

Tanzanian Cassette Digitization Project

Project Results, Recommendations, and Documentation

June–August 2011

[Finalized 9 September 2011]

Table of Contents

Part I. Results and Recommendations

| | |
|--|---|
| 1. Executive Summary of Project Results, August 2011 | 2 |
| 2. Project Goals..... | 3 |
| 3. Recommendations | 4 |

Part II. Documentation and Technical Apparatus

| | |
|--|----|
| 4. Preservation and Metadata Documentation Notes | 6 |
| 5. Workflows..... | 9 |
| 6. Appendices | 13 |

This report contains the interim results and project documentation for the Tanzanian Cassette Digitization Project (TCD). The first section contains results of and recommendations for the project as of August 2011. The second section contains the workflows and technical apparatus that documents current procedures and preserves information about the organization of the digital objects.

This document was written and compiled by Jesse Johnston over the summer of 2011. The notes are meant to provide a snapshot of the project at the end of August, as well as to facilitate the continuation of the project and provide documentation of the cassettes. This will help to ensure that individual songs can be located for research purposes over a long-term period. Tom Bray set up and checked most of the lab equipment on 24 June 2011.

Part I. Results and Recommendations

1. Executive Summary of Project Results, August 2011

1.1 Results

Since June, the project has come into its main digitization phase. Along the way, a comprehensive inventory of all cassettes in the collection has been created, planning for long-term digital access and preservation has been undertaken, and a digitization workflow has been put in place. As of mid August, the work may be summed up as follows:

- Reorganization and ordering of all Tanzanian cassettes in the collection (alphabetized by primary performer and tape title).
- Comprehensive inventory of the collection (425-line .xls spreadsheet), including cross-references to song lyrics transcribed in notebooks, note of the cassette locations in the cabinet, and validated data fields for group names cross-referenced with the discography in Kelly Askew's *Performing the Nation* (U. Chicago Press, 2002).
- Printing and binding of 5 hard copies of the inventory to provide access and catalog control.
- Digitization of over 80 micro cassettes.
- Creation of over 95 gigabytes of digital audio files (.wav, .aif) and images (.tif) that include file master copies for each complete cassette, each unique song, and each cassette liner.
- Documentation and preservation information, including this report (.doc) and two major spreadsheets (.xls), containing metadata about the digital files.
- Creation and labeling of over 80 archival-quality CD-Rs to serve as stable backups and copies of the master cassette files. (Files on these CD-Rs are .aif format.)
- File naming standardized to link all files associated with each cassette.
- Recommendations for continuing the project and enhancing access for future research.

1.2 Navigating the Collection

To find individual cassettes, consult the spreadsheets or printed inventory organized by group name, song title, and tape number. These are in a separate printed document that inventories the collection, and also viewable on the lab desktop in an Excel spreadsheet. Physical tapes are organized alphabetically by group and title in the lab cabinet.

As of August 2011, it is recommended that listening for research and transcription purposes be done using the QuickTime player on the workstation for all cassettes that have been digitized. If desired, individual CDs (.wav or .mp3) may be created on the workstation to provide access copies.

2. Project Goals

The primary goal of the project is to create usable, accessible and sustainable digital copies of selected cassettes for research purposes. (Given sufficient resources, the majority of the collection may be digitized.) The collection is comprised of compact audiocassettes gathered by Kelly Askew in Tanzania. Ultimately, the digitization project aims to make the cassettes easily accessible for research, and to ensure that these copies remain accessible for research over a long-term period.

In order to maintain access to the collection, multiple copies have been created. A digital backup of each cassette will be created at the initial time of digitization, which will be stored on archival grade compact discs. These discs will be kept in a cabinet separate from the original cassettes. Next, digital files of each tape will be created and stored on the desktop computer in the lab as well as on an external hard drive. Finally, each unique song on each cassette will be isolated into a single file that will be stored on the desktop computer and an external hard drive for access and backup copies. In future, these files may be processed into smaller formats more easily moved and accessed, such as .mp3.

The project planning prioritized the development and rigorous implementation of a consistent digitization workflow and file-naming scheme so that the many digital objects created in the process can be usefully linked to each other and remain accessible in the long term. The main tool for this work is the adoption of a regularized naming convention for files that will link them to individual tapes, as well as the creation of a spreadsheet to inventory the cassette collection. Although there is a good chance that the collection will not be digitized in entirety, the comprehensive inventory gives an overview of the collection contents, makes it possible to search for and find tapes and files, and provides a means for administrative and intellectual control of the collection. It is also a necessary tool to find individual tapes, groups (and, ultimately, songs) quickly so as to allow for more efficient location and prioritization of tapes in the digitization process.

3. Recommendations

In order to increase the value of the collection, and particularly to facilitate access and long-term preservation for the future, I would suggest that the following be implemented before the project is completed.

3.1. Database

In order to simplify and facilitate the use, browsing, and searching of the digitized files, it would be ideal to input the current files into a database. This could allow the option of a graphical interface to search and browse the files, and ideally also include the possibility to play the files from within an interface rather than searching the printed index of files, opening an audio player on the computer, and then playing the file. In addition, a database would allow for easier file handling. For example, a browser window or database report could be designed that, for a given tape, displays a song list, the album cover, the major artists and group on the recording, and links to the transcribed texts of each song, perhaps in both Swahili and English (see recommendation 3.4).

Although this would be a significant further investment (e.g., purchase of FileMaker Pro or the time to initiate another management tool), it would give the collection a much higher research value. As of the end of August, I began exploring database options with Tom Bray. The current spreadsheet organization is designed to offer metadata and fields that can easily be ingested into a database program. I have also contacted Catherine Foley, digital librarian at the MATRIX center at MSU, to explore the possibility of creating a database using the KORA software that has been developed at MATRIX.

3.2. Broadcast WAV metadata completion

Digital collections require rigorous metadata for long-term preservation, particularly to remain identifiable and accessible. Therefore, in order to provide as much information about each sound files, it would be best to fill in BWF metadata for each of the current .WAV files. This may offer provide for future identification should the files become slightly corrupted, disorganized, or be lost from a database. At the very least, the minimum required information for the BEXT chunk, as suggested by the Federal Digitization Guidelines (FADGI), should be filled in. Some of this information is possible to create en masse using the batch processing with the BWF Metaedit tool from FADGI.¹ At the time of this document's August 2011 revisions, it was not possible to input BWAV data using BWF Metaedit because it will not process files created by Peak 4, possibly due to a coding anomaly in Peak or version incompatibility. Nonetheless, it would be desirable to collect this information in future.

For proposed BWF metadata fields, see Appendix 3.

3.3. Update software

At the last revision of this document, the desktop computer in the lab (a Macintosh PowerPC G5, Dual 2.5 GHz) was running the Mac OS X version 10.4.11. The aging software, including the Microsoft Office 2004 suite and OS, which is over three versions behind the current Mac OS, should be brought up to date. This is a best practice that would ensure long-term file accessibility and data stability. It is important to periodically migrate files from old formats in order to keep them viable and check for data corruption. At present, the major software used to complete the digitization (Peak 4, Office 2004) appears not to be fully compatible with newer versions. For

¹ See "Embedding Metadata in Digital Audio Files: Guideline for Federal Agency Use of Broadcast WAVE Files," Federal Agencies Audio-Visual Working Group (September 2009), available at <http://www.digitizationguidelines.gov/audio-visual/>.

example, the Office suite does not support Microsoft's XML file formats; likewise, the shortcoming in the audio sampling rates may be due to the limited processing capacity of the computer and the versions of software used. This problem does not pose an immediate threat to the viability of the digital files; however, it has already limited the quality and usability of the resulting files.

In addition, if a database is implemented, which would significantly increase the accessibility of the files in the project, it may not be compatible with these versions of the programs. FileMaker Pro, a widely used database for cultural heritage applications, should be able to interface with Peak (according to the documentation for the newest version of Peak, www.bias-inc.com) but a current version of FileMaker would not communicate with Peak 4. In order to get the most usable files with the highest integrity, I would suggest updating the software and hardware as soon as it is feasible.

3.4. Digitize notebooks

As of August, the song texts for the first batch of digitized songs are contained in bound paper notebooks. These notebooks appear to be of newsprint-quality paper, though the paper is not highly brittle and does not appear to be at a stage of immediate risk. The paper may degrade in future, however. In addition, the notebooks are handwritten transcriptions of the song texts, which are not searchable, possible to reprint, or possible to ingest into a database. Therefore, scanning the notebooks to create digital copies would offer useful access and preservation options. This might be undertaken with scanners in the lab at CAAS, or it may be possible to have done in the digitization lab of the University Library. Ultimately, it would be desirable to create electronic text versions of these song texts; since they are handwritten, this would likely require hand transcription into electronic form. An electronic version may speed translation and may speed the future publication or conversion of the texts into another printed format.

3.5. Relabel CD-Rs with archival labeling and Rehouse compact cassettes

Labels for the backup CD-Rs can be printed on archival-quality labels, perhaps obtained from Gaylord Bros (see contact information in the Appendix). This will create a uniform appearance and will provide a check of the current information listed on each CD-R. This may facilitate further digital access copies of the cassettes. In addition, it is good preservation practice to provide clean, consistent, and archival-quality materials to house the original copies of the collection. The old labels are at risk of falling off as the adhesive dries out, and it is unknown whether materials in the existing labels and plastic cases for the cassettes may affect the tape.

This step would require an investment in labels and time for printing, but it is likely that the information for the labels would not have to be retyped. Instead, the process could be implemented by printing labels from the already entered data from the cassette inventory and file metadata spreadsheets. Once a database containing these tables is created, the information fields can be output in various formats to fit a labeling template.

Part II. Documentation and Technical Apparatus

This section provides documentation for the Tanzanian Cassette Digitization Project (TCD) with an aim of communicating the major workflows and technical concerns that were important at the outset of the project. This section contains technical notes and apparatus for the TCD.

4. Preservation and Metadata Documentation Notes

4.1. Hardware Chain

At the end of June 2011, the audio signal proceeded along the following equipment chain:

Playback: Tascam 122 MK III (Cassette Player/Recorder)
out: XLR and/or RCA cables to:

CD Recorder: Marantz CD Recorder/Player CDR 500
out: RCA cables to:

USB converter: Behringer UCA202
out: USB to:

Computer: Mac PPC G4

Monitoring of signal levels: initially on Tascam deck, then on Marantz, finally via headphones (connected to Behringer USB control box).

All possible signal manipulation was avoided when creating the initial files.

No test targets or test tones were used in calibrating or assessing the scanners or audio equipment.

4.2.1. Software

The following software was used:

- Audacity 1.2.6 (open source) (for documentation and updates: audacity.sourceforge.net)
- Peak 4 (G4+) (for documentation: bias-inc.com)
- Microsoft Excel 2004 (version 11.1 for Mac)
- Microsoft Word 2004 (this document)
- Preview 3.0.9
- HP Scan Pro 6.1.3
- SQLite Database Browser 2.0

4.2.2. File Types

In choosing file types, the goal was to choose the most widely compatible files as well as file types that entailed the least amount of compression. Even though all files were not always created at the highest resolutions, uncompressed formats were desirable so as not to limit future file processing options. The .wav format was chosen as it is a de facto standard and offers the possibility to edit and embed metadata.

The following file types were preferred: .XLS, .TXT, .TIF, .WAV, .DOC; in addition, some files may be found in .JPG, .DB.

4.3.1. File Processing

Following the standards recommended by IASA TC-03 and TC-04, no initial processing was carried out on the files.² Tapes were played back without Dolby noise reduction. Once transferred to hard drive, a raw .WAV file of the entire tape side was recorded and saved using Audacity at 44100 kHz and 16-bit depth. These files were deemed to be of sufficient quality for future use in soundtracks or for films. As of August 2011, the software on the computer was not able to record at higher resolutions, possibly because of incompatibility in software versions or the outdated operating system on the desktop computer. (The cause has not been determined, but when other conditions were tried the capture program invariably crashed.)

In creating files for the individual songs on each cassette, the following operations were carried out using the batch processor functions in Peak 4 (G4+): Normalization (95%), Fade In and Fade Out (0.5–3 seconds, depending on each tape).

Cover images were scanned with an HP Scanjet 4670 using the bundled HP Scan Pro software. Image files were scanned with 24-bit color, 200 dpi, with no sharpening, and saved as .TIF files.

4.3.2. File Naming Convention

The idea behind the file naming convention is to create unique and regular file names that will provide an indication of what the file contains. Since individual compact cassettes comprised the basic unit in the physical collection, the digital collection and file-naming conventions were designed around the cassette as a basic unit. Each cassette is given a three-digit number when it enters the digitization workflow. In order to match files to individual cassettes, each file name must contain the same three-digit number as a prefix to match it to the cassette. Thus, for cassette 032, the following files would comprise elements of the digital objects corresponding to the cassette:

- 032-a-full.wav (wave audio file of the entire A side)
- 032-b-full.wav (wave audio file of the entire B side)
- 032-a1-songtitle.wav (wave audio file of the first song on side A)
- 032-cover.tif (tiff image file of the cassette cover)

Since there was no existing catalog or other database at the beginning of the project, it was essential to associate every file with a particular cassette so that individual files and elements can be matched later on. It is important to make sure at the beginning of the workflow that the three-digit number is not duplicating any extant files, and to ensure that all elements matched to the cassette at the end of the process have matching index numbers. As a final step in the digitization workflow, it should be confirmed that the file ID recorded in the spreadsheets and other metadata (inventory, etc) match; otherwise, files may be easily lost.

In future, it would be desirable to encode this information, which includes file information for multiple files that make up a coherent digital object, in a METS (Metadata Encoding and Transmission Standard) record format.

² These are the Technical Communications issued by the International Association of Sound Archives (available at <http://www.iasa-web.org/tc03/ethics-principles-preservation-strategy> and <http://www.iasa-web.org/tc04/audio-preservation>).

4.3.3. File Taxonomy

The files are organized hierarchically in the folder “Cassette Project.” As of late August 2011, the updated folder is located on the external hard drive with the volume name FWD_513_500. This drive was loaned from Tom Bray at the Duderstadt Center in August 2011.

The file tree, including basic descriptions of important folders and files, is as follows:

- / Cassette Project
 - / catalog-data
 - / groups-full.xls = *spreadsheet with names of each group represented in the tape collection (names standardized to follow Askew's Performing the Nation discography).*
 - / inventory.xls = *full spreadsheet of groups, tapes, titles, and physical location (among other data) represented in the cabinet.*
 - / tapes-digitized-full.xls = *spreadsheet containing tape_id numbers and notes about the files created for each full side of each tape.*
 - / tapes-digitized-songs.xls = *spreadsheet containing tape_ids and metadata for each individual song.*
 - / covers /
 - contains .tif image files of the cassette liners, identified by the unique tape id number.*
 - / documentation.doc = *this document*
 - / tapes-complete /
 - contains .wav and .aiff files for each side of each cassette, identified by unique tape id number, side indication, and designated "full".*
 - / tapes-songs /
 - contains .wav files for individual songs listed on each cassette's liner or according to the transcribed texts in the notebooks.*

5. Workflows

Each subheading refers to separate steps in the digitization process. Many, however, can be combined during the digitization process. For example, metadata can be gathered while a tape is playing and being recorded.

Rule of thumb: the total time to transfer each cassette should not take more than twice the time of the original! Most of the cassettes are 60 minutes long, so do not spend more than 120 minutes processing materials for each tape. The obvious minimum is the time of the original cassette, and ideally the process will take only 1.5 times the original once the workflow is established.

5.1. Tape Transfer: Analog to Digital Steps (June 2011 workflow)

Turn on the system: power strips, cassette deck, CD recorder, scanner/printer, speakers, computer. (Note on speakers: ON LAST, OFF FIRST!)

Select the next cassette from the storage cabinet. Insert the cassette into the cassette deck. (If the cassette is not rewound, do so at this time. Note: to determine which side of the cassette is ready to play, turn the cassette so that the full reel is on the left, then put the cassette in the player with that side facing out.)

Assign a unique three-digit number (e.g., 032) to the cassette. This number should be matched to the CD backup, the cassette liner, and all files that pertain to the cassette.

Prepare CD Recorder. Set levels.

Prepare computer for sound capture.

1. Open Audacity.
2. If no blank sound window opens, type APPLE + N to begin a new file.
3. Files should be captured in .WAV format at 44100 kHz sample rate and 32-bit float bit depth. (These are the default settings, but if you need to check or modify them, select the Audacity menu, then “Preferences...”; the sample rate and bit depth may be set in the “Quality” tab of the dialog box that opens.) Note that when 32-bit float bit depth is selected, the file will still be recorded at 16-bit depth.
4. When the entire chain is ready, press the record button (the one with the large red dot). You will see a wave form take shape once the sound begins.

Begin recording.

1. Press “Play” on the cassette deck. The green light should be on.
2. Before sound begins, press “Record” on the CD recorder.
3. Check levels on the CD recorder once sound begins, and confirm that the sound is registering on the computer. While the tape is transferring, it is a good time to scan the cover, prepare the CD label, and to enter metadata in the spreadsheets. (See steps 3.2 and following)

Export the file of the complete side to .WAV format and save it of the cassette in the “tapes-complete” folder. The prefix of the file name should match the cassette number and be followed by the indication of the side (“a” or “b”) and the word “full.”

Repeat these steps for the other side of the cassette. (Note: it’s best to save the file to WAV format immediately in order to avoid losing work if the computer crashes or Audacity quits.)

5.2. Label the Cassette Liner

Use the assigned three-digit number to the label the cassette and name all the files. Each cassette must have a unique, three-digit number.

1. Insert a label inside the case that contains the number of the cassette written in marker and the name of the performer (for alphabetic filing).
2. Enter the number of the cassette, the group or main performer, and the title of the cassette into the spreadsheet ("inventory.xls").
3. If the lyrics are transcribed in one of the notebooks, enter the first page number that the lyrics for the cassette in the "Notebook" column.
4. As a quality check, make sure that the three digit number you assign to the cassette matches the number on the recorded CD and that every file created from this cassette contains the same number in the file name. Finally, make sure that this number is accurate in the spreadsheet (inventory.xls, tapes-digitized-full.xls, and tapes-digitized-songs.xls).
5. Return the cassette to the storage cabinet.

Follow the steps below for scanning the cover, labeling the CD, and saving and processing the files.

Before leaving, turn off the system. (Note on speakers: LAST ON, FIRST OFF!)

5.3. Cover Copying and Scanning

These steps can be done while the first side of the cassette is playing.

Create a color copy.

1. Use the Epson copier / scanner (tape table) to copy the insert onto plain paper.
2. Trim this paper and insert it inside the jewel case of the CD.

Scan the cover.

1. Use the HP Scanner (computer desk) to scan the cassette insert.
2. Use the HP Scan Pro software on the computer with the following settings:
 - 24-bit color
 - 200 dpi resolution
 - NO sharpening or other processing
3. Save the image as a .TIF file in the "covers" folder. The prefix of the file should match the cassette number and be accompanied by the word "cover" for the insert; if an image is on the back of the insert, scan it too and use the name "inside." (See file naming convention notes in this document.)

5.4. File Processing: Creation and Storage

These steps can be done only after the raw wave files have been created.

Open raw .WAV file in Peak 4 (G4+)

Create regions for each track.

Format the Batch processor. (Select the File menu and select "Batch Processor...")

1. If there is a saved batch script, select it. If not,
2. Choose the Processes you want: Normalize... (95%), Fade In... (apply to first second), Fade Out... (apply to last 2 seconds).
3. In the Save Changes... fields, select "Set..." and select .WAV 16-bit and your desired destination folder ("tapes-songs").
4. Turn the batch processor on (the choice at the top of the dialog box).

Select all [APPLE+A] of the track.

Export the Regions.

1. On the File menu, choose “Export regions...”
2. Indicate that “none” of the extra information (regions, markers) should be contained in the resulting files.
3. Verify that the files will be saved to the hard drive, and that the destination folder (“tapes-songs”) is correct.
4. In the prefix area, input the three-digit cassette number, followed by a hyphen. If the regions are named according to side (“a” or “b”), song number (1, 2), and short title, these will be automatically appended. (Note: it is best to choose just the first word of the title since Peak will not write files with names longer than 32 characters.)
5. Select “Begin.”
6. Check in the Finder to ensure that these files have indeed appeared. If in doubt, check the files using Quicktime or another program to check quality.

Save and Close the file in Peak.

If the above does not work, Audacity can also be used to save .WAV versions of each song. However, as the version of Audacity currently running on the computer (most up-to-date for OSX 10.4) does not support batch processing, this adds significant time to the transfer procedure and should be avoided.

5.5. Metadata Gathering

As of the last update of the document, BWF metadata was not recorded. However, see Appendix 5.3 for an explanation of the proposed BWF fields and contents, which can be updated using the BWF Metaedit program.

5.5.1 Metadata for Spreadsheets

Since BWF metadata is not being gathered, it is important to gather as much information about each digital file as possible. This information is kept in Excel spreadsheets. The important spreadsheets, followed by requisite metadata fields, are:

- *inventory.xls* (an inventory of all cassettes in the Tanzanian Cassette collection);
 - tape_id: the three-digit number
 - cross reference for lyrics to notebook if applicable
- *tapes-digitized-full.xls* (a list of all files recorded for an entire cassette side, and files from cover scans, all matched to the three-digit cassette number; this is the record of information at the cassette level);
 - tape_id
 - tape title and performer
 - date digitized
 - information about the liner image file, if one was created
 - notes about the digitization process (order of songs, whether tape levels were exceptionally high or low, or other relevant information), including the exact image file name
 - the bit rate and depth of the audio capture
 - audio file format (.wav or .aiff)
 - exact file names of the audio files associated with the cassette
- *tapes-digitized-songs.xls* (a list of all files for individual songs on each cassette, matched to notes about each file, a reference to the notebook of song texts if applicable, and the three-digit cassette number);
 - tape_id

- song title
- notes about the song file (live, unlisted on liner, out of order, or other information)
- record any performers named on the cassette liner in the “vocalist (muimbaji) – default” column
- the exact file name of the associated audio file
- file format
- the page number of the transcribed song lyric, if applicable.

Note about the tapes-digitized-songs.xls file. To ensure that the file name in the spreadsheet matches the files exactly, it is possible to have the computer generate a list. To do this, open the Terminal application. In Terminal, navigate to the path (folder) containing the audio files; currently, type `cd /Volumes/FWD_513_500/`. To get a list of the filenames in the folder output to a spreadsheet, type `ls > files.csv`; in this example, “files.csv” will be the name and format of the resulting file. This creates a .csv file (open in OpenOffice) with a list of the files.

Each spreadsheet requires different metadata, but as in previous steps it is highly important to ensure that the file_id field records the same number as the three-number tape_ID. This allows for files to be identified at the tape level.

In case of errors or problems in the digitizing process, make a .txt file that briefly describes the problem and give it the name of the missing file. This file should be noted in the appropriate spreadsheet.

5.5.2. Update External Metadata: CD Liner, Notebook Contents Page

Each CD-R should be labeled by hand (using the DVD pens) and placed in an individual jewel case with a liner (see Appendix 6.2).

In addition to updating the spreadsheet metadata with cross-references for the lyrics, write down the three-digit tape ID number in the table of contents containing transcribed lyrics for the cassette.

5.6. File Backup

All files should be backed up onto the external hard drive, at least once every week. This should be done every Friday afternoon. A backup may also be done if an important point in the project is reached (e.g., finishing the digitization of an entire series), but since this takes significant time away from digitizing, do not back up files overzealously!

6. Appendices

6.1. Frequent Decisions or Challenges

Many issues will arise during the transfer process. Many will trigger decision points that require one-time solutions, but there are other situations that will arise again and again. In order to respond to recurrent issues, the following questions and responses are presented. Please add to this list when you encounter further recurring issues. They are presented in question form:

- In doubt about which track you're hearing? Try to listen for lyrics containing one of the titles listed on the cassette liner.
- The computer crashed, a file was corrupted, or something else prevented the file from being saved! What should I do? Try to determine if this was an important cassette for research purposes. (Have the lyrics been transcribed and translated?) If not, then create a text file (.TXT) with the three-digit cassette number in the file name, and make a note in the file to the effect that the files for this portion of the project were corrupted. If the file is necessary, then see if it is possible to recover an audio file from the CD that was recorded. If this is not possible, then you may have to re-record that side of the cassette (!). As a rule of thumb, always save every raw .WAV file as soon as it is imported to the computer; this will limit the amount of material lost if a file is corrupted or unable to save after a later operation.
- There are more songs on the tape than are listed on the liner or in the notebook! In this case, most of the extra material duplicates other songs on the tape, or it is material from a previous recording that was incompletely overdubbed, *or* it is an "extra" song added by the cassette vendor for added value. In any of these cases, it is not necessary to create individual song files for these extra snippets. Leave them as part of the full file, but do not create individual song files for them. This can be recorded as a note in the digitized songs spreadsheet.
- The levels are very low! In this case, adjust the levels for output from the tape player and also check the "Rec Level" setting on the CD recorder. It is best not to adjust these significantly while recording. In most cases, low levels or distortion can be fixed in batch processing when you create individual song files from Peak. It is best to leave the full .WAV files unmodified.
- The tape will not play! Try to rewind or fast forward both sides completely. If this is a tension problem, this is an easy fix. If the player will still not function, then it may be necessary to dismantle the cassette housing to look for further problems. If the cassette has a 5-screw case (most of the collection does), then this is a simple matter. Use the precision screwdriver, and be extremely careful! (If the tape falls off the reel, it is near impossible to recover unless you have the patience and time to wind the reel by hand.)

6.2. Information on the CD-ROM covers (handwritten)

Each cassette is transferred to a CD-ROM. The idea is to create a physical backup for each tape. In addition, files may be ripped from these copies if they are corrupted or lost on the hard drive. The following figure presents the information and layout in a standard schema used on the CD liners created in summer 2011. (In future, it would be nice to type these.)

| | |
|------------------------------------|---|
| CD-74-GUJ | Gold Archival Grade |
| Muungano Cultural Troupe Vol. 2 | |
| ① Side A | Dereva Ubaya Kwako Hatorudi Kibiriti |
| ③ Side B | Limbukeni Fundi Msichokoze Mapenzi Limbukeni (remix) |
| Notebook: A 22-35 | |
| 7/2011 | 014 |

From top to bottom (example content from above):

- Group Name (*Muungano Cultural Troupe*)
- Tape Title (*Vol. 2*)
- Track number on CD, tape side, and song list (for all tape sides on the CD)
- Cross-reference to pages in notebook of song texts (if applicable; eg, *Notebook: A 22-35*)
- Month/Year digitized (*7/2011*)
- Three-digit Tape ID (*014*)

If the labeling recommendation is implemented (4.5), then this information should be duplicated onto the new labels.

6.3. Proposed Broadcast .WAV metadata (BEXT and INFO chunk)

These proposals for the field descriptions were derived from the following guidelines:

- “Specification of the Broadcast Wave Format (BWF): A Format for Audio Data Files in Broadcasting,” version 2.0, by the European Broadcasting Union (Geneva, May 2011), available at <http://tech.ebu.ch/webdav/site/tech/shared/tech/tech3285.pdf>.
- “Embedding Metadata in Digital Audio Files: Guideline for Federal Agency Use of Broadcast WAVE Files,” version 1.0, by the Federal Agencies Audio-Visual Working Group (September 2009), available at <http://digitizationguidelines.gov/guidelines/digitize-embedding.html>.

Broadcast Audio Extension Chunk (BEXT)

| <i>Originator</i> | |
|-------------------|--|
| Description | Text string that identifies the entity responsible for archiving the digital item. |
| Schema | Country, University/Department |
| Example | US, UMIC/DAAS |

| <i>OriginatorReference</i> | |
|----------------------------|--|
| Description | Text string that provides or points to the best, most specific identifier at the file level and used by the entity responsible for archiving the digital item. |
| Schema | See Description for identifiers [this matches NARA example] |
| Example | [See Description for identifiers; eg, id-tapeside-title] 003-b1-chozi |

| <i>Description</i> | |
|--------------------|---|
| Description | Text string that describes the sequence, contains a unique identifier used by the originating entity, and can also point to external, non-embedded metadata. In our case, this will be the same or similar to the file name since this contains the number that groups each file by cassette and also contains a basic title. |
| Schema | UMICH, Tanzanian Cassette Project, id=file_id-side-sequence-shorttitle |
| Example | UMICH, Tanzanian Cassette Project, id=003-b1-chozi |

| <i>Origination Date</i> | |
|-------------------------|------------------------------------|
| Description | Identifies the file creation date. |
| Schema | YYYY-MM-DD |
| Example | 2011-07-01 |

| <i>Originator</i> | |
|-------------------|--|
| Description | Text string that identifies the entity responsible for archiving the digital item. |
| Schema | University, Department, Supervisor unique name |
| Example | UMICH, DAAS, kaskew |

| <i>Coding History</i> | |
|-----------------------|--|
| Description | Group of text strings that describes the coding processes applied to the audio data in creating the digital version. This is easily standardized using the EBU presets that can be chosen in the BWF Metaedit software. Most important here is the T= element, which can be used to describe the source and software. Each coding receives a new line in the chunk, thus, each song BWF file should contain two lines (analog source and digitized version). More specificity could be possible, but is prohibitive given time constraint. |

| | |
|---------|--|
| Schema | Follow FADGI guidelines (as below) and use EBU presets in BWF Metaedit software that prompt for input except in the text (T=) element. |
| Example | A=ANALOG,M=stereo,T=Tascam 122MKIII; compact cassette A=PCM,F=44100,W=16,M=stereo,T=Behringer UCA202; A/D |

Note: BWF provides for recording origination time, but this seems unnecessary for the project.

Information Chunk (INFO)

| | |
|-------------------|--|
| <i>Originator</i> | |
| Description | Text string that identifies the entity responsible for archiving the digital item. |
| Schema | University, Department, Supervisor username |
| Example | UMICH, DAAS, kaskew |

Note: BWF allows for more information in the INFO chunk, but much duplicates the BEXT chunk. For the purposes of the project, it seems most important to record the technical metadata. In addition, much of the INFO is being captured in the metadata gathered in the spreadsheets (especially tapes-digitized-songs.xls, see 3.6 above). In future, it may be possible to join these various formats into unified digital items using a METS framework (Metadata Encoding and Transfer Standard, www.loc.gov/standards/mets/).

6.4. Useful Contacts

Kelly Askew: kaskew@umich.edu
Questions about songs, individual tapes, and musical issues

Tom Bray: (734) 417-0218
tbray@umich.edu
Questions about media, computers, audio

Jesse Johnston: (734) 657-9733
jajohnst@umich.edu
Questions about this document

Gaylord Bros. Inc.
www.gaylord.com
Replacement labeling pens:
IdentiPen (for labeling CD covers)
Kaiser CD/DVD pen (for marking CDs)

Notes