

# Making Accessible **MEDIA**

## Humber College Captioning Style Guide



Version 3.0  
May 24, 2019

# Table of Contents

- Captioning at Humber..... 1**
- Differences Between Captions & Subtitles..... 1**
- Transcription Guidelines ..... 2**
  - Spelling..... 2
  - Punctuation ..... 2
  - Filler, False Starts, and Discourse Markers..... 3
  - Lectures & Tutorials..... 4
  - Slang & Contractions..... 5
  - Numbers ..... 5
  - Profanity ..... 5
  - Bad Quality or Missing Audio..... 6
- From Transcript to Captions ..... 7**
  - Screen Placement & Appearance..... 7
  - Breaking Lines ..... 7
  - Timing..... 9
  - Compressing ..... 9
  - Identifying Speakers ..... 9
  - Expression & Emotion..... 12
  - Foreign Languages & Accents ..... 12
  - Sound Effects ..... 13
  - Indiscernible or Inaudible..... 15
  - Music ..... 15
- Conclusion ..... 17**

# Captioning at Humber

## Fostering inclusivity

Captioning is one of the tools Humber College uses to help students prosper in an inclusive, barrier-free environment. By captioning audio-visual material – including lectures, classroom materials, Humber-produced media, and student projects – Humber ensures that people with sensory, information-processing, or language or literacy challenges can fully engage in the college learning experience. Additionally, search engines such as Google and Internet Archive use captions to organize and index videos for easy retrieval.

Captioning has been available on televisions since the mid-1960s, and has been expanding in the media marketplace ever since. The Canadian Radio-television Telecommunications Commission (CRTC) imposed captioning requirements on television broadcasters in the 1990s, and is pressing for increased captioning of online video. Other regulations, such as the Accessibility for Ontarians with Disabilities Act (AODA), necessitate captioning on video shown by any public or private institution in the province, whether broadcast over the air, shown to a live audience, or posted online.

At Humber, the Accessible Media Department primarily provides closed captioning. This allows the viewer to choose whether or not the captions appear on the screen. When a video is meant to be played in an area where the volume is permanently muted, such as a public hallway or the gym, open captions – captions permanently “burned” to the video that cannot be turned off – may be utilized.

## Differences Between Captions & Subtitles

### Application & intention

The word “caption” is sometimes used interchangeably with the word “subtitle,” but there are subtle differences between the two.

Captions are designed for an audience that cannot hear any of the audible content in a video. Speakers need to be clearly identified, either by the caption’s placement and timing, or by speaker identification markers. Sound effects, music, and other audible cues also need to be captured and presented. In most cases, foreign language content is identified but not translated.

The assumption with subtitles is that the audience can hear what’s happening on screen, but the dialogue is spoken in a foreign language, drowned out by background noise, or obscured in some other way. The audience is otherwise able to identify speakers and on-screen sounds without assistance.

# Transcription Guidelines

## From sound to screen

The first step to captioning a video is creating a transcript of the audio. The goal is to capture the audio as verbatim as possible while making the speakers' meaning clear. However, written language and the spoken word don't always correlate; what makes sense coming out of a person's mouth can seem unintelligible when transcribed to text. In order to maintain clarity, the "fully verbatim" rule has a few exceptions. Also, different types of programming present different transcription challenges. Below are some of the issues that arise when transcribing, and how to resolve them.

### Spelling

---

**Spelling should always be done to Canadian standards** – colour, neighbour, centre, theatre, etc. Exceptions are made for proper names, such as The Kennedy Center. Exceptions to spelling rules also come into play when transcribing lectures and tutorials, described in this guide.

### Punctuation

---

Conventions of Canadian English grammar should be followed consistently when transcribing text. However, if following a grammar rule sacrifices clarity in a transcript, go for clarity. Since a transcript must represent a speaker's words as closely as possible, it cannot be rewritten into clean copy; we must work with what we are given.

Serial or "Oxford" commas should be employed for clarity, paying special attention to the speaker's meaning. For instance, "The colours of the flags are red, and green and blue," describes two flags, one of one colour and one of two colours. "The colours of the flags are red, and green, and blue," talks about a red flag, a green flag, and a blue flag. A sentence like, "I like peanut butter and jam, and tuna sandwiches," conveys a much different meaning than: "I like peanut butter and jam and tuna sandwiches."

Parentheses should not be used when transcribing text as they are used in captions to denote sound effects or other audio content that is not narration or dialogue. In parenthetical statements where commas would not be clear, double-dashes should be used.

**"The colours you see, red, green, and blue, are often found on flags."**

**"The colours you see -- red, green, and blue -- are often found on flags."**

Dashes should be used to end a sentence that is cut off or interrupted. But when speakers interrupt themselves or trail off at the end of a sentence, use ellipses to mark the interruption.

Breaking up run-on sentences is something that you will have to do often. When someone is speaking, it can be difficult to determine where one sentence ends and the next should begin. Don't be afraid of starting sentences with conjunctions like "and," "but," and "or" in order to break things up. If you need to, go back and read the sentence without listening to the audio to ensure it still makes sense.

Semicolons can also cause confusion in captions. Whenever possible, avoid their use. Where a semicolon would be used to separate clauses, often a full sentence break would be appropriate. Though semicolons can also be used as "super-commas" to separate items in a complex list, try and see if those list items can be broken into separate sentences. If the use of semicolons is unavoidable in the project you're working on, aim for consistency to maintain clarity.

## Filler, False Starts, and Discourse Markers

---

Much of our spoken language – more than most of us would care to admit – is taken up by filler. "Um," "uh," "er," and other sounds that pepper the edges of words are generally meaningless. When we speak and listen to others, our brains rarely process filler, ignoring it in favour of the "real" words that convey meaning.

Discourse markers – words and phrases like "so," "well," "I mean," "you know," "okay," and others – are, in an academic sense, used to manage flow and structure in speech. In spoken language, they're often used so much that they become filler – "like" is a prominent example of this. If discourse markers cause confusion and removing them will not change the meaning of the text, they can be omitted.

**"So, you know, if we take a look at this example..." → "If we take a look at this example..."**

False starts are common in speech, and present a challenge when transcribing. As a rule of thumb, single-word false starts or stutters can be left out.

**"When, when you consider..." → "When you consider..."**

**"I, I, I think..." → "I think..."**

Longer false starts, when a speaker says a few words, stops, then continues speaking, sometimes on a completely different thought, are more challenging. Whether what follows the false start is capitalized or not depends on whether it can be considered the beginning or the continuation of a sentence.

**"So when we look at... let's take a look at this."**

**If "So when we look at..." was to be removed, "Let's take a look at this," becomes the beginning of the sentence; "Let's" is capitalized.**

**“This is what I mean when I say... when we say this is a theory.”**  
**“when we say this is a theory,” can’t stand on its own as a sentence without,**  
**“This is what I mean”; “when” is not capitalized.**

When transcribed, filler, false starts and overused discourse markers can make captions difficult to understand and synchronize. Unless these words seem deliberate or convey aspects of the character speaking, they are removed if their removal doesn’t alter the speaker’s meaning.

## Lectures & Tutorials

---

Much of what we caption consists of lectures and on-screen tutorials. When an instructor is describing what’s happening on the screen, it can be difficult to create a transcript that makes sense. We have a few standards in place meant to help keep transcripts consistent and easy to understand.

Alphanumeric keys are presented as capital letters, numbers, or symbols (A, X, 1, 0, ?, /). Cursor and navigation keys are named and capitalized (Left Arrow, Page Up, Backspace, Insert, Home, End, etc.). The modifier keys on a Windows keyboard are Ctrl, Alt, Shift, Caps Lock, and Windows (or Win) Key. Mac keyboards have Command and Option keys.

Key combinations are presented with a “+” between the modifier and the key being pressed: Ctrl+S, Ctrl+Alt+Del, Option+X. For clarity, certain symbol keys are spelled out in this context, such as Plus, Minus, and Period: Alt+Plus, Shift+Period.

Menu navigation is another challenging aspect of tutorials. Different speakers handle menu navigations differently, but it is common for them to describe what they’re clicking as they go. Capitalize and spell menu selections as they appear on screen, separated by commas: To save your document, click on File, Save As.

On-screen elements should be rendered in the transcript as they are on the screen, but only when those elements are being referred to specifically: To create a black-and-white photograph, go to Image, Adjust, Black & White. Also remember that Canadian spelling is to be used throughout a tutorial except when the speaker is specifically referring to an on-screen element – in Photoshop, you pick a colour using the Color Picker.

“Okay” is used when transcribing speech, but click OK for the on-screen action as that is how OK is rendered in most computer programs.

Tutorials that involve solving math equations or typing computer code are especially difficult, and developing standards of transcribing coding demonstrations is challenging. Often, what a speaker is saying while typing or writing can’t be transcribed in a way that makes sense on its own. In these cases, simply try to be as consistent as possible throughout the video.

## Slang & Contractions

---

How slang and contractions are rendered depends on the type of program being captioned. For dramatic programming, informal contractions like “gonna,” “wanna,” “kinda,” and “till” can be left as-is if that is how the character speaks. With materials such as lectures and tutorials, they should be rendered as “going to,” “want to,” “kind of,” and “until” respectively for clarity and consistency.

Contractions should be rendered as spoken whenever possible: “wouldn’t,” “isn’t,” “could’ve,” even “ain’t.” Exceptions are made with particularly difficult contractions. For instance, “I’d’ve” should be rendered as “I’d have” for clarity.

“All right” is always spelled out as two words.

## Numbers

---

Numbers from one to nine are written as words unless they are part of an equation in a lecture: We have two equations, one is 1 plus 2. Note that the mathematical symbols in equations are spelled out: 45 divided by 9, 3 minus 1, 2 times 2, X equals 3, etc.

Words are also used for uncertain or estimated amounts: Thousands of people paid hundreds of dollars to see dozens of animals.

Numerals for low numbers are also used for school grades if the number is at the end: “Grade 7,” but “Seventh Grade.”

Use the word “per cent” (two words) rather than the % symbol.

## Profanity

---

Any profanity or “swear words” should be captured as they are spoken. Don’t try to replace words with less offensive terms. If it’s said, it’s captioned. If you’re working on a show with language or subject matter that you find offensive, you can ask for a different assignment.

In cases where profanity is censored, capture it as it is presented in the audio.

<b>“Who the (bleep) do you think you are?”</b>	<b>→</b>	<b>Profanity obscured by bleeping sound.</b>
<b>“Who the f--- do you think you are?”</b>	<b>→</b>	<b>Profanity partially obscured, with the first sound of the word being distinct.</b>
<b>“Who the (no audio) do you think you are?”</b>	<b>→</b>	<b>Profanity completely silenced from audio.</b>

## Bad Quality or Missing Audio

---

Sometimes, no matter how many times you listen to a sentence, you just can't figure out what a speaker is saying. In these cases, you would either mark the audio as (indiscernible) or (inaudible).

(indiscernible) is used in cases where the audio is there, but you just can't make it out. Maybe a speaker is mumbling, the audio level drops for a moment, or the speaker's accent makes a word unclear.

A word is marked as (inaudible) when it is entirely obliterated from the recording and there is no way to make out what is being said. Either the audio levels drop out completely, or an incidental sound fully overpowers what the speaker is saying.

**“When we look at (indiscernible) chart in the assignment...”**  
**It's possible to hear a word, but it can't be made out clearly.**

**“Looking at (inaudible). Then we move on to...”**  
**A word or group of words can't be heard at all.**

If you are uncertain about what you hear when transcribing, it is important to mark a word as either indiscernible or inaudible rather than making guesses or assumptions. Wrong guesses cause confusion.

### Further Reading

These are some basic rules for transcribing audio. Refer to The Canadian Press Style Book and Canadian Press Caps and Spelling for more complex grammar questions, and to ensure words are spelled to Canadian standards.



# From Transcript to Captions

## Presentation and timing

After completing a transcript, it must be imported into our captioning software, broken up into captions, and timed to the video. Captions are a maximum of 32 characters per line containing no more than two lines of speech. A third caption line can be added at the top for speaker identification or other information if necessary.

## Screen Placement & Appearance

---

Due to technical limitations of video platforms such as Panopto, YouTube, and Facebook, captions must be placed centre-justified at the bottom of the screen. This makes the need for clear and consistent speaker identifications extremely important as a caption cannot be placed right beside a speaker as seen in television captions and movie subtitles.

## Breaking Lines

---

When breaking lines of text, both within and between captions, every effort should be made to avoid breaking up phrases that work as “syntactic wholes,” words that work together to form a complete unit of speech.

- 1 – Keep articles (a, an, the) and nouns together

INCORRECT	CORRECT
<b>We link back to the raw footage</b>	<b>We link back to the raw footage</b>

- 2 – Keep adjectives and adverbs with the words they modify.

INCORRECT	CORRECT
<b>Is this green frog poisonous or not?</b>	<b>Is this green frog poisonous or not?</b>

- 3 – Keep prepositions with the words they refer to.

INCORRECT	CORRECT
<b>The car is parked beside the house.</b>	<b>The car is parked beside the house.</b>

**4** – Avoid splitting proper names

INCORRECT	CORRECT
Today, we'll look at Albert Einstein's theories.	Today, we'll look at Albert Einstein's theories.

**5** – Clauses should be kept together

INCORRECT	CORRECT
We visited Paris first, and then went to Rome.	We visited Paris first, and then went to Rome.

**6** – Also, try to maintain a balance between the lengths of the top and bottom lines.

INCORRECT	CORRECT
You want video tracks and audio tracks to be included.	You want video tracks and audio tracks to be included.

A caption must end at the end of a sentence, even if that caption ends up being only one line. Adjust captions above so that the end-of-sentence caption contains enough words to stay on screen long enough to be readable while maintaining synchronization with the audio.

INCORRECT	CORRECT
Companies use social media more and more in order to reach out to audiences. It's also about starting to think as content creators.	Companies use social media more and more in order to reach out to audiences. It's also about starting to think as content creators.

Other examples of caption line division can be found in the DCMP Captioning Key

If it's hard to break up a caption properly, look at it within the context of the other captions around it. You may be able to make adjustments to adjoining captions in order to create clean breaks.

Sound effects and music descriptors should be presented in their own caption. See the sections on Sound Effects and Music in this guide for details.

## Timing

---

Each caption should have a minimum duration of 1:15 (one second and 15 frames) and a maximum of four seconds. Captions of four seconds or more should be broken up into smaller captions to sync more comfortably with the audio.

When timing captions for lectures, tutorials, and other such videos, it's enough that the captions sync as closely to the audio as possible. Timing captions for more cinematic programs, such as dramas and documentaries, should match shot changes as closely as possible. This makes captions look more professional and less tiring to read. If shot changes are quick, a caption can stay up through multiple shot changes in order to be comfortably read.

If speech is very fast, some of the rules for breaking caption lines may need to be broken in order to have as many words as possible in a caption so that it remains on screen long enough to read. Captions that have less time on screen can “steal” frames from previous or following captions as long as all the affected captions are up for at least 1:15 and are still reasonably in sync with the audio. Usually, up to 15 frames from a caption above or below can give a caption enough display time without a noticeable effect on audio synchronization. If captions go by too fast for these adjustments to work, edit their content for time. Be mindful of maintaining the speaker's meaning when editing.

When a speaker is very slow, it's best to present what's being said in smaller, shorter captions, even if that means creating one-line captions in the middle of sentences in order to maintain a comfortable display rate. Try to avoid blanks in between captions of continuous speech if possible. If there is a pause long enough to warrant a blank between words, end the first caption with ellipses, continuing the sentence when the speaker starts again after the blank.

## Compressing

---

Although it's something we try to avoid, there will be times when editing the content of a caption to maintain a comfortable reading speed is inevitable. As mentioned in the transcription guidelines, first look at cleaning up filler, discourse markers, and false starts. After that, make whatever edits that are possible without altering the speaker's meaning.

## Identifying Speakers

---

In a lecture or tutorial where there's only one speaker, no speaker identification is necessary. If there are multiple speakers who are not on the screen, such as a software tutorial, speakers will need to be identified.

If the names of the speakers are known, put the name of who is speaking in the top line of the caption in capital letters, followed by a colon. Use first and last name for the speaker's first appearance, only the last name subsequently.

**BOB SMITH:**  
I'd like to introduce  
Jane Jones to the discussion.

**JANE JONES:**  
It's great to be here, Bob.

**SMITH:**  
It's great to have you.

If speakers' names are unknown, the captioner can choose suitable descriptors for each speaker.

**SPEAKER 1:**  
This place hasn't changed.

**SPEAKER 2:**  
I don't know how long the building  
has been around.

**SPEAKER 3:**  
I'm sure it's been  
at least 100 years.

**SPEAKER 2:**  
That's impressive!

Other common speaker IDs can be MAN, WOMAN, STUDENT, PROFESSOR, etc. Choose speaker IDs that are appropriate to the content, and be careful to maintain consistency throughout the video.

In the case of only two unknown speakers, a hyphen followed by a space to signal a speaker change may be sufficient.

- My biggest project  
is my series of regular  
Photoshop tutorials.

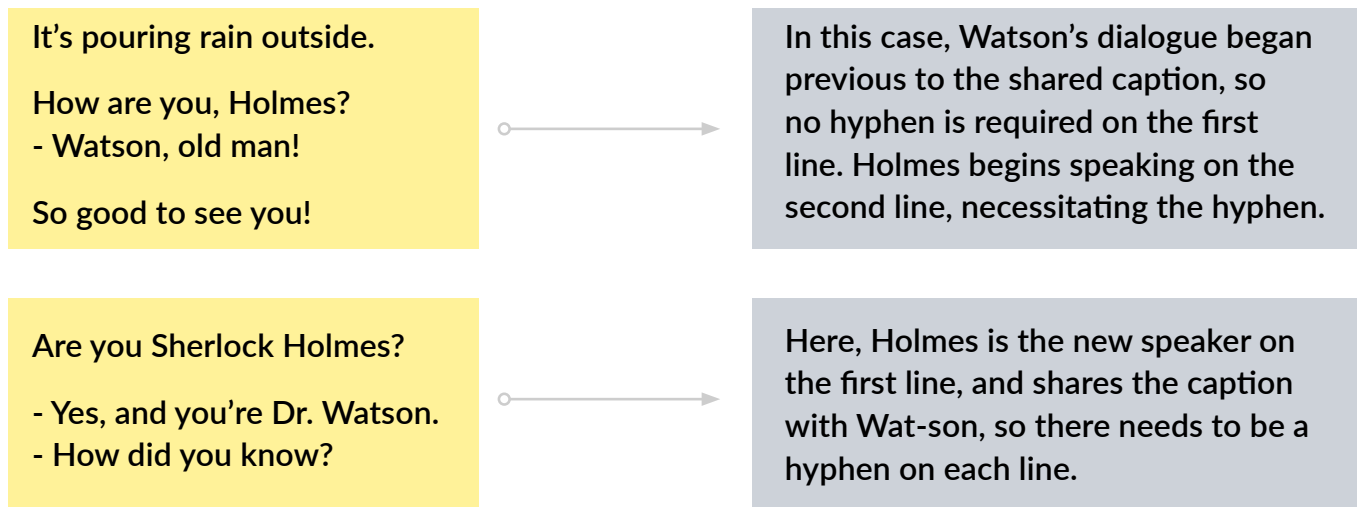
- How often do you do those?

- About once a month.

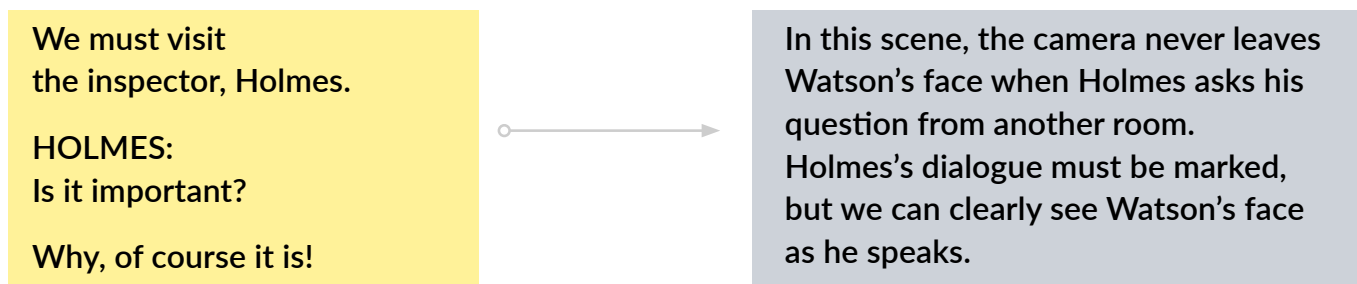
Cinematic programming is more of a challenge. There are several ways to ensure the proper identification of who is saying what, and different methods will work better with certain programs than others. To limit confusion, try to stick with no more than two different methods of identifying new speakers, and remain consistent throughout the project.

In situations where only one speaker appears on the screen at a time and speaker changes are visibly obvious, identification does not need to be added to the caption.

If timing is too close to give each speaker on the screen a separate caption during an exchange, a hyphen should be used to differentiate speakers with dialogue in the same caption. Each speaker should have a separate line in the caption. Be sure to use the hyphen when signalling a change in speaker.



Off-screen characters who are interacting with characters who are on the screen should have their dialogue identified with their name in capitals followed by a colon on the top line of the caption. The on-screen speaker's dialogue doesn't require identification if it is obvious when they are the one speaking.



Voice-over and narration is identified by the speaker's name in capital letters, followed by (V.O.) and a colon in the first caption of the narration. If the name of the narrator isn't known, ANNOUNCER, NARRATOR, or other appropriate identification can be used.

**WATSON (V.O.):**  
**The city you consider as home**  
**is never so attractive**  
**as when you return to it**  
**after long and difficult times.**

## Expression & Emotion

---

For clarity or dramatic effect, it may be necessary to add the expression or emotion a speaker uses to say something. For on-screen speakers, the expression or emotion is written in lowercase, in parentheses, on the top line of the caption above the dialogue as if it were a speaker ID.

**(sarcastically)**  
**I've never had**  
**so much fun!**

For off-screen speakers, the expression or emotion shares the same line as the speaker identification, before the colon.

**CHILD (whispering):**  
**Could someone please**  
**open the door?**

## Foreign Languages & Accents

---

Untranslated foreign language content in a video is tagged the same way as sound effects. Do not try to translate what is being said when captioning, even if you are familiar with the language being spoken.

If the language is known, include that in the description.

**(speaking Russian)**  
**(speaking Elven language)**

If the language is unknown or indiscernible, simply tag it as a foreign language.

**(speaking foreign language)**

When subtitles are used on-screen to provide a translation, no captions are necessary. Be sure to time captions before and after subtitles so that subtitles are not obscured.

Accents and dialects do not need any special treatment in lectures or tutorials. The same goes for documentary programs. Except for accepted dialect spellings such as “y’all” and “guv,” stick with standard English spellings.

In dramatic programs, you may come across a character whose accent is so central to the narrative that ignoring it makes no sense. As spelling heavily accented words phonetically is subjective and would cause confusion, tag a character’s accented dialogue the first time they speak, but stick with standard spelling.

**(in Russian accent)**

**We make big trouble for moose  
and squirrel.**

If an accent is too thick to understand, mark it in a similar way you would mark a foreign language. If you are unfamiliar with the accent and it hasn’t been identified in the program, simply tag it as an accent.

**(speaking in thick accent)**

If the speaker’s accent has been identified in the program, include it in the description.

**(speaking in thick Spanish accent)**

## Sound Effects

---

As the focus of captioning is to make a video presentation understandable to an audience that cannot hear any of the audio, sound effects must be described. Capturing every non-speech sound in a video, however, is often neither possible nor practical. Inserting sound effect markers for mouse clicking in a tutorial can distract from the lecture. If a character in a movie is having a drink by a window, there are traffic noises outside, the ice in her glass clinks, a cell phone rings, a bird chirps, there’s jazz playing in the background, someone beside her laughs... the screen can become cluttered with competing sound effect descriptions. It is up to the captioner to decide which non-speech elements are significant to the video’s narrative, and then present them accordingly.

A sound effect is presented in lowercase letters, enclosed in parentheses, in its own caption.

**(splashing)**

**(engine revving)**

It should stay up for at least two seconds if possible. Otherwise, if the sound is quick but vital to the narrative, the minimum 1:15 duration is acceptable.

Abrupt sounds are described in simple present tense.

**(snaps)**  
**(horn honks)**

For ongoing sounds, use present continuous tense.

**(dog barking)**  
**(sirens wailing)**

If the source of an onscreen sound is obvious, a description of the sound itself is sufficient.

**(meows)**  
**(cheering)**

If the source is not obvious, it should be identified.

**(kitten meows)**  
**(distant crowd cheering)**

If a described background sound continues through several captions, this should be indicated where time permits.

**(siren wailing continues)**  
**(crowd cheering continues)**

Sometimes, sounds can overlap. If it doesn't make sense to separate out sounds into their own captions, multiple sound effects can be tagged in one caption. Be careful not to overfill the caption. Capture the significant sounds, but maintain clarity.

**(gunfire and screaming)**  
**(engines revving and horns honking)**

Always present sounds as a source (if needed) and as objective a description of the sound as possible. Onomatopoeia can be hard to understand for someone who has never heard a particular sound. It can also be culturally biased; while a native English speaker may immediately recognize "woof-woof" as the sound of a barking dog, there are dozens of ways the sound of a barking dog is rendered around the world. Sticking to simple English descriptions of sounds is the best way to ensure clarity for the widest audience.



## Indiscernible or Inaudible

---

As with transcription, words that can't be made out should be marked as (indiscernible) or (inaudible) accordingly.

**When we look at  
(indiscernible) chart...  
Look at (inaudible).**

Longer periods of indiscernible or no audio should be marked following the rules of sound effects. If the audio is indiscernible for a specific reason, describe as appropriate.

**(no audio)  
(indiscernible chatter)**

## Music

---

There are many ways music is used in video presentations, leading to many different ways the captioning of music is handled. Use judgment to decide whether or not a musical interlude should be captioned at all. If a span of music is less than five seconds, or if captioning the music will interrupt dialog or cause confusion, leave it out unless it seems significant to the narrative.

If appropriate, incidental background music can be identified by a music note displayed for as long as the music plays without interruption. This is the only time a caption length should exceed four seconds.



When the title of a piece of music isn't known or lyrics are not discernible, a description of the mood and style of the music can be used. Descriptions of music, like sound effects, are presented in lower case in parentheses.

**(mournful orchestral music playing)  
(bright jazz playing)**

If the musical piece or the artist is known, identify it.

**("Ode to Joy" by Beethoven playing)  
(Louis Armstrong singing "It's a Wonderful World")**

Whether identified generically or with an artist and title, music descriptors should be left up for two seconds, followed a music note if time permits.

During musical passages when a song's lyrics are clear for an extended period of time and captioning them will not interfere with narration or dialogue, caption the lyrics with musical notes at the beginning and end of each caption. If time permits, each line of lyrics should have its own caption.

♪ I see trees of green ♪  
♪ Red roses, too ♪  
♪ I see them bloom ♪  
♪ For me and you ♪

If the lyrics are too fast to time to single lines, then use two lines.

♪ Say your prayers,  
little one ♪  
♪ Don't forget,  
my son ♪  
♪ To include everyone ♪

# Conclusion

This is a basic overview of how to caption effectively. Any changes to captioning procedures, as well as further examples and answers to frequently asked questions will be added to this document regularly.

There are many other sources of captioning information online, including

## **Closed Captioning Standards and Protocol for Canadian English Language Television Programming Services**

Canadian Association of Broadcasters, 2008

<http://www.cab-acr.ca/english/social/captioning/captioning.pdf>

## **TED Subtitling Resources**

<https://www.ted.com/participate/translate/subtitling-resources>

## **The CBC Captioning Style Guide, CBC, 2003**

<https://www.dcmp.org/caai/nadh218.pdf>

## **Which Sounds are Significant? Towards a Rhetoric of Closed Captioning**

Sean Zdenek, 2011

<http://dsq-sds.org/article/view/1667/1604>

Created with generous support from:



BROADCASTING ACCESSIBILITY FUND  
FONDS POUR L'ACCESSIBILITÉ DE LA RADIODIFFUSION



HUMBER