INTERNATIONAL STANDARD

ISO/IEC 20071-23

First edition
2018-09

Information technology — User interface component accessibility —

Part 23: Visual presentation of audio information (including captions and subtitles)

Technologies de l'information — Accessibilité du composant interface utilisateur —

Partie 23: Présentation visuelle d'informations sonores

Reference number
ISO/IEC 20071-23:2018(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 35, *User interfaces*.

A list of all parts in the ISO/IEC 20071 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Captions and/or subtitles for audio content provide visual alternatives for audio information in audiovisual content. This document provides requirements and recommendations on the production and design of the visual presentation of audio information (including captions and subtitles) that supports users who cannot make use of the audio content.

The use of this document helps to support universal and inclusive media content production practices. It provides guidance for producers, exhibitors, or distributors of audio content (including the medium of distribution and the medium of delivery) to support the accessibility and usability of visual alternatives of audio content.

Different jurisdictions have different expectations of what belongs in a caption or a subtitle. From the point of view of the user, what is important is receiving the information in an accessible design, not the technological means of its delivery. This information can include text conveying speech, sound information, verbatim transcription of the spoken word content, translations of the spoken word content, etc. This document uses "visual presentations of audio information" to include all audio information needed to be made accessible for some users.

Standardized guidance for producing visual presentations of audio information is important to meet a variety of needs. For example, it is important to recognize acceptable values for specifying typography variables such as the letter size and/or number of characters in visual alternatives that rely on text.

Providing visual presentations of audio information (including captions and subtitles) can be beneficial to all, and in particular to diverse users who cannot hear or understand the audio content in diverse contexts, including: persons with hearing loss, persons who are deaf or hard of hearing, persons with learning difficulties or cognitive disabilities, persons watching a movie in a non-native language, persons who need the content to be in another language, persons who cannot hear the audio content due to environmental conditions, or circumstances where the sound is not accessible (e.g. noisy surroundings), the sound is not available (e.g. muted, no working speakers), or the sound is not appropriate (e.g. a quiet library). Although this guidance acknowledges the need of visual presentations of audio information to provide non-visual presentations for diverse users, it does not include guidance for producing non-visual presentations, such as spoken captions/subtitles (see ISO/IEC TS 20071-25 for further reference) and tactile displays (e.g. Braille). The production, delivery, and exhibition of visual presentations of audio information based on this standard are not intended to interfere with or change the meaning of the audio content.

The production, delivery and exhibition of visual presentations of audio information vary according to the time and methodology of production, the technology used for its production, the system of delivery, and the display (including the brand and model of the display).

Information technology — User interface component accessibility —

Part 23: Visual presentation of audio information (including captions and subtitles)

1 Scope

This document provides guidance for producers, exhibitors, and distributors on the visual presentation of alternatives to audio information in audiovisual content, such as captions/subtitles.

This document provides requirements and recommendations that are intended to support users who are not able to use the audio information, prefer to use a visual representation of audio information, or prefer both audio and visual presentations.

NOTE Many users do not have a choice, for instance, when in a noisy environment (e.g. bar, restaurant, etc.). In these situations, the user does not select a visual presentation of audio information but is offered the content with captions/subtitles.

This document acknowledges the various needs and preferences of viewers (end users) as well as the different approaches to visual presentation of audio information. It applies to all presentations of visual alternatives to audio information intended to be presented as captions/subtitles.

This document does not apply to the presentation devices or transmission mechanisms used to deliver the content or visual presentations of audio information. These devices could include, but are not limited to: televisions, computers, wireless devices, projection equipment, DVD and home cinema equipment, video game consoles, and other forms of user interfaces technology. This document does not apply to transcoding files and formats for the various video outputs.

This document gives guidance on visual presentations which are delivered in the same language as in the audio (i.e., intra-lingual captions/subtitles) and visual presentations which are translated into a different language (i.e., inter-lingual captions/subtitles). This document does not apply to the specific process of language translation.

This document helps to improve accessibility. This document does not establish requirements on specific industries (e.g. television broadcasting, motion pictures) nor is it intended to supersede specific international standards within their domain.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at http://www.electropedia.org/

— ISO Online browsing platform: available at https://www.iso.org/obp

3.1
information
knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning

Note 1 to entry: Although information will necessarily have a representation form to make it communicable, it is the interpretation of this representation (the meaning) that is relevant in the first place.

[SOURCE: ISO/IEC 2382:2015, 2121271]

3.2
content
interactive or non-interactive object containing information represented by text, image, video, sound, or other media

[SOURCE: ISO/IEC/IEEE 23026:2015, 4.6]

3.3
caption/subtitle
transcription or translation of audio content, visually presented together with the content

Note 1 to entry: Transcriptions or translations include speech and/or non-speech information.

Note 2 to entry: Transcriptions or translations are often suitable for use as an alternative or a complement to the audio content.

3.4
open caption
open subtitle
caption/subtitle visually presented regardless of user preference

Note 1 to entry: Open captions/subtitles do not include visual elements that are a part of the original video contents.

3.5
closed caption
closed subtitle
caption/subtitle visually presented only in response to user preference

Note 1 to entry: Closed captions/subtitles are usually presented by a specialised device or decoder.

3.6
non-speech information
NSI
part of the audio content, other than spoken words

Note 1 to entry: NSI can convey information about: plot, humour, mood, or meaning of a spoken passage.

EXAMPLE Speaker identification information (e.g. off-screen speakers and multiple on-screen speakers), sound effects, music (e.g. singing, background music, instrumentation), manner of speaking (e.g. whispering, emotion, word emphasis), audience reaction (e.g. laughing, groaning, booing).

3.7
visual alternative container
VAC
opaque or translucent area visually presenting alternative content

Note 1 to entry: While VACs are largely used to provide alternatives to audio content, they can also be used to provide alternatives to other content.

Note 2 to entry: There can be multiple VACs presented at the same time.

Note 3 to entry: VACs can be displayed to indicate where the visual presentation of content will appear in the future or has appeared in the past.

EXAMPLE Caption/subtitle-boxes, -stripes or –lines are common examples of VACs.

3.8
audiovisual content
content that includes audio and visual components

Note 1 to entry: Only the audio or the visual components might be active at some times within the presentation of audiovisual content.

3.9
video
combination of audio and visual content presented together in a synchronized manner via Information and Communication Technology

Note 1 to entry: While the visual content is often presented using a screen, it might also be presented via other technologies e.g. a projected hologram.

[SOURCE: ISO/IEC TS 20071-25:2015, 2.1.2, modified – Note 1 to entry has been added.]

3.10
content category
<audiovisual> classification of audiovisual content

Note 1 to entry: Content categories are not necessarily mutually exclusive.

Note 2 to entry: When content category is considered from an artistic perspective, it is often referred to as genre.

EXAMPLE Content categories include: dramas, museum and art gallery exhibits, heritage tours, comedies, documentaries, video users' guides and manuals, university lectures, meetings, sporting events, etc.

3.11
importance
level of need for users to know information in the content

3.12
essential (information)
<importance> information that is necessary for users to understand the content and/or its function

3.13
significant (information)
<importance> information that provides a more detailed understanding of the content for most users most of the time

3.14
helpful (information)
<importance> information that provides a thorough understanding of the content for some users

3.15
unhelpful (information)
<importance> information that does not help users understand the content and/or might interfere with that understanding

4 Framework for the visual presentation of audio information

4.1 Purpose

Visual presentations of audio information should aim at providing viewers with alternative or complementary visual information that meets users' needs and contexts of use (e.g. noisy environments).

It is important that visual presentations of audio information present information contained in speech and other audio content.

4.2 Motivation

Audio content conveys information through verbal and non-verbal sounds. People who might not be able to fully access the content include those who cannot access the audio components such as:

a) persons with sensory disabilities such as the deaf or hard of hearing;

b) persons who cannot hear the sound for other reasons (for instance, not having the sound on, or having difficulty to hear the sound in a noisy environment);

c) persons with difficulties to access the oral verbal content.

NOTE Persons with difficulties understanding oral language include those with cognitive diversity as well as people learning a new language.

Not being able to access the meaning of sound used in the audio content has a direct impact on the understanding and enjoyment of the content. It also implies that certain people are excluded from educational, cultural and social contexts (e.g. when an audio content is discussed by colleagues in informal contexts).

A visual presentation of audio information should be perceived as equitably as possible to the auditory perception of the content.

Facilitating access to the sounds used in the audio content improves the experience in terms of comprehension and enjoyment, and guarantees access in critical emergency situations where information is provided auditorily.

Providing visual presentations of the audio information enhances access to audio content.

4.3 Locations of presentations

There are three locations of visual presentations of audio information; they can be:

a) superimposed onto the visual content;

b) displayed on the same screen but outside the visual content;

c) displayed on a separate (second) screen or display device.

4.4 Modes of presentations

There are two modes of visual presentations of audio information which can be presented alone or in combination:

a) Text presentations are visual presentations of audio information that rely on text to represent audio content. They are encoded separately from the audio content and presented to the viewer with the content (e.g. closed captions).

b) Figure/graphic presentations are visual presentations of audio information that rely on static or dynamic graphics to represent audio content. They are encoded as a figure and presented to the viewer with the content (e.g. emoticons, avatars, animations, pictures, etc.).

4.5 Modes of access

There are two ways to access visual presentations of audio information:

a) Visual presentations of audio information prepared separately from the content. The viewer needs to use some device or software to access the visual presentation of audio information (e.g. closed captions/subtitles, a display device at live contents such as theatre or opera).

b) Visual presentations of audio information included together with the content. The presentation is independent from the characteristics of modes of presentation (e.g. open captions/subtitles).

NOTE Multiple channels might be made available to the viewer, one or more with visual presentations of audio information and one or more without any visual presentations of audio information. Viewers select a channel according to their needs.

4.6 Modes of display

There are four modes to display text-based visual presentations of audio information, based on how they are cued-in (appear) and cued-out (disappear):

a) Pop-on (or “block”): Visual presentation of audio information where all information appears at once (as a block), remains for a period of time, and then disappears at once (as a block).

b) Scrolling (or "roll-up"): Visual presentation of audio information rolls onto and off the screen in a continuous motion. Usually two or three lines of text appear at one time. The presentation appears to “roll”; as a new line of text appears on the bottom of the VAC, the other lines on the screen move up and the line at the top is removed.

c) Word-by-word: Visual presentation of audio information is displayed on the screen according to the writing direction of the language used (i.e., in a left-to-right or right-to-left manner). The words appear one after the next. Word-by-word can be cued-out as a block or by scrolling.

d) Line-by-line: Visual presentation of audio information is displayed on the screen according to the writing direction of the language used (i.e., in a left-to-right or right-to-left manner). The text appears one line after the next. Line-by-line can be cued-out as a block or by scrolling.

4.7 Levels of importance

4.7.1 General

There are four levels of importance of audiovisual content (i.e., essential, significant, helpful, unhelpful) to support the understanding of the visual components of the audiovisual content.

Levels of importance depend on the context of use of the audiovisual content, including the use, purpose, and content category of the audiovisual content.

NOTE 1 Level of importance largely changes whether audiovisual content is consumed for entertainment purposes or information purposes. To have an engaging entertainment experience, information about audio content such as sound effects, music, an actor’s tone of voice, and so on needs to be available in a non-audio modality that supports those who cannot access the audio content.

NOTE 2 Determine the levels of importance from the perspective of consideration to match the intended meaning (see 6.1), consideration of output devices (see 6.3), evaluation by viewers (see 6.8 and 6.9), and other specific evaluation methods (see Clause 15 and Annex A).

4.7.2 Essential information

Essential information shall be displayed in visual presentations of audio information.

NOTE 1 Essential information in visual presentation of audio information ensures that all viewers will have access to this information.

NOTE 2 Viewers might be confused as to what the audiovisual content is presenting without essential information.

NOTE 3 Viewers have no idea why the audio content is there or what the audio content is for without essential information.

NOTE 4 Essential information might include the essence, purpose, function, or intent of the audiovisual content.

4.7.3 Significant information

Significant information should be displayed in visual presentations of audio information. Significant information goes into more details about the essential information.

NOTE The amount of significant information to be displayed depends on the amount of essential information that is already available.

4.7.4 Helpful information

Helpful information may be displayed in visual presentations of audio information.

NOTE 1 Helpful information is specific details that might be of interest to some who are the viewers of the audiovisual content.

NOTE 2 Helpful information can provide the viewer with a better understanding of audiovisual content when the viewer is not familiar with the content.

NOTE 3 Helpful information might reassure the viewers that they have not missed something of greater importance.

NOTE 4 Without helpful information, viewers have a fairly complete understanding of what the audiovisual content is about but might have some things that they still want to know.

4.7.5 Unhelpful information

Unhelpful information should be avoided in visual presentations of audio information.

NOTE 1 Unhelpful information is not important enough to mention.

NOTE 2 Unhelpful information might result in unintended confusion or misunderstanding of the audiovisual content.

EXAMPLE In a video of a tennis match, the sound of the ball being hit is unhelpful information.

5 Applicability of requirements and recommendations

5.1 Predictable audio contents

When audio content is predictable (e.g. content was recorded, or live but planned, or scripted), all the requirements and recommendations in Clauses 6 - 15 should be evaluated for their applicability to visual presentation of audio information.

5.2 Unpredictable audio contents

When audio content is unpredictable (e.g. content is spontaneous, unscripted, or live without plan or script, and unexpected during the production of visual presentations of audio information), particular requirements and recommendations in 7.3, 7.4, 7.5, 8.2, 9.8, 9.12, 9.13, 10.10, 10.11, 11.7, 11.8, 11.9, 12.4, 14.7 and 14.8 may be not applicable.

6 Production of visual presentations of audio information

6.1 Consideration to match the intended meaning

Visual presentations of the audio information shall convey the intended meaning of the audio information.

NOTE 1 Intended meaning is established by the original developer of the audio content.

NOTE 2 It is important that visual presentations of audio information closely match the intended meaning of the audio content. While the experience without the audio content or with any replacement to the audio content is unlikely to be exactly the same as it would be with the audio content, conveying an experience and understanding of the audio content that matches the intended meaning is desirable.

6.2 Ease of understanding of visual presentations of audio information

Visual presentations of audio information should enable the viewer to easily understand as much as possible the meaning of the audio information of the content.

6.3 Consideration of output devices

It is recommended to consider the following:

a) Where the receiving technology is able to be controlled by the viewers, producers and distributors of visual presentations of audio information should design presentations that utilise the device presentation features that will directly improve the usability of the presentation.

   NOTE It is important that elements such as the placement, colour, text size, font face, character leadings, half-width characters, speaker-identification information, etc. are made appropriate to the characteristics and capabilities of the display device.

b) Where the receiving technology is not able to be controlled by the viewers, producers and distributors of visual presentations of audio information should take into account the features of the displaying device(s).

6.4 Verification of visual presentations of audio information with the intended output devices

Producers and distributors of visual presentations of audio information should test the user interaction for accessibility and usability of the presentation on the range of device(s) where it can be expected to be displayed.

NOTE It is important that the devices tested represent the variations in characteristics (e.g. screen sizes) that can be expected.

6.5 Connecting visual presentations of audio information data with content data

Visual presentations of audio information should be usable together with the content data.

NOTE A visual presentation of audio information is usable by meeting requirements mentioned in this standard such as being appropriately synchronised with the audio content and not obstructing important video content.

6.6 Combining multiple visual presentations of audio information

Visual presentations of audio information should be usable together with other visual presentations of audio information for the same content.

EXAMPLE Graphical icons used to indicate certain audio information appear together with text to show the speech content. The two visual presentations of audio information are usable by, for example, not obstructing each other, or drawing attention away from one another.

6.7 Update of visual presentations of audio information data

If the audio content is revised, visual presentations of audio information shall be updated to reflect the changed content.

6.8 Evaluation

During the production and distribution processes, the content with the visual presentations of audio information should be reviewed to confirm that the visual presentations are accurate, are as easy to understand as possible, and provide an equitable alternative to the audio contents from the perspective of viewers who cannot access the audio content. See also Clause 15.

6.9 Evaluations including contribution of typical users

As part of the evaluation process, typical users of visual presentations of audio information should be invited to comment on the accessibility and usability of the visual alternative and its use with the content.

7 Visual design

7.1 General

Visual presentations of audio information shall be displayed in a manner that distinguishes them from background visual content.

Visual presentations of audio information should be visible and easy to perceive and recognize in space and time irrespective of the circumstances of viewers.

7.2 Personalization

Viewers should have the ability to set their own preferences. The word individualization is also used for personalization.

NOTE ISO 9241-129 and ITU-T Recommendation H.702 provide guidance on individualization.

Where viewers have the ability to set their own preferences, the producer and/or distributor should not override or modify the setting. The default settings for visual presentations of audio information should apply the design considerations in 7.

Where viewers do not have the ability to set their own preferences, and the producer and/or distributor does, visual presentations of audio information should apply the design points in 7.

7.3 Engagement

The design of visual presentations of audio information should consider understanding, enjoyment and engagement of viewers for both audio and background visual content.

7.4 Synchronization of presentations

When audio content is predictable (i.e., content was recorded, or live but planned, or scripted), visual presentations of audio information shall be synchronized with the content.

NOTE 1 Slightly different editions of the content might not correctly synchronise with the same visual presentations of audio information. It might be necessary to create different versions of visual presentations of audio information to be used with different versions of the content.

NOTE 2 In recorded production it is expected that captions/subtitles appear instantly (or slightly before) with the sounds they represent.

NOTE 3 If speech is fast and visual presentation is presented at the same speed, it is not accessible to read. In such cases, it is not necessary to design the timing of the visual presentation exactly at the same time (or frame) with the audio information. However, it is still important to design the timing of the visual presentation to easily understand the synchronization or association of the visual presentation with the content.

When audio content is unpredictable (i.e., content is spontaneous, unscripted, or live without plan or script, and unexpected during the production of visual presentations of audio information), visual presentations of audio information shall be closely synchronized with the content.

NOTE 4 In live without plan production, viewers ask for less than 3 s delay.

NOTE 5 In some languages, the delay of the visual presentation of audio information is more than 6 s after the sounds they represent for a technical constraint. Maximum lag times can vary between writing systems, and in the case of inter-lingual translation.

7.5 Avoidance of information obstruction

When presented on the same screen or superimposed onto the visual content (see 4.3), visual presentations of audio information shall not obstruct the most important areas of visual information in the content.

NOTE Important areas of visual information in the content includes features of the speakers, such as faces or lip movements, and lower thirds.

EXAMPLE 1 Captions/subtitles presented with a live football or baseball game, are placed so that they do not obstruct the ball.

EXAMPLE 2 Captions/subtitles used in a foreign film to translate a sign into the viewer's language are placed so they do not obstruct the sign.

EXAMPLE 3 Captions/subtitles presented with a live press conference are placed so that they do not obstruct graphics displaying identifying information about the speaker and important objects or items, and the crawlers/banners offering information at the bottom of the screen.

7.6 Font size

For visual presentations of audio information that include text, the font size shall be legible and readable under typical viewing conditions.

NOTE 1 In some cases, actual font size depends on the decoding device being used.

NOTE 2 Where possible, it is preferable to allow viewers to adjust the font size.

NOTE 3 ISO DIS 24509[1)] provides guidance on the minimum legible font sizes that support accessible design. The Annex provides resources on font size.

1) Under preparation. Stage at the time of publication: ISO/DIS 24509:2018.

7.7 Font type

For visual presentations of audio information that include text, the font type used shall provide readability and consistency.

NOTE 1 In most cases, sans-serif fonts are used as they have better screen readability and legibility than serifed fonts.

NOTE 2 In some cases, actual font type depends on the decoding device being used.

NOTE 3 Where possible, it is preferable to allow viewers to adjust the font type.

7.8 Font face

For visual presentations of audio information that include text, a different font face may be used to draw the viewer's attention to specific information.

EXAMPLE North American captions/subtitles often use italics for emphasis, product names, and the titles of albums, books, movies, plays, newspapers, or magazines.

7.9 Upper, lower, and mixed case letters

For visual presentations of audio information that include text in writing systems with case distinction:

a) Content should appear in sentence case (upper and lower-case letters).

NOTE Text presented in ALL CAPS (all capitalized) is less legible and readable than lower-case text.

b) Descriptive text, web site addresses, and e-mail addresses should appear in all lower case.

EXAMPLE [applause], www.example.com, info@example.com

c) The spelling of a name may appear in mixed case.

EXAMPLE The orthographic styling of names appearing in the Roman alphabet, is usually capitalizing the first letter (e.g. Smith, Obama). However, some names appear in mixed case (e.g. MacDonald) and others are not spelled with initial capitalization (e.g. de Havilland).

d) The capitalisation conventions of the language being displayed should be followed.

EXAMPLE 1 In English, an uncapitalised proper noun that can also be mistaken for a verb, will impact the viewer's comprehension. For instance, compare "Hey Cook!" with "Hey cook!". The first sentence is a greeting, while the second is an order. "Cook" is capitalised to show that "cook" is used to address a person not as the verb "cook". See also Clause 9.

EXAMPLE 2 A visual presentation of audio information presented in German capitalises all nouns.

7.10 Contrast and use of colour

a) There shall be sufficient luminance contrast between the visual presentations of audio information and the background visual content. See ISO/IEC 40500 (WCAG 2.0).

NOTE It is effective to use a black edge around bright text or put a VAC in the background of the text to reduce the brightness of the background on which text is displayed (see Clause 8).

b) The colour of visual presentations of audio information shall be significantly different from the background visual content.

c) The colour of visual presentations of audio information should enable persons with colour vision conditions (colour blindness) to distinguish between visual presentations of audio information and the background visual content.

d) When possible, colours should be chosen with consideration of the different properties of contrast and colour between display devices.

NOTE In some countries, the colours to be used are specified within a national standard.

7.11 Speed

Visual presentations of audio information shall be displayed for a minimum of 500 milliseconds.

NOTE This is a minimum for short words or a few simple graphics. Longer presentations will need more time.

The design of visual presentations of audio information should consider the reading speed.

7.12 Number of lines

For visual presentations of audio information that include text, the number of lines of text displayed at the same time should be a maximum of three.

If a visual presentation is designed to use four or more lines, special care should be taken into account for the time and space of the visual presentation to achieve accessibility of both the visual information and the background visual content.

NOTE Contexts of use that impact how many or how few lines of text can be used depend on many factors, such as: whether the user is using magnification, the screen size of the display, the use of a second screen or device to display the visual presentation, if the VAC appears within or outside the visual content, the density of visual information in the visual content, the presentation style of the visual presentation (e.g. roll-up), and the characteristics of the intended users (e.g. children vs. young adults).

EXAMPLE In Japan, usually two lines are used for captions/subtitles, but two additional half-height lines are necessary for printing kana or rubi (for reading Chinese characters). These lines are used to support readability and do not include extra information which is not already presented in the main lines of text. Therefore, these additional half-height lines do not count as lines of text.

7.13 Spacing between characters and lines (kerning and leading)

For visual presentations of audio information that include text, the spacing between characters and lines should be neither too narrow nor too wide to assure easy reading.

NOTE 1 Leading is a metric of how text is spaced vertically in lines. For multiple lines of readable text, the distance from the bottom of one line and the top of the next line needs appropriate spacing. See ISO/IEC 40500 (WCAG 2.0).

NOTE 2 Kerning is a metric of the distance between two letters on the same line. When letters are set too closely together, words are indecipherable. When letters are set too far apart, words are awkward to read.

7.14 Correct punctuation

Visual presentations of audio information that include text shall have correct punctuation, when appropriate to the language.

NOTE 1 This is important to support visual readability and legibility, for languages that use punctuation to aid the understanding and correct reading of text.

NOTE 2 Proper punctuation supports text segmentation because, once the text is correctly broken into sentences, the sentences can be sub-divided into multiple lines of visual presentations of audio information as needed.

EXAMPLE In English, punctuation can improve the meaning of a poorly written sentence. For instance, compare “She enjoys painting her family and her dog.” with “She enjoys painting, her family, and her dog.”. The added commas makes “painting” understood as one of several things rather than an act imposed upon others (i.e., putting paint on the dog) or an activity (i.e., painting a portrait of the family and dog). See also 7.17.

7.15 Spacing between words and phrases

For visual presentations of audio information that include text, the spacing between words and/or phrases shall be consistent and neither too narrow nor too wide to assure easy reading.

NOTE 1 In alphabetical writing systems, the size of spacing between words and phrases is often determined by the character set used to display the text.

NOTE 2 In alphabetical writing systems, the space between words is usually one letter space.

EXAMPLE A caption/subtitle uses a blank space the same width as the font's lowercase "r" character to denote a space between words.

NOTE 3 In Japan and China, a space between each word is not common. But a space is inserted between phrases in most cases. (e.g. "今日は いい天気だから 公園に行こうよ" ["Today | Nice weather so | Let's go to the park."]).

7.16 Transitions between presentations

Transitions between visual presentations of audio information should be clearly indicated.

NOTE For visual presentations of audio information using a pop-on display mode, inserting a short blank is an easy way for viewers to recognize the change of information and to reduce the possibility of confusion.

7.17 Sentence segmentation

For visual presentations of audio information that include text, a long continuous sentence should be broken over multiple lines or multiple presentations.

A long sentence presented over multiple lines and/or multiple presentations shall be segmented according to grammatical breaks and natural phrasing to indicate continuity of the sentence.

NOTE 1 Line length is sometimes determined by the features of the display device (e.g. limitations such as the maximum number of characters depends on the display size).

NOTE 2 Proper segmentation is important because it makes the text easier to read quickly. Improperly divided text stops the viewer and forces them to spend time re-reading instead of allowing the viewer to quickly read the text in the visual presentation of audio information and then pay attention to the visual information in the content.

NOTE 3 Dividing captions/subtitles into individual phrases can be preferable to dividing captions/subtitles per word.

EXAMPLE 1 In English, if an article, preposition or conjunction begins a phrase, it is used at the beginning of a new line or new caption/subtitle.

EXAMPLE 2 In English, line breaks do not: break a modifier or auxiliary verb from the word it modifies; break after a conjunction; or break a person's name or title.

7.18 Indication of sentence breaks over multiple visual presentations

For visual presentations of audio information that include text, continuation of a long sentence that has been divided into several presentations may be visually indicated.

NOTE 1 Symbols such as "→", '–', '- -' are used as a sign to describe continuity of a sentence. The ellipsis '…' can be used; however, it could be misinterpreted because it has other meanings.

NOTE 2 Diverging practices, including the absence of a sign, can be found in the usage of signs to indicate sentence continuity between multiple visual presentations.

7.19 Additional duration for location change

If the position of visual presentations of audio information changes within content, the information in this changed location should be displayed for a longer duration than usual, carefully taking into account the need of the viewer to adjust their reading.

NOTE When captions/subtitles change their position, viewers need time to respond to the change, locate the new caption/subtitle, and read the text.

7.20 Modes of display

Special attention should be paid to modes of display since it affects legibility and readability (see also 4.6).

8 Visual alternative container (VAC)

8.1 General

The purpose of the VAC is to provide contrast between visual presentations of audio information and other visual content.

Colour and transparency of the VAC shall be consistent within the same content and across related contents.

Colour and transparency of the VAC shall be distinguished from the colour of visual presentations of audio information.

NOTE Having sufficient luminance contrast between the visual presentations of audio information and the VAC is one way to ensure their colours are distinguishable. See ISO/IEC 40500 (WCAG 2.0).

The colour of visual presentations of audio information should enable persons with colour vision conditions (colour blindness) to distinguish between visual presentations of audio information and the background visual content including the VAC.

8.2 VAC position and area

VACs should be positioned where the containing visual presentation of audio information are easily perceived, while attention remains focused on the background visual information in the content.

Depending on the features of the display device, the area of VACs can cover all or part of the display. Visual presentations of audio information can use all or part of a VAC.

Depending on the features of the display device, more than one VAC can be presented at a time.

9 Describing speech

9.1 Describing verbal content

Visual presentation of intra-lingual speech (i.e., in the same language) should be as verbatim as possible.

NOTE 1 This is only for intra-lingual captions/subtitles, and cannot apply to inter-lingual captions/subtitles.

NOTE 2 There exist specific contexts of use where visual presentations of audio information are edited.

If it is difficult to follow the visual presentation of audio information because of the presentation speed (see 7.11), it:

a) should not be verbatim;

b) may be limited to the presentation of the essential information in the audio content.

9.2 Grammar

Verbal content should be presented accurately as heard, even when grammatically incorrect.

NOTE In some cases, it is more important to present a caption/subtitle as close to verbatim as possible than correcting poor use of language in audio content.

9.3 Vulgar verbal content and slang

Verbal content should be presented accurately as heard, even if the verbal content includes slang, vulgar language, discriminatory language, culturally unsuitable expressions, or other potentially inappropriate expressions.

9.4 Language variation

Visual presentations of audio information for speech that is uttered in a strong dialect or uses words only understood by local groups should:

a) present the verbal content as spoken, or

b) provide a standard language alternative with an indication of the presence of a dialect.

NOTE 1 Presenting the verbal content as spoken is the preferred approach.

NOTE 2 It is not sufficient to provide a standard language alternative only. Accents are often a deliberate part of characterization in fictional contexts. Not including this information will impact accuracy.

9.5 Foreign accents

Verbal content should be presented accurately as heard, even when a person has an accent. Visual presentations of audio information should be used to make note of an accent when necessary.

NOTE 1 Information about a speaker's accent can be essential or significant in importance when it can explain why a visual presentation of audio information has issues of quality such as long pauses, incorrect tenses, etc. Care should be taken to not insult the speaker or viewer community when providing this information. See 4.7.

NOTE 2 Inter-lingual visual presentations of audio information indicate the presence of a person's foreign accent in the original language content only if it has significant importance to the character, setting, or story context. See 4.7.

NOTE 3 Using a descriptive caption/subtitle to indicate that a speaker has some difficulty speaking the language is one way to achieve this guidance.

EXAMPLE 1 "[Russian accent] I go for walk now."

NOTE 4 It is important to note in a descriptive caption/subtitle if a person speaks with an accent for comic or dramatic effect.

EXAMPLE 2 "[over the top Italian accent] We-a go for pizza later, eh?"

9.6 Indiscernible audio content

Where verbal content is not clear, visual presentations of audio information should use a descriptive caption/subtitle to note that what is being spoken is unclear.

EXAMPLE "[indiscernible conversation]"

9.7 Spelling

Visual presentations of audio information containing text:

a) shall be correctly spelled;

a) should use the spelling conventions of the expected viewers.

EXAMPLE Captions/subtitles using UK English spelling conventions are presented to a European viewer.

9.8 Abbreviations

Unless an abbreviated word is spoken, words should always be spelled out in full.

EXAMPLE The verbatim "Any info you can give me would be great." is correctly presented using "info" and incorrectly presented using "information" because the speaker did not say the full word "information".

9.9 Homophones, homonyms, homographs, heteronyms, and heterographs

Errors in spelling homophones, homonyms, homographs, heteronyms, and heterographs can mislead the viewer of the meanings of the audio content. All such words should be checked carefully to ensure correct spelling.

Words that are pronounced differently but spelled the same often have different meanings (i.e., heteronyms and some homographs). Such words may need to be presented with additional information to support viewers' understanding of the meanings of the audio content.

Words that are pronounced the same but spelled differently (i.e., heterographs and some homophones) are a source of error for creators of visual presentations of audio information.

Such words should be checked carefully to not mislead the viewer of the meanings of the audio content.

Words that are pronounced the same and spelled the same but have different meanings (i.e., homonyms):

a) should be checked carefully to not mislead the viewer of the meanings of the audio content, and

b) may need to be presented with additional information to support viewers' understanding of the meanings of the audio content.

9.10 Long speech

An edited presentation of long speech may be used where the time available does not allow long verbal content to be simultaneously presented as both audio and visual components while also still readable.

NOTE Long speech might be impossible to be displayed verbatim. The time available might not allow the whole speech to be simultaneously presented as both audio and visual components while also still readable.

9.11 Describing multiple simultaneous information

When using visual presentations of audio information to display a variety of information simultaneously, care should be taken to ensure that each piece of information can be read, recognized, and understood.

9.12 Confirmation by content producers when producing visual presentations of speech

When some important speech is difficult to explain using words or graphics, clarification and assistance should be sought from the producers or directors of the audiovisual content.

NOTE A script or screenplay might suffice, if contacting the producers or directors is impossible.

9.13 Sources of information

When producing visual presentations of audio information, information provided by the producers of the content may be used as a reference.

NOTE 1 This can include pre-production or post-production scripts, screenplays, or other materials, if they are available.

NOTE 2 The speech of the audio content is likely different from that written in the original screenplay or script.

10 Non-speech information (NSI)

10.1 General

The levels of importance contained in 4.7 shall be used to determine which NSI is to be presented.

NOTE Music, emotion, and silence are types of NSI which have additional guidance. See Clause 11, 12, and 13.

10.2 Describing NSI

Words, pictorial representations or signs can be used to visually present NSI (e.g. sound effects, ringtone, doorbell, etc.).

When NSI is presented in words, parentheses or brackets shall be used to distinguish NSI from speech.

EXAMPLE Captions/subtitles of audio content recorded in front of a live audience uses text in square brackets to describe the reaction of the audience (e.g. [audience laughs]).

10.3 Correct description of NSI

Words, pictorial representations or signs used to visually present NSI shall correctly convey the sound, which can be achieved by conveying the sound itself, its source, and/or its function.

EXAMPLE 1 "[barking]"

Words, pictorial representations or signs used to visually present NSI should correctly convey the source.

EXAMPLE 2 "[dog barking]"

Words, pictorial representations or signs used to visually present NSI may convey the function or purpose.

EXAMPLE 3 "[dog barking at intruder]"

10.4 Well-known sound descriptions

Care should be taken to use words, pictorial representation or signs that are commonly and widely recognized to represent the particular sound.

10.5 Onomatopoeia

Care should be taken to use onomatopoeic expressions easy to understand for viewers who are not familiar with the sound (e.g. deaf or hard of hearing people).

When onomatopoeic expressions have a particular meaning, they should be used.

Whenever possible, onomatopoeia should be accompanied by a descriptive explanation.

EXAMPLE Captions/subtitles describing noise made by a dog includes a description of the sound effect (i.e., [dog barks]) and the onomatopoeic description of the sound (i.e., "woof-woof").

10.6 Sound effects in speech

When speech includes the use of NSI such as sound effects, the onomatopoeic expression may be used.

EXAMPLE An animated movie with a dog character includes speech where the dog changes from oral speech to barking. The captions/subtitles show the sound effect as the onomatopoeic expression "woof" rather than the sound effect "[bark]".

10.7 Censored language

If a vulgar word has been bleeped out or simply silenced, visual presentations of audio information shall note this.

EXAMPLE "That's [bleep] stupid!"

10.8 Paralinguistic sound effects

When speech includes paralinguistic sounds, they should be presented in visual presentations of audio information.

NOTE Paralinguistic sounds include laughter, coughs, grunting, crying, etc.

EXAMPLE "[man coughs]"

10.9 Discrete and sustained sound effects

Words, pictorial representations or signs used to visually present NSI shall correctly convey whether the sound is discrete or sustained.

NOTE 1 Non-speech sounds that have a clear beginning and end are discrete sounds.

EXAMPLE 1 "[laughs]", "[shouts]"

NOTE 2 Non-speech sounds that have an on-going action, repetition, or overlapping on-going action are sustained sounds.

EXAMPLE 2 "[laughing]", "[shouting]"

NOTE 3 When a sustained sound lasts longer than the original caption/subtitle describing its presence, a new caption/subtitle can be used to show it is on-going.

EXAMPLE 3 "[laughing continues]"

NOTE 4 When a sustained sound stops, a new caption/subtitle can be used to show it has stopped.

EXAMPLE 4 "[laughing stops]"

10.10 Confirmation by content producers when producing visual presentations of NSI

When some important sound(s) is difficult to explain using words or graphics, clarification and assistance should be sought from the producers or directors of the audio content.

NOTE A script or screenplay might suffice, if contacting the producers or directors is impossible.

10.11 Sources of information

When producing visual presentations of audio information, information provided by the producer or director of the content may be used as a reference.

NOTE 1 This can include pre-production or post-production scripts, screenplays, or other materials, if they are available.

NOTE 2 The sounds of the finalized audio content are likely different from those written in the original screenplay or script.

11 Music

11.1 General

The levels of importance contained in 4.7 shall be used to determine which music related information is to be presented.

NOTE Music is a type of NSI. Further applicable guidance can be found in Clause 10.

11.2 Describing presence of music

Visual presentations of music shall be distinct from presentations of speech and other types of NSI.

Presence of music should be indicated using a visual presentation which will be understandable for intended viewers.

NOTE In some countries, the presence of music is indicated using specific characters as per convention.

EXAMPLE 1 In North American captions/subtitles, the presence of music is shown with one or more single eighth note characters (♪).

EXAMPLE 2 In Japanese captions/subtitles, the presence of music is shown with two different symbols either a single eighth note character (♪) which indicates the presence of music or musical sound (guitar etc.) being used in the scene; or a beamed eighth note character (♫) which indicates the presence of background music or a theme song added in post-production.

EXAMPLE 3 The presence of music is shown in Australian subtitles with a number sign (#) because it looks similar to a musical sharp sign (i.e., ♯).

11.3 Describing the reason or purpose for the music

The reason or purpose for the music may be presented.

NOTE 1 This description might need to reference cultural information, historical information, and/or the emotion evoked to explain the purpose of the music.

EXAMPLE 1 In the movie "Jaws",[2] viewers need to know that the shark theme is playing to know that the shark is about to appear. The caption/subtitle might be "♪shark theme♪" or, "♪scary shark theme♪".

EXAMPLE 2 In the "Star Wars"[3] film series, viewers need to know that the Jedi theme is playing to show that the story is shifting from the Empire to the Jedi worlds.

NOTE 2 Including information on why the music is being used supports deaf and hard of hearing viewers because they might only have an eighth note character on the screen to convey the presence of the music and no further information whatsoever. This is basically the rule "if you think it is important enough to add it to the content, tell me what it is."

11.4 Provide information that identifies the music

Visual presentations of audio information should provide the name, artist, and other identifying information for all music.

NOTE This information would normally be recognized by a person who can hear the music.

2) Jaws (1975). Director: Steven Spielberg. Producer: David Brown, Richard D. Zanuck, Universal Pictures. https://www.imdb.com/title/tt0073195/.

3) Star Wars (1977). Director: George Lucas. Producer: Gary Kurtz, George Lucas, Rick McCallum, Lucasfilm, Twentieth Century Fox https://www.imdb.com/title/tt0076759/.

11.5 Clarification of music descriptions

Care should be taken to use words, pictorial representations, or signs commonly and widely understood to describe the music.

Every attempt should be made so that viewers who have difficulty in hearing or accessing the music are supported by the description.

EXAMPLE 1 When the guitar music is slowly, softly, and gently picked, a caption/subtitle is presented as "♪ [slowly and quietly played guitar] ♪".

EXAMPLE 2 The musical description "♪ [soft guitar] ♪"is not appropriate for the description of music because it is not clear to a viewer that has never heard a guitar what "soft" might mean in this context.

EXAMPLE 3 For visual presentations of audio information that rely on text, music styles are indicated in the same manner as sound effects, with square brackets or parentheses (e.g. "♪ [violins played poorly] ♪", "♪ [bluesy jazz] ♪").

11.6 Presentation of lyrics

It is not enough to merely indicate that music is playing.

The levels of importance contained in 4.7 shall be used to determine which lyrics are to be presented.

NOTE Use a descriptive caption/subtitle (e.g. "♪ [singing in Russian] ♪") when lyrics are not available or not clear (e.g. "♪ [very fast rap, words unclear] ♪").

If foreign language lyrics are presented, they should be presented in the expected character set of the viewer.

EXAMPLE 1 A film where all speech is in English, and thus captions/subtitles are in English, contains a song sung in Russian. The captions/subtitles show that the singer is singing in Russian (e.g. [singing in Russian]), the name of the song if known, the name of the artist if known, and the lyrics of the song in Latin alphabet transliteration (not Cyrillic) if available.

The visual presentation of lyrics should follow the phrasing of the lyric.

EXAMPLE 2 Three separate captions/subtitles of a song's lyrics appear as:

♪ Happy birthday to you ♪

♪ Happy birthday dear John ♪

♪ Happy birthday to you ♪

11.7 Distinction of lyrics from speech

Visual presentations of lyrics should be designed to be visually distinct from those used to indicate speech. See also 10.2.

a) Lyrics are accompanied by some visual indication that the visual presentations of audio information are lyrics, and

   NOTE Using distinct text, or icons such as musical notes are two ways to visually distinguish lyrics from speech.

b) The beginning and ending of the song is clearly indicated.

EXAMPLE 1 In Canada, captions/subtitles of music in a version of the musical "The Pirates of Penzance" are presented with a single eighth note character (♪) on the left and right of each line of lyrics to distinguish lyrics from speech, e.g.:

No, Frederic, I shall live and die a Pirate King ...

♪ Oh, better far to live and die♪

♪ Under the brave black flag I fly, ♪

EXAMPLE 2 In Canada, the end of the song is often indicated by two eighth notes, without a space between them (i.e., "♪♪"), in the caption/subtitle after the last lyric.

11.8 Confirmation by content producers when producing visual presentations of music

When producing visual presentations of audio information for music, clarification and assistance should be sought from the producers or directors of the content.

11.9 Sources of lyrics

Since music in the content might differ from the original source of the lyrics, any visual presentations of audio information of lyrics should be edited to ensure it accurately reflects the lyrics clearly heard in the completed version of the audio content.

12 Emotions

12.1 General

The levels of importance contained in 4.7 shall be used to determine which intended emotional nuance related information is to be presented.

NOTE Emotion is a type of NSI. Further applicable guidance can be found in Clause 10.

12.2 Describing intended emotional nuance

When emotion is presented in words, parentheses or brackets shall be used to distinguish emotional nuance from speech. See also 10.2.

Visual presentations of audio information should convey as much as possible the intended emotional nuance.

EXAMPLE 1 The sound of a chainsaw in the film "The Texas Chainsaw Massacre"[4] has a very different significance than the same sound in the cartoon short "Box Office Bunny"[5], a Bugs Bunny cartoon where Daffy Duck and Elmer Fudd are also chased by a character with a chainsaw. The first uses the sound effect to strike fear into the viewer, the second uses a similar sound effect for humour. In the first case the caption/subtitle could read "[scary revving chainsaw]" while the second case might read "[chainsaw motor]".

EXAMPLE 2 If someone screams and it is difficult to perceive the intended emotional nuance only by the visual information provided, the emotional purpose of the scream (e.g. terror, fear, joy, glee, surprise or other emotional purpose) needs to be described in the captions/subtitles (e.g. [joyful scream]).

12.3 Describing the reason or purpose for the emotional nuance

The reason or purpose for the emotion may be presented.

4) The Texas Chain Saw Massacre (1974). Director: Tobe Hooper. Producer: Tobe Hooper, Kim Henkel, Vortex. http://www.imdb.com/title/tt0072271/.

5) Box-Office Bunny (1990). Director: Darrell Van Citters. Producer: Warner Bros. Animation, Warner Bros. http://www.imdb.com/title/tt0099171/.