REMOTE INTERPRETING FOR LIVE EVENTS & BROADCAST – INCLUSION IN MULTIPLE WAYS

Robin Ribback

VerbaVoice GmbH, Germany

ABSTRACT

Roughly 360 million people worldwide are affected by significant hearing loss. Without the help of sign language or written text, they are unable to understand spoken language. Their full participation in the society around them depends on complete access to spoken communication. This paper examines the current state of captioning on television, the internet and live events and presents different approaches and solutions currently available to media companies and businesses for delivering their spoken communication via video or live text. In addition to the systems currently available, the paper also offers a look ahead to solutions – such as automatic and semi-automatic speech recognition – which are currently being developed to improve the working conditions of speech to text reporters and interpreters as well as broadcasters in the near future. Another aspect is the chance to employ interpreters who have sensory impairments themselves and thus foster a society providing inclusion in multiple ways through technology.

INTRODUCTION

Live content gathers the highest audience share on TV today – since pre-produced content can already be viewed on demand. Live sports, political events, and popular shows like the Eurovision Song Contest, Royal Weddings, the UEFA cup, the Oscars, the World Cup and the Olympics are live – and hardly accessible for people with hearing impairments. The access services provided by broadcasters today are very limited in terms of user experiences. Access services cannot be customized by the individual user although the needs of the audience vary a lot. Also, broadcasters increasingly offer their live programs via IP-streaming in addition to TV broadcast (in a 24/7 fashion, specifically also for live events e.g. Olympic Games). Support for accessibility of media using this means of distribution is, however, still lacking sufficient solutions. There are currently approx. 15 million people living in Germany who suffer from some form of hearing loss. Just fewer than one million of these people are so limited in their hearing capacity that they require visual aid in order to understand spoken content.

Hard of hearing people need unlimited access to spoken content in order to fully participate in the society around them. A recent regulation enacted by the German Federal Ministry of Labour and Social Affairs requires all websites and online services to be free of communication barriers. For the time being, this requirement only applies to the domain of German federal administrative agencies, but an extension of its applicability appears likely – since the German Association of the Deaf (Deutscher Gehörlosenbund, DGB) has, among other measures, called for the creation of a legal basis for the immediate implementation of full captioning in all television programming presented by the entire spectrum of publicly and privately owned television networks. This DGB request is also aimed at all forms of online media.

To accommodate the demands of both the government and associations representing the interests of the deaf and hard of hearing, innovative and cost-effective solutions are necessary in order to remove barriers to spoken communication on television and the internet. This includes solutions for (live) captioning and for visualizing content through sign language.

BACKGROUND – A GROWING DEMAND THROUGHOUT EUROPE

According to a report by the European Broadcasting Union (EBU), access service provisions are regulated at the national or even regional level. Regulation varies widely from country to country. In several countries, laws are being reinforced to protect disabled people from discrimination. Although disability organisations are lobbying for change, there is currently no Europe-wide regulation on the provision of access services. The EU Audiovisual Media Service Directive (AVMSD) leaves legislation to each EU country to implement. The spirit seems to be that the market should decide what services to provide. This means that access service reception equipment is not widely available, because the potential individual markets may be regarded as too small by manufacturers, and standards are not mandated.

In contrast, "must carry" is partly regulated at the European level. This requires distribution parties (such as cable distributors) to transport the complete broadcast content, including any access services. Without such a requirement, distributors may choose to 'strip out' some service components (e.g. the subtitle stream) to free bandwidth for other services or to fit the reduced service in available bit-rate. There is one caveat: the European legislation is currently optional. Broadcasters therefore are encouraged to lobby for mandatory legislation on 'must-carry' in their own countries.

TV and the Internet are media based on visual and auditory communication. People with hearing loss can only take advantage of these offerings in a limited capacity. In Germany, the number of television channels that provide these persons with the option of subtitles or an interpreter feed is relatively small. According to the German Association for the Deaf, only 8.4 percent of the 22 most watched TV networks caption their content for the hearing impaired.

This small percentage is mostly accounted for by publicly owned broadcasting stations. By the end of 2013, ARD, the main publicly owned television channel in Germany, was providing nearly all of its initial broadcasts with captions for the deaf and hard of hearing (a total of 97%). However, with only 67% of its content featuring captions, ZDF, another major publicly owned channel in Germany, still has a long way to go towards providing captioning for all of its programming.

The situation is even more severe in the case of privately owned broadcasting stations. These channels rarely, if ever, caption their programs. Some stations even have a rate of 0%. Compared to other countries, TV stations in Germany have much work to do in catching up to the accessibility levels in the U.S. or in Great Britain, for example. With an overall captioning rate of 100%, these two countries are the leaders in this area, followed by the Netherlands, where 80% of TV stations subtitle their content.

YouTube, currently the primary online video portal with more than 100 hours of video content uploaded every minute, already offers an automatic captioning function for all videos on its platform. Users can add automatic subtitles to any video by activating the "Captions" function. However, the quality of these automatic captions in German is still far

from adequate. In the future, whenever media companies or "conventional" businesses publish videos on their own websites or through Internet video platforms, the content will need to be fully inclusive and include captioning – ideally also visualized in sign language – so that it can also be enjoyed by persons with hearing loss.

LIVE BROADCASTING THROUGH THE INTERPRETER TELEPRESENCE SYSTEM

In comparison with the pre-produced program, the captioning of live content poses a particularly big challenge – since live programming and events include unexpected conversational situations and background noises, such as music, applause, etc.

TV programmes with bad speech intelligibility

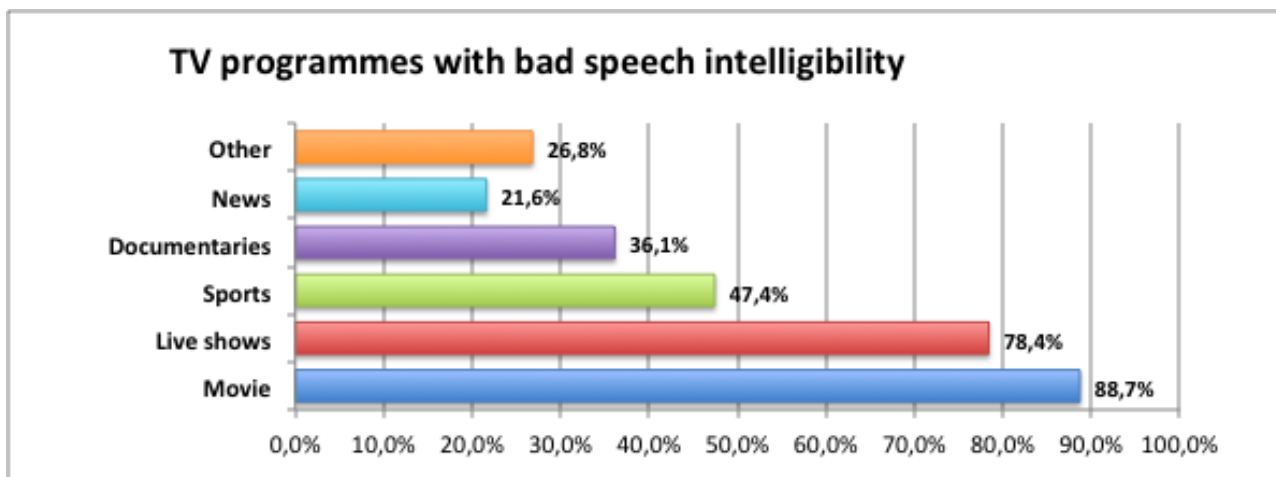| Programme	Percentage
Other	26,8%
News	21,6%
Documentaries	36,1%
Sports	47,4%
Live shows	78,4%
Movie	88,7%

Figure 1 – TV programmes with problematic speech intelligibility for hearing impaired people

In light of this, the hard of hearing need access not only to automatic captioning, but also to sign language interpreters, who are better equipped to handle unexpected situations and therefore more adept at communicating whatever transpires during the broadcast. This is the only way to ensure that the deaf can also share in the live experience in real time. The technical solution of the Interpreter Telepresence System (ITS) developed by VerbaVoice is already used for e.g. live broadcasts from the legislative assemblies of several German states, for events and conferences (for example, the "re:publica" web 2.0 conference held in Berlin annually) for international soccer matches and for educational content. In addition, ITS offers media companies, businesses and cultural institutions a number of options for live language visualization. In 2014, the ITS was used to deliver roughly 100,000 hours of live interpreting services. With the aid of the ITS, which is used throughout Europe, television programs and events can be broadcast with an online interpreter feed or in combination with an interpreter/ captioner working on-site.

Simultaneous interpreting is used very frequently in education, an area where the service is in high demand and can give students the ability to actively participate in classes, for example. In this case, the interpreter listens to the speaker's words (on a feed using low-latency codecs) and dictates the audible content into a speech recognition system in real time. Dragon Naturally Speaking from Nuance is the speech recognition software currently used to handle this task. In simultaneous translation, the interpreter can re-phrase content slightly (which is not always desirable) and also dictates punctuation marks in order to formulate complete sentences. ITS then displays the recognized text for the user on a

laptop or other mobile device in real time – nearly word-by-word. This method is also referred to as re-speaking. For the captioner (similar to a speech-to-text reporter), it's a task marked by a level of difficulty comparable to the simultaneous translation of content into another language.

The amount of practice and experience with the speech recognition system, as well as the captioner's interpreting qualifications and – most importantly – speech training, are all critical factors for achieving high recognition rates. To create an individual speaker profile, each captioner must read a large number of specially selected sentences aloud. The resulting recorded speech data allows the language recognition program used to learn the voice of each interpreter. With this method, it is possible to achieve speaker-specific recognition error rates well below 5 percent.

It does, however, have two critical disadvantages:

a)   -The system is installed locally on the interpreter's computer and is typically also used for other administrative tasks such as e-mail or word processing. As a result, various different programs are in constant competition for the same computing resources. This often places significant strain on the computer's performance. In addition, the systems require a great deal of maintenance to prevent their data from fragmenting.

b)   -Nuance/Dragon is considerably overloaded by its front-end. Profiles for individual jobs must be managed separately and manually.

To resolve these issues, a centralized voice recognition system is currently being developed. On one hand, this system handles language profiles more intelligently by managing them centrally using a big data approach. On the other hand, a data centre provides large amounts of dedicated computing capacity for the individual speech recognizers.

The core system is designed as a "true" cloud system. It is set up across multiple data centres, each of which features a switched connection to the Internet, and is composed of three system groups:

Workflow system: This system is used to book and coordinate all assignments. Communication with the participants is based on of different job steps defined for a particular area. It is also where the language profiles are managed.

Telepresence system: This system provides the interpreter with real-time feeds for audio, phone calls, video, PowerPoint presentations and other background information. It also handles the distribution of any real-time transcripts that are created. Another big advantage of ITS is that interpreters located around the world can be assigned specific tasks, making it no longer necessary for them to be on-site. Equally beneficial is the fact that the system enables transparent interpreter switching, i.e., that different interpreters can rotate during a broadcast.

Interpreter assistance systems: These systems provide functions that make the interpreter's job easier. Similar to driving a car, users must still operate the system themselves, but the assistance systems increase responsiveness and quality. The diverse range of support systems includes glossaries for specific subject areas and the advance availability of automatic transcriptions. These features lead to improved quality and also make the service more affordable for customers.

Numerous adapter systems with access to the centralized system allow the system to be used in various different situations, enabling both unidirectional and bidirectional access. Some examples of adapter systems are fully accessible web players, mobile second-screen apps, live caption inserters, HbbTV apps, smart TV apps, second-screen apps, caption editors, quality measurement editors, etc.

MORE FLEXIBILITY THROUGH HBBTV

For television broadcast stations, today the main focus is on live captioning and the subsequent captioning of archived material. The areas of application are many and varied and include news, journal programs, talk shows, sitcoms and sporting events. The following options are available for television stations to choose from: subtitles (for translation), captions (for the hard of hearing), closed-captions (display/activation for the hard of hearing only), live captioning, or open/burnt-in captioning (keyed onto video, visible on the screen at all times).

Transmitting audible content via teletext has proven to be a successful method for many television stations. However, applications such as HbbTV or smart TV apps and second-screen apps offer more options. Due to the huge increase in the use of new devices also referred to as second-screen devices, such as smartphones and tablets, these types of devices will also be an important transmission channel alongside HbbTV. The transmitted content is displayed in a second screen app that the hard of hearing can install on their mobile devices.

Several areas of application for HbbTV are available to television stations, and the options can also be used across multiple channels. Sign language closed captioning and subtitle translations can all be displayed simultaneously. It is also possible to display the transcript of an entire program. This makes it easier for the hard of hearing to place the content in its proper context. Linking the second screen to television content via QR codes is another new option. After scanning the code, users with hearing loss immediately receive all of the services available for the current program.

If a station already provides a teletext option, a special caption inserter can be used to display live subtitles without the need to replace existing infrastructure. To do this, the live caption inserter is connected to the cloud system via VPN access. The live inserter then merely replaces the subtitle pages (e.g. 777, 160, 150) in the transport stream.

A fully accessible web player is another interesting method that removes communication barriers for not only the deaf and hard of hearing, but also for the blind (using keyboard control). The advantage of a fully inclusive web player lies in the fact that text, sign language and original video content are also time-synchronized for content that is broadcast live.

The caption text and sign language streams are transmitted as clean feeds (i.e. every audio-visual element like the original sound, sign language video or audio description for the blind, is transported to the player individually). This allows the user to select individual settings. Object size, position and appearance can be adjusted as desired. Currently, rtmp (the de facto industry standard at the moment) is used as a transmission protocol for audio and video (H.264). For the purpose of text transmission, a special real-time text-streaming protocol (VerbaVoice Live Text (VLT)) was developed which has a technical delay of just a few milliseconds.

In keeping with the options made available by the HbbTV standard, the plan is to gradually integrate all of the web player functions into the HbbTV app as well.

QUALIFIED JOBS FOR INTERPRETERS WITH SENSORY IMPAIRMENTS

The combined use of the ITS for live events and TV broadcast poses further opportunities to provide full inclusion for people with hearing, visual or mobility impairments. The spoken content of e.g. a plenary debate can be transferred to a speech-to-text reporter (STTR) or sign language interpreter online who can work from anywhere through the cloud-based and secure online platform of the ITS. The produced live text or sign language video is then displayed on screens, through the accessible web player and the VerbaVoice mobile apps including augmented reality devices like the Sony SmartEyeglass which the VerbaVoice iCap solution turns into live subtitling glasses to be used on-site.

Inclusion thus works in more than one dimension: Deaf and hard of hearing people get access to information and political events while the web player and HbbTV app enable people with mobility impairments to follow online. At the same time, the system creates qualified job opportunities for deaf sign language interpreters and blind speech-to-text-reporters. Deaf sign language interpreters base their translation on the live text provided by the STTR and thus interpret into their native language (sign language). At the same time, speech recognition and a braille display connection enable blind STTR to produce the transcription of what they hear. Moreover, the location-independent online system gives mobility impaired STTR the chance to work from home. The system is already in use in several German state parliaments, making all public debates accessible through live text and sign language.
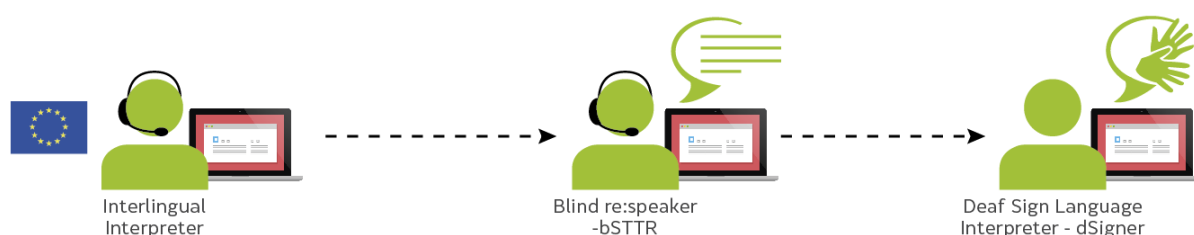

Figure 2 – The live workflow of blind respeakers and deaf sign language interpreters

OUTLOOK: AUTOMATIC SOLUTIONS AND MONETIZATION

Since for providing live captioning via re-speaking the logistics and practical costs are currently prohibitive broadcasting companies need to invest in an automated system that can also be used in live scenarios.

A consortium of international technology and research partners is currently working on a project to deliver fully automated, high-quality captions for television programs in real time. These Automatic Captioning Systems (ACS) will have a very positive impact in terms of both fully inclusive accessibility and added possibilities for using content across different forms of media (for example, MAM searches or for SEO purposes).

Figure 3 – Cross-media solutions for broadcasters

The solution will be integrated into an innovative and scalable workflow and into a software environment for producing and distributing transcriptions. The System will provide the following access services:

Audio descriptions for blind and visually impaired people

Same language subtitles for deaf and hearing impaired people

Subtitles in foreign languages to overcome international language barriers

It will be built in an ecosystem with constant iteration of results allowing continuous inclusion of advances in the quality of the generated access services. The project also aims to monetize access services in different segments, to make them profitable rather than costly for content providers. The project is based on results from partners' previous research & development projects like DTV4ALL, HBB4ALL, HBB-NEX, SAVAS, SUMAT, TV-RING or LiveCaption. The new project aims to bring the results of these R&D projects into operation in the real world.

The services shall be provided both in automatic and human (semi-automatic) modes. The speech recognition and audio description engines will generate live text. Directly after a text element (e.g. a word) is created, humans will be able to edit it. This workflow will allow text quality output to be high while error rates will remain low. The text can then be automatically translated into a different language. From the generated live text or the translated text, a text to speech (TTS) engine can generate spoken output (voicing) live.

Depending on the chain of access services that need to be performed the delay varies. Whenever possible, the original corresponding audio / video stream will be delayed (e.g. in

the internet or on HbbTV) to re-synchronize voicing, subtitles, etc. with content for a better user experience.

REFERENCES

Hadmut Holken, Vincent Bullich, 2012. Accessibility as Driver for Innovation, Content production, eInclusion, Connected TV, Market Perspectives in the Value Chain, LabSIV, University of Paris 13.

Weitnauer, M., Bohne, M, 2014. Automatic and customizable improvement of the speech intelligibility from TV signals for hearing impaired people, Proceedings of Tonmeistertagung 2014.

Simon, C. and Fassio, G., 2012. Optimierung audiovisueller Medien für Hörgeschädigte – Hörtest zur Überprüfung der Wirksamkeit im Bezug auf die Sprachverständlichkeit.

Fuchs, H., Tuff, S. and Bustad, C., 2012. Dialogue Enhancement – technology and experiments, EBU Technical Review.

Martínez Pérez, Juan and Romero-Fresco, Pablo, 2011. Apples and Oranges: A Comparative Study of Accuracy in Respeaking. 4th International Media for All Conference, Imperial College London.

Romero-Fresco, Pablo 2011. The NER Model to Assess Accuracy in Respeaking. ITU Workshop on Telecommunications Relay Services for Persons with Disabilities. [Electronic version available at: http://www.itu.int/ITU-T/worksem/trs/programme.html]

Linder, Gion and Martínez, Juan 2013. Access the Accuracy for Live Subtitles in a Quick and Effective Way. ITU Workshop on Making Media Accessible to all: the options and economics. [Electronic version available at: http://www.itu.int/en/ITU-T/Workshops-and-Seminars/mmaa/201310/Pages/programme.aspx]

ONLINE REFERENCES

http://www.verbavoice.de/live-caption

http://www.who.int/mediacentre/factsheets/fs300/en/

http://www.who.int/mediacentre/factsheets/fs300/en/

http://ec.europa.eu/digital-agenda/en/audiovisual-media-services-directive-avmsd

http://www.ebu.ch/en/legal/index.php

https://tech.ebu.ch/docs/i/i044.pdf

http://www.eeurope-standards.org/e_accessibility.htm

http://www.hbb4all.eu

https://www.irt.de/de/publikationen/technische-richtlinien.html