LANGUAGE DOCUMENTATION & CONSERVATION

Vol. 14 (2020), pp. 503–519
http://nflrc.hawaii.edu/ldc
http://hdl.handle.net/10125/24965
Revised Version Received: 24 Aug 2020

Keeping it real: Video data in language documentation and language archiving

Mandana Seyfeddinipur
SOAS University of London

Felix Rau
Cologne University

Working with video data is on its way to becoming standard practice in language documentation. However, documenters looking on the web for guidance on standards and best practices for archiving audio-visual data encounter a vast and potentially confusing diversity of information. Unfortunately, a lot of information on archiving video is concerned with digitized film stock and not with the type of video data produced in language documentation. This paper presents relevant standards and established community best practices in a short and realistic manner, pledging to keep things real.

1. Introduction Documentary linguistics is concerned with the creation of a long-lasting, multipurpose record of the language use of speakers/signers. This multipurpose record consists of a collection of audio and video recordings with transcriptions, translations, and annotations which are archived. These are in turn then made available for use by speakers/signers, the public, and scholars from different disciplines like linguistics, anthropology, and history, and in many years to come for uses we may not be able to foresee now (cf. Himmelmann 1998, 2006; Woodbury 2003). To be true to this agenda, a crucial part of the task is to record as rich information as possible so it can be used by different parties. The rich information mandate calls for the use of video as the main recording device.

More importantly, from a linguistic theoretical perspective: language use is grounded in face-to-face interaction where information is distributed in different modalities (cf. Clark 1996). All other forms of language use are deviations and compensate for the lack of face-to-face through, for example, emoticons in texting or changing description techniques when describing a route on the phone. Manual gesture is key in all types of language use (cf. Kendon 2004). For signed languages, where the auditory modality is mostly irrelevant, the use of video is mandatory. For spoken languages this still seems to be an issue. Take procedural text, a fundamental text type linguists record. Here speakers describe and demonstrate how to make things. Speakers are showing how to “turn it and then twist it” and “put it through here” and “attach it there” and “pull it this side like that”. This type of discourse is grounded in demonstrations, in showing how objects are handled and made. When this is only

E-ISSN 1934-5275

recorded on audio, the crucial semantic content of the recording is not recoverable, while of course syntax and phonology is available for analysis. We do not know what “it” is and we do not know how “it” is turned or attached or pulled as there are many ways of doing this. While the situation has improved, sadly, the archives are still being filled with procedural text only recorded on audio. These recordings are being made recently, when digital video is widely available, and not back in the day when linguists only had pen and paper and maybe a cassette recorder at hand.

Despite the fact that cameras have become cheap and sophisticated and are easily available, we still see way too many projects today where documenters go to the field with only audio recorders capturing language as a purely auditory phenomenon. A fundamental perspective shift on the object of documentation still needs to happen from a narrow view of language as a system of rules to language use as a fundamentally multimodal negotiated joint action (cf. Seyfeddinipur 2012). Note that we assume it is common sense not to video record when, for example, speakers do not want to be filmed,[1] there are taboos around capturing images, or capturing images could endanger the speaker or the community and the like. At the same time there are situations in which the documenters feel they should not film (coming from their own cultural perspective), but the local people wish them to do so, and the documenter needs to be sensitive to that too. We will not go into detail here as the focus of the paper is on the technological side of video recording, processing, and archiving for language documentation.

The second issue around “as rich information as possible” is to clearly differentiate between video in linguistic documentation and making a documentary film (cf. Ashmore 2008 for different uses of video). National Geographic style documentaries and many ethnographic films in general are made to tell a story of a community, a family, or a single person and to create an emotional effect in the audience. The film maker is transmitting a message with the filming and editing, an interpretation of the state of affairs using artistic means like framing, camera angle, lighting, and so on. Shallow depth of field foregrounds the face of a person and blurs the background, and with that the person in focus is isolated from their surroundings. The goal in using these aesthetic means is to direct attention, create interest or suspense, and foster empathy with the key actors. Documentaries tell stories that have been scripted and composed with specific interpretations and goals to be reached by watching.

This is not the case for language documentation. Artistic measure may make a clip pleasurable to watch but, as a consequence, much information is lost (cf. Margetts & Margetts 2012). In order to understand why a speaker is using a particular register, a type of syntactic construction, or ellipses or an accent or a particular name instead of another, we need to be able to see who they are talking to when and where and how. The location, time, situation, social context, and common ground between speakers determine every aspect of how language is used. This means that the video needs to capture as much information as possible. This already shows why aesthetic

[1] See ELAR blog post about change of community attitude and increased interest and participation once they saw the videos by Joey Lovestrand: https://blogs.soas.ac.uk/elar/2017/07/27/video-documentation-of-the-barayin-language/.

measure should not be applied or be kept at a minimum. This especially means that basic filming ground rules still are in place: taking well-framed shots of the conversational interaction, no panning and zooming, long shots, deciding what the shot is and staying still on it (see Seyfeddinipur 2012).

Imagine that a community wants to use the recordings for language maintenance purposes. Here vernacular language use situations like greetings, farewells, invitations, or market negotiations should be recorded in their natural context if possible, so the recordings can be more easily utilized for teaching materials. If the recordings were made within a documentary approach bringing aesthetical measures to bear – like shallow depth of field focusing on the face of a single speaker – the shot might be beautiful, but we are not able to see what is going on in the background, who is talking to whom and where, or other measures necessary to understand linguistic phenomena, thus rendering the recording not as useful for the community itself.

A second aspect also needs to be taken into consideration. Documentaries are usually made by professionally trained film teams. Their goal is to tell a story through film. In other words, their goals are different from the goal of a documentary linguist who creates a multipurpose record for posterity. They do not write grammars, compile word lists, transcribe an oral language, translate it, and create a multimedia digital collection for posterity, nor are they concerned with ethics, informed consent, and sensitivities of making data publicly available in the same way language documenters are. Importantly, they work in teams where one person is responsible for filming, one for sound and one for interviewing, in addition to the massive amounts of work going into post-processing like editing, sound design, music selection, and so on. In the majority of PhD and postdoc documentation projects, this is all done by one linguist who is all by themselves and is working with prosumer equipment. Even in larger linguistic documentation projects, it is the linguists working together doing all of this work, while being evaluated on the paper/book publication output and not yet on the documentation collection. This challenge is exaggerated by factors like lack of reliable electricity, remoteness of documentation location, lack of urban infrastructure, political sensitivities, and community relations, just to name a few. The fact that video training is mostly lacking in documentary linguistics training does not help the situation. It is important be clear that the tasks of the documenter or the documentation team are manifold. At the same time, funds and time available are limited as is the supposedly supporting infrastructure at the home institution. The only viable option to manage is all in the field is keeping it real.

This paper is based on many years of teaching video in language documentation at ELDP trainings and other summer schools where we have learned about a variety of fieldwork situations, recording situations, and technical abilities of the recording person. Additionally, years of advising and helping depositors in managing video for archiving and analysis and discussions with our colleagues at other archives like AILLA, PARADISEC, and TLA have shaped our views of what works and what does not, reinforcing the pragmatic stance to keep things simple. There are international standards for audio-visual data creation in broadcasting and archiving. The language archives use a subset of these standards that are the most useful for both our user community and their requirements as well as the requirement for long term preservation and use of our data sets.

There are many technical possibilities, but they do not make a difference in terms of the recording quality, the archiving and processing for analysis procedures other than often complicating them, because of, for example, size or because they cannot be played on the average computers we use. Of course, it is always possible to come up with a scenario where recording at a frequency that the human ear cannot perceive or a resolution the human eye does not notice could potentially make sense. Here we are capturing the majority of project types we have come across in consulting, granting, archiving, and discussions at conferences.

We will address the main issues around the technical aspects of the use of video in language documentation from choosing a camera, to settings, recording, and processing for archiving and analysis. These are the different choice points in a documentation project and this paper provides a guide of what to consider when making the relevant choices and the consequence of a given choice for the workflow and the end product. We will base our discussion on the most general documentation project set up and context to cover most ground. What we mean here are projects which will record a range of language use scenarios in the field, transcribe and translate the materials, and archive the recordings. One necessary disclaimer is that all of the technical information we are providing here is likely to change in the next five years, and thus this guide only represents a snapshot for the next 5–10 years. The considerations detailed here are based on best practice in language documentation, taking into account the reality of documenters on the ground.

Given that there are many uses for video recording, there are also many different formats, sizes, data carriers, and recorders. These range from private use, such as making home videos of birthday parties or weddings, recording with phones for social media channels, action camera recording of skate or snowboarding sessions, to artistic experimental films and full blown professional broadcast or cinematographic film making. Each one of these uses has implications for output formats, size options, resolutions, audio qualities and formats, and processing and archiving. Thus, when looking for standards and best practices, the information available is vast and potentially confusing. The consumer video market has changed drastically over the last twenty years. There have been changes in the carriers as well as in the recording format. We went from analog VHS to Hi8 tapes to digital video tapes to recording on minidisk, on hard drives and then to recording on SD cards. What can be seen in the developments of the last thirty years is not only an advancement of media technology but also campaigns by technology companies for market leadership. These campaigns are not driven or decided by logic or technological superiority of the products but by economic power. That notwithstanding, over the past ten years the prosumer market has become quite stable with cameras recording file-based digital video to SD cards. In this paper we will lay out the relevant information for a linguistic documentation project which will allow you to make informed decisions about what camera to buy, what settings to choose, and how to process the recordings for archiving and for analysis (transcription, translation, and glossing in tools like EXMARaLDA Partitur Editor or ELAN).

A word about websites with advice. There are many websites and video tutorials offering advice on how to film, what mic to use, and which format to create. They do not agree on many points. For all of them, keep in mind that the advice is goal-driven in that it is written for a particular audience. Most of them are made for hobby videographers or aspiring YouTube stars who do not have to take many of the points into account that a field linguist has to consider, such as travel to a field site and hauling around luggage, no regular power supply in the field, compatibility with the computer you are using and the software you need to analyze the data, and crucially, producing a documentary record suitable for scientific analysis.

2. Purchase Once the documentation project is set and the location of the community and the focus of the project are determined, the particular equipment needs to be chosen which can capture the data in the best possible way. When choosing a video camera, a few main issues need to be considered according to your budget: what you want to film, how it has to be filmed and the relevant light and sound situation. Of course, it would be ideal if you could get a set of cameras for different purposes, like you get different microphones, for instance a stereo mic for recording conversations versus a head mounted mic for more phonetic analysis. For cameras, you would want an action camera set for hunting sessions or bush walks with a radio mic system, a 360° camera for multiparty conversations and dances in circles, and a night shot camera for filming story telling at night around the fire. Given the financial restrictions and the remoteness of many fieldwork locations, one camera is probably the limit, so you will need a versatile camera which can be used in most situations. You can then find a workaround for the set ups which are more difficult to capture.

A video camera can **see** but cannot **look** and it can **hear** but cannot **listen** (which holds for audio recorders generally as well). What we mean here is that you need set the camera such that it can focus and record (look) what you want it to record, and place the microphone and adjust the recording level such that what the speakers are saying is recorded well. For the visual side this means you need to manage the light conditions in a room through placement of the camera (e.g., avoid filming against the light) and adjustments accordingly. Three functions are relevant here: backlight, exposure and white balance. These functionalities allow you to adjust to extreme light situations with managing the exposure (too bright), back lighting (when speaker is filmed against light) and white balance (color cast of the image). Another very useful functionality that cameras tend to have nowadays is image stabilization which reduces blurriness due to movement.

Video recording of language use is only useful if the audio is recorded well and we can hear what the speakers say. Audio recording functionality on video cameras is based on simple built-in microphones which will pick up camera noise, lens noise, and even the sound of the operator. This is why your voice behind the camera is very loud but the voice of the speaker who is 3m in front of the camera is not captured. In other words, you have to make sure that the audio recording captures the voices of the speaker(s) and not the noise of the chickens around or the children playing next to the speaker. For that you need the right microphones and they need to be placed as close as feasible to the source of sound, the speaker(s). We will explain the two dimensions and the technical affordances one by one.

Throughout our lifetime experience, our brains have learned to listen to what we are interested in. We are able to talk to the person standing next to us in a big crowd because our brain filters out irrelevant noise – we are able to listen to a particular source. With a video recorder this can only be achieved through the use of the appropriate microphone and its respective placement close to the source. This means that you need to consider how you will connect microphones to your camera. You then need to monitor the recording with closed back headphones, so you can carefully listen to the audio. Additionally, you need to be able to adjust the recording level to avoid recording a low signal close to the noise floor or vice versa with clipping of loud peaks.

2.1 Camera types There are three general price ranges that are available on the prosumer market which you can consider. The cameras at the different price points listed below are quite similar and we just give some examples for reference models (Table 1).

Table 1. Camera types, price ranges, and required features

Example Model	Price Range	Audio Options	Video Options
Zoom Q8	£300–400	XLR input Headphone input Manual sound Control	none
Canon Vixia/ Legria HF G26 Sony FDR-AX53	£600–1000	Mini jack input Headphone input Manual sound control	Back light Exposure White Balance Stabilization
Canon XA11 Sony PXW-X70	£1500–2000	XLR input Headphone input Manual sound control	Back light Exposure White Balance Night Shot Stabilization

In general, the Zoom Q8 or Q4 is a good camera for a very small budget but ideally it should not be the main camera for a documentation project if possible. It can function as a second camera or as a back-up option. It has the advantage of excellent quality for audio recording with XLR inputs and manual sound controls, but it has no adjustment options for handling different light situations. It has an ultra-wide 160-degree lens – also called a fisheye lens – with wide angles of view but also noticeable visual distortion. While it allows you to film in very small spaces it distorts the image substantially.

The next price range is between £600–100. In this price range at the time of writing the new 4K cameras have hit the market which record in a very high resolution (3840 x 2160 pixels as compared to 1920 x 1080 pixels of HD cameras), but it is not clear if the human eye can even perceive the difference under normal viewing conditions. Cameras at this price range usually all have good image stabilization and allow to adjust to extreme light situations with managing the exposure, back lighting and white balance.

A disadvantage of cameras in this price range is that they usually have a 3.5mm mini-phone connector (mini jack) input for microphones which is not ideal. Using microphones with mini jack cables is heavily error prone. The unbalanced mini jack cable is susceptible to external noise caused by electromagnetic interference. Mini jack cables, plugs, and sockets are far less robust and reliably built than standard XLR cables. This property can be particularly detrimental in a remote field location. A better option is to use XLR microphones which are generally of better quality and are connected through much sturdier cables and connectors. You can connect the XLR with a connector cable to the mini jack input, but here you run into the danger again that the mini jack connection breaks. An XLR audio adapter – like a Beachtek box – which can be screwed between camera and tripod is a safer and sturdier solution for the connection problem.

The semi-professional camera range of about £1500-2000 overcomes all of these restrictions. They allow connecting XLR mics, have manual sound control, excellent image stabilization, and sophisticated technology to manage different light conditions and the adjustment functionalities. A bonus with these cameras is that they often are also good to shoot at nighttime. This still means that you will have a night shot which casts the image green and lets the pupil shine bright. Models like the Canon XA20 or XA30 have an option to choose a pure black-and-white in night shot mode and that also reduces pupil reflection. However, this may not be aesthetically pleasing but that is not a measure that should drive your recording in the first place. It allows you to video record story-telling which often happens in the night around a fire, rituals, dancing, spirit possession, as well as ordinary conversations over dinner.

2.2 Action cameras Action cameras are compact, rugged, and to a certain degree, waterproof digital cameras that are designed to capture sportive activities from an immersed perspective. These cameras can have a purpose in language documentation, in particular in body-worn video applications such as in bush walks, hunting expeditions, or other highly mobile activities. These cameras are designed to be small and rugged while putting less emphasis on video recording quality, setting options, and in particular, audio quality. These design principles make action cameras less than ideal general-purpose language documentation cameras.

2.3 DSLR It is important to comment on the use of digital single-lens reflex camera (DSLR) and digital single-lens mirrorless (DSLM) cameras for video recording. DSLR cameras are not primarily video cameras. They are designed for taking professional still pictures. They offer full control over aperture, focus, and exposure time. As a consequence, they allow one to take pictures with complete artistic control, including shallow depth of field, a shot that isolates the subject from its environment. This is a photographic and cinematographic technique where one part of the image is in focus while the rest is out of focus. DSLRs can also mount different types of lenses including telephoto lenses which have naturally shallow depth of field. This means the very strength of the DSLR cameras goes against the basic goal in documentation to record as much information as possible as neutral as possible, because something that might not be interesting to the documenter may be of interest to another potential user. Furthermore, most devices classified as photo cameras stop recording at 29 minutes 59 seconds to avoid the higher EU taxation on video cameras, meaning that you have to be aware that the camera just stops recording after half an hour.

2.4 Phone Cell phones are also used for video recording especially for social media purposes and are tempting as they are so small. Again, if this is the only option you have and you have a unique opportunity to record something special, you should use your mobile phone but this is about the only reason why to use a phone to record video. You should not record vertical as the sides are cut off, as you will have seen many times on social media channels. This option does not translate well to other media. You need to record the audio separately with a microphone as you will have to move the phone away to capture the interlocutors.

2.5 Batteries A word about batteries and power management in the field. All cameras come with small batteries which do not last long. Therefore, you need to purchase extra batteries with long recording times. Take into account that the bigger the camera the more battery power it will need. This has implications for your power management and the amount of batteries you need to bring along and keep charged.

2.6 External Mic options We strongly advise against using the camera internal mics as they are not designed for high quality audio recording but for capturing sound sources that are close to the camera. They also pick up camera noise, lens noise, and even the sound of the operator and other background noise. You are aiming for the best signal-to-noise ratio and you can reach that by placing your microphone as close as possible to the sound source, the speaker(s). In this sense, a camera-mounted external microphone is better than the built-in microphone, but it is often still too far away because the camera has to be at a distance to the speakers to be able to have the interlocutors with their entire upper bodies in the frame. As an aside, do not worry about a microphone being in the shot, as we are filming to document a language and do not need to pretend that we are not present. What is important is not to occlude anything with a microphone that is important to see, for example the mouth or the hands.

The next option is to have microphones connected to the camera with cables. This is a safe option if you are working with sturdy XLR cables and microphones. Here you need to have effective cable management in place, so no one trips over the cable.

There is also the option to use radio mics. These microphones work via narrow frequency range transmitting the signal they pick up. They are very nice as there are no cables running between camera and microphone. Their downside is that they compress audio for transmission and the range between transmitter and receiver is limited. As they work with radio frequencies, they are subject to interference as they work on narrow frequency range which is often crowded.

3. Camera Settings Depending on the camera brand and model, a camera will come with preset parameters and parameters that can be set by the user. The recording format and file container are mostly fixed for the device. Virtually every camera will allow the user to modify certain recording settings. This usually includes resolution, frame rate, and bitrate. These parameters determine the video quality and as a consequence the file size.

Resolution specifies the number of pixels that make up the frame. The names of the standard formats are based on the number of horizontal lines (720p or 1080p) or on the number of pixels on the diagonal (4K or 8K). Modern cameras record progressive video, indicated by the small “p” after the number of horizontal lines. Older cameras may still have an option of recording interlaced video – indicated by a lower case “i” as in 720i. Interlaced video is useful with older display technologies such as cathode-ray tube (CRT) displays. For modern displays and for most cases of language documentation, progressive video is the better option.

Older cameras might record the pixels in an interlaced process. Most modern cameras will offer two or three video resolution settings. The most common resolution is 1080p. It was standardized as ITU-R Recommendation BT.709 and is also known as HD or Full HD. Some cameras also have the option to record 720p (also known as HD or Standard HD). Many modern cameras also feature a recording resolution of 2160p (more commonly called 4K or 4K UHD).

4K – often 3840 × 2160 pixels and approximately 4000 pixels in the diagonal – is the current high-end video standard and is used in motion pictures and high-quality online videos. Files with a 3840 × 2160 are considerably larger than 1080p videos and require more storage space and powerful computers for processing. Even if a camera offers 4K video, recording 4K can push some of these cameras to their thermal or processing limits. This can result in cameras running hot and switching off when recording long 4K videos. Some cameras will have reduced image quality in combination with camera motion – resulting in rolling shutter or motion blur.

Additionally, some models might not offer the higher frame rates available for 1080p when recording 4K video. The larger file size also means that recording media will hold videos of considerably shorter length than videos shot in 1080p. If your camera does perform well in 4K mode, you are carrying enough large (≥ 128GB) SD cards, and your computer can process 4K video without problems, recording 4K video is a good and future-proof option. However, 1080p is a very good compromise between image quality and workload in processing.

Another fundamental parameter for video recording is frame rate. Frame rate specifies the frequency by which the sensor will record the (picture) frames that in sequence make up the video. Worldwide, there are two major frame rate systems. One is based on 25fps (frames per second) and generally includes 50fps and 100fps and potentially higher multiples of these numbers. This system was used in the analog PAL and SECAM TV systems. The other system is based on 24fps and 30fps and generally includes 60fps and 120fps and potentially higher multiples of these numbers. Due to historical technical reasons 24fps are in reality approximately 23.9760 (i.e., 24000/1001) frames per seconds and all other frame rates in this system are equally odd numbers derived through division by 1001. This system was used in the analog NTSC TV system.

The 25fps system is common in Europe, Africa, most of Asia (including China and India), parts of South America (Brazil, Argentina, Uruguay, and Paraguay) as well as in Australia and Oceania. The 24fps system is used in North America, most of South America as well as in Japan, South Korea, Taiwan, the Philippines, and Myanmar.

Up to the point where the human eye cannot distinguish the difference between high frame rates anymore, it is generally true that the higher the frame rate the more realistic the depiction of movement. However, modern eyes are accustomed to motion pictures in TV (24fps, 25fps, or 30fps), cinema (24fps), and online movie streaming settings (mostly 24fps, 25fps, or 30fps), so that more realistic movement does not necessarily convert to better aesthetic evaluation.

Choosing a frame rate for language documentation recordings depends on the expected use and re-use of the recording. Higher frame rates allow for a more precise analysis of movement. If time duration is calculated on frame base, the 25fps system offers easier fractions of seconds without rounding errors. If a re-use in broadcast media is expected, the frame rate system of relevant local broadcast can be the best choice.

A higher frame rate will increase the accuracy of movement recording and consequently the fluency of its display, but it will also increase the file size. The principle behind the choice of the recording frame rate should be the highest quality that is practical with the available hardware and processing pipeline. The choice between 25fps or 24/30fps frame rate systems is dependent on the expected and envisioned use and re-use of the data. Higher frame rates – i.e., multiples of 25fps or 30fps – can have benefits, but as with higher resolutions, higher frame rates increase the file size considerably and can push the hardware and processing pipeline beyond their limits.

Some cameras allow a choice of two or more bitrates for the same setting of resolution and frame rate. The bitrate determines the amount of information that is available to encode a single video frame. A higher bit rate – normally given in million bits per second (Mbps or simply M in some setting interfaces) – will result in better video quality, and again, larger file size.

Some cameras feature an automatic white balance (AWB). In most recording environments, automatic white balance will result in well-balanced colors, so that for most researchers enabling AWB will be the right choice.

In the audio settings, some cameras offer a surround sound (5.1) audio option. For stereo signals, surround sound is never necessary and will cause problems in the processing pipeline.

4. Formats and processing The principles of dealing with digital video data are quite simple. Re-encoding video (and audio) data will never improve the quality, but it will lead to an inferior video quality in most cases. Changing the wrapper of video data will not affect the data quality and can be done more freely.

Standard formats and in particular non-proprietary open standards are better from a preservation point of view than non-standardized or proprietary formats. Uncompressed encoding is generally better than lossy compressed encoding. On the other side, format support by hardware and software is important to ensure future renderability. Therefore, a universally-supported proprietary format can have better long-term preservation prospects than an obscure open format.

There are countless data and file formats for video data and the landscape can be confusing for novices and seasoned researchers alike. For the consumer and prosumer market, the situation is far more manageable. Current consumer and prosumer cameras will nearly exclusively encode video data as H.264/MPEG-4 AVC, which is a joint standard of the ITU-T Video Coding Experts Group (standardized as H.264) and the ISO/IEC JTC1 Moving Picture Experts Group (standardized as MPEG-4 AVC). It is not an open standard and contains patented proprietary components. It is however widely implemented in software as well as hardware and can be rendered by virtually every modern system that can play video. Audio data encoding shows some variation between manufacturers and models, but consumer and prosumer camcorders encode audio data either as Advanced Audio Coding (AAC), as Dolby Digital AC-3, or as an uncompressed linear pulse-code modulated signal (LPCM).

Advanced Audio Coding (AAC) is an ISO and IEC standard that is part of MPEG-4 Part 3 (ISO/IEC 14496-3), and Dolby Digital AC-3 has been standardized by the Advanced Television Systems Committee as ATSC A/52. Linear pulse-code modulated signal (LPCM) is itself not a standard, but is defined and included in numerous standards, including the standards specifying Audio CD, DVD, Blu-ray, and HDMI. LPCM is free of patents while the patents covering Dolby Digital AC-3 expired in March 2017.

Manufacturers differ in what options they implement and how they call their format choices (Table 2). There are four formats that differ in details of the video encoding and in their choice of audio formats, as well as in their choice of the wrapper format. Sony, Panasonic, and JVC use the AVCHD format for HD video. Newer Sony cameras record HD and 4K video as XAVC S, while newer cameras from Canon, Panasonic, and GoPro will save video as MP4 (MPEG-4 Part 14). Some JVC and Zoom cameras, such as the Zoom Q8, store video as QuickTime File Format (extension: .mov).

XAVC S with uncompressed LPCM audio will be stored in a not standard conform MP4 wrapper. This may cause problems when trying to render the file in software that validates formats.

Increasingly, cameras are encoding 4K video in HEVC (H.265/ MPEG-H Part 2), the successor of H.264/MPEG-4 AVC. This encoding format is currently less well supported than H.264, but we would recommend the same treatment for HEVC video data as we recommend for H.264/MPEG-4 AVC.

Table 2. Common recording formats in consumer and prosumer video cameras

Format	Extension	Video format	Audio format
AVCHD	.MTS or .m2ts	H.264/MPEG-4 AVC (level 4.1 or 4.2)	Dolby Digital AC-3
XAVC S	.mp4	H.264/MPEG-4 AVC (level 5.2)	AAC or LPCM
MPEG-4 Part 14	.mp4	H.264/MPEG-4 AVC (any level)	AAC or Dolby Digital AC-3
QuickTime	.mov	H.264/MPEG-4 AVC	LPCM

There are better and more high-quality video formats than H.264/MPEG-4 AVC or HEVC, such as JPEG2000 in a Motion JPEG2000 or Material Exchange Format (MXF) container. These formats are the right choice for digitizing film stock and many other analog videos. For the digital videos from cameras used in language documentation, conversion into these formats does not offer any advantages and will impede renderability and further processing. The same is true for re-encoding data into open standards such as with VP8 encoded video with Opus audio in Matroska or WebM containers. While open standards offer benefits in respect to long-term preservation, the re-encoding from one lossy format to another lossy format will deteriorate the data quality.

4.1 Processing for archiving The goal of archiving is always to preserve the highest feasible media quality in a format that can be guaranteed to be long-term viable and renderable. Since re-encoding will never improve, but will often decrease, the media quality, there is always a trade-off between keeping data in an original non-optimal format, and re-encoding data in a better format but losing quality in the process.

Video data will be encoded as H.264/MPEG-4 AVC in most current cameras. Unless an archive requires very specific video profile settings that differ from the encoding of the camera, re-encoding video for archiving should be avoided. For audio data, there are more options of proper treatment. The main difference is whether you have recorded uncompressed LPCM audio or lossy compressed AAC or AC-3 audio. Uncompressed audio should always be preserved for archiving, either as part of the video file or as a separate audio file.

Some archives will allow archiving of QuickTime File Format which is a combination of H.264/MPEG-4 AVC video with LPCM audio in QuickTime File Format from Zoom cameras or converted from the non-standard XAVC S MP4 from Sony. Other archives might allow other media formats wrapped in either a Matroska Multimedia container (.mkv) or in a Material Exchange Format container (.mxf). If an archive only allows MPEG-4 Part 14 containers (.mp4), the only acceptable solution to preserve audio and video in its original quality is to extract the LPCM audio in Waveform Audio File Format, also known as a WAV file (.wav), and to re-encode the audio in the MP4 file as AAC, so that the resulting H.264/MPEG-4 AVC with AAC audio is a valid MPEG-4 Part 14 file.

For MP4 files with Dolby Digital AC-3 audio, most archives will recommend re-encoding the audio as standard AAC which converts a lossy format into another lossy format. As AC-3 is a standard compliant MPEG-4 Part 14 audio format that is patent-free as of March 2017, archiving the MP4 file with Dolby Digital AC-3 can be another option, if the archive can guarantee long-term preservation of MP4 with Dolby Digital AC-3.

MPEG-4 Part 14 files with AAC audio data are generally accepted by most, if not all, language archives. While this audio format features a lossy encoding, MP4 files with this format combination are generally considered to have good long-term viability and renderability, and normally do not require any modification for archiving.

If recording includes separate audio from an external audio recorder, it is always advisable to archive the audio as a separate WAV audio file (.wav). Most video editing software allows for automatic and manual syncing of audio and video and the replacing of the original audio track with externally recorded audio. However, as long as the video recording contains passable audio, the extra effort of syncing the external audio with video recording and replacing the audio track of the video file is not worth the effort.

The current best practice in language archiving is to preserve the video and audio data as closely as possible as it was recorded by the video camera in a MP4 file with H.264/MPEG-4 AVC (Table 3). The audio is stored in the video file as AAC audio, and if applicable, as an extracted separate LPCM WAV file. Optionally, it is sometimes possible to archive an additional QuickTime file that preserves the close association of the H.264/MPEG-4 AVC video stream with the LPCM audio stream in a single media file.

Table 3. Recommended video formats for archiving of language documentation data

Format	Extension	Video format	Audio format
MPEG-4 Part 14	.mp4	H.264/MPEG-4 AVC (any level)	AAC (Separate WAVE audio file)
QuickTime	.mov	H.264/MPEG-4 AVC	LPCM

4.2 Processing for analysis The principles for processing video for analysis are based on the objective to have video files with a reasonable quality and file size in a format that will play in standard annotation software – in particular ELAN and EXMARaLDA – on all standard desktop operating systems (Windows, MacOS, Linux). Currently, this can be achieved by the combination of H.264/MPEG-4 AVC with AAC audio in a MPEG-4 Part 14 container (.mp4).

The more detailed parameter settings depend on the original quality and the acceptable file size. A video resolution of 720p is nearly always enough for analysis, as is a standard bitrate. In general, video with constant bitrate (CBR) is preferred by annotation software, even though variable bit rate (VBR) can produce a lower file size. Reducing the quality of the AAC audio will, in most cases, not result in significant size reduction.

To reduce the file size of a video file, reducing the resolution from 4K or 1080p to 1080p or 720p respectively can already reduce the file size considerably, as can reducing the frame rate from multiples of 25fps or 30fps to these base frame rates. While re-encoding the video from constant bitrate (CBR) to variable bitrate (VBR) can also reduce the file size, this may cause issues for the annotation software and should be generally avoided.

Table 4. Recommended format for working copies of language documentation data

Format	Extension	Video format	Audio format
MPEG-4 Part 14	.mp4	H.264/MPEG-4 AVC (any level)	AAC
			(Separate Audio file)

The best practice is to generate a standard compliant MPEG-4 Part 14 with a .mp4 extension. The file should contain H.264/MPEG-4 AVC video in the original frame rate – or if the original frame rate was high, optionally reduced to the 25fps or 30fps divisor. The resolution should be set at 1080p or 720p ideally with a constant bitrate (CBR). Audio should be encoded as AAC with the original bitrate (e.g., 128 kBit/s). A file with these parameters can be played in virtually all software on all relevant operating systems.

5. Conclusions

In this paper we have triangulated best practice in video documentation and preservation with the available technology and market options with actual demands on the linguist like budget, fieldwork conditions, and academic tasks. The basic stance we have taken is pragmatic: what is the safest and easiest option you can choose that will allow you to record good documentation video, process it for analysis, and archive it smoothly. The framework we have assumed here covers you for recording and, importantly, for archiving. There is a basic standard that all archives adhere too, and then there are variations due to the focus of the archive, its budget, and mandate.

When dealing with technology it is always tempting to go for bigger, better, and more functionality but often the standard is the best option for simple use in language documentation. Using broad standards has many advantages like a community of users and problem solvers. Of course, one can always use more sophisticated cameras but it does not help when there is no archive which will accept the output when it is in a non-standard format, and so to be able to archive and preserve it you would have to reduce the quality. And maybe that format will be discontinued after a short time. For example, there are many projects whose data is sitting inaccessibly on minidisks that can no longer be read.

There is a benefit to not being at the forefront of development and to letting other more sophisticated users be the beta testers who do not have to write grammars and scholarly papers and create dictionaries on top of creating video documentation materials. We often do not have the time and the resources to be using something that is not widely tested, to learn about all the downsides the hard way, and to lose time in our endeavor to document a language before it is too late.

We have based our recommendations on the largest intersection of project types we have seen as funders of documentation projects (ELDP), as archivists (ELAR, LAC, and the DELAMAN archives, especially AILLA, PARADISEC, and TLA), and as scholars in a network of documenters. There are always exceptions and we are happy to discuss any of them and to again take a pragmatic view on what is the safest and easiest solution.

The first goal in a documentation project is to create a multipurpose record of a language in use. This can be achieved by capturing as much information of language use in situ as possible as neutrally as possible. With this approach many stakeholders will be able to use the recordings now and in the future, which allows for maximal usability. When edited according to film-making principles, aesthetically pleasing clips can be created for promotional use or for the community. The ELAR vimeo channel contains such clips which have been used in exhibitions and outreach events. Saloumeh Gholami describes in the ELAR blog the positive community response to the 3-minute clip edited from her raw documentation recordings by Chouette Films.[2]

A last note on using other devices than video cameras for recording. It is true that the phones, the action cameras, and the music recorders are getting better and better, but they are not yet where we need them to be to make things easy in the field for the entire processing chain from recording to archiving. Maybe in five years the verdict will be different, but for now, in our experience what we have presented here are the most solid options. The future holds 360° cameras, action cameras, drones, contact lens video recording, and 6K is already being marketed. The future is bright and exciting and confusing … so watch this space.

References

Advanced Television Systems Committee. 2012. Digital Audio Compression (AC-3, E-AC-3) (ATSC A/52:2012). https://www.atsc.org/wp-content/uploads/2015/03/A52-201212-17.pdf.

Ashmore, Louise. 2008. The role of digital video in language documentation. In Austin, Peter K. (ed.), *Language Documentation and Description*, vol. 5, 77–102. London: SOAS. http://www.elpublishing.org/PID/064.

Clark, Herbert. 1996. *Using Language*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620539.

Floyd, Simeon. 2016. Modally hybrid grammar? Celestial pointing for time-of-day reference in Nheengatú. *Language* 92(1). 31–64. doi: 10.1353/lan.2016.0013.

[2] https://blogs.soas.ac.uk/elar/2018/01/04/zoroastrian-dari-community-reaction-to-chouette-films-video-installation/.

Himmelmann, Nikolaus. 1998. Documentary and descriptive linguistics. *Linguistics* 36:161–195.

Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In Gippert, Jost, Nikolaus P. Himmelmann and Ulrike Mosel (eds.) *Essentials of language documentation*, 1–30. Berlin, New York: Walter de Gruyter.

International Organization for Standardization. 2007. *Information technology – JPEG 2000 image coding system: Motion JPEG 2000* (ISO/IEC 15444-3:2007). https://www.iso.org/standard/41570.html.

International Organization for Standardization. 2014. *Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding* (ISO/IEC 14496-10:2014). https://www.iso.org/standard/66069.html.

International Organization for Standardization. 2016. *Information technology – JPEG 2000 image coding system: Core coding system* (ISO/IEC 15444-1:2016). https://www.iso.org/standard/70018.html.

International Organization for Standardization. 2017. *Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 2: High efficiency video coding* (ISO/IEC 23008-2:2017). https://www.iso.org/standard/69668.html.

International Organization for Standardization. 2018. *Information technology – Coding of audio-visual objects – Part 14: MP4 file format* (ISO/IEC 14496-14:2018). https://www.iso.org/standard/75929.html.

International Telecommunication Union Radiocommunication Sector. 2015. *Parameter values for the HDTV standards for production and international programme exchange* (ITU-R Recommendation BT.709). https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.709-6-201506-I!!PDF-E.pdf.

International Telecommunication Union Telecommunication Standardization Sector. 2005. *Information technology – JPEG 2000 image coding system: Motion JPEG 2000* (ITU-T Recommendation T.802). https://www.itu.int/rec/T-REC-T.802/en.

International Telecommunication Union Telecommunication Standardization Sector. 2019. *Advanced video coding for generic audiovisual services* (ITU-T Recommendation H.264:2019). Series H: Audiovisual and multimedia systems. Infrastructure of audiovisual services – Coding of moving video. https://www.itu.int/rec/T-REC-H.264/en.

International Telecommunication Union Telecommunication Standardization Sector. 2019. *High efficiency video coding* (ITU-T Recommendation H.265:2019). Series H: Audiovisual and multimedia systems. Infrastructure of audiovisual services – Coding of moving video. https://www.itu.int/rec/T-REC-H.265/en.

Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511807572.

Lovestrand, Joey. 2017. Video documentation of the Barayin language. ELAR blog, SOAS London, July 27, 2017. https://blogs.soas.ac.uk/elar/2017/07/27/video-documentation-of-the-barayin-language/.

Margetts, Anna & Andrew Margetts. 2012. Audio and video recording techniques for linguistic research. In Thieberger, Nicholas (ed.), *The Oxford handbook of linguistic fieldwork*, 13–53. Oxford: Oxford University Press.

Seyfeddinipur, Mandana. 2012. *Reasons for documenting gestures and suggestions for how to go about it*. In Thieberger, Nicholas (ed.), *The Oxford handbook of linguistic fieldwork*, 147–165. Oxford: Oxford University Press.

Society of Motion Picture and Television Engineers. 2011. *SMPTE Standard – Material Exchange Format (MXF) – File Format Specification* (ST 377-1:2011). doi:10.5594/SMPTE.ST377-1.2011.

Valin, Jean-Marc, Koen Vos, & Timothy B. Terriberry. 2012. Definition of the Opus Audio Codec. *RFC* 6716: 1–326. doi: 10.17487/RFC6716.

Wilkins, Paul, Yaowu Xu, Lou Quillio, James Bankoski, Janne Salonen, & John Koleszar. 2011. *VP8 Data Format and Decoding Guide*. (RFC 6386). https://-tools.ietf.org/html/rfc6386.

Woodbury, Anthony C. 2003. Defining documentary linguistics. In Peter K. Austin (ed.), *Language Documentation and Description*, Vol. 1, 35–51. London: SOAS.

Mandana Seyfeddinipur
ms123@soas.ac.uk

Felix Rau
f.rau@uni-koeln.de
orcid.org/0000-0003-4167-0601