uk/mb3/archives/about-us/.

PET Archives and Special Collections (2021). *The Joseph Berke Archive, 1960-2003.* Retrieved June 27, 2022, from https://archives.mulberrybush.org.uk/search/all:records/0_50/all/score_desc/%20berke.

Schiel, Florian, ClarinD Supportsystem (2021, February 25). Text und Audio verknüpfen (Email).

Sherman, S. (2007). *America's child: A woman's journey through the radical sixties* (1st ed.). Willimantic, Conn.: Curbstone Press.

sugarydelight. *How Much $ Can You Make in1 Hour on Rev.com? Transcription Challenge.* Retrieved June 05, 2021, from https://www.youtube.com/watch?v=vYyB5mzskK4.

Text Encoding Initiative (2021). *Transcriptions of Speech.* Retrieved May 04, 2021, from https://tei-c.org/release/doc/tei-p5-doc/en/html/TS.html.

ubiqus (2021). *Automatic Speech Recognition (ASR).* Retrieved May 04, 2021, from https://www.ubiqus.com/en-us/technologies/asr-automatic-speech-recognition/.

WorldCat. *International Dialectics of Liberation congress, London, July, 1967.* Retrieved June 05, 2021, from https://www.worldcat.org/title/international-dialectics-of-liberation-congress-london-july-1967/oclc/6134144&referer=brief_results.

Zajechowski, M. (2020). *Automatic Speech Recognition (ASR) Software – An Introduction.* Retrieved May 04, 2021, from https://usabilitygeek.com/automatic-speech-recognition-asr-software-an-introduction/.