International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

FG AVA TR

Version 1.0
(10/2013)

Focus Group on Audiovisual
Media Accessibility

Technical Report

Part 4: Final report of activities:
Working Group A "Captioning"

FOREWORD

The procedures for establishment of focus groups are defined in Recommendation ITU-T A.7. The ITU-T Focus Group on Audiovisual Media Accessibility (FG AVA) was proposed by ITU-T Study Group 16 for creation in-between TSAG meetings and it was established on 22 May 2011. The Focus Group was successfully concluded in October 2013.

Even though focus groups have a parent organization, they are organized independently from the usual operating procedures of ITU, and are financially independent. Texts approved by focus groups (including Technical Reports) do not have the same status as ITU-T Recommendations.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Technical Report may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU-T Focus Group participants or others outside of the Technical Report development process.

Table of Contents

Summary

This Technical Report of FG AVA was prepared by Working Group A "Captioning" and outlines the conclusions of the work of this working group.

This Technical Report is divided in two parts:

–   Part 1-Introductory report of WG A

–   Part 2-Complement of information to the WG A report

Part 1 – Introductory report of the WG A work

1 Introduction

Captioning (also known as subtitling) is one of the oldest access services for audiovisual media and dates back to 1929. It has been used with same-language and foreign language content (intralingual and interlingual communication).

For same-language captioning, the focus is on helping persons with hearing impairments including deafness as a complement to visual signing for persons born deaf.

Foreign language captioning, however, is an alternative to dubbing or lecturing to help viewers understand a programme in a language which is not their own.

Spoken captioning (also known as audio subtitles (AST)) is not included here as it is covered by the work of Working Group B (FG AVA Technical Report Part 5).

This Technical Report describes the scope and objectives of WG A and it is done by addressing the following four questions:

2 Question one: What is the state of the art of captioning?

– The value network and the key stakeholders (looking at the planning, production, exchange, distribution and use of captioning).

– Existing work practices in the area/domain (societal differences on the extent to which same-language captioning aims to offer verbatim transcriptions or an access service tailored to the needs and capabilities of its users; the distinction between prepared and live subtitles; genre-related differences; the coding of non-linguistic data in the captions including aural events of importance for comprehension and [colour] coding for different speakers).

– Examples of good practice (including the work of NHK (Japan Broadcasting Corp.). NHK has been providing live closed-captioning for broadcast by using Japanese speech recognition with a direct method or a respeaking method. Steno keyboards and the speech recognition are used in different types of TV programmes. Language models and dictionaries for speech recognition are trained and updated separately in each programme category such as baseball, football, information and news. Accuracy of recognition is around 90%-95% in such programmes and any errors are manually corrected in real time with a touch panel screen and a keyboard.

2 Question two: What is the vision for 2015 and 2020 for captioning?

– A scenario which explains how the needs and interests of persons with disabilities (in its broadest sense) would be addressed.

– The design and production processes (clarifying what constitutes quality).

– The key technologies (distribution platforms that allow viewers to change the appearance of caption text in size, position, colour or display timing; buffering in the receiver to deal with the delay in presenting live captions; the impact of hybrid solutions (e.g. integrated broadcasting broadband (IBB)); being able to differentiate captioning to cater for a variety of user requirements including those of persons with reading disabilities).

– The business models being used to assure captioning.

– The legislative and regulatory frameworks needed to support visions of media accessibility using captioning.

3   **Question three: What are the barriers in 2011 that currently prevent these visions from becoming a reality?**

–   Barriers related to current design and production processes (standardization and legacy issues).

–   Barriers related to current technologies (the transition from text to markup languages).

–   Barriers related to current business models for accessible digital media[1].

–   Barriers related to current legislative and regulatory frameworks needed to support these visions (unhelpful or ambiguous targets specifying the proportion of programmes that should be broadcast with captions leading to night-time repeats with captioning to meet targets).

–   Other barriers:

    o   The quality of speech recognition: To train the language models and dictionaries of the speech recognition for each TV programme takes so much work such as collecting relative scripts and checking vocabularies by human, although news programmes with electronic manuscripts are exception of such troubles. Performance of the speech recognition is not good enough for any kind of TV program or topic without the preparation of proper language models and dictionaries. Not only language models but also acoustic models should be trained with large speech and text database for a target language, and it is rather expensive to have such a language- and topic-dependent database.

    o   The costs of respeaking systems using speech recognition: Although the respeaking method helps avoid background noise and overlapping among multiple speakers, an ideal method is a direct method, which requires no cost for respeakers. Recognition errors are inevitable and manual error correction is necessary, but the capital and operating costs are considerable. Big urban or public TV stations can afford to maintain such captioning systems, but small local stations and many commercial stations cannot. One reason for the high-cost systems is that closed-captions are regarded only for the hearing impaired and the elderly, resulting in a small market.

         A strategy to promote the development of language models for a given language by sharing the development costs between public and private stakeholders will be needed.

4   **Question four: What actions are needed to break down current barriers in order to make the 2015 and 2020 visions a reality?**

–   Actions related to current design and production processes.

–   Actions related to current technologies.

–   Actions related to current business models for accessible digital media.

–   Actions related to current legislative and regulatory frameworks needed to support these visions.

–   Other actions:

    o   Commercially available speech recognition products beyond TV station usage should be developed in any language by supporting manufacturers and university researchers with funds. As demands of closed-captions and any other speech-to-text technologies increase, research and development advance rapidly and the production cost decreases. In terms of multilingual speech recognition, technology easily porting a system in one

---

[1] E.g. the case of falling revenue streams for teletext advertising that in the past paid for captioning on TV New Zealand.

language to another language should be developed. As for the flexible display of captions, another way using the Internet such as IBB should be developed. The idea that closed-captions are not only for the hearing impaired and the elderly but also for all the viewers should be widely accepted in the world because it is helpful for all viewers to understand the programme under muting situation. Legislative and regulatory frameworks help to do so.

Part 2 – Complement of information to the WG A report

1   General

Captioning (North America)/subtitling is the interlingual translation of dialogues or spoken commentary of an audiovisual production, displayed as one or two lines of text, usually at the bottom of a TV or cinema (or other) screen. There is a more detailed explanation of the terminology in the 'vocabulary' document (FG AVA Technical Report Part two).

Note that there are different traditions and practices when it comes to both interlingual and intra-lingual displaying of dialogues or spoken commentary. North America uses quality criteria in which the aim is to provide a verbatim transcription wherever this is possible.

In Europe, linguistic condensation − reducing the amount of text without distorting the underlying meaning-is more common. The rationale is to take into consideration the reading ability of the viewer and allow time for 'decoding' of the picture content.

2   Environment

2.1   Legal framework

The provision of captions/subtitles to promote accessibility involves stakeholders from content commissioning, production, distribution to audiovisual media use. Captions/subtitles are still predominantly a national activity for broadcast media, whereas content on the Internet is more global in nature. In a world with multiple distribution networks (broadcast, fixed net and mobile Internet), the main challenges facing caption/subtitle provision are linked to interworking/cross-platform delivery and doing this effectively and efficiently.

Legislators and regulators have to determine:

–   Which content requires captioning/subtitling (channels, content genre, series/programmes or also spots, trailers and advertising).

–   Which platforms are covered (broadcast channels only or also all delivery platforms-satellite, cable, on-demand, Internet, etc.).

–   What proportion of first-time and repeat broadcasts are covered.

–   The criteria against which captioning/subtitling provisions will be measured.

Recommendations

–   Where possible, co-regulation with the active involvement of all the key stakeholders is a good first approach for implementing audiovisual (AV) accessibility.

–   Legislators and regulators:

o   Should take into consideration the recommendations of international bodies as regards the production, exchange and interworking of captioning/subtitling.

o   Should encourage the multiple use of subtitles on various devices on the whole value chain.

o   Should determine that broadcasting content must be subtitled on all devices and therefore must determine that access services are an integral part of the television signal.

o   Should harmonize the way of technical dissemination (e.g. digital video broadcasting (DVB)-subtitles) in order to simplify the reception.

o   Should determine that access services must be part of the budget of subsidized audiovisual contents.

2.2   Target groups

Persons who are:

–   deaf or have become deaf;

–   hard-of hearing;

–   have cognitive impairments;

–   parents and the immediate family of children with disabilities;

–   immigrants.

3   Preparation

3.1   Scheduling

–   Determining what exactly is to be subtitled and what is exempt:

o   the content (the programme itself);

o   metacontent like commercials or trailers, etc.;

o   content not suitable for subtitles.

–   Defining in which manner the content has to be subtitled:

o   live;

o   near-live;

o   pre-recorded.

–   Defining which technology is to be applied:

o   regular keyboard input;

o   voice recognition with respeaking;

o   automatic voice recognition;

o   stenotype or other shorthand keyboard input.

–   Arranging materials:

o   videos;

o   scripts or keyword directories;

o   access to a content management system.

Recommendations

–   Live content: Prepared content should not be respoken as this usually results in delays.

– Near-live content: Those preparing captions/subtitles are required to prepare the content in advance in order to subtitle the content without delay during the transmission.

– Broadcasters should be encouraged to disseminate their live signal with a short delay in order to sync subtitles.

3.2 Production, proofreading and quality control

Production

– ordered (in-house or outsourced); or

– automatically generated subtitles; or

– use of free subtitles (crowd sourced production).

Application of a style book

– intra-, inter- or multilingual subtitles;

– defining a language level (assumption: customized subtitles with different language levels will remain wishful thinking):

   o verbatim;

   o reformulated subtitles (linguistic condensation to take into consideration the range in viewer reading speeds);

   o simple language.

– positioning on the screen:

   o horizontal: bottom, top, upscaled subtitles;

   o vertical: left-aligned, centred, right-aligned;

   o cadence: block subtitles, word-by-word subtitling.

– Frame space between two subtitles;

– multiple use:

   o usable on all kinds of devices;

   o usable on the whole value chain (cinema, pay-TV, video on demand (VOD), free TV, etc.);

Proofreading and correction

Quality control

– pre-recorded subtitles (according to the style book);

– for live subtitles. Application of the NER-Model[2], at least some live subtitles should be analyzed.

Recommendations

– Reading speed should be adapted to the reading skills of the target group and thus can differ from content to content.

– In order to reuse the subtitles on the whole value chain, the style books should be harmonized for any language.

---

[2] The NER is a model used to calculate accuracy in live subtitling.

– On-screen display should be applied according to the latest research results.

– Quality control should be outsourced in order to increase credibility.

– A minimum of two frames should be left between subtitles to register the appearance of a new subtitle[3].

3.3 Broadcasting or online distribution

Distribution

– analogue and digital distribution;

– closed or open captions.

Recommendations

– None.

3.4 Reception

Recommendations

– The activation of subtitles should be very easy either by pushing a single button or by switching them on in the options menu of the device.

– It should be possible to set the font size of subtitles individually as many hearing impaired users also have bad eyesight.

– Sellers at electronic retail stores should have the know-how to advise hearing impaired customers. In addition, social service staff should be able to switch on access services on the devices of the handicapped people.

3.4 Broadcasting or online distribution

None.

_____

---

[3] See discussions in EBU-TT-D where this may not always be the case.