

PAMMTV

Best practices guide for captioning media art

By Mariana Dias

Table of Contents

[PAMMTV Statement](#)

[What are closed captions?](#)

[Legal precedents](#)

[DCMP Style Guide](#)

[Font](#)

[Line Division](#)

[Caption Duration](#)

[Caption Placement](#)

[Spelling, capitalization, punctuation and grammar](#)

[Presentation rate](#)

[Captioning sound effects](#)

[How to identify speakers](#)

[Captioning at PAMMTV](#)

[Process overview](#)

[Caption display and file formats](#)

[Closed captions](#)

[File formats](#)

[Subtitling tools](#)

[Preparing materials](#)

[Video specifications](#)

[Source and target languages](#)

[Production](#)

[Display](#)

[Transcribing](#)

[Creating a glossary](#)

[Editing](#)

[Proofreading](#)

[Translation](#)

[Outsourcing](#)

[Subtitled videos](#)

[Describing sound in video artworks](#)

[Location sound](#)

[Designed sounds \(folly or digital\)](#)

[Developing or repeated sound](#)

[Music, melodies, songs, singing](#)

[Dialogue](#)

[Next steps](#)

[Focus group testing](#)

PAMMTV Statement

This first iteration of PAMMTV's best practices guide delineates our developing process for captioning and translating video artworks for in-person and online museum experiences. It introduces closed captions and captioning standards; our internal captioning process; and addresses the particular challenge of describing sound in artworks.

With this guide as a starting point, our **long-term goals** are:

- To cultivate a presentation, interpretation, study and care of contemporary media arts that reaches and reciprocally engages audiences across the Caribbean, Latin America and African Diaspora across language and ability
- To nurture collaborative processes between artists, museums, and the audiences the captions are intended for
- To advance accessibility in contemporary media arts by collectively identifying captioning and translation practices that extend the expressive potential of artworks

What are closed captions?

Closed captions are subtitles that contain descriptions of sounds. Also known as subtitles for the d/Deaf and hard-of-hearing (SDH), closed captions make video accessible by conveying important information from the audio track that may not be perceived by a d/Deaf or hard-of-hearing person. In addition to transcribed dialogue and sound descriptions, captions identify speakers, tone of voice, music and the sources of sounds.

The difference between subtitles and closed captions is that subtitles translate dialogue only, while closed captions transcribe dialogue and sounds. For hearing audiences, captions can also be beneficial to those who rely on text for comprehension.

Legal precedents [1](#)

Together, these entities set a legal precedent for making online audiovisual media accessible and set baseline standards for closed captioning practices:

- The ADA (Americans with Disabilities Act) does not contain specific rules for closed captions, but it does set a precedent for defining the internet as a public space.

- The FCC (Federal Communications Commission) defines standards for closed captions in broadcast television that have been applied to online media platforms.
- W3C (Web Accessibility Initiative) specifies best practices for captioning audiovisual media on the internet.
- DCMP (Described and Captioned Media Program) provides resources and services for audio description and closed captioning of educational content for students. They also offer a research-based captioning style guide that serves as PAMMTV's primary reference.

DCMP Style Guide 2

PAMMTV references the [DCMP Closed Captioning Key, Guidelines and Best Practices for Captioning Educational Video](#). Upholding accuracy, consistency, readability and preserved meaning and intention as elements of quality captioning, the DCMP provides specific instructions for how to write, style and format captions.

Below is an overview of their guidelines. For detailed instructions, the entire guide is available on the DCMP's website.

Font

Use white, medium weight, sans serif font with a drop or rim shadow. A translucent box background can be added for greater visibility.

Line Division

Each set of captions should contain 1-2 lines, with a maximum of 32 characters per line including spacing. Within this limit, sentences should be split in a manner resembling speech.

Caption Duration

Each caption should remain on screen long enough to be read, but synchronized with the audio, with a minimum duration of 1 second and a maximum duration of 6 seconds.

Caption Placement

While typically placed on the bottom center, captions can be adjusted to avoid overlap or indicate sound sources.

Spelling, capitalization, punctuation and grammar

Standard **capitalization and punctuation** rules apply for dialogue.

These elements complement descriptions by conveying information about the location of a speaker or sound, and tone of speech:

All caps are reserved for screaming or shouting. Captions for sounds should be **all lowercase** (example). **Italics** indicate that a sound or speaker are off-screen.

Ellipsis indicate a pause in speech. **Quotation marks** indicate when an on-screen speaker is reading, and **quotation marks with italics** indicate when an off-screen speaker is reading.

Presentation rate

Refers to the ideal rate of words per minute (WPM) or captions per second (CPS) needed to read and comprehend captions. Sometimes, verbatim dialogue is shortened or adapted in order to adhere to the presentation rate.

Captioning sound effects

All sounds deemed essential for understanding the content or narrative should be described in the captions. This may include background sounds that convey contextual information.

How to identify speakers

Captions can be placed under each speaker. If placement is not an option, include the speaker's name in parenthesis in the first line. Avoid gendered language and disclosing a speaker's identity before they are introduced.

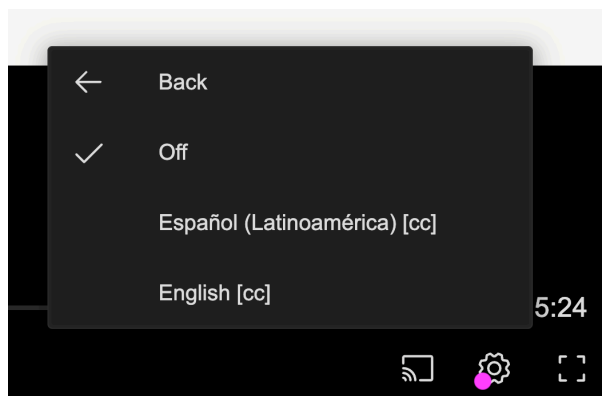
Captioning at PAMMTV

Process overview

Captions for PAMMTV are currently produced and translated in-house. This is a lengthy, detailed process that involves preparing materials, creating a glossary, transcribing the audio, syncing the text with the video, editing, formatting, proofreading and finally, translation and publishing. Although the steps are outlined in an order, they are not set in stone. As we continue to develop our workflow, our long-term goal is to bridge the artists, the museum and the audience in the captioning production process.

Caption display and file formats

Captions can be displayed in two ways: “closed” or “open.” **Closed captions can be turned on or off** by the user in the video player. **Open captions are embedded onto the video** and cannot be turned on or off. On PAMMTV, we offer closed captions that can be turned on or off in the player’s settings.



Both can be created in a video editing tool or a subtitling tool, but the formatting options and final outcome differ:

Closed captions

- Outcome of the captioning process is a subtitle file (the most common file formats are .srt or .vtt)
- Font style and caption placement are preset by the video player
- Files can be opened or played by video editing or subtitling tools, or by a video player

Open captions

- Outcome of the captioning process is a video file
- Captioning process occurs in a video editing or subtitling tool, where text style and caption placement can be customized
- Customized captions are embedded onto the video

File formats

- **.srt** - The most common format. Can be opened by all subtitling tools and most video editing tools, by plain text editors, or played as a subtitle track in media players.
- **.vtt** - Suited for streaming platforms and is the format we upload onto PAMMTV.

Subtitling tools

While captions could be created in a plain text editor, using a subtitling tool, or a video editing tool that supports subtitling, is highly recommended. There are various options for all operating systems and budgets. The main advantages of using a subtitling tool is that it facilitates the process of syncing the captions with the video, sets parameters for character count and reading speed, and allows for formatting and styling text.

Preparing materials

Gather information and materials before creating the captions. This helps determine turnaround time, parties involved, and the final outcome.

Video specifications

- What is the length of the video, or the “running time”?
- Does the video contain copyrighted music?

Source and target languages

- Is there dialogue in the video? In what language(s)?
- If there is no dialogue, what language will the captions be created in?
- Will the captions be translated?

Production

Will the captions be created in-house or be outsourced to a subtitling service provider?

Display

How will the captions be displayed: as closed captions, or embedded into the video?

Transcribing

The transcript is the foundation of captions (and subtitles). Transcribing and time-syncing the text with the audio can be done simultaneously. At this time, sounds are also described and can be edited later. The goal of this step is to form the general structure, or rough draft, of the captions.

Creating a glossary

Defining key terms ensures consistency and a certain degree of parity between the content of the captions and the subject matter of the artwork. Various elements can inform the glossary: perhaps the artist can suggest a vocabulary for describing certain sounds, or the curatorial text about the artwork may contain keywords that can be worked into the captions. The glossary can be created before, during, or after the transcript, but should be completed

before the translation. The goal of this step is to define specific language that will be used in the captions.

Editing

From the draft, the text is then edited, styled and formatted to conform to caption standards – this is when the DCMP Style Guide is referenced. This step also involves editing and polishing the sound descriptions. The goal of this step is to create the final draft of the captions. If the captions will be embedded, formatting and styling should also be completed.

Proofreading

The goal of this step is to make any final corrections before exporting the caption or video file. Recreate the watching experience by playing the captions with the video and search for these errors:

- Typos
- Inconsistent language and phrasing
- Sync and timing issues (lines that appear too fast, or stay on too long)
- Overlap between captions and text in the video

Translation

We recommend translating captions once the primary captions (or steps above) are complete. Having a glossary will ensure key terms are translated consistently. PAMMTV translates captions into Latin American Spanish and soon, into Haitian Creole.

In Latin American Spanish, language may be adapted to be gender-neutral. In this case, we use the (-e) suffix (e.g. Latine).

Outsourcing

Another option is to outsource most of the captioning and/or translation processes to a subtitling service vendor. They can create subtitles, captions (SDH) or a transcript that can be then edited in-house.

Outsourced captions must ALWAYS be reviewed before publishing. Never publish captions without reviewing them first. Depending on the service, the vendor can make the corrections.

Vendors usually charge a rate per video minute which varies by language. Be prepared to provide them with following information and materials:

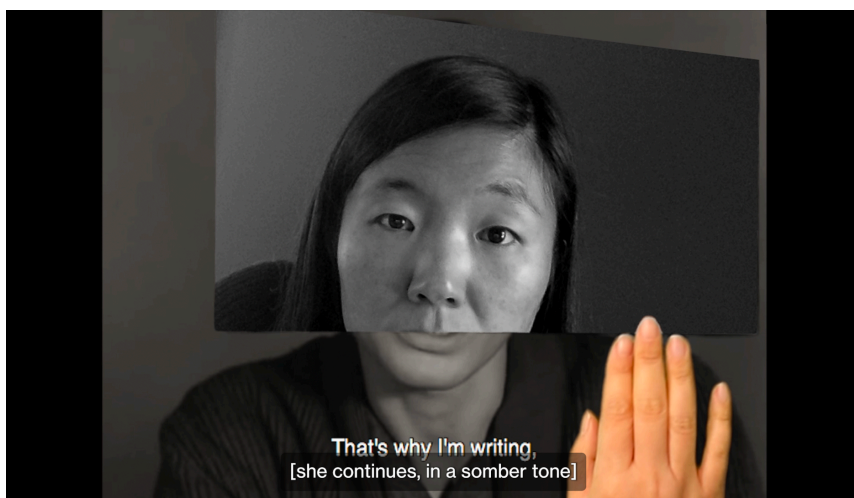
- Video file

- Source and target language(s)
- Video length (also known as the “running time”)
- If translating, the primary caption (SDH) or subtitle file
- Glossary and/or specific instructions

Subtitled videos

A video may already have embedded subtitles in its original edition. To make the video accessible to a non-hearing audience, consider options for modifying the subtitles into captions. Options may involve adapting the subtitles to captions, and then creating a new version of the video with the embedded captions, or creating a closed caption file with only the sound descriptions that can be played over the subtitled video.

Example: “Dear Tomas” by Bang Geul Han (2018)



The video contains embedded subtitles for dialogue in Swedish. Captions with sound descriptions only can be played over the video with minimal interference with the subtitles.

Describing sound in video artworks

The unconventional narratives present in video art can make describing their sounds a challenge. Like describing what we see in a painting, describing sound can be a subjective, interpretive act.

While PAMMTV primarily references the DCMP Style Guide to ensure that captions meet standards for accessibility, we search for innovative approaches to captioning that could be better suited for video art’s unconventional narratives.

Captioning considers visible (on-screen) or non-visible (off-screen) elements, and the source of a sound. These elements, in turn, determine how the sound is described and how the text is formatted. In this initial phase of the captioning program, we've identified different sound "profiles" that could inform approaches to describing sound.

Location sound

Sound that has been recorded with the image in real time.

- Common in live-action performance videos
- Sounds produced by main props or objects are amplified
- Sounds from the surrounding environment are also recorded

→ When and how should background sounds be described? How do they provide contextual information?

"Ligature/Signature" by Kamari Carter (2021)

"This suspenseful video displays a close-up scene of hands grasping a strand of rope and trying to spell a name. The sound is stark—one hears the friction of rope rubbing against skin and the sound of heavy rope dropping onto a surface. As two unidentified hands struggle with the piece of rope, off-screen someone throws more twine, increasing the level of duress."



a) The "unseen person" throws ropes on the table.



b) As more ropes are thrown, the impacts become louder and louder.

“Table Manners: Season 1 (2014-2016): Barisuka Eats Ice Fish and Mu” by Zina Saro-Wiwa (2016)

“...the eater’s performance punctuates silence with the sounds of appetite: savoring, sucking, swallowing, smacking. Audiences are forced to confront the uniformity of globalized food culture and attune to cultural distinctions between food and the body, as well as the self and others.”



a) The microphone captures a buzzing fly, a common sound if eating outdoors.



b) “Loudly” adds emphasis to the action.

Designed sounds (folly or digital)

Sounds that have been designed for the video, either through digital software or folly recordings, and then added over the video image.

- Common in digital animation videos
- Can mirror visual qualities in the video, like textures and shapes
- Can act as the imagined sound made by a character, object or place
- Can be sustained throughout the entire video, like “ambient music”
- Sustained sounds that develop and transform

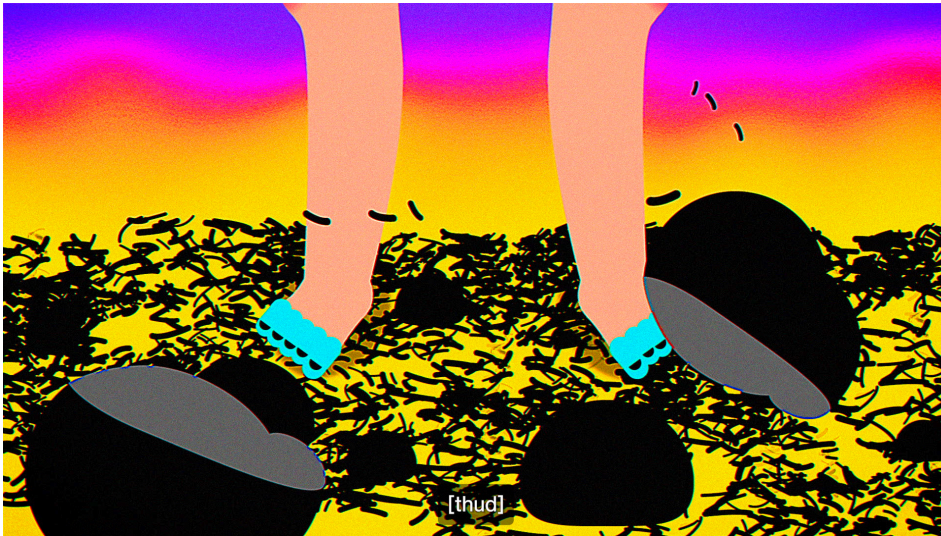
→ How can descriptions reflect qualities in the image?

→ Certain sustained, ambient sounds may not be intended as musical but could be interpreted by a hearing person as such. What are alternatives to describing the tone or mood of this type of sound, without referring to it as music?

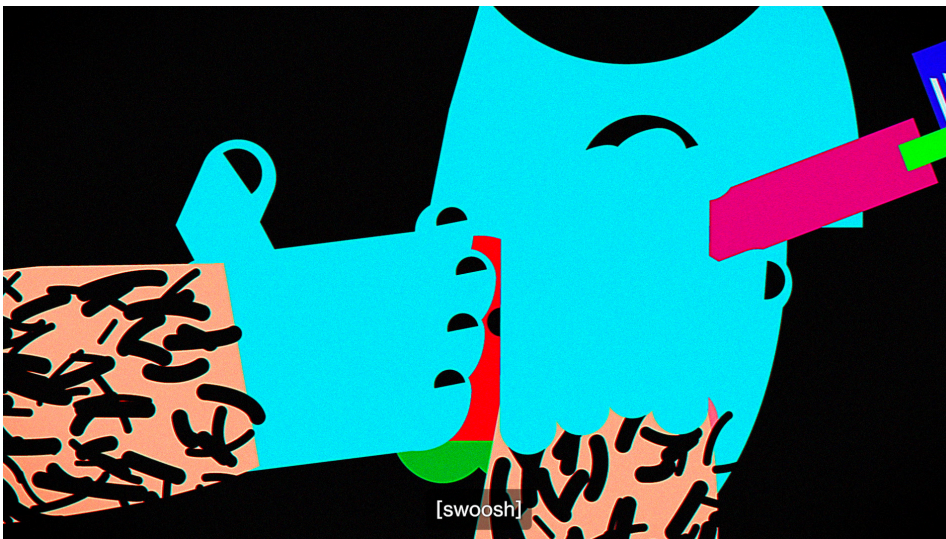
→ Can captions function as a parallel narrative?

“The Screen Shaver” by Wong Ping (2014)

Blocky shapes and bold colors are sounded by sharp, clear sounds and were captioned with onomatopoeias (finding equivalents for onomatopoeias in Spanish was a challenge).



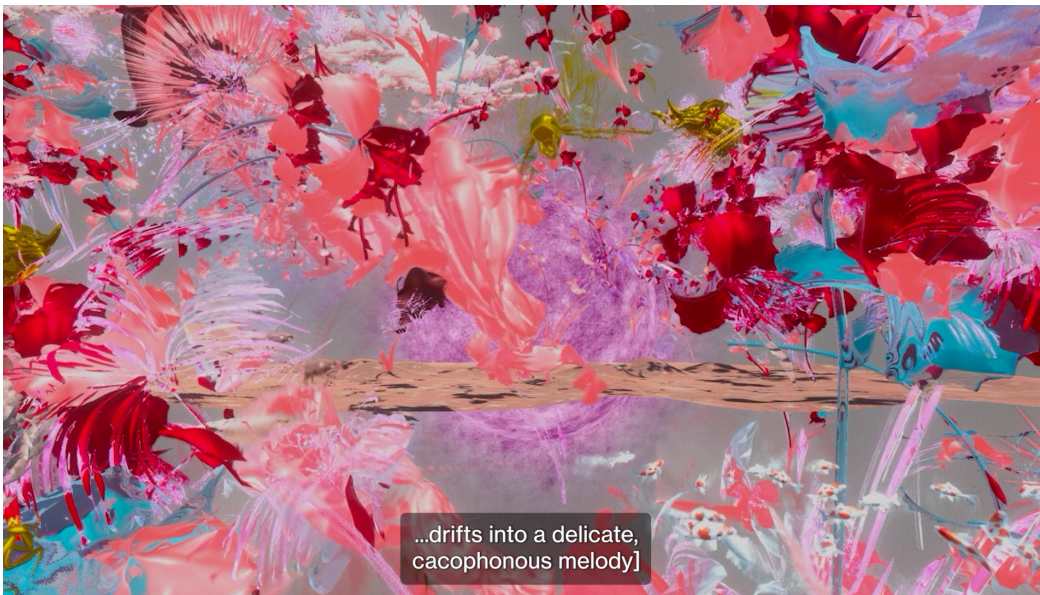
a) Chunks of hair make a blunt “thud” when they hit the ground.



b) The motion of a shave makes a swift “swoosh.”

“Love Birds, Night Birds, Devil-Birds” by Cassie McQuater (2019)

The combination of sounds can be perceived as a continuous musical store that morphs in the same unpredictable quality as the image.



“Devenir” by Tania Ximena (2021)

“...In Devenir, Ximena documents the process of the ice melting, following its journey to where it meets the sea in the Gulf of Mexico in Boca del Río, as a strategy to resist the glacier’s disappearance.”

Throughout the video, the sounds trace the cycle of water from ice to ocean.



a) The video begins with the sounds of melting ice.



b) The melted ice becomes a gentle flow of water.



c) Loud splashing contrasts with the initial low rumbling of the ice.

Music, melodies, songs, singing

Aspects of the music like change in pitch or speed, emotional quality, or instruments could all potentially be included in the description.

→ What are ways to describe music in an accessible way (without it being too technical for audiences who aren't versed in music theory)? In how much detail could music be described?

“The Night Watchers” by Federico Solmi (2020)

The video is soundtracked by piano music, which changes slightly with the addition of instruments played by the characters.



a) The video begins with piano.



b) Drums in the animation are reflected in the music.

“La fuerza que emancipa al cuerpo (The force that emancipates the body)” by Colectivo Ixqcrear (2022)

“...a woman sleeps in clay, personifying pepem ixq (butterfly woman), the spirit of all subjugated women [and is] awoken by the voices of emancipated women...”

In the video, the protagonist dances to a sound of whistling that changes in pitch and echo, giving it an otherworldly quality.



a) Whistling begins, sounding unaltered.



b) Whistling in a lower pitch signals a shift in spiritual frequency.

Dialogue

Describing dialogue is widely addressed in captioning standards and is achievable through text placement, speaker identifiers and text formatting.

In the film “Onde Está Mymye Mastroiagnne?” by biarritzzz (2023), one of the characters has an AI, text to speech generated voice. Compared to another character who speaks who is a human, the AI voice perfectly pronounces and enunciates speech in a flat monotone.

→ With the increasing use of AI generated voices in video art, how could captions reflect the difference between human and artificial voices?

Next steps

Focus group testing

PAMMTV is committed to testing its captions with focus groups of d/Deaf and hard-of-hearing audiences, and its caption translations with Spanish and Haitian Creole speaking audiences.

These are some preliminary questions we aim to address:

- Did the word choice and phrasing of the captions match thematic qualities perceived in the artwork?
- Did the captions tell a coherent story, or did they awkwardly break the experience apart?
- What makes sense to be described and how? Why? How could something be described instead?
- Was anything described unnecessarily? Was anything missing?

Links & Resources

1. [3PLAYMEDIA's “Ultimate Guide to Closed Captioning”](#)
2. [DCMP Captioning Key](#)